

## **Coronavirus: Abschätzung der Testquote mittels Modellierung der tatsächlichen Infektionskurve**

Der nachfolgende Beitrag erläutert einen Ansatz zur Abschätzung der Testquote. Diese sagt aus, wie hoch der Anteil der Corona-Infizierten ist, die positiv auf eine Infektion mit dem Coronavirus getestet werden. Während allerdings die Zahl der täglichen positiven Befunde bekannt ist – sie wird vom Robert-Koch-Institut (RKI) regelmäßig veröffentlicht – liegt der Nenner, die Gesamtzahl der Infizierten, im Dunkeln. Um diese Dunkelziffer abzuschätzen, wird nachfolgend eine Kurve der tatsächlichen täglichen Neuinfektion aus den Daten des Robert-Koch-Instituts heraus mit Stand 6.5.2020 modelliert.

Dem Modellierungsansatz liegt dabei die Annahme zugrunde, dass im Verlauf der Epidemie bei einem konstanten Anteil der Infizierten eine Erkrankung diagnostiziert wird. Eine solche konstante Diagnoserate dürfte jedoch nicht der Realität entsprechen, weil sich zum einen das Verhalten der Menschen im Verlauf der Epidemie vermutlich geändert hat, der Arzt wird also häufiger oder vielleicht umgekehrt auch seltener aufgesucht, und zum anderen weil inzwischen durch angeordnete Reihenuntersuchungen zusätzliche Erkrankte gefunden werden, die wegen schwacher Symptome von sich aus nie zu einem Arzt gegangen wären. Dies wird hier aber vernachlässigt.

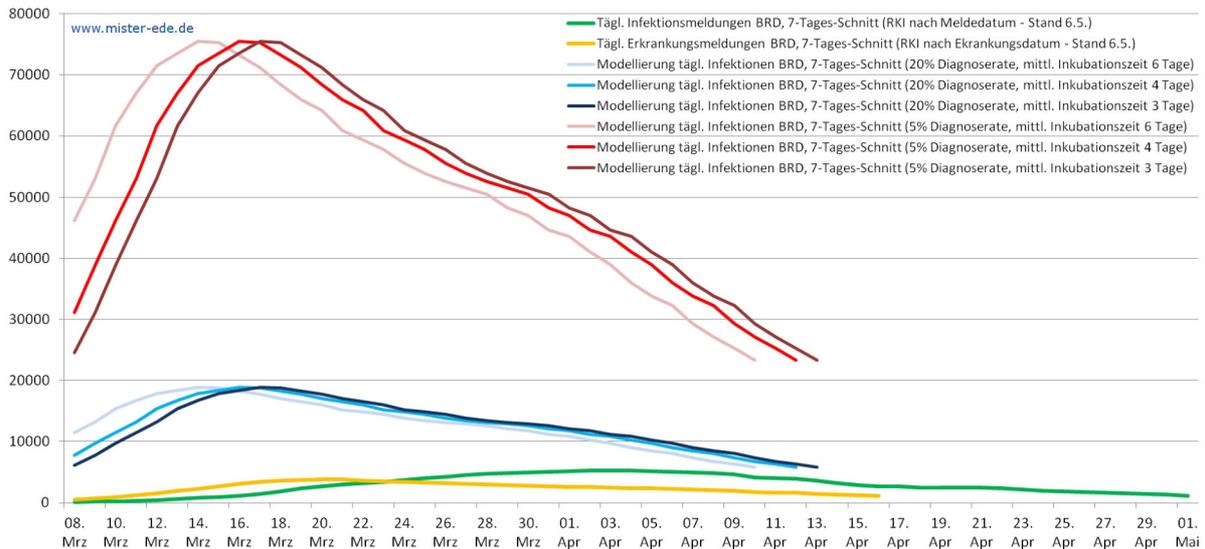
Zur Abschätzung der Testquote werden pro Tag zwei RKI-Daten verwendet, zum einen die Infektionsmeldungen nach Meldedatum und zum anderen die Erkrankungsmeldungen nach Erkrankungsdatum. Das Meldedatum spiegelt dabei das Datum wieder, an dem ein meldepflichtiger Verdachtsfall vorliegt, bei dem später eine Coronavirus-Infektion bestätigt wurde. Das Erkrankungsdatum spiegelt hingegen jenen Zeitpunkt wieder, an dem aufgrund von Symptomen eine Erkrankung an COVID-19 festgestellt werden konnte. Somit liegen beispielsweise für den 9. April die Zahl der an diesem Tag neuen und später positiv auf das SARS-CoV-2 getesteten Verdachtsfälle vor (4.923) sowie die Zahl derjenigen, die an diesem Tag erste Symptome einer COVID-19-Erkrankung verspürt haben (1.683).

### **Modellierung der Kurve der tatsächlichen täglichen Neuinfektionen:**

Um die Daten des RKI für die Abschätzung der Testquote von Tages- und Wochenendeffekten zu bereinigen, wird jeweils der 7-Tages-Schnitt der Infektionsmeldungen (grüne Kurve) und der 7-Tages-Schnitt der Erkrankungsmeldungen (gelbe Kurve) verwendet. Weil es bei den jüngeren Daten noch eine gewisse Unsicherheit gibt, die auf verzögerter Übermittlung bzw. erst später festgestellten Erkrankungen beruhen, werden die Meldedaten hier nur bis zum 1.5.2020 und die diagnostizierten Erkrankungen nur bis zum 16.4.2020 ausgewiesen. Unter den zuvor getroffenen Annahmen kann dann zunächst eine Kurve der tatsächlichen täglichen Neuinfektionen modelliert werden, indem die Parameter Inkubationszeit und Anteil der Erkrankungen unterschiedlich gesetzt werden.

Bei einer Diagnoserate von 5% und einer Inkubationszeit von 4 Tagen ist die modellierte Kurve der tatsächlichen Neuinfektion 20-mal höher und verläuft 4 Tage vor der Kurve der Erkrankungsmeldungen. Bei einer Diagnoserate von 20% und 6-tägiger Inkubationszeit ist die Kurve der tatsächlichen Neuinfektion hingegen nur 5-mal höher und verläuft dafür 6 Tage früher.

Infektions- und Erkrankungsmeldungen sowie Modellierung der Gesamtzahl der tägl. Neuinfektionen (7-Tages-Schnitt)



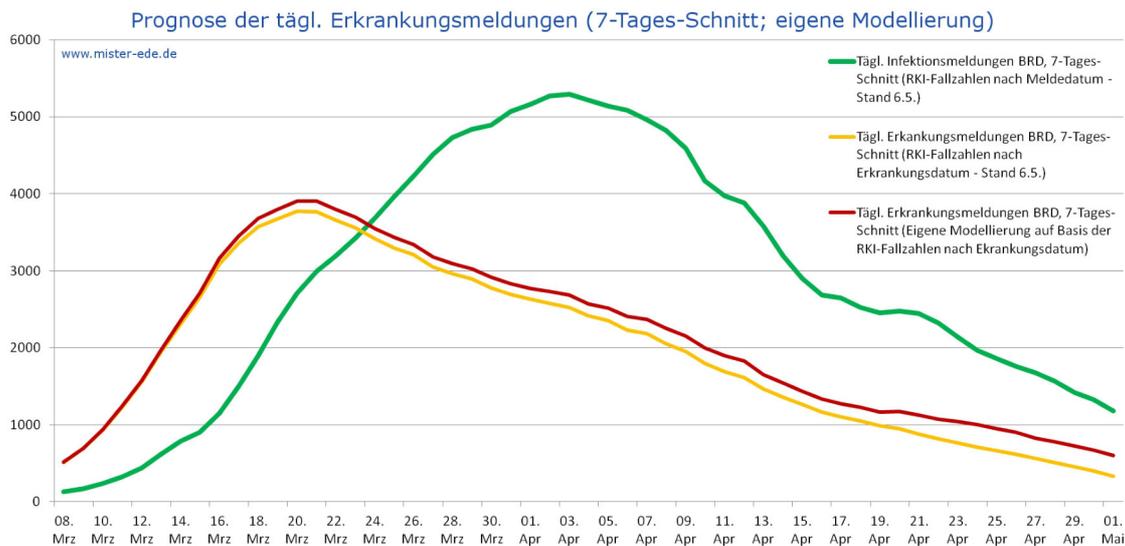
Deutlich wird, dass die Inkubationszeit keine sonderliche Rolle für die Abschätzung der tatsächlichen Neuinfektionen spielt, sehr wohl jedoch die Frage, wie viele der SARS-CoV-2-Infizierten eine positive COVID-19-Diagnose gestellt bekommen. Es wäre daher sinnvoll, durch eine repräsentative Studie einen näherungsweisen Wert für die aktuelle Diagnoserate in Deutschland insgesamt und den verschiedenen Regionen zu ermitteln. Denn der Vorteil einer Modellierung der tatsächlichen Infektionen anhand der Erkrankungsfälle besteht darin, dass diese zu Beginn einer Epidemie eher korrekt erfasst werden als die Gruppe der symptomlosen Infizierten. Außerdem ist die Zahl der Erkrankungsfälle nicht so stark von der Anzahl der Tests pro Einwohner in einem bestimmten Zeitraum oder einer bestimmten Region abhängig, wie das bei den Infektionsmeldungen der Fall ist.

#### Abschätzung der Testquote:

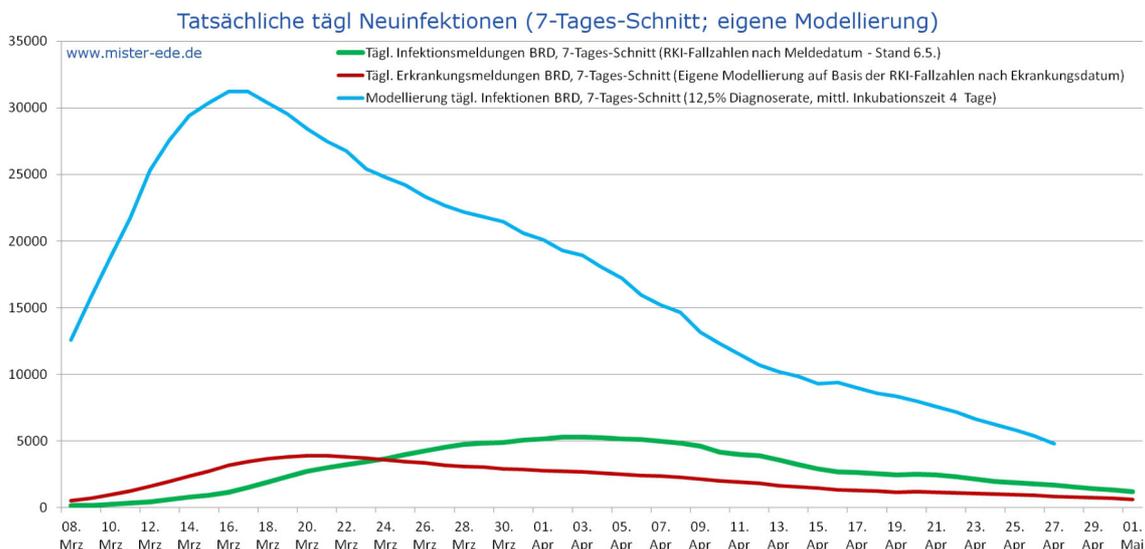
Vergleicht man die modellierte Kurve der tatsächlichen Neuinfektionen mit der Entwicklung der bestätigten Corona-Meldungen, kann in etwa abgeschätzt werden, wie viele der Corona-Infizierten über die durchgeführten Corona-Tests gefunden werden. Am Anfang der Epidemie waren das verständlicherweise nur wenige bis sich ab Ende März, Anfang April die beiden Kurven annähern. Um besser abzuschätzen, ob die Testquote zu diesem Zeitpunkt nur 15% - 20% oder doch schon deutlich mehr betragen hat, bräuchte es allerdings einen engeren Korridor für die Diagnoserate als die sehr groben 5% bis 20%. Schlussfolgern kann man aber schon aus dieser Modellierung, dass wir in Deutschland zumindest nicht mehr gänzlich im Dunkeln stochern.

### Abschätzung der Testquote auf Basis prognostizierter Erkrankungsmeldungen:

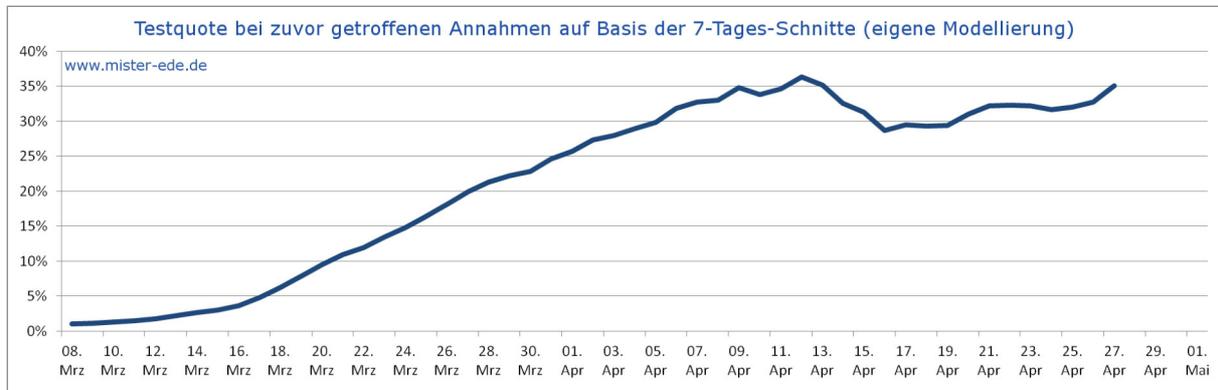
Mit einer in Studien ermittelten Diagnoserate lässt sich die Genauigkeit der Kurve erhöhen. Die Aktualität lässt sich hingegen nur mit Hilfe von Prognosen verbessern. Das RKI verwendet hierzu ein eigenes Nowcast-Modell, um die aktuelle Zahl der Erkrankungsmeldungen zu prognostizieren. Diese Zahlen könnten als Basis für eine Modellierung verwendet werden. Für die nachfolgende Modellierung wird allerdings auf einen eigenen Prognose-Ansatz zurückgegriffen, um getroffene Annahmen und Schätzfehler (z.B. bei zusätzlichen Feiertagen) besser überblicken zu können. Während die gelbe Kurve die aktuellen Rohdaten des RKI zur Zahl der Erkrankungen im 7-Tages-Schnitt darstellt, ist die rote Kurve die rudimentär prognostizierte Gesamtzahl der Personen, deren Erkrankungsbeginn für diesen Tag irgendwann diagnostiziert worden sein wird.



Für die anschließende Modellierung der Gesamtzahl der Infizierten werden eine Diagnoserate von 12,5% und eine mittlere Inkubationszeit von 4 Tagen angenommen. Die blaue Kurve der tatsächlichen Infizierten verläuft damit 4 Tage früher und 8-mal höher als die rote Kurve der prognostizierten Erkrankungsmeldungen. Die auf Basis dieser Annahmen und Prognosen modellierte Kurve der täglichen Neuinfektionen würde dann wie folgt aussehen:



Die jeweilige Testquote zu einem bestimmten Zeitpunkt errechnet sich dann als Bruch aus den täglichen Infektionsmeldungen des RKI durch die modellierte Gesamtzahl der tatsächlichen täglichen Neuinfektionen:



Während der Kurvenverlauf bis Mitte April einigermaßen der Realität entsprechen dürfte – danach nehmen die Unsicherheiten durch die Prognose der Erkrankungszahlen erheblich zu – ist die tatsächliche Höhe der Testquote ungewiss. So führen schon kleine Änderungen der angenommenen Diagnoserate zu erheblich höheren oder niedrigeren Testquoten. Mit geeigneten Annahmen dürfte dies allerdings ein durchaus probater Ansatz sein, um einen Überblick zu erhalten, wie gut es gelingt, neue Infektionsfälle aufzuspüren.

Nikolai Edinger

06.05.2020

[www.mister-ede.de](http://www.mister-ede.de)