



Universität Hamburg

Abteilung Kommunikation und Marketing

Referat Medien- und Öffentlichkeitsarbeit

Tel.: +49 40 42838-2968

E-Mail: medien@uni-hamburg.de

27. Oktober 2022

66/22

WIE HOCH IST DIE TATSÄCHLICHE INFektionsZahl?

## **FORSCHENDE BESTIMMEN CORONA-DUNKELZIFFER MIT SUCHANFRAGEN-ANALYSE**

**Die Corona-Fallzahlen sind in der Pandemie ein sehr wichtiger Parameter. Die Dunkelziffer der Covid-19-Infektionen in Deutschland wird dabei immer als hoch eingeschätzt. Doch wie kann man sie bestimmen? Christina Maaß aus dem Fachbereich Volkswirtschaftslehre der Universität Hamburg hat einen neuen Ansatz entwickelt, sich dieser Zahl durch die Analyse von Suchanfragen bei Google anzunähern. Die Studie wurde im Fachmagazin „PLOS ONE“ veröffentlicht.**

Um die Dynamik der Corona-Pandemie zu verstehen und die ergriffenen Maßnahmen zu bewerten, ist eine möglichst genaue Bezifferung der Anzahl der Erkrankten erforderlich. Doch die Zahlen sind oft unvollständig, etwa weil es zu Anfang der Pandemie noch keine umfangreichen Testkapazitäten gab, weil Erkrankte trotz Symptomen nicht zum Testen gehen oder keine Symptome zeigen und die Infektion nicht bemerken.

Die Analyse großer Datenmengen auf Basis von Suchmaschinen oder Social Media hat sich bereits bei anderen Erkrankungen, etwa der Grippe, als sinnvolle Analysebasis für das Infektionsgeschehen bewährt – auch weil die Daten schnell und frei verfügbar sind. Ein Team unter Leitung von Christina Maaß von der Universität Hamburg hat nun erstmals untersucht, ob diese digitalen Nutzungsdaten geeignet sind, die Informationslücke zu den tatsächlichen Corona-Fallzahlen für Deutschland zu schließen.



Dafür brachten die Wissenschaftlerinnen und Wissenschaftler die offiziellen Angaben des Robert Koch-Instituts (RKI) mit den Corona-Anfragen bei der meistgenutzten Suchmaschine Google zusammen. Diese als „Google Trends“ bekannten Daten geben die Entwicklung des Suchvolumens für einen Begriff in einem bestimmten Zeitraum an. Berücksichtigt wurden die Suchbegriffe „Geruchsverlust“, „Geschmacksverlust“, „Testcenter“, „Quarantäne“, „Coronatest“ und „Lungenentzündung“.

In der Untersuchung konnte auf Bundes- und Landesebene ein signifikanter Zusammenhang zwischen dem Suchvolumen einzelner Begriffe und den vom RKI herausgegebenen Zahlen während der ersten drei Infektionswellen (März bis Mai 2020, Oktober 2020 bis Januar 2021 und Februar bis Mai 2021) nachgewiesen werden. Mit den offiziellen Zahlen stiegen und sanken auch die Suchen. Auffällig war, dass sich die Suchbegriffe mit dem stärksten Zusammenhang zur offiziellen Zahl während der Wellen unterschieden: In der ersten Welle war es der Begriff „Quarantäne“, in der zweiten „Geruchsverlust“ und in der dritten „Coronatest“.

Darüber hinaus konnten die Forschenden durch verschiedene statistische Verfahren zum einen zeigen, dass eine kausale Beziehung zwischen Suchanfragen und registrierten Infektionen besteht, und dass zum anderen die Entwicklung des Suchvolumens für die Schätzung von Infektionsraten genutzt werden kann. Auf dieser Basis entwickelten sie eine Berechnungsmethode für die Dunkelziffer, nach der die Zahl der Infizierten in der ersten Welle um 31 Prozent höher lag als die offizielle Zahl. Im Winter 2020/21 lag die Dunkelziffer bei 43 Prozent und im Frühjahr 2021 – mit Beginn der weit verbreiteten Testmöglichkeiten – bei 28 Prozent.

„Mit unserer Methode können wir die offizielle Zahl vor allem um die Erkrankten ergänzen, die leichte Symptome haben, sich aufgrund der Testkapazitäten oder -vorgaben aber nicht testen lassen können oder es nicht wollen“, erklärt Christina Maaß, wissenschaftliche Mitarbeiterin im Team der Professur „VWL, insb. Internationale Wirtschaftsbeziehungen“ von Prof. Dr. Thomas Straubhaar. Damit sei diese Herangehensweise vor allem für die Phasen einer Pandemie geeignet, in denen es keine uneingeschränkten Testangebote gibt.

„Zusätzlich können die asymptomatisch Infizierten mit einem Mittelwert aus den Ergebnissen deutscher und internationaler Studien berücksichtigt werden, der bei rund 20 Prozent liegen würde. In Kombination mit unseren Ergebnissen würde die tatsächliche Zahl der Infizierten bis zu 72 Prozent höher liegen als die offizielle Zahl“, so Maaß.

Das Forschungsteam sieht in dem Big-Data-Ansatz eine wichtige Ergänzung bestehender Instrumente: „Wir konnten zeigen, dass die Analyse der Suchanfragen Aussagen über das tatsächliche



Infektionsgeschehen zulässt“, resümiert Maaß. Durch die weitere Differenzierung der untersuchten Begriffe, etwa durch maschinelles Lernen, könne die Methode ein fester Baustein in der Bestimmung des Infektionsgeschehens werden.

**Originalpublikation:**

Maaß, Christina (2022): Shedding light on dark figures: Steps towards a methodology for estimating actual numbers of COVID-19 infections in Germany based on Google Trends. PLoS ONE 17(10): e0276485. <https://doi.org/10.1371/journal.pone.0276485>

**Für Rückfragen:**

Christina Maaß  
Universität Hamburg  
Fakultät für Wirtschafts- und Sozialwissenschaften  
Fachbereich Volkswirtschaftslehre  
Tel.: +49 40 42838-4025  
E-Mail: [christina.maass@uni-hamburg.de](mailto:christina.maass@uni-hamburg.de)

