

LongCovid Einzelwahrscheinlichkeiten Kanada

Michael Schlüter*

19. Februar 2024

Inhaltsverzeichnis

1 Studie aus Kanada	1
2 Auswertung	2
2.1 Fragebogen CCAHS-FQ	2
2.2 Prozessmodell	3
2.3 Zusammengefasste Daten	4
2.4 Datenaufbereitung	4
2.5 Variationsanalyse	5
2.6 Ergebnisse	10
3 Fazit	11

Zusammenfassung

Dieses Dokument analysiert eine Studie aus Kanada zu LongCovid in den Jahren 2020 - 2022. Die dort berichtete Zunahme mit der Anzahl an Covid19-Infektionen lässt sich gut durch einen Wahrscheinlichkeitsbaum mit konstanter Chance auf LongCovid je Infektion beschreiben.

1 Studie aus Kanada

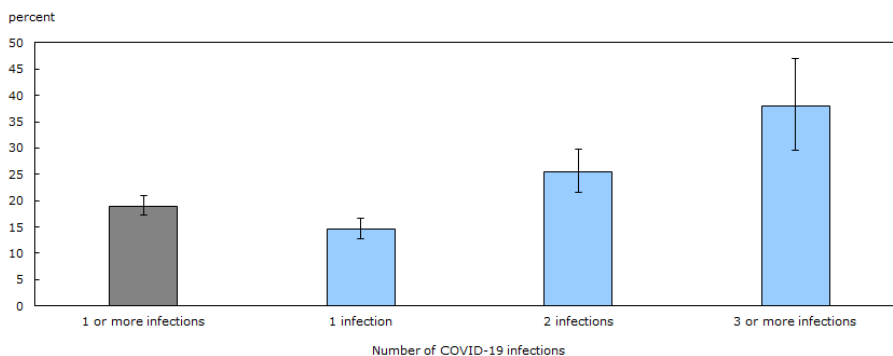
Studie [1] aus Kanada trägt Daten über LongCovid aus den dort genannten Fragebögen zusammen. Die Häufigkeitsverteilung für LongCovid nach 1, 2, 3 oder mehr Infektionen ist interessant, siehe Abbildung 1 auf der nächsten Seite.

Dieser Artikel nimmt eine prozesshafte Sicht ein und schätzt aus jener Verteilung die Einzelwahrscheinlichkeiten, je Covid19-Infektion wieder gesund oder mit LongCovid hervorzugehen.

Einzelwahrscheinlichkeiten erweitern den Blick:

*Michael Schlüter System- und Produktoptimierung, <https://www.ms-spo.de>

Chart 2
Percentage of Canadian adults with long-term symptoms, by number of self-reported COVID-19 infections, June 2023



Source: Statistics Canada, Canadian COVID-19 Antibody and Health Survey - Follow-up Questionnaire, 2023.

Abbildung 1: Langzeit Symptome, abhängig von der Anzahl der Covid19-Infektionen mit 95 % Konfidenzintervall, aus [1]

- als Konsistenzcheck (Wie stimmig ist das?)
- zur Folgenabschätzung (Wie geht es wohl weiter?)

2 Auswertung

2.1 Fragebogen CCAHS-FQ

Für die weitere Auswertung ist es entscheidend, wie die Anzahl der Infektionen aus Abbildung 1 zu verstehen sind. Anhaltspunkte aus dem Fragebogen, [2]:

- INF_Q05: Selbsteinschätzung über die vermutete Anzahl Covid19-Infektionen
- INF_Q10: Wann hatten Sie Ihre *erste* Covid19-Infektion ? (z.B. Anfang März 2021)
- INF_Q15: Wann hatten Sie Ihre *letzte* Covid19-Infektion ? (z.B. Ende July 2022)

Der Fragebogen erlaubt daher beispielsweise bei 3 Infektionen keine Unterscheidung zwischen Szenarien wie:

- 1. Infektion+ LongCovid , 2. Infektion, 3. Infektion
- 1. Infektion, 2. Infektion, 3. Infektion+ LongCovid

sondern liefert die Aussagen:

- Person *hat* LongCovid und *hatte* insgesamt n Infektionen
- oder als kumulierte Häufigkeitsverteilung eben genau $P(n)$ (s.u.)

Ein möglicher Genesungserfolg nach einer LongCovid Episode wird nicht abgefragt.

2.2 Prozessmodell

Seit Beginn der Pandemie 2020 wurden folgende Eigenschaften einer Covid19-Infektion als Gesamtbild bekannt, siehe [3] und insbesondere [4]:

- man kann sich mehrfach anstecken
- einige Menschen entwickeln eine LongCovid Symptomatik
- bisher muss LongCovid für die Mehrheit als unheilbar gelten¹
- bei bereits vorhandenem LongCovid macht eine erneute Covid19-Infektion die Sache nicht besser
- insbesondere sind dem Autor bisher keine heilende Wirkung von Covid19-Infektionen auf LongCovid bekannt
- als Multiorganerkrankung scheint jede Covid19-Infektion den Körper des scheinbar Genesenen in schlechterer Verfassung zu hinterlassen (progressive Vorschädigung)
- die Möglichkeit, nach einer Covid19-Infektion auch zu versterben, wird hier vernachlässigt, denn die Studie erfasst diese Ereignisse nicht

Diese Beobachtungen können im Modell aus Abbildung 2 auf der nächsten Seite vergrößert nachgebildet werden, mit folgenden Regeln:

- je Covid19-Infektion gibt es eine Wahrscheinlichkeit $0 < p < 1$, LongCovid zu entwickeln
- sowie die Gegenwahrscheinlichkeit $q = 1 - p$, wieder zu genesen
- beide Wahrscheinlichkeiten sind für jede Infektion konstant
- wer bereits an LongCovid erkrankte, scheidet aus der Kette aus
- infektionsweise sind die LongCovid Erkrankten fortlaufend aufzusummieren

Abbildung 3 auf der nächsten Seite zeigt dazu ein Rechenbeispiel. Wie also ermittelt man nun den Parameter p für die kumulierte LongCovid Wahrscheinlichkeit $P(n) = 1 - q^n = 1 - (1 - p)^n$ als Funktion der n -ten Ansteckung aus den Daten der kanadischen Studie [1], s.a. Abbildung 1 auf der vorherigen Seite?

¹Es gibt zum Glück immer wieder Ausnahmen mit schrittweiser, oft begrenzter, Rehabilitation. Zu diesem komplexen Krankheitsbild, Linderungs- und Genesungseffekten siehe [5]. Zur Visualisierung des Linderungspotentials bei LongCovid siehe [6].

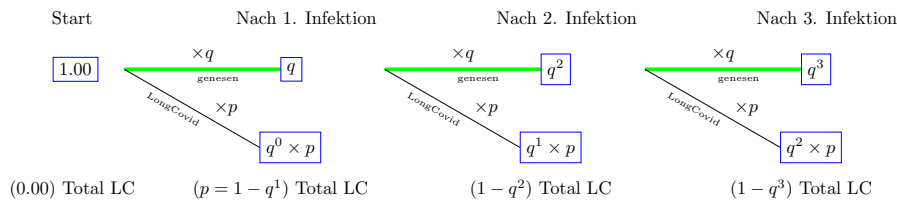


Abbildung 2: Prozessmodell mit konstanten Pfadwahrscheinlichkeiten

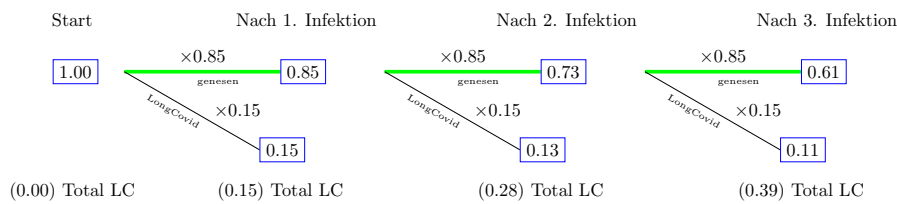


Abbildung 3: Beispiel $p = 0.15 = 15\%$ für LongCovid je Covid19-Infektion

2.3 Zusammengefasste Daten

Die im Folgenden gezeigte Auswertung funktioniert am besten mit Rohdaten: Diese enthalten sogenanntes Rauschen (Noise, [7]), das alle Fehler, alle nicht erfassbaren Einflüsse usw. enthält. Diesen Beitrag zu kennen ist unerlässlich, will man belastbare Schlüsse ziehen.

Leider bietet die Studie nur einige wenige zusammengefasste Daten, so dass viele interessante Schlüsse daraus nicht zu ziehen sind.

2.4 Datenaufbereitung

Rückblickend zeigt Abbildung 1 auf Seite 2 in den blauen Balken die kumulierten Wahrscheinlichkeiten, LongCovid in 1 Infektion zu bekommen, in 1 oder 2 Infektionen, oder in 1, 2 oder mit der 3. Infektion.

Das entspricht jeweils der unteren Total-Reihe in Abbildung 2 oder in Abbildung 3. Ihr Gegenereignis, zu genesen, folgt in der oberen Reihe einem einfacheren Gesetz: $y_G = q^n$.

Durch logarithmieren auf beiden Seiten folgt daraus eine lineare Abhängigkeit. Im Sinne von [7] ist das eine dynamische Charakteristik mit dem Eingangssignal M (n -te Infektion) und der Antwort $Y(M)$, siehe Tabelle 1 auf der nächsten Seite.

Tabelle 2 auf der nächsten Seite zeigt die in diesem Sinne vorbereiteten Daten aus [1]. Da jene Studie auch den Einfluss des Infektionsjahres betrachtet, wird es hier als Indikator A mitgeführt. Vereinfachend erhält der letzte Balken in Abbildung 1 auf Seite 2 den Wert $M = 3$. Abbildung 4 auf Seite 6 zeigt diese Verhältnisse als Parameterdiagramm nach [7]:

- das n -te Infektions-Signal ruft eine Erkrankung hervor, mit der Reaktion

Ereignis	Wahrscheinlichkeit	Ideal	Kommentar
LongCovid	p	konstant	
Genesene	$q = 1 - p$	$y_G = q^n$	n -te Infektion
<i>logarithmiert</i>		$Y(n) = \lg(q) \times n$	Steigung: $\beta := \lg(q)$
<i>Dynamisch</i>		$Y(M) = \beta \times M$	Signal: $M := n$

Tabelle 1: Analyseansatz: lineare Idealisierung über Genesene

der Genesung

- diese Reaktion ist nach der Modellerwartung idealisiert linear $Y = \beta \times M$ (Tabelle 1)
- das Jahr (Indikator A) könnte helfen, diese Reaktionen zu unterscheiden
- Unwägbarkeiten (Noise) stören die Reaktion
- deren Effekt quantifiziert später die Variationsanalyse

i	n. Infektion Signal M	LongCovid <i>gemittelt</i>	Genesene	logarithmiert Reaktion Y	Jahr <i>Indikation</i> A
1	1	0.142	0.858	-0.0665	A1 2020
2	1	0.178	0.822	-0.0851	A2 2021
3	1	0.133	0.687	-0.0620	A3 2022
4	2	0.341	0.659	-0.1811	A1 2020
5	2	0.188	0.812	-0.0904	A2 2021
6	2	0.258	0.742	-0.1296	A3 2022
7	3	0.453	0.547	-0.2620	A1 2020
8	3	0.331	0.669	-0.1746	A2 2021
9	3	0.339	0.661	-0.1798	A3 2022

Tabelle 2: Auszug aus Table 1 in [1], vorbereitet für die Analyse

Abbildung 5 auf der nächsten Seite zeigt die Daten aus Tabelle 2. Eine hohe Streuung, also ein deutlicher Durchgriff aller Unwägbarkeiten (Noise), ist klar zu sehen.

2.5 Variationsanalyse

Die Variationsanalyse nach [7] meint Folgendes:

- nach Abbildung 5 auf der nächsten Seite erhalten wir Messwerte Y_1, \dots, Y_n mit der (totalen) Variation $S_t = \sum_{i=1}^n Y_i^2$
- die unterstellte lineare Abhängigkeit wird darin einen großen Variationsanteil S_β haben (Abbildung 5 auf der nächsten Seite)

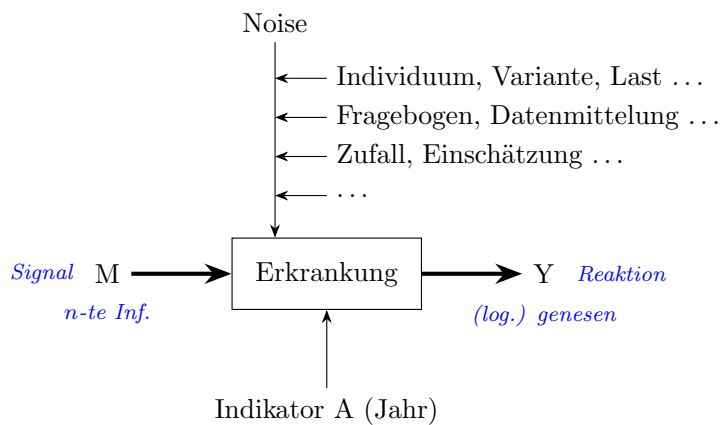


Abbildung 4: Parameterdiagramm zu Tabelle 2 auf der vorherigen Seite

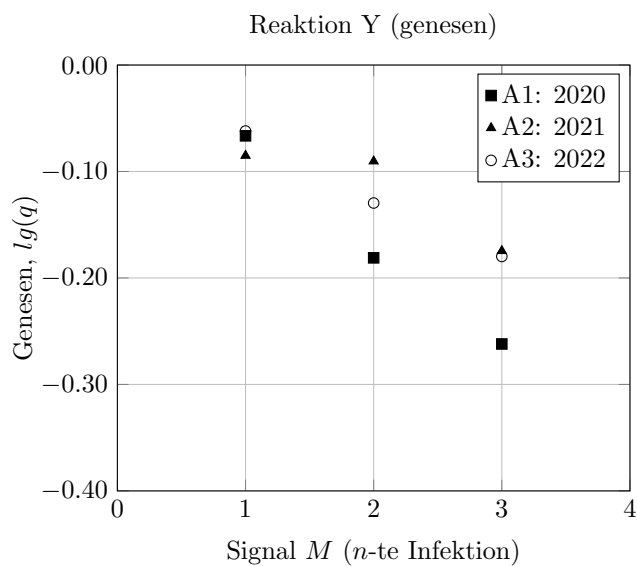


Abbildung 5: Aufbereitete Daten

- Unterschiede durch den Indikator A könnten einen bedeutsamen Anteil S_A an der Variation S_t haben
- der verbleibende Variationsanteil $S_e = S_t - S_\beta - S_A$ wird durch Restfehler verursacht (Noise in Abbildung 4 auf der vorherigen Seite)
- daraus lassen sich u.a. quantifizieren: β_{pand} , β_{2020} , β_{2021} , β_{2022} , σ_e

Anzahl	n		9	
Lineare Summe	$L = \sum_{i=1}^n M_i Y_i$		-2.8651	
Signalvariation	$r = \sum_{i=1}^n M_i^2$		42	
				f Anteil
Totale Variation	$S_t = \sum_{i=1}^n Y_i^2$	0.2047	9	
Variation durch β	$S_\beta = L^2/r$	0.1954	1	95.5 %
Variation durch A (Jahr)	S_A , s.u.	0.0074	2	3.6 %
Restvariation	$S_e = S_t - S_\beta - S_A$	0.0019	6	0.9 %
Restvarianz	$V_e = S_e/f_e$	0.0003		<i>vorläufig</i>
Restfehler für Y	$\sigma_e = \sqrt{V_e}$	0.0178		<i>vorläufig</i>

Die durch den Indikator A (Jahr) verursachte Variation könnte mit 3.6 % bedeutsam sein. Um das zu prüfen, hier die Variationen und einzelnen Steigungen nach [7]:

Jahr:	2020	2021	2022
$L_j = L(A_j)$	$L_1 + L_4 + L_7$ -1.2148	$L_2 + L_5 + L_8$ -0.7897	$L_3 + L_6 + L_9$ -0.8606
$r_j = r(A_j)$	$1^2 + 2^2 + 3^2$	14	14
$S(A_j) = L_j^2/r_j$	0.1054	0.0445	0.0529
$S_A = \sum_{j=1}^3 S(A_j) - S_\beta$	0.0074		
$\beta_j = L_j/r_j$	-0.0868 β_{2020}	-0.0564 β_{2021}	-0.0615 β_{2022}

Abbildung 6 auf der nächsten Seite zeigt:

- im Rahmen des gesamten Restfehlers σ_e in Y weitestgehende Übereinstimmung mit den dazugehörigen Ausgleichsgraden
- damit weitgehende Übereinstimmung für die Jahre 2021 und 2022
- damit einen klaren Unterschied zu 2020
- und insgesamt eine gute Modellanpassung in $q = 10^\beta$

Die Studie [1] benennt diesen Unterschied auch und bringt ihn mit der Covid19 -Impfung in Verbindung. Dazu gibt es zwei systematische Kritikpunkte:

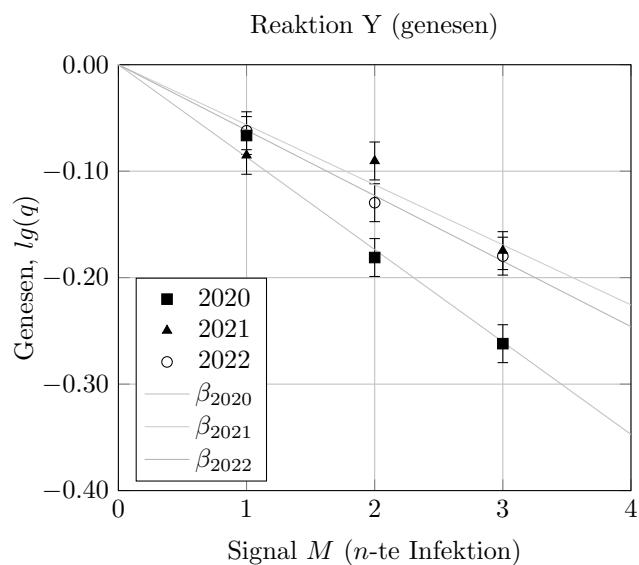


Abbildung 6: Restfehler σ_e in Y mit den drei Ausgleichsgraden

1. Die indicative Einteilung nach Jahren wirkt willkürlich. Nach dem Abtasttheorem [8] wäre die halbe LongCovid Spanne, also 3 Monate, angemessener gewesen, um den zeitlichen Verlauf rekonstruieren zu können.
2. Ein Fallenlassen des Indikators A (Jahr) unterstützt einen etwas pessimistischeren Ansatz der Art: „Dieses Virus barg und birgt viele Überraschungen. Wir sollten eher den mittleren Effekt dieser 3 Jahre betrachten.“

Damit modifiziert sich die obige Betrachtung etwas, visualisiert in Abbildung 7 auf der nächsten Seite sowie Abbildung 8 auf der nächsten Seite:

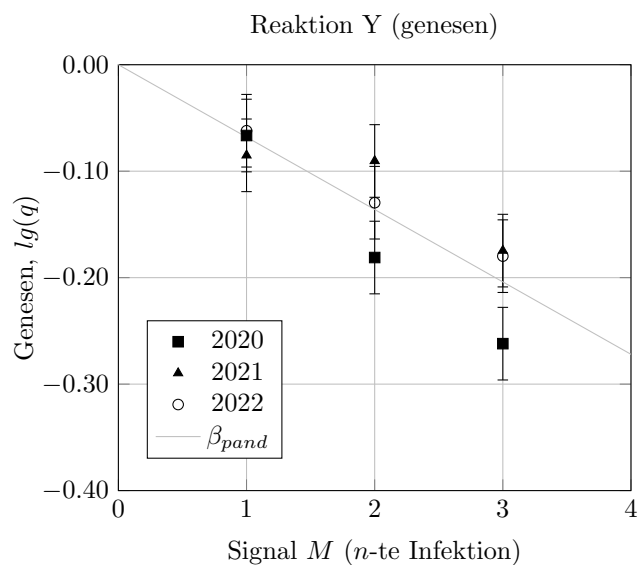


Abbildung 7: Pessimistische Restfehler σ_e in Y und Ausgleichsgerade für die Pandemiejahre 2020 - 2022

			f	Anteil
Totale Variation	$S_t = \sum_{i=1}^n Y_i^2$	0.2047	9	
Variation durch β	$S_\beta = L^2/r$	0.1954	1	95.5 %
Restvariation	$S_e = S_t - S_\beta$	0.0093	8	4.5 %
Restvarianz	$V_e = S_e/f_e$	0.0012		
Restfehler für Y	$\sigma_e = \sqrt{V_e}$	0.0341		<i>pessimistisch</i>
Bereinigte β Variation	$S_\beta^* = S_\beta - V_e$	0.1943		
Damit	$\beta^{*2} = S_\beta^*/r$	0.0046		
Bereinigte mittlere Steigung	$\beta^* = \text{sign}(L) \times \sqrt{\beta^{*2}}$	-0.0680		<i>pessimistisch</i>

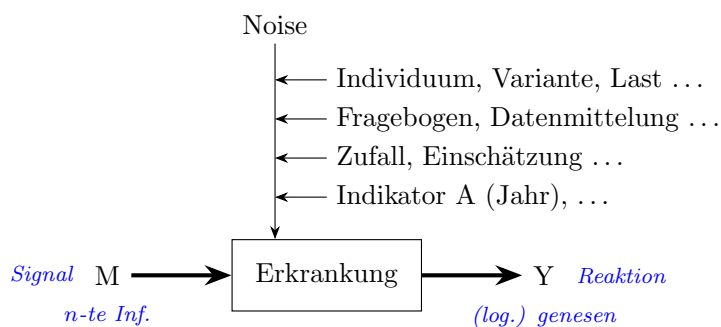


Abbildung 8: Parameterdiagramm für die pessimistischere Betrachtung

2.6 Ergebnisse

Tabelle 6 stellt die Ergebnisse der Variationsanalyse, ihre Umrechnung in die Pfadwahrscheinlichkeiten q und p sowie deren Streuung zusammen.

	Pessimistisch	2020	2021	2022
β	-0.0680	-0.0868	-0.0564	-0.0615
σ_e	± 0.0341	\rightarrow	± 0.0178	\leftarrow
$q = 10^\beta$	0.855	0.819	0.878	0.868
($+\sigma_e$)	0.925	0.853	0.915	0.904
($-\sigma_e$)	0.790	0.786	0.843	0.833
$p = 1 - q$	0.145	0.181	0.122	0.132
($+\sigma_e$)	0.075	0.147	0.085	0.096
($-\sigma_e$)	0.210	0.214	0.157	0.167

Tabelle 6: Pfadwahrscheinlichkeiten je Ansteckung zu Abbildung 2 auf Seite 4

Abbildung 9 auf der nächsten Seite zeigt die Konsequenzen, die sich aus wiederholter Covid19-Infektion für LongCovid nach dem Modell aus Abbildung 2 auf Seite 4 mit den aus der Studie [1] ermittelten Werten Tabelle 6 ergeben:

- $P(n) = (1 - (1 - p)^n)$ ist die kumulative Wahrscheinlichkeit für LongCovid
- Beispiel: Mit der 5. Covid19-Infektion hätten sich 2021 oder 2022 insgesamt ca. 50 % der Menschen LongCovid entwickelt (und behalten es bei)
- die Hälfte der Population ist ca. mit der 3. - 6. Covid19-Infektion von LongCovid betroffen (LD_{50}), am wahrscheinlichsten mit der 4.

Als Populationen kommen beliebige Gruppen in Frage, etwa die gesamte Bevölkerung, eine Bürogemeinschaft, eine Klassenstufe in einer Schule usw. Populationen mit unterschiedlich vielen Covid19-Infektion sollten für Vergleiche nach Anzahl der erfolgten Infektionen aufgeteilt werden.

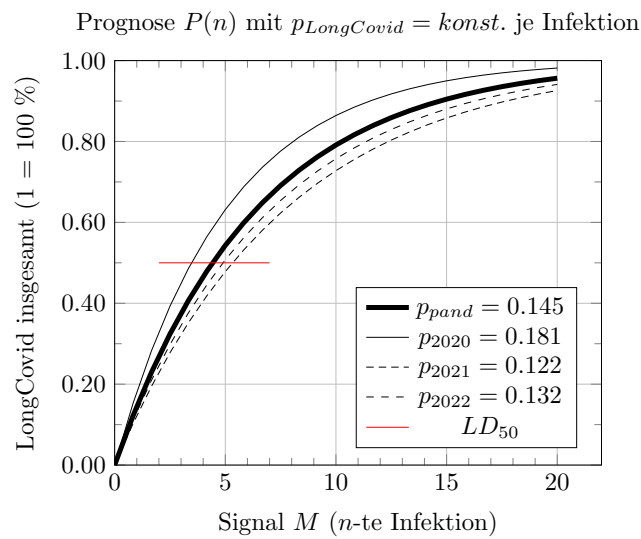


Abbildung 9: Modellprognose für eine Population

3 Fazit

Für die Studie aus Kanada [1] gilt:

- die beobachtete LongCovid Verteilung aus Abbildung 1 auf Seite 2 lässt sich gut mit einem Prozessmodell beschreiben
- die konstante Wahrscheinlichkeit für LongCovid je Covid19-Infektion ist recht hoch: $p_{LongCovid} \approx 15\%$
- damit ist zu erwarten, dass die Hälfte der Bevölkerung ca. mit der 4. Covid19-Infektion dauerhaft LongCovid entwickelt haben wird

Literatur

- [1] Kuang, Sianne, Steven Earl, Janine Clark, Dianne Zakaria, Alain Demers and Samina Aziz. 2023. “Experiences of Canadians with long-term symptoms following COVID-19”. Insights on Canadian Society. December. Statistics Canada Catalogue no. 75-006-X. URL: <https://www150.statcan.gc.ca/n1/pub/75-006-x/2023001/article/00015-eng.htm>
- [2] Canadian COVID-19 Antibody and Health Survey - Follow-up Questionnaire, May 2023, URL: https://www23.statcan.gc.ca/imdb/p3Instr.pl?Function=assembleInstr&lang=en&Item_Id=1505693
- [3] “Wissenschaftliche Literaturempfehlungen für Forschende und Behandelnde”, Fortlaufend aktualisierte Auswahl an Studien, Übersichtsarbeiten etc., BMG Initiative, 2024, URL: <https://www.bmg-longcovid.de/service/wissenschaftliche-studien-und-publikationen>
- [4] Ralf Wittenbrink, Kurzzusammenfassungen neu erscheinender Studien, X ehemals twitter, URL: <https://twitter.com/RWittenbrink>
- [5] Gez Meding & Professor Danny Altmann, “The Long Covid Handbook”, Penguin, 2022
- [6] Suprion Verlag, Zusammenfassung zu Linderung, Informationsdefiziten und Versorgungsforschung bei LongCovid aus der Pressekonferenz des BGM vom 12.07.2023, X ehemals twitter, 2023, URL: https://twitter.com/suprion_verlag/status/1681917751705600000
- [7] Genichi Taguchi, Subir Chowdhury, Yui Wu, “Taguchi’s Quality Engineering Handbook”, WILEY, 2005
- [8] “Nyquist-Shannon-Abtasttheorem”, Wikipedia, URL: <https://de.wikipedia.org/wiki/Nyquist-Shannon-Abtasttheorem>