

## Karzinom-Screening durch Analyse von Daten aus Internet-Suchmaschinen?

Internet-basierte Suchmaschinen gehören zum globalen Alltagsleben. Mehr als 3 Mrd. Suchanfragen werden täglich vom Marktführer (Google) bearbeitet (1, 3). Die Suchdienste registrieren bei jeder einzelnen Anfrage die Suchbegriffe, die Suchergebnisse, den Zeitpunkt und (anonymisiert) den Rechner, von dem aus die Anfrage eingegeben wurde. Diese Daten können 18 Monate rückverfolgt werden. Die daraus kalkulierbaren zeitlichen und geographischen Trends der eingegebenen Suchbegriffe sind inzwischen Gegenstand statistischer Analysen. Diese können (in eingeschränktem Ausmaß) von jedem Anwender auch selbst durchgeführt werden (2). Korrelationen der Suchbegriffe wurden nachgewiesen mit sozioökonomischen Parametern, z.B. Arbeitslosigkeitsraten, und epidemiologischen Phänomenen, z.B. Grippeausbrüchen (3).

Eine aktuell publizierte Studie weist auf ein theoretisch mögliches medizinisches Einsatzgebiet solcher "Big Data"-Analysen hin, nämlich das Screening auf maligne Erkrankungen, bei denen die Prognose von einer frühzeitigen Diagnose abhängt (4).

**Methode:** Es wurden retrospektiv Millionen von US-amerikanischen (englischsprachigen), anonymisierten Internet-Suchanfragen an die Microsoft-Suchmaschine Bing.com nach bestimmten Kriterien analysiert. Anwender, bei denen offensichtlich mit sehr hoher Wahrscheinlichkeit ein Pankreaskarzinom (PK) diagnostiziert wurde, wurden als "Positivfälle" betrachtet. Dies wurde dann angenommen, wenn eine spezifische vordefinierte Suchanfrage vorlag (z.B. "Warum habe ich ein PK bekommen?"; "Mir wurde gesagt, ich habe ein PK; was erwartet mich?" etc.) und bis eine Woche zuvor zumindest einmal nach vordefinierten typischen Symptombegriffen sowie deren Synonymen gesucht wurde (z.B. "gelbe Haut", "dunkler Urin", "weicher Stuhl", "plötzlicher Gewichtsverlust" etc.). Die Suchanfragen-Trends der "Positivfälle" wurden dann mit "Negativfällen" verglichen. Als solche wurden all jene Anwender angesehen, die nach denselben Symptombegriffen gesucht hatten, aber innerhalb eines definierten Intervalls (dreifache Standardabweichung des Zeitintervalls zwischen Symptombeginn und "Diagnosestellung" in den "Positivfällen") offenbar kein PK entwickelten. Ein Teil der so gefundenen "Negativ- und "Positiv-Fälle" wurde aufgrund bestimmter Charakteristika im Suchverhalten von der Analyse ausgeschlossen; so wurde z.B. bei einem Anteil von > 20% medizinischen Suchanfragen davon ausgegangen, dass es sich nicht um "Patienten", sondern um Angehörige von Gesundheitsberufen handelte). Die Autoren entwickelten aus den Suchaktivitäten mathematische Algorithmen, die es erlauben, künftige "Positivfälle" aus den Suchaktivitäten einer großen Zahl von Benutzern vorherzusagen und herauszufiltern. Schwerpunkt war, das Verhältnis von falsch positiven Vorhersagen (FP) zu richtig positiven Vorhersagen (RP) möglichst klein zu halten.

**Ergebnisse:** Es wurden 1.072 “Positivfälle” und 3.025.046 “Negativfälle” gefunden. Die Autoren errechneten, dass es bei Beschränkung auf einen FP:RP-Bereich von 0,0001 bis 0,00001 möglich sei, zwischen 5 und 30% der Positivfälle frühzeitig aus den Suchdaten zu erkennen. Durch Kombination mit anderen Daten (Zahl und Kombinationen von Suchbegriffen und Synonymen, Alter, sonstiges Suchverhalten etc.) könne die Vorhersagekraft noch weiter verbessert werden. Je kürzer der Zeitabstand zur “Diagnosestellung”, desto genauer sei auch die Vorhersage.

**Einschränkungen der Aussagen und Diskussion:** Die Analyse beruht ausschließlich auf Internet-Suchdaten. Es wurde nicht überprüft, ob bei den “Positivfällen” tatsächlich ein PK diagnostiziert wurde. Suchanfragen (sogenannte “Web Search Logs”) sind außerdem nur einem bestimmten Rechner und nicht einem bestimmten Anwender zuzuordnen. Es bestehen somit erhebliche Unschärfen im Vergleich zu medizinischen Studien. Unter den Autoren der Studie sind Forscher von “Microsoft Research” und ein Ph.D.-Student der Columbia University (auch er ein ehemaliger Microsoft-Mitarbeiter). Es handelt sich um rein theoretisch-mathematische Analysen, wie sie sonst zu anderen Zwecken, z.B. Marktforschung oder Fahndung, angewandt werden. Mit gängigen medizinischen “aktiven” Screeninguntersuchungen, bei denen die Spezifität gegenüber der Sensitivität eine untergeordnete Rolle spielt, ist dieses Verfahren eines “passiven Bevölkerungsscreenings” jedenfalls nicht vergleichbar. Die Autoren sagen selbst, dass die Information und Beratung der Anwender mit auffälligem Suchverhalten eine “Herausforderung” darstelle. Unseres Erachtens zeigen solche Analysen vor allem das grundsätzliche Potenzial, das die zunehmende Digitalisierung und Vernetzung des Alltagslebens mit sich bringt. Je mehr gesundheits- und krankheitsbezogene Informationen von Anwendern selbst (direkt oder indirekt über Softwareanwendungen und Hardwarezubehör; vgl. 5) im Internet preisgegeben werden, desto mehr wird diese Entwicklung auch den Gesundheitssektor betreffen. Dieses Potenzial birgt sicher Chancen, aber auch zahlreiche Gefahren. So wie “Big-Data“-Analysen keine randomisierten kontrollierten Studien ersetzen können (6), so können sie auch nicht Eigenverantwortung der Patienten und eine sorgfältige Anamnese und Untersuchung durch den Arzt ersetzen. Ob sie möglicherweise zusätzlich einen positiven Effekt haben können, ist höchst zweifelhaft. Es bleibt vielmehr der unangenehme Eindruck, dass das Risiko des (kommerziellen) Missbrauchs von Anwenderdaten wesentlich größer ist als der potenzielle Nutzen.

**Fazit:** Eine aktuelle, industriennahe Untersuchung zeigt, dass es möglich ist, durch eine bevölkerungsbasierte Analyse des Suchverhaltens von Anwendern im Internet statistische Hinweise auf Frühsymptome und dadurch möglicherweise Frühstadien maligner Erkrankungen zu entdecken. Ob damit tatsächlich – wie von den Autoren postuliert – ein sinnvolles, kostengünstiges und großflächiges Monitoringinstrument zur Verfügung steht, muss stark bezweifelt werden. Einer praktischen Anwendung stehen zahlreiche ungeklärte methodische, ethische und juristische Fragen entgegen.

## Literatur

1. [de.wikipedia.org/wiki/Suchmaschine](https://de.wikipedia.org/wiki/Suchmaschine) [Link zur Quelle](#)
2. [www.google.com/trends/](https://www.google.com/trends/) [Link zur Quelle](#)
3. [en.wikipedia.org/wiki/Google\\_Search](https://en.wikipedia.org/wiki/Google_Search) [Link zur Quelle](#)

4. Paparrizos, J., et al.:J. Oncol. Pract. Juni2016. [Link zur Quelle](#)
5. AMB 2013, **47**, 87b. [Link zur Quelle](#)
6. AMB 2016, **50**, 32DB01. [Link zur Quelle](#)