

Wo sind die Virologinnen? Geschlechterungleichheiten bei der Autorenschaft von COVID-19-Papern und in der Medienberichterstattung

Eine Datenanalyse von 1995 Studien mit Covid-19-Bezug und mindestens einer Autorin oder einem Autor einer Forschungsinstitution aus der Schweiz zeigt, dass der Anteil an Frauen insgesamt 38 Prozent beträgt. Jedoch publizierten die Forscherinnen insgesamt öfters als Erstautorin: Der Anteil hierzu beträgt 43 Prozent.

Der Artikel

Noch unpubliziert (siehe PDF des Textes im Anhang). Die Publikation einer erweiterten Variante des Textes mit den dann aktuellen Daten erfolgt im Frühling nach der Veröffentlichung der neusten Untersuchung des FÖG zur Medienberichterstattung über die Corona-Pandemie.

Ausgangsthese

Forscherinnen bei wissenschaftlichen Publikationen zu COVID-19 sind gegenüber Männern unterrepräsentiert, sowohl als Mitautoren als auch als Erstautorinnen. Diese Tendenz widerspielt sich in der Untersuchung des FOEG zur Berichterstattung während der ersten Welle der Corona-Pandemie die zeigte, dass vorrangig Wissenschaftler in der Corona-Berichterstattung zu Wort kamen. Ich möchte diese Tendenz in der Medienberichterstattung mit der Häufigkeit der Autorenschaft in wissenschaftlichen Papern vergleichen.

Idee

Expertinnen und Experten prägen die Sichtweisen, wie wir ein komplexes Phänomen wahrnehmen. Fachwissen wird heruntergebrochen und vereinfacht, so dass eine breite Öffentlichkeit einen wissenschaftlichen Sachverhalt besser verstehen kann. Für ein Ausnahmeereignis wie die Corona-Pandemie, kommt Virologen, die gegenüber der Öffentlichkeit die Funktionsweise des Virus und der Pandemie erklären, eine Schlüsselrolle zu. Obwohl es in der Schweiz mehrere Virologinen in führenden Forschungspositionen gibt, etwa Isabella Eckerle, Leiterin der Abteilung für Infektionskrankheiten am Universitätsspital Genf oder Emma Hodcroft, Epidemiologin an der Universität Bern, kamen bei der Berichterstattung während der ersten Welle vor allem männliche Epidemiologen zu Wort. Davon ausgehend wollte ich untersuchen, wie es sich mit der Gewichtung von Forscherinnen bei COVID-19 Forschungspapern verhält. Gibt es auch hier ein Ungleichgewicht der Geschlechterverteilung?

Einschätzung von Aufwand/Ertrag

Die internationale PUB-MED Forschungsdatenbank, mit der ich arbeiten möchte, ist gut strukturiert. Ich kann die Suchergebnisse als csv-Datei abspeichern. Allerdings sind ausgerechnet bei den Vornamen, die für eine Analyse des Geschlechts der Forscher_innen entscheidend sind, die Daten abgekürzt. Sie bestehen nur aus dem ersten Buchstaben. Auch die Informationen zu den Instituten der Forscher_innen sind nur in den separat abrufbaren Metadaten vorhanden. Dies macht die Zuordnung zum bestehenden Datensatz in der csv-Datei umständlich.

Knackpunkte

Bei den Metadaten zu den Forschungspapern auf PUB-MED fehlen leider bei einigen Forschern_innen immer noch die Angaben zu den Vornamen. Diese Vornamen, die nur

aus einem einzelnen Buchstaben bestehen, dürfen bei der Analyse der statistischen Häufigkeit von Forscherinnen und Forschern nicht berücksichtigt werden, weil es sonst zu Verzerrungen bei der Berechnung kommt. Weiter weist der Genderclassifier des Python Moduls NTLK, mit dem ich arbeite, leider nur eine Genauigkeit von 80% bei der Zuordnung des Geschlechts auf. Dies macht die kritische Überprüfung der Resultate, zumindest bei der Analyse der Erstautorenschaft von Schweizer Forscherinnen und Forschern, notwendig. Eine zusätzliche Analyse der Berücksichtigung von Forscherinnen bei der Medienberichterstattung sprengt den zeitlichen Rahmen des Projektes. Zudem wird diese Analyse bereits vom Medienforschungsinstitut FOEG gemacht.

Briefingperson

Zu Beginn des Projektes habe ich mit Stephanie Schnydrig, Wissenschaftsredaktorin von Keytone-SDA, mit der zusammen ich das Projekt realisiere, detaillierter über das Thema gesprochen. Schnydrig hat sich bereits genauer mit der Thematik auseinandergesetzt. So wurde im Schweizer Forschungsmagazin «Horizonte» darauf hingewiesen, dass es eine Geschlechterdiskrepanz bei der Autorenschaft von Papern über COVID-19 gibt. Auch der Wissenschaftsartikel «Where are the women? Gender inequalities in COVID-19 research authorship» der Publikation BMJ Global Health hat sich bereits dem Thema angenommen. Wir beschliessen, diese Analyse für die Schweiz vorzunehmen. Für den Artikel spricht Schnydrig mit Daniel Vogler, Medienwissenschaftler am Forschungszentrum Öffentlichkeit und Gesellschaft (FÖG) der Universität Zürich. Vogler ist an der Auswertung der Schweizer Berichterstattung zur zweiten Corona-Welle beteiligt.

Datensatz/Programmiercode/Visualisierung

URL: <https://cutt.ly/Zk7Tv9L>

Arbeitsprotokoll /Vorgehen

Von Mitte Dezember 2020 bis Mitte Februar 2021 arbeite ich – unter Mithilfe der Dozenten – an 10 bis 12 Tagen an folgenden Arbeitsschritten:

1. Ich suche nach allen Publikationen zu "Covid-19" und "Switzerland" auf PUB-MED, lade die Resultate runter und exploriere die Daten
2. Ich baue mit Hilfe des PUB-MED-Metadatenfiles auf der PUB-MED Website (PUB-MED; All Results; Format: PUB-MED) zu allen Suchergebnissen einen Dictionary mit den vollständigen Namen der Autoreninnen, mit Hilfe dessen ich die unvollständigen Vornamen der Forscher_innen in der csv-Datei ergänzen kann. Dies mache mit Hilfe einer List Comprehension.
3. Ich baue den Gender-Identifizierer "Gender Identification by name using NLTK", mit dem ich das Geschlecht der Autoren_innen analysieren kann. Die Zuweisung des Geschlechts geschieht mit Hilfe von Testlisten, mit denen das Programm lernt, männliche und weibliche Namen zu erkennen.
4. Ich kombiniere den Gender-Identifizierer mit einer Listen-Funktion, welche die männlichen, weiblichen und wegen fehlenden Angaben nicht zuordenbaren Namen zählt. Hiermit kann ich die ersten Auswertungen machen.
5. Für die Auswertung der Erstautorinnen baue ich nochmals eine List Comprehension, um die Autorennamen zu vervollständigen. Mit Hilfe des Gender-Identifizierers und mit einer Listen-Funktion werde ich die Aufteilung der Erstautorinnen und Erstautoren aus.

6. Mit Hilfe der Metadaten und einer if/exept-Funktion bauch ich einen neuen Dataframe, um die Anzahl Forschern an Schweizer Instituten berechnen zu können. Diesen durchsuche ich nach den Erstautoren und den Instituten, die ich aneinander zuweise und dann, um mich auf die Schweizer Forschungsinstitute einzuschränken, nach den Keywords «Switzerland, Zurich, Lucerne, Basel, Berne, Geneva, Lausanne, ETH, EPFL» eingrenze.