

Meldeverzug der Neuinfektionen in den RKI Corona-Daten

Anfängerpraktikum, BSc Statistik, LMU München, Institut für Statistik



Vortragsgruppe: Quang Le, Qingyan Lu, Wolfgang Stefani, Joshua Weinert
Projektpartner: StaBLab (Maximilian Weigert)
Betreuer: Patrick Kaiser

Beschreibung des Projekts

Für dieses Projekt wurden uns von unserem Projektpartner, dem Statistischen Beratungslabor (StaBLab), 698 Tagesdatensätze an vom RKI publizierten Coronadaten im Zeitraum vom 27. März 2020 bis zum 22. Februar 2022 übermittelt. Diese Datensätze enthielten Informationen zu aktuellen Infektionszahlen, sowie mit Coronainfektion in Verbindung stehenden Todes- und Genesungsdaten. Konkret wurden uns von unserem Projektpartner hierfür 3 Aspekte genannt, anhand derer wir die Daten analysieren sollten: **(1)** Die Analyse der Veränderungen beim Meldeverzug im Zeitverlauf, **(2)** der Darstellung von Unterschieden zwischen verschiedenen Bundesländern und Altersgruppen in Bezug auf den Meldeverzug, und **(3)** die Erstellung eines zusammengeführten Datensatzes aus den täglichen Datenbeständen des RKI.

Beschreibung der Methoden

Wir haben bei der Zusammenführung der Einzeldatensätze festgestellt, dass die Daten für 7 Tage punktuell fehlen. Aufgrund der geringen Anzahl und ihrer Verteilung über den Beobachtungszeitraum haben wir diese Daten in unserer Analyse nicht berücksichtigt. Für die Abgabe wurden diese noch entsprechend ergänzt. Anschließend haben wir den Meldeverzug als die angepasste Differenz zwischen dem Meldedatum einer Neuinfektion und dessen Publikationsdatum definiert. Für die Analyse haben wir uns hauptsächlich die relativen Anteile der mit Meldeverzug gemeldeten Fälle, sowie die absolute Verspätung des Falles im 90., 95., und 99. Perzentil angeschaut. Für diese Werte haben wir für die Veranschaulichung jeweils den 7-Tage rückblickenden gleitenden Mittelwert berechnet, welchen wir durch die wöchentliche Zyklizität des Meldeverzugs begründen konnten.

Ergebnisse

Als Resultat unserer Analyse halten wir **9** Hauptergebnisse fest. **(1)** Die meisten Neuinfektionen (ca. 65%) werden ohne Meldeverzug gemeldet (wobei dies größtenteils von den hohen Fallzahlen der Omikronwellen dominiert wird). **(2)** Wie auch bei der Meldung von Neuinfektionen, unterliegt der Anteil an mit Meldeverzug gemeldeten Fällen einem wöchentlichen Zyklus. **(3)** Insgesamt fällt der Anteil an mit Meldeverzug gemeldeten Fällen über den gesamten Betrachtungszeitraum von anfangs etwa 70% auf schlussendlich um die 30% ab. **(4)** An den meisten Tagen werden 90% bzw. 95% der Fälle innerhalb von 0 bis 1 Tagen nach Meldedatum publiziert. **(5)** 99% der Fälle sind oft erst nach erheblicher Verspätung gemeldet, teilweise erst nach weit über 100 Tagen. Diese Spätmeldungen sind zeitlich meist zwischen Phasen der Pandemie mit hohen Fallzahlen (Wellen) angesiedelt mit einem Cluster zwischen dem Mai und Oktober 2020, sowie um den Juli 2021. **(6)** Wenn diese Spätmeldungen berücksichtigt werden, korrigiert sich die tägliche 7-Tage Inzidenz pro 100.000 Einwohner, wenn man einem Tag wartet um 5-15% nach oben und insgesamt um bis zu etwa 20% (vereinzelt auch 30%+) nach oben. **(7)** Es keine großen Unterschiede im Anteil an mit Verspätung gemeldeten Fällen, sowie der absoluten Verspätung des Falles des 90. Perzentil zwischen den Altersgruppen. **(8)** Bei der Betrachtungsweise pro Bundesland fällt auf, dass der Anteil an mit Meldeverzug gemeldeten Fällen in den meisten Bundesländern gemäß dem generellen Trend über die Zeit abfällt, in Hamburg jedoch konstant bei über 95% bleibt. Absolut werden 90% der Fälle in den Bundesländern (inkl. Hamburg) bis auf einige Ausreißer, meist zum Anfang der Pandemie, jedoch mit sehr wenig Tagen Meldeverzug gemeldet. **(9)** Eine ähnliche Analyse der Todesfälle wie zu den Infektionszahlen ist aufgrund der Datenlage nicht möglich. Bei der Betrachtung der Zeit zwischen Coronainfektionsdatum und der Publikation des Todesfalles fällt auf, dass dieser in Fallintensiven Phasen der Pandemie bei um die 25 Tage liegt, zwischen den „Wellen“ jedoch vereinzelt stark ansteigt. Hierbei ist nicht klar, ob diese größere Streuung auf die niedrigeren Fall- und somit auch Todeszahlen zurückzuführen ist (weniger Fallzahlen vergrößern die emp. Varianz), oder es dort eine strukturelle Erklärung gibt.