



VON PROF. DR. RALPH BRINKS, DR. BETTINA BERGER, PROF. DR. GERD ANTES

## Schwierige Prognosen

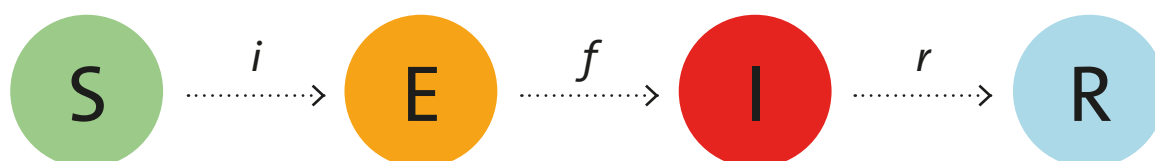
Wie funktionieren Infektionsmodelle, die zur Vorhersage des Corona-Pandemieverlaufs eingesetzt werden? Wo liegen die Fehlerquellen?

Im Laufe der SARS-Cov-2-Pandemie haben Infektionsmodelle eine herausragende Stellung in der öffentlichen Berichterstattung und politischen Entscheidungsfindung eingenommen. Es lohnt sich also, die Grundlagen der Infektionsmodellierung etwas genauer zu betrachten, um zu verstehen wie solche Modelle funktionieren und wo mögliche Fehlerquellen für Vorhersagen des Pandemieverlaufs liegen.

Im Wesentlichen ist ein Infektionsmodell eine strukturierte Denkweise, wie die Ausbreitung einer Pandemie beschrieben und in ihren wesentlichen Aspekten verstanden werden kann. Wie bei allen Modellen in den Wissenschaften sollen wesentliche Eigenschaften des zu modellierenden Sachverhalts betrachtet werden. Ein Infektionsmodell hat jedoch nicht den Zweck, alle Eigenschaften einer Pandemie mit seinen mannigfaltigen Interaktionen zwischen Erregern, Menschen und Umwelt zu beschreiben oder gar zu erklären. Diese Fokussierung eines Modells auf die wesentlichen Teilaspekte anstatt eine ganzheitlich erklärende Perspektive anzunehmen, wird durch den Aphorismus des Statistikers George E. P. Box zusammengefasst, dass grundsätzlich alle Modelle falsch, einige aber nützlich seien.

Die älteste verwandte Technik bei Infektionsmodellen nennt man Compartment-Modelle. Jedes Individuum der betrachteten Bevölkerung (Population) wird danach je nach Expositionsstatus in genau eine von mehreren Klassen (Compartments) eingeteilt, z.B. Suszeptibel oder Genesen. Es ist nicht möglich, dass ein Individuum keiner oder mehreren Klassen zugeteilt wird. Im Zeitverlauf der Epidemie können die Individuen der Population die Klasse beibehalten oder ändern. Ein häufig verwendetes Compartment-Modell ist das SEIR-Modell, das in Abbildung 1 gezeigt ist.

Gemäß den vier Klassen des SEIR-Modells wird angenommen, dass eine mögliche Infektion bei jedem Individuum in vier wesentlichen Phasen abläuft. Jemand der für eine Infektion empfänglich ist, befindet sich in der Klasse ‚suszeptibel‘, wird möglicherweise von einer infektiösen Person (in der Klasse ‚infektiös‘) angesteckt und befindet sich dann für eine gewisse Zeit im Compartment ‚exponiert‘. Während dieser Zeit ist die infizierte Person selbst noch nicht ansteckend, wird aber zwangsläufig nach dem Ablauf der Latenzzeit ansteckend werden und die Klasse von ‚exponiert‘ nach ‚infektiös‘ ändern. Nach der anste-



**Abbildung 1:** SEIR-Modell. Jedes Individuum der zu modellierenden Population wird genau einer der vier Klassen Suszeptibel, Exponiert, Infektiös oder Genesen zugeordnet. Die Anfangsbuchstaben der englischen Entsprechungen Susceptible, Exposed, Infectious, Removed sind namensgebend für das Modell.

ckenden Phase, wo eine infizierte Person eine oder mehrere andere suszeptible Personen anstecken kann, gelangt die infektiöse Person in die Klasse ‚removed‘, die neben den Genesenen auch die Verstorbenen umfasst.

Die Anzahlen der Personen in den Klassen und die Interaktionen zwischen den Klassen werden bei Compartment-Modellen durch mathematische Gleichungen beschrieben, was hier für ein einfaches Beispiel ausgeführt werden soll. Mit Blick auf Abbildung 1 sollen die Anzahlen der Personen in den zugehörigen Kompartments mit den Großbuchstaben S, E, I, und R bezeichnet werden. Wechsel zwischen den Kompartments werden durch Wahrscheinlichkeiten beschrieben. In Abbildung 1 sind die Bezeichnungen dieser Wahrscheinlichkeiten durch Kleinbuchstaben über den Pfeilen gegeben, die die Kompartments verbinden. Andere Wechsel zwischen Klassen als diejenigen, die durch Pfeile gekennzeichnet sind, sind nicht möglich. Im gezeigten SEIR-Modell ist also beispielsweise kein Übergang aus der Klasse ‚genesen‘ zurück in die Klasse ‚suszeptibel‘ möglich. Das impliziert, dass eine Person höchstens einmal infiziert werden kann.

Da bei SARS-CoV-2 täglich Fallzahlen, Zahlen von Infizierten und Genesenen berichtet werden, soll im Folgenden tagesaktuell gerechnet werden. Die Anzahlen S, E, I und R bekommen entsprechend einen Subindex  $k$ , der angibt, auf welchen Tag sich die Angabe bezieht. Die Angabe  $S_{10}$  bedeutet also die Anzahl der Suszeptiblen in der Population am Tag  $k = 10$ . Der Tag  $k = 0$  bezeichnet den Tag, an dem das Virus zum ersten Mal in der Population auftaucht (oder die Zählung der Infizierten beginnt).

Auch die Übergangswahrscheinlichkeiten  $i$ ,  $f$  und  $r$  in Abbildung 1 sind mit dieser Festlegung tagesaktuell zu interpretieren. Die Angabe  $i_{10}$  bedeutet also die Wahrscheinlichkeit für einen Suszeptiblen während des zehnten Tages angesteckt zu werden, also am zehnten Tag aus der Klasse S in die Klasse E zu wechseln.

Will man nun die zeitliche Dynamik einer Pandemie mit dem SEIR-Modell beschreiben, betrachtet man aus Sicht der einzelnen Kompartments, was beim Übergang von einem Tag  $k$  zum nächsten Tag  $k+1$  geschieht. Betrachtet man beispielsweise das Kompartment der Infizierten, so kommen von einem zum nächsten Tag Personen hinzu, die den Zustand ‚exponiert‘ verlassen, und es gehen Personen ab, die genesen oder versterben. Am Tag  $k+1$  gibt es also die infektiösen Personen vom Vortag (Tag  $k$ ), die ihren Zustand nicht verändern (Anzahl  $I_k$ ), es kommen Personen hinzu (Anzahl  $f_k \times E_k$ ) und es gehen Personen ab (Anzahl  $r_k \times I_k$ ). Folglich ergibt sich die Gleichung

$$I_{k+1} = I_k - r_k \cdot I_k + f_k \cdot E_k$$

Mit ähnlichen Überlegungen ergeben sich die Gleichungen für die anderen Kompartments:

$$S_{k+1} = S_k - i_k \cdot S_k \quad (1a)$$

$$E_{k+1} = E_k - f_k \cdot E_k + i_k \cdot S_k \quad (1b)$$

$$I_{k+1} = I_k - r_k \cdot I_k + f_k \cdot E_k \quad (1c)$$

$$R_{k+1} = R_k + r_k \cdot I_k \quad (1d)$$

Im Folgenden gehen wir zur Vereinfachung der Betrachtung von einer selbst-immunisierenden Erkrankung aus, d.h. die Wirkung von etwaigen Impfungen oder Kontaktbeschränkungen wer-



den nicht betrachtet. In diesem Fall ergibt sich die Inzidenz  $i_k$  der Infektion an Tag  $k$  als

$$i_k = R_0 \cdot I_k / (N \cdot d), \quad (2)$$

wobei  $d$  die (mittlere) Dauer der infektiösen Phase und  $N$  die Populationsgröße sein soll. Der Einfachheit halber wird angenommen, dass die Populationsgröße  $N$  konstant bleibt, d.h. es gibt weder Migration in oder aus der betrachteten Population noch sollen Todesfälle auftreten. Will man Todesfälle berücksichtigen, kann das durch einfache Anpassungen der Gleichungen (1a bis 1d) geschehen. Für die erste Betrachtung kann man mit diesen Annahmen aber gut arbeiten. In Gleichung (2) kommt eine neue Größe  $R_0$  hinzu, die Basis-Reproduktionszahl. Die Basis-Reproduktionszahl  $R_0$  beschreibt, wie sich der Erreger in einer Population ausbreitet, die nur aus Suszeptiblen besteht.

Für die Wahrscheinlichkeiten  $f_k$  und  $r_k$  ergeben sich folgende Zusammenhänge

$$f_k = 1/\tau \quad (3)$$

$$r_k = 1/d \quad (4)$$

wobei  $\tau$  die (mittlere) Dauer der Latenzphase und  $d$  wie gehabt die Dauer der infektiösen Phase ist.

Mit dem SEIR-Modell kann man den Verlauf einer selbst-immunisierenden Pandemie sehr gut beschreiben. Als Beispiel sei die Spanische Grippe in den zehn größten Städten der USA genannt [Mills et al 2004].

Wir nehmen nun an, dass die Population, die wir mit dem SEIR-Modell beschreiben wollen, eine Größe von  $N = 200.000$  hat. Die Basis-Reproduktionszahl  $R_0$  soll den Wert 3 haben. Die infektiöse Phase und die Latenzphase haben eine angenommene Dauer von 5 beziehungsweise 2 Tagen. Zusätzlich stellen wir uns vor, dass die Infektion von einer Person in die Population gebracht wird.

Ziel ist es nun, den Verlauf der Pandemie in den ersten 100 Tagen zu beschreiben. Mit diesen Angaben lassen sich die Gleichungen (1) bis (4) einfach in ein Tabellenkalkulationsprogramm übertragen. Wir wählen anstelle der Tabellenkalkulation das kostenlose, quelloffene Statistikprogramm R (The R Foundation for Statistical Computing).

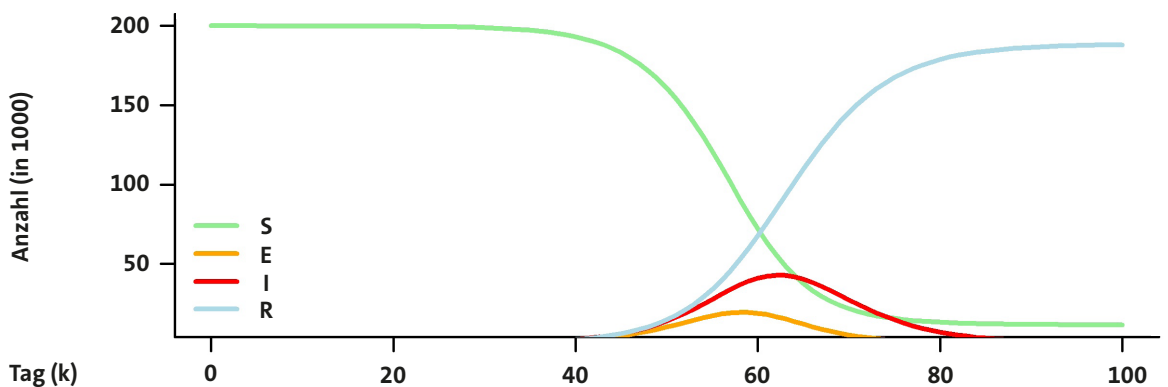
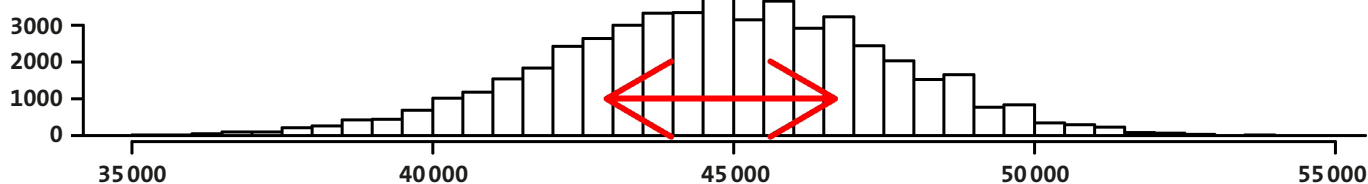


Abbildung 2: Epidemische Kurve. Zeitlicher Verlauf der hypothetischen Epidemie.



**Abbildung 3:** Verteilung der maximalen Anzahl Infektiöser ( $I_{max}$ ). Wenn  $R_0$  gemäß einer Normalverteilung mit Mittelwert 3 und Standardabweichung 0,15 schwankt, ergibt sich die dargestellte Häufigkeitsverteilung von  $I_{max}$ . Der zentrale Bereich, in dem sich 50% der Ausgabewerte befinden, heißt Interquartilsbereich und ist durch einen roten Doppelpfeil gekennzeichnet.

Es ergibt sich der in Abbildung 2 gezeigte Verlauf der Zahlen in den Kompartments S, E, I, R. Das Maximum der Anzahl der Infektiösen ergibt sich an Tag 60 mit einem Wert von  $I_{max} = 44.755$ , was 22% der Populationsgröße beträgt.

Je nach Art der Infektion und deren Symptomen kann ein so großer Anteil an Infektiösen die Gesundheitsversorgung an ihre Grenzen bringen. Während der SARS-CoV-2-Pandemie wurde deswegen zeitweise die Strategie "flatten the curve" ausgerufen, was bedeuten sollte, das Maximum der roten Kurve ( $I_{max}$ ) in Abbildung 2 bei Tag 60 weniger hoch werden zu lassen. Wichtig für unser Beispiel soll die Beobachtung sein, dass die maximale Anzahl  $I_{max}$  an Infektiösen eine wichtige Größe im zeitlichen Verlauf der Pandemie sein kann. Ihre Vorhersage kann damit als ein wichtiges Ziel bei der Modellierung einer Epidemie sein.

Wir hatten gesehen, dass neben den Gleichungen (1) bis (4) auch die Zahlenwerte für die "Parameter"  $R_0$ ,  $\tau$ ,  $d$  und  $N$  wichtig sind. Die Bestimmung der Basis-Reproduktionszahl  $R_0$  eines neuen Erregers ist oft nur mit Unsicherheiten möglich – gerade in der Anfangszeit der Epidemie. Wenn  $R_0$  anstatt des Wertes 3 den geringfügig kleineren Wert 2,9 hat, wird

die maximale Anzahl an Infektiösen erst zwei Tage später am Tag 62 angenommen, wobei die maximale Anzahl dann  $I_{max} = 42.919$  beträgt, mehr als 4% weniger als bei dem ursprünglichen Wert von 3.

Man erkennt also, dass eine Änderung der Basis-Reproduktionszahl um 3% zu einer Verringerung der maximalen Anzahl der Infektiösen von mehr als 4% führt. Die Veränderung im Modellparameter  $R_0$  führt also zu einer Veränderung im Ausgabewert  $I_{max}$  des SEIR-Modells. Die Gesetzmäßigkeit, dass Änderungen in den Parametern (Eingangswerten) zu Änderungen in den Ausgabewerten führen, nennt man Fehlerfortpflanzungsgesetz. Typischerweise hat man die Eingangswerte in Form von Messwerten mit einem Fehlermodell gegeben, z.B. wird  $R_0$  als Mittelwert einer Normalverteilung mit einer vorgegebenen Standardabweichung gegeben. In diesem Fall führt die Verteilung der Eingangswerte zu einer Verteilung der Ausgabewerte. Eine Normalverteilung von  $R_0 = 3$  mit einer Standardabweichung von 5%, also 0,15, führt zu einer Verteilung von  $I_{max}$  wie in Abbildung 3 gezeigt.

In Abbildung 3 ist der zentrale Bereich, in dem sich 50% der Ausgabewerte von  $I_{max}$  befinden, durch einen roten Doppelpfeil gekennzeichnet.



Die Breite dieses Bereichs nennt man Interquartilsabstand (IQA). In unserem Beispiel beträgt der IQA, also die Breite des roten Pfeils 3802.

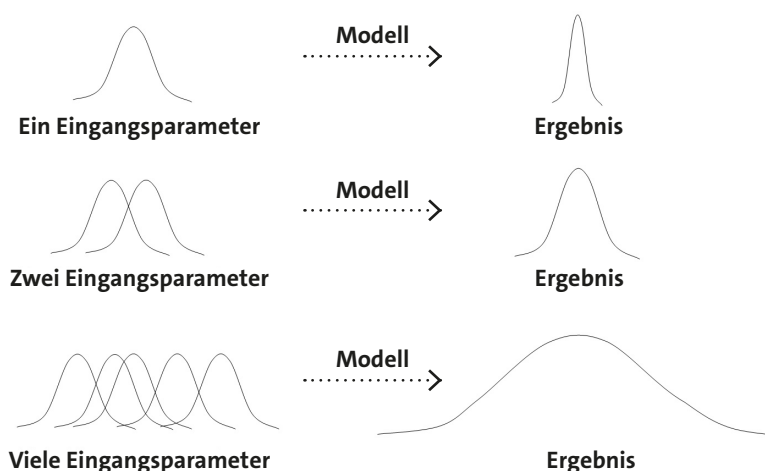
Unser SEIR-Modell hängt jedoch nicht nur von einem Parameter  $R_0$  ab, sondern auch von den Parametern  $\tau$ ,  $d$  und  $N$ . Mit der Anzahl der Parameter akkumulieren sich die Unsicherheiten in den Parametern zu immer breiteren Schwankungen in den Ausgabewerten aus. Das ist schematisch in Abbildung 4 gezeigt.

Nun wollen wir genauer untersuchen, wie sich Unsicherheiten in diesen Parametern zu Schwankungen in den Ausgabewerten  $I_{max}$  des Infektionsmodells gemäß des Fehlerfortpflanzungsgesetzes ergeben. Hierzu untersuchen wir den IQA, wenn wir sukzessive die Parameter  $R_0$ ,  $\tau$ ,  $d$  und  $N$  um 5% schwanken lassen. Das Ergebnis ist in Tabelle 1 gezeigt. Schwankt nur der Parameter  $R_0$  um 5% be-

trägt der Interquartilsabstand 3802, kommt eine Schwankung des Parameters  $\tau$  von 5% hinzu, beträgt der IQA schon 3877. Schwanken schließlich alle vier Parameter um 5%, beträgt der IQA schon 5000. Verglichen mit dem ursprünglichen IQA, wo nur  $R_0$  geschwankt hat, hat der IQA um mehr als 30% zugenommen.

**TABELLE 1:**  
**IQA UND UNSICHERHEITEN IN DEN EINGANGSPARAMETERN**

Parameter mit je 5% Schwankung	Interquartilsabstand (IQA)	Zunahme des IQA (%)
$R_0$	3802	Basis
$R_0, \tau$	3877	2,0%
$R_0, \tau, d$	3976	4,6%
$R_0, \tau, d, N$	5000	31,5%



**Abbildung 4:** Schematische Wirkung von Unsicherheiten in Eingangsparametern. Je mehr unsicherheitsbehaftete Parameter ein Modell hat, desto größer ist die Unsicherheit in den Ergebnissen.

Das Verfahren, das wir hier angewendet haben, nämlich alle Parameter gemäß ihren Verteilungen zu variieren und die Schwankungsbreite der Ergebnisse zu analysieren, nennt man probabilistische Sensitivitätsanalyse (PSA). Sie wird auch bei komplizierteren Modellen benutzt [Oakley & O'Hagan 2004].

Das Beispiel zum SEIR-Modell verdeutlicht die Wirkung des Fehlerfortpflanzungsgesetzes bei Infektionsmodellen mit mehreren Parametern. Unser Beispiel hängt nur von vier Parametern  $R_0$ ,  $\tau$ ,  $d$  und  $N$  ab. Durch Hinzunahme der letzteren drei Parameter hat sich der Interquartilsabstand IQA um mehr als 30% gegenüber dem IQA erhöht, der sich ergäbe, wenn ausschließlich  $R_0$  mit einer Unsicherheit von 5% behaftet wäre. Das Fehlerfortpflanzungsgesetz impliziert, je mehr Parameter

ein Infektionsmodell hat, um so größer ist die Unsicherheit in den erzielten Ergebnissen. Infektionsmodelle aus der Praxis haben Parameter-Anzahlen, die in die Hunderte gehen. So haben Modelle, die verschiedene Altersgruppen betrachten, oftmals Parameter für jede der Altersgruppen. Das Impfmodell des RKI hat mit seinen zwölf Altersklassen [Scholz et al 2021] mehr als 200 Parameter, so dass man von enorm breiten Schwankungsbreiten der Ausgabewerte ausgehen kann. Eine diesbezügliche Anfrage bei einem der Modellierer ergab folgende Antwort: "Wenn wir die Verteilungen aller Eingangsparameter stochastisch berechnen akkumuliert sich die Unsicherheit und wir bekommen KIs oder CIs [gemeint sind Unsicherheitsintervalle] die eigentlich keinerlei Bedeutung mehr haben, da sie einen sehr großen Wertebereich abdecken." Mit anderen Worten: Würde man eine probabilistische Sensitivitätsanalyse berechnen, wären die Ergebnisse des Modells ohne jegliche Aussage.

Weil das Fehlerfortpflanzungsgesetz ein kompliziertes Zusammenspiel der Unsicherheiten in den Eingangsparametern auf die Ergebnisse einer Modellierung haben können, empfehlen Leitlinien zur Modellierung bei der medizinischen Technikfolgenabschätzung probabilistische Sensitivitätsanalysen [Claxton et al 2005].

Die Betrachtungen hier verdeutlichen die Wirkung des Fehlerfortpflanzungsgesetzes bei Infektionsmodellen. Infektionsmodelle können neben dem Aufaddieren der Schwankungsbreiten in den Ausgabewerten auch andere Schwierigkeiten haben: Rechenfehler durch Computer-Arithmetik, sensitive Abhängigkeit von den Anfangswerten (Schmetterlingseffekt). Dass Infektionsmodelle chaoti-

sches Verhalten aufweisen können, wurde im Zusammenhang von nachlassender Immunität bei Keuchhusten von Hethcote vor mehr als 20 Jahren schon gezeigt. Die nachlassende Impfwirkung der Vakzine gegen SARS-CoV-2 könnte eine ähnliche, möglicherweise chaotische Dynamik bei den Infektionswellen hervorrufen. ■

---

**PROF. DR. RER. NAT. RALPH BRINKS**

Lehrstuhl für Medizinische Biometrie und Epidemiologie an der Universität Witten/Herdecke  
Fakultät für Gesundheit/Department für Humanmedizin  
Alfred-Herrhausen-Straße 50 / 58448 Witten

**DR. PHIL. BETTINA BERGER**

Lehrstuhl für Medizintheorie, Integrative und Anthroposophische Medizin  
Universität Witten/Herdecke  
Fakultät für Gesundheit/Department für Humanmedizin  
Alfred-Herrhausen-Straße 50 / 58448 Witten

**PROF. DR. RER. NAT. GERD ANTES**

Cochrane Deutschland  
Universitätsklinikum Freiburg  
Breisacher Straße 153 / 79110 Freiburg

Literaturstellen

Claxton K, Sculpher M, McCabe C, Briggs A, Akehurst R, Buxton M, Brazier J, O'Hagan T (2005) Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. *J Health Eco* 14 (4): 339-47

Mills C, Robins J, Lipsitch M (2004) Transmissibility of 1918 pandemic influenza. *Nature* 432: 904-6

Oakley JE, O'Hagan A (2004) Probabilistic sensitivity analysis of complex models: a Bayesian approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 66: 751-769.

Scholz S, Waize M, Weidemann F, Treskova-Schwarzbach M, Haas L, Harder T, Karch A, Lange B, Kuhlmann A, Jäger V, Wichmann O (2021): Einfluss von Impfungen und Kontaktreduktionen auf die dritte Welle der SARS-CoV-2-Pandemie und perspektivische Rückkehr zu prä-pandemischem Kontaktverhalten, *Epid Bull* 13:3-22