Vorhersage von Krebsrisiken auf der Basis nationaler Gesundheitsdaten

Wissenschaftlerinnen und Wissenschaftler vom Deutschen Krebsforschungszentrum (DKFZ) und vom European Bioinformatics Institute EMBL-EBI, Hinxton, UK, nutzen die dänischen Gesundheitsregister, um die individuellen Risiken für 20 verschiedene Krebsarten mit hoher Treffsicherheit vorherzusagen. Das Vorhersagemodell lässt sich auch auf andere Gesundheitssysteme übertragen. Es könnte helfen, Menschen mit hohen Krebsrisiken zu identifizieren, für die man gezielt individuelle Früherkennungsangebote im Rahmen von Studien erproben könnte.

Werden Krebserkrankungen früh erkannt, sind die Heilungschancen in der Regel höher und die Betroffenen müssen weniger intensiv behandelt werden. Doch Screeningprogramme zur Krebsfrüherkennung gibt es bisher nur für wenige Tumorerkrankungen – und längst nicht alle Menschen nehmen diese Angebote wahr.

Gäbe es eine einfache Möglichkeit, Menschen mit sehr hohen Krebsrisiken herauszufiltern, so könnten für diese gefährdeten Personen gezielt Früherkennungsangebote entwickelt werden. Eine Machbarkeitsstudie dazu haben Forschende um Moritz Gerstung vom DKFZ und vom European Bioinformatics Institute EMBL-EBI, Hinxton, UK, nun veröffentlicht. Die Informatiker nutzten die umfassenden Daten der dänischen Gesundheitsregister, in denen alle klinischen Diagnosen der Bevölkerung hinterlegt sind, um für 20 verschiedene Krebsarten die individuellen Erkrankungsrisiken zu quantifizieren.

Die Forscher trainierten zunächst ein Vorhersagemodell an den Daten von 6,7 Millionen erwachsener Dänen aus den Jahren 1995 bis 2014. In den Trainingsdatensatz flossen mehr als 1000 unterschiedliche Vorerkrankungen ein, sowie Krebserkrankungen bei Familienmitgliedern, Alter und – wo vorhanden – grundlegende Körperdaten sowie Risikofaktoren wie Tabakkonsum oder Übergewicht.

Anschließend wurde das Modell an den Datensätzen von 2015 bis 2018, die 4,7 Millionen Dänen umfassen, validiert und lieferte eine hohe Vorhersagegenauigkeit. Das Modell ermöglicht eine Aussage über die individuellen Risiken, an 20 verschiedenen Tumorarten zu erkranken. Über den Lauf des Lebens erreichte das Modell eine Genauigkeit von 81 Prozent. Unter Berücksichtigung von Alters- und Geschlechtseffekten ergab sich eine Genauigkeit von 59 Prozent. Die höchste Präzision erzielte das Modell für Krebserkrankungen des Verdauungssystems, sowie für Schilddrüsen-, Nieren- und Gebärmutterkrebs.

Um zu prüfen, ob sich diese Vorhersageleistung auch in den Gesundheitsdaten anderer Länder bestätigte, validierten die Forschenden ihr Modell auch an Daten der britischen UK Biobank und erzielten eine vergleichbare Treffsicherheit. Die Analysen ermöglichen keine exakte Vorhersage, bei welcher Person Krebs auftreten wird. Sie stellen aber das individuelle Risiko fest und ermöglichen einen Vergleich mit Personen vergleichbaren Alters.

„Mit der Arbeit wollten wir demonstrieren, dass es grundsätzlich möglich ist, individuelle Krebsrisiken auf der Basis nationaler Gesundheitsdaten zu modellieren", erklärt Moritz Gerstung. Eine solche Risikostratifizierung könnte dabei helfen, weiterführende Untersuchungen gezielt denjenigen Personen anzubieten, die am meisten davon profitieren würden. Neben etablierten Früherkennungen könnten dies beispielsweise in Zukunft blutbasierte Krebstests sein, an denen weltweit intensiv geforscht wird und die teilweise bereits in klinischen Studien überprüft werden. Die Hoffnung: Künftig könnten nach einer Risikostratifizierung mit einer bestimmten Anzahl von Tests mehr Krebserkrankungen entdeckt, Personen mit geringem Risiko unnötige Tests erspart und falsch-positive Ergebnisse und Überdiagnosen vermieden werden.

Allerdings, so stellt Moritz Gerstung klar, ist dafür eine geeignete Datenbasis unverzichtbar. „Die dänischen Gesundheitsdaten sind einzigartig, da sie einen großen Zeitraum abdecken und miteinander verknüpft werden können. Vergleichbares bieten nur wenige europäische Länder, etwa Finnland und Schweden oder spezielle Forschungskohorten in Großbritannien.

Auch in Deutschland laufen Bemühungen, nationale digitale Gesundheitsinfrastrukturen aufzubauen. „Sinnvoll wäre es, bereits bei der Planung zu berücksichtigen, welche Art von Daten für die Bewertung des Krebsrisikos am besten geeignet sind", so Gerstung. In seiner aktuellen Arbeit haben sich die ICD-10-Diagnosecodes, die auch in anderen europäischen Gesundheitssystemen verwendet werden, als nützlich erwiesen.

Da auch grundlegende Angaben zu Körpermaßen und bekannten Risikofaktoren wie etwa Tabakkonsum wichtige Informationen lieferten, erscheint es ratsam, die Erfassung solcher Informationen auf Bevölkerungsebene zu erleichtern. „Wären diese Angaben in den dänischen Gesundheitsregistern flächendeckend verfügbar gewesen, hätte unser

Vorhersagemodell wahrscheinlich eine noch deutlich höhere Treffsicherheit erzielt", resümiert Gerstung.

Publikation:

Jung, Alexander W., et al. "Multi-cancer risk stratification based on national health data: a retrospective modelling and validation study." The Lancet Digital Health 2024, DOI: 10.1016/S2589-7500(24)00062-1

Pressemitteilung

23.05.2024
Quelle: Deutsches Krebsforschungszentrum (DKFZ)

Weitere Informationen

- Deutsches Krebsforschungszentrum