

# Statistische Prognosemethoden – Top oder Flop?

Christel Weiß

Medizinische Fakultät Mannheim, Universität Heidelberg

**Zusammenfassung** Das Ziel von statistischen Prognosetechniken besteht im Allgemeinen darin, basierend auf empirischen Daten zukünftige Entwicklungen vorherzusagen oder eine Auftrittswahrscheinlichkeit zu ermitteln für ein Ereignis, das in der Zukunft eintreten wird. Wichtige zukunftssträchtige Entscheidungen und Planungen – sei es in der Politik, in der Medizin oder in der Wirtschaft – stützen sich meist auf Prognosen. So kann beispielsweise in der Medizin aufgrund einer individuellen klinischen Diagnose, genetischer Informationen und anderer patientenspezifischer Daten eine Entscheidung darüber getroffen werden, welche Therapie den größten Nutzen verspricht; Epidemiologen entwickeln mathematische Modelle, um künftige Entwicklungen einer Pandemie unter diversen Modellszenarien vorherzusagen; Meteorologen erstellen mittels hochkomplexer Datenanalysen Vorhersagen und warnen, falls erforderlich, vor extremen Wetterereignissen. Prognosen können demnach sowohl für einzelne Individuen als auch für eine gesamte Population überlebenswichtig sein. Statistische Methoden erlauben es, basierend auf umfangreichem Datenmaterial unter Berücksichtigung diverser Einflussfaktoren Wahrscheinlichkeiten für künftige Ereignisse zu quantifizieren oder den wahrscheinlichsten Verlauf einer Entwicklung vorherzusagen. Irrtümer und Überraschungen sind dabei jedoch keineswegs ausgeschlossen. In diesem Beitrag wird anhand von Beispielen erläutert, welche statistischen Verfahren für Prognosen zur Verfügung stehen, welche Anforderungen dabei zu beachten sind und wie die Ergebnisse eines prognostischen Modells zu interpretieren sind. Schließlich wird der Frage nachgegangen, wie es zu Fehleinschätzungen kommt und welche Konsequenzen daraus zu ziehen sind.

## 1 Einleitung

„Prognosen sind schwierig, besonders wenn sie die Zukunft betreffen.“ Dieses humorige Bonmot, das dem dänischen Physiker Niels Bohr (1885–1962) zugeschrieben wird, verleitet zum Schmunzeln. Er suggeriert, dass es auch Prognosen geben könnte, die nicht die Zukunft betreffen und die deshalb weniger schwierig seien. Das ist freilich – und darin liegt der Wortwitz – ein Widerspruch in sich.

Prognosen beziehen sich immer auf die Zukunft, und sie sind deshalb immer riskant. Wie sollte es anders sein? Prognose bedeutet Voraus-Wissen. Dieses gründet sich auf bereits erworbenes Wissen und auf empirische Erfahrungen, die in der Vergangenheit gesammelt wurden. Allerdings ist es kein Naturgesetz, dass Entwicklungen kontinuierlich ohne große Sprünge absehbar verlaufen; Überraschungen sind keineswegs ausgeschlossen. Da niemand von sich behaupten kann, mit hellseherischen Fähigkeiten gesegnet zu sein, und da auch kein Hochleistungsrechner und kein komplexes statistisches Modell Abweichungen von bislang üblichen Verläufen einzukalkulieren vermag, ist es nicht verwunderlich, dass hin und wieder Vorhersagen getroffen werden, die sich später als Fehlprognosen erweisen.

Welche Konsequenzen sind daraus zu ziehen? Ein retrospektiver Blick in die Geschichte zeigt, dass Menschen zu allen Zeiten und in allen Kulturen den Wunsch verspürten, einen Blick in die Zukunft zu werfen und sich bei wichtigen Entscheidungen an Prognosen zu orientieren – sei es in der großen Politik oder im persönlichen Umfeld. Daran hat sich prinzipiell bis heute nichts geändert. Insofern wäre es unangemessen, eine Prognose generell als nutzlos zu erachten.

Allerdings haben sich die Methoden zum Erstellen einer Prognose im Laufe der Zeit grundlegend gewandelt: Während Menschen in früheren Zeiten Orakeln oder sonstigen obskuren Medien vertrauten, stehen heute formalisierte Techniken zum Erstellen einer Prognose zur Verfügung. Diese ermöglichen, basierend auf umfangreichem Datenmaterial mittels komplexer Analysemethoden in nachvollziehbarer Weise Vorhersagen über ein künftiges Ereignis zu treffen. Zwar können auch diese Methoden nicht garantieren, dass eine Prognose in jedem Fall korrekt ist. Allerdings sind die zugrunde liegenden Annahmen überprüfbar, und deren Entstehungsprozess ist nachvollziehbar. Fehlprognosen bedeuten nicht zwangsläufig, dass die Annahmen unrealistisch oder die verwendeten Methoden ungeeignet waren. Eine andere Ursache könnte beispielsweise dadurch gegeben sein, dass Menschen angesichts einer düsteren Prognose ihr Verhalten ändern und so aktiv dazu beitragen, dass sich die Prognose nicht erfüllt. Insofern leisten auch Fehlprognosen einen wichtigen Beitrag zur Gestaltung der Zukunft.

## 2 Das Erstellen von Prognosen

### 2.1 Historische Betrachtungen

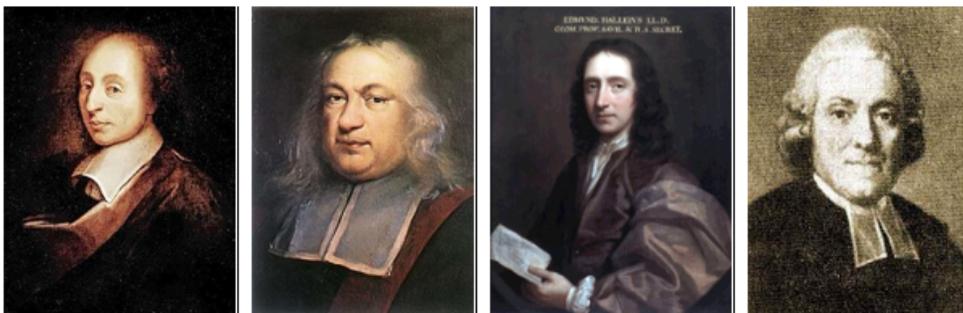
Seit jeher waren Menschen beseelt von dem Wunsch, die Zukunft vorhersagen zu können – sei es aus purer Neugierde oder um sich gegen drohendes Unheil zu wappnen. Vielfältige Methoden wurden entwickelt, um plausible Zukunftsszenarien zu entwerfen [4]. Im klassischen Altertum vertraute man auf Orakel und Propheten. Bekannte Beispiele sind das Orakel des Ammon im alten Ägypten, das Orakel von Delphi im antiken Griechenland oder biblische Propheten. Diese genossen hohes gesellschaftliches Ansehen; bei wichtigen politischen Entscheidungen vertraute man ihrem Rat und glaubte ihren Prophezeiungen. Im Orient und im alten Rom war es üblich, die Zukunft anhand der Eingeweide von Opfertieren zu deuten. Mit dem Aufkommen des Christentums verloren diese Methoden ihre Bedeutung. Man glaubte, dass alles Kommende allein in Gottes Hand läge und dass es Menschen nicht zustünde, einen Blick in die Zukunft zu werfen.

Auch Wissenschaftler waren bis weit ins 18. Jahrhundert hinein sehr autoritätsgläubig. Dabei galt das sichere Wissen, das seit den Zeiten des griechischen Universalgelehrten **Aristoteles** (384–322 v. Chr.) Leitvorstellung des Forschens gewesen war, als hehres Ideal. Wissenschaftler waren vor allem bemüht, den Status quo zu ergründen und Gesetze zu formulieren, denen das Weltgeschehen unterworfen war. Sie lebten und forschten für die Gegenwart; bei Entscheidungsfindungen orientierten sie sich an den Lehrmeinungen anerkannter Autoritäten, ohne diese kritisch zu hinterfragen. Prognosen hatten bei dieser Denkart keinen Raum.

Dieses autoritätsgläubige Denken machten sich auch Ärzte zu eigen: Sie beriefen sich bei ihren diagnostischen Befunden oder therapeutischen Verfahren auf alte Lehrmeister wie **Hippokrates von Kos** (460–370 v. Chr.) oder **Galen von Pergamon** (128–131 bis 199–216 n. Chr.). Zweifel waren bei diesem dogmatischen Ansatz quasi ausgeschlossen. Wenn der Patient nach einer Therapie gesundete, sah man darin eine Bestätigung der Theorie. Ansonsten suchte und fand man Argumente, um Misserfolge zu negieren. Prognosen waren überflüssig, da man a priori von der Richtigkeit einer Diagnose oder die Wirksamkeit einer therapeutischen Maßnahme überzeugt war.

Trotz dieser vorherrschenden Grundeinstellung war der Glaube nicht auszurotten, dass es irgendwie möglich wäre, die Zukunft zu erkunden. In Europa bediente man sich dazu seit dem Mittelalter der Astrologie, die bis ins 18. Jahrhundert hinein weit verbreitet war (und sich als Pseudowissenschaft bis heute gehalten hat). Deren Anhänger glaubten, anhand von Gestirnkongstellationen am Himmel sowohl individuelle Schicksale als auch globale Ereignisse wie Kriege, Natur- oder Hungerkatastrophen vorhersagen zu können. Freilich fehlte der Nachweis, dass diese Prophezeiungen kritischen Prüfungen standhalten würden.

Seltsamerweise bot das Glücksspiel Anlass dafür, dass Prognosen Einzug in die Wissenschaft hielten [8]. Die französischen Mathematiker **Blaise Pascal** (1623–1662) und **Pierre de Fermat** (1607–1665) befassten sich in ihrem berühmt gewordenen Briefwechsel aus dem Jahr 1654 mit Gewinnchancen bei Würfelspielen. Freilich konnten sie nicht sicher prognostizieren, ob ein Spiel gewonnen oder verloren werden würde. Sie waren jedoch in der Lage, ein Spiel formal darzustellen, den zu erwartenden Ausgang durch einen sogenannten Erwartungswert zu quantifizieren und so die Gewinnchancen zu maximieren. Dieser Ansatz zeigte, dass Gewinnen oder Verlieren keine reine Glückssache ist, sondern teilweise beeinflussbar ist. Dieser Gedanke übertrug sich auf andere Bereiche. Der englische Mathematiker **Edmond Halley** (1656–1742) schätzte basierend auf Geburts- und Sterbedaten Lebenserwartungen und Wahrscheinlichkeiten, ein bestimmtes Alter zu erreichen. Dem preußische Feldprediger **Johann Peter Süßmilch** (1707–1767) gelang es, mit Hilfe von Geburts- und Sterbeeintragungen in Kirchenbüchern Bevölkerungsentwicklungen vorherzusagen. Versicherungen konnten Risiken ermitteln und passende Policen anbieten. Auch in der Medizin setzte sich allmählich die Erkenntnis durch, dass systematische Datenerhebungen und -analysen zum Fortschritt beitragen können und dass man Risiken und Krankheiten nicht machtlos ausgeliefert ist. So konnte der französische Arzt **Pierre Alexandre Louis** (1787–1872) nachweisen, dass der Aderlass nutzlos oder gar schädlich war. Was Wissenschaftler bis dahin für undenkbar gehalten hatten, erwies sich nun als evident: Die Zukunft war nicht mehr komplett im Dunkeln verborgen, sondern schien berechenbar zu sein. Der Mathematiker **Pierre-Simon Laplace** (1749–1827) mutmaßte gar, dass sie eines Tages vollständig vorhersehbar wäre, wenn hinreichend viele Informationen gepaart mit hoher Intelligenz zur Verfügung stünden.



**Abbildung 1:** Pioniere der Prognostik. 1a. Blaise Pascal, 1b. Pierre de Fermat, 1c. Edmond Halley, 1d. Johann Peter Süßmilch

Das Fortschreiten der Statistik sowie das Aufkommen leistungsfähiger Computer und benutzerfreundlicher Software im 20. Jahrhundert haben maßgeblich dazu beigetragen,

dass Prognosen mit formalisierten Methoden in den unterschiedlichsten Bereichen erstellt und als Entscheidungshilfen herangezogen werden: bei der Wettervorhersage, in der Klimaforschung, im Versicherungsgewerbe, im Finanzwesen, in der Unternehmensbranche, der Medizin, der Demoskopie und der Wahlforschung.

## 2.2 Anforderungen an eine Prognose

Bei einer Prognose sind zwei Seiten involviert: Der Produzent, der sie erstellt, und der Konsument, der darauf basierend eine Entscheidung trifft. Die Güte einer Prognose muss aus Sicht beider Seiten beurteilt werden.

Die Basis einer empirisch gehaltvollen Prognose bilden überprüfbare Daten oder Fakten sowie begründbares Erfahrungswissen und eine effiziente Analyseverfahren. Das Erstellen einer Prognose ist mit einem Risiko verbunden: Sie kann sich als zuverlässig oder als falsch erweisen. Umso wichtiger ist es, dass dieser Prozess nachvollziehbar ist. Die Güte einer Prognose lässt anhand folgender Kriterien beurteilen [2,3]:

1. **Nicht-Trivialität:** Eine Prognose darf keine allgemein gültige Aussage darstellen. Eine triviale Prognose ist zwar mit keinerlei Risiko für deren Produzenten verbunden, aber in keiner Weise hilfreich für den Konsumenten. So ist beispielsweise die alte Bauernregel „Wenn der Hahn kräht auf dem Mist, ändert sich das Wetter oder es bleibt, wie es ist“ allenfalls als humoristische Belanglosigkeit, keinesfalls jedoch als ernstzunehmende Wetterprognose anzusehen.
2. **Objektivität:** Die Prognose sollte nachvollziehbar und unabhängig von der Person sein, die sie erstellt. Das Kriterium der Objektivität verhindert, dass allein basierend auf Intuition oder unrealistischen Annahmen unplausible Prognosen gestellt werden. Dazu sind die Annahmen, denen die Prognose unterliegt, sowie die verwendeten Verfahren zu spezifizieren. Außerdem sollte dargelegt werden, aus welchen Gründen und unter welchen Voraussetzungen die Annahmen sinnvoll erscheinen.
3. **Validität:** Die Prognose sollte eine Aussage über den Sachverhalt beinhalten, der im Fokus des Interesses steht. Dieses Kriterium ist nicht unbedingt leicht umsetzbar. Nicht alles, was prognostiziert werden soll, ist auf einfache Weise quantitativ messbar oder in absehbarer Zeit erfassbar. In solchen Fällen ist es naheliegend, einen leichter zu erhebenden Surrogatparameter zu verwenden. Dabei sollte sichergestellt sein, dass das Ersatzmerkmal inhaltlich annähernd dasselbe wie der zu prognostizierende Parameter wiedergibt und eng mit diesem assoziiert ist.
4. **Überprüfbarkeit:** Letztlich sollte überprüfbar sein, ob das prognostizierte Ereignis eingetreten ist oder nicht. Dies ist leicht möglich, wenn ein konkretes Ereignis in naher Zukunft vorhergesagt wird. Dagegen sind Aussagen, die einen weit in

der Zukunft liegenden Zeitraum betreffen, oder vage formulierte Prognosen diesbezüglich problematisch.

Jedes dieser Kriterien scheint sinnvoll zu sein und sollte sowohl vom Produzenten bei der Erstellung als auch vom Konsumenten im Entscheidungsprozess bedacht werden. In der Praxis stößt man dabei jedoch hin und wieder auf Schwierigkeiten. Dies sei an einigen Beispielen erläutert.

Für wissenschaftliche Prognosen ist das erste Kriterium in aller Regel erfüllt: Kein Forscher würde sich die Blöße geben, eine Trivialität als Prognose zu formulieren. Dagegen sind mehrdeutige Aussagen in Horoskopen wie etwa „In der nächsten Woche haben Sie ein Erfolgserlebnis“ der Kategorie „trivial“ zuzuordnen.

Für einen Wissenschaftler ist es außerdem selbstverständlich, die verwendete Methodik detailliert darzulegen. Nur so kann der Konsument beurteilen, inwieweit die zugrunde liegenden Annahmen zutreffen und ob sie auch für seine spezielle Situation gelten. Allzu banale Annahmen (z. B. „Das sagt mir mein Bauchgefühl“) sind nicht überprüfbar; darauf basierende Prognosen sind wertlos.

Das Kriterium der Validität erfordert, dass das zu prognostizierende Ereignis präzise definiert ist, und dass eine geeignete Erfassungs- oder Messmethode zur Verfügung steht. Das ist bei weitem nicht immer der Fall. Beispiele finden sich bei klinischen Studien: Um zu prognostizieren, ob ein neu entwickeltes Medikament geeignet ist, einem (möglicherweise weit in der Zukunft liegenden) Schlaganfall vorzubeugen, liegt es nahe, statt die Inzidenz von Schlaganfällen die Senkung des Blutdrucks zu evaluieren (der schnell und einfach zu messen ist). Allerdings muss dabei bedacht werden, dass eine Senkung des Blutdrucks nicht automatisch mit einer Verringerung des Schlaganfall-Risikos einhergeht.

Auch die Überprüfbarkeit der Korrektheit einer Prognose ist mitunter schwieriger als es den Anschein hat. Am ehesten überprüfbar sind kurzfristige Prognosen, seien sie sehr allgemein („Morgen wird es regnen“) oder sehr präzise formuliert („Um 11 Uhr wird eine Temperatur von 10 Grad Celsius verbunden mit Schauern erwartet“). Schwieriger sind dagegen Prognosen zu beurteilen, die eine Wahrscheinlichkeit beinhalten, wie beispielsweise „Morgen wird es mit 90 %-iger Wahrscheinlichkeit regnen“. Man wird bei einer solchen Ankündigung nicht ohne Regenschirm das Haus verlassen und geneigt sein, die Prognose als bestätigt anzusehen, sobald der erste Tropfen fällt und sie bei ausbleibendem Regen zu bezweifeln. So einfach ist es jedoch nicht: Um die Richtigkeit der Prognose zu überprüfen, müssten alle Tage, für die mit 90 %-iger Wahrscheinlichkeit Regen prognostiziert wurde, herangezogen und der Anteil der Regentage ermittelt werden. Außerdem gilt: Je langfristiger eine Prognose angelegt ist, desto schwieriger ist sie zu überprüfen. Dies trifft beispielsweise zu für Prognosen bezüglich der Klimaänderung (auch wenn sie auf soliden Annahmen basieren und valide sind).

Gänzlich unseriös erscheinen apokalyptische Prophezeiungen wie „Die Menschheit wird aussterben“, da diese in aller Regel auf unbegründeten Annahmen basieren und somit nicht objektiv sind und da deren Wahrheitsgehalt niemals überprüfbar sein wird.

## 2.3 Methoden zum Erstellen einer Prognose

Formalisierte Prognosemethoden lassen danach unterscheiden, ob sie eher qualitativ oder eher quantitativ orientiert sind.

### 2.3.1 Qualitative Methoden

Zu den qualitativen Verfahren zählen die Delphi-Methode und die Szenariotechnik.

**Delphi-Methode:** Damit werden mehrere Experten in aufeinander folgenden Phasen bezüglich eines zukünftigen Szenarios befragt mit dem Ziel, einen Konsens zu finden. Nach jeder Fragerunde wird jeder Beteiligte über das Ergebnis informiert und hat die Gelegenheit, seine Einschätzung zu revidieren. Diesem Verfahren liegt der Gedanke zugrunde, dass mehrere Fachleute eine bessere Prognose abgeben als ein einzelner Experte. Die Delphi-Methode bietet sich an, wenn keine validen Daten vorliegen.

**Szenariotechnik:** Damit werden aufgrund unterschiedlicher Annahmen mehrere zukünftige Szenarien als hypothetische Abfolge von Ereignissen entworfen. In der Regel werden dabei neben Extremszenarien („*Worst Case*“ und „*Best Case*“) auch typische Szenarien (Trendszenario) simuliert. Die Extremszenarien beschreiben eine optimale bzw. eine extrem ungünstige zukünftige Situation, ausgehend von optimistischen bzw. pessimistischen Grundannahmen, während das Trendszenario („*Most Likely Case*“) als das plausibelste angesehen wird. Szenariotechniken finden Anwendung in der Psychologie und in der Wirtschaft, wo sie dazu dienen, Markt- und Technologieentwicklungen zu simulieren und Handlungsmöglichkeiten aufzuzeigen.

### 2.3.2 Quantitative Methoden

Für Statistiker sind freilich quantitative Prognosetechniken interessanter. Diese reichen von einfachen Schätzverfahren über multivariable Regressionsanalysen bis hin zu Methoden der Künstlichen Intelligenz.

**Schätzungen:** Die einfachste Methode stellt die Schätzung eines Parameters aus den Daten einer vorzeitig vorhandenen Stichprobe dar. Dieser Schätzwert soll das das

Ergebnis der gesamten Population prognostizieren (das erst zu einem späteren Zeitpunkt vorliegt). Ein typisches Beispiel ist die Hochrechnung von Wahlergebnissen kurz nach Schließung der Wahllokale. Die Stichprobe gewinnt man, indem man einen Teil der Wähler unmittelbar nach deren Stimmabgabe befragt oder indem man sich auf Auszählungsergebnisse stützt, die zu einem frühen Zeitpunkt vorliegen. Ein solches Schätzverfahren ist keineswegs trivial: Zum einen ist darauf zu achten, dass die Stichprobe repräsentativ für die Population aller Wähler ist; ansonsten ist die Prognose unzuverlässig. Zusätzlich sollte für jeden Schätzwert ein Konfidenzintervall berechnet werden, das mit einer vorgegebenen Wahrscheinlichkeit (üblich sind 95 %) den zu prognostizierenden Wert überdeckt. Je schmaler dieses Intervall, desto präziser ist die Schätzung. Dessen Breite wird maßgeblich durch den Stichprobenumfang bestimmt. Es erscheint einleuchtend, dass eine große Stichprobe eine präzise Schätzung mit einem schmalen Konfidenzintervall ermöglicht, während ein geringer Stichprobenumfang mit einem eher unzuverlässigen Schätzwert einhergeht.

**Regressionsanalysen:** Diese Methoden eignen sich für die Analyse umfangreichen Datenmaterials (das aus Registern rekrutiert wird oder im Rahmen von Studien erhoben wird), um den Einfluss eines oder mehrerer prognostischer Parameter zu untersuchen. Die geeignete Analyse hängt von der Prognosevariablen (Zielgröße) ab. Bei einer quantitativen Prognosevariable  $y$  bietet sich eine lineare Regression an. Dabei wird der Zusammenhang zwischen  $y$  und einem oder mehreren Prädiktoren  $x_j$  durch eine lineare Funktion beschrieben. Wenn beispielsweise im Rahmen einer Therapiestudie zum Vergleich zweier blutdrucksenkender Medikamente Patienten dahingehend untersucht werden, um welchen Wert ihr Blutdruck sinkt, stellt das quantitative Merkmal „Blutdrucksenkung“ (gemessen in mmHg) die Zielvariable  $y$  dar, während als Prädiktoren die Therapieform, das Alter der Patienten, das Geschlecht, der Body-Mass-Index, der Blutdruckwert zu Beginn der Therapie und eventuell Risikofaktoren oder Komorbiditäten untersucht werden. Bei Kenntnis dieser Einflussfaktoren für einen individuellen Patienten lässt sich dann ein Wert für die zu erwartende Blutdrucksenkung berechnen (Punktprognose). Da eine exakte Vorhersage nicht möglich ist, wird üblicherweise ein Prognoseintervall konstruiert, das die zukünftige Realisierung der Zielvariablen mit einer hohen Wahrscheinlichkeit (meist 95 %) enthält. – Bei binären Prognoseparametern wird eine logistische Regression verwendet. Damit lässt sich die Wahrscheinlichkeit für das Eintreten eines Ereignisses (z. B. das Auftreten einer Komplikation) quantifizieren. – Für die Analyse von Zeitdauern eignet sich eine Cox-Regression. Dabei können auch zensierte Zeiten berücksichtigt werden. Diese entstehen, wenn das interessierende Endereignis (etwa der Tod eines Patienten oder der Nachweis eines Rezidivs) während der Beobachtungszeit der Studie nicht eintritt. Mit dem Cox-Regressionsmodell lässt sich zwar nicht die zu erwartende Zeitdauer (etwa die Überlebenszeit eines Individuums) schätzen; sie ermöglicht jedoch als semi-quantitatives

Verfahren den Vergleich zweier Individuen oder Subgruppen in Abhängigkeit von spezifischen Prädiktoren in Form von *Hazard Ratios*. Die Überlebensraten lassen sich für mehrere Subgruppen zu unterschiedlichen Zeitpunkten berechnen und graphisch anhand von Kaplan-Meier-Kurven darstellen.

**Zeitreihenanalysen:** Sie eignen sich, um die Entwicklung eines Merkmals in seinem zeitlichen Verlauf zu beschreiben, die Form der Zeitreihe zu untersuchen (auf Trends, Zyklen, saisonale oder irreguläre Schwankungen) und den Einfluss von äußeren Faktoren zu überprüfen. Dabei wird vorausgesetzt, dass Beobachtungswerte an zahlreichen Zeitpunkten (mindestens 30) vorliegen [6]. Im einfachsten Fall wird die gesamte Zeitreihe über ein lineares Polynom beschrieben (lineare Trendanalyse). Damit werden basierend auf Erfahrungswerten aus der Vergangenheit mittels Extrapolationen zukünftige Entwicklungen vorhergesagt. Mit autoregressiven Modellen lässt sich untersuchen, inwieweit der Wert zu einem bestimmten Zeitpunkt von vorangegangenen Zeitpunkten abhängt. Anwendungsbeispiele finden sich bei der Wettervorhersage und in der Demographie, um Bevölkerungsentwicklungen vorherzusagen.

**Künstliche Intelligenz (KI):** Methoden der KI erlauben es, mittels komplexer mathematischer Algorithmen bislang unbekannte Zusammenhänge zu entdecken, differenzierte Vorhersagemodelle zu entwickeln und diese aktuellen Gegebenheiten anzupassen (Maschinelles Lernen). Anwendungsbeispiele finden sich in allen Bereichen, in denen umfangreiches Datenmaterial aus unterschiedlichen Quellen (*Big Data*) bei sich ständig wechselnden Rahmenbedingungen schnell und effizient zu verarbeiten ist: bei der Wettervorhersage, in der Wirtschaft sowie bei der Energieversorgung.

## 2.4 Das Erstellen eines statistischen Modells

Das Erstellen eines statistischen Prognosemodells umfasst drei Phasen: die Modellspezifikation, das Generieren des Modells und schließlich dessen Validierung [2].

**Modellspezifikation:** In der ersten Phase werden die Zielgröße und geeignete Prädiktoren definiert. Diese Wahl ist überwiegend inhaltlich begründet und obliegt dem Konsumenten (z. B. einem klinischen Forscher, der eine Therapiestudie plant). Die Vorgaben sind bei der Studienplanung zu berücksichtigen. Der Statistiker ist dann vor die Aufgabe gestellt, in Abhängigkeit der Fragestellung eine geeignete Analyseverfahren auszuwählen und eine adäquate Fallzahl zu schätzen.

**Modellgenerierung:** In der zweiten Phase wird ein statistisches Modell generiert, das die Zielgröße in Abhängigkeit von mehreren Prädiktoren mittels einer mathematischen Gleichung erklärt. Meist wird das finale Modell nur statistisch signifikante Variablen enthalten. Bei dem allgemein üblichen Signifikanzniveau von  $\alpha=0,05$  werden also nur Prädiktoren berücksichtigt, deren p-Wert unter 0,05 liegt. Damit soll erreicht

werden, dass die durch das Modell dargelegten Zusammenhänge unter den gegebenen Rahmenbedingungen statistisch abgesichert sind.

**Modellvalidierung:** Es liegt in der Natur der Sache, dass das finale Modell nicht eindeutig bestimmt ist. Insbesondere bei Studien, in denen zahlreiche Prädiktoren erfasst werden, sind mehrere Modelle mit unterschiedlichen Funktionsgleichungen und Variablen-Kombinationen denkbar. Um einen objektiven Vergleich zu ermöglichen, ist es wichtig, die Güte eines statistischen Modells zu quantifizieren. Dafür stehen diverse Gütemaße zur Verfügung. Eine detaillierte Erläuterung würde den Rahmen dieses Beitrags sprengen. Deshalb seien sie an dieser Stelle nur kurz erläutert; für interessierte Leser sei auf Fachliteratur verwiesen [1, 7].

Bei einer linearen Regressionsanalyse bietet sich der Determinationskoeffizient  $R^2$  an: Dies ist das Quadrat des Korrelationskoeffizienten, der aus den beobachteten Werten  $y_i$  und den durch das Modell geschätzten Werten  $\hat{y}_i$  berechnet wird. Je näher  $R^2$  bei 1 liegt, desto besser ist das Modell. Andere Gütemaße basieren auf den individuellen Residuen ( $y_i - \hat{y}_i$ ). Aus diesen Differenzen (die jeweils negativ oder positiv sein können) lassen sich der mittlere quadratische und der mittlere absolute Prognosefehler schätzen. Dabei stehen kleinere Werte für eine bessere Modellgüte.

Auch bei einer logistischen Regression lässt sich der mittlere quadratische Fehler (der sogenannte „Brier-Score“) berechnen, indem man für den wahren Wert der Zielvariablen 1 oder 0 einsetzt (je nachdem, ob das Ereignis eingetreten ist oder nicht). Ein häufig verwendetes Maß ist außerdem die AUC („Area under the Curve“). Dieses Maß gibt an, wie gut das statistische Modell die Subgruppen (Ereignis „eingetreten“ oder „nicht eingetreten“) diskriminieren kann. AUC = 1 steht für ein perfektes Modell; AUC = 0,5 besagt, dass die Unterscheidung durch das Modell nicht besser ist als der Zufall.

Darüber hinaus existieren spezielle Gütemaße wie das AIC und das BIC (*Akaike* bzw. *Bayesian Information Criterion*). Je höher diese Werte, desto schlechter ist das Modell. Diese Maße sind bei allen Regressionsmodellen berechenbar; sie berücksichtigen explizit – anders als die oben erwähnten Gütemaße – die Anzahl der im statistischen Modell enthaltenen Parameter; in das BIC fließt zusätzlich der Stichprobenumfang ein. Je höher die Anzahl der prognostischen Parameter (und je höher der Stichprobenumfang beim BIC), desto größer wird der „Strafterm“. Einfache Modelle mit weniger Variablen sind – bei ähnlich guten Vorhersagen – gegenüber aufwendigen Modellen mit zahlreichen Variablen zu bevorzugen. Ein statistisches Modell sollte so einfach wie möglich und so komplex wie nötig sein.

A priori muss festgelegt werden, welchen Anforderungen das statistische Modell genügen soll, um als intern valide angesehen zu werden. Mögliche Gründe für ein nicht-valides Modell können ungeeignete Prädiktoren, das Nicht-Beachten wesentlicher Aspekte bei der Modellgenerierung oder eine nicht adäquate funktionale Beziehung

bei der Formulierung der Regressionsgleichung sein. Auch nicht valide Messmethoden, fehlerhafte, ungenaue oder veraltete Daten können die Ursache für unbrauchbares statistisches Modell sein.

## 2.5 Qualitätsanforderungen an ein Prognosemodell

Die im obigen Abschnitt erwähnten Güteparameter beschreiben nur die Validität des Modells. Sie besagen wenig über dessen prognostische Güte und sind insofern nur eingeschränkt aussagekräftig. Ein statistisches Prognosemodell sollte jedoch auch stabil sein. Das bedeutet: Es beschreibt nicht nur die Zusammenhänge innerhalb der Stichprobe („*In-Sample-Prognose*“), sondern ist auch für das Schätzen zukünftiger Messwerte und Ereignisse anwendbar („*Out-of-Sample-Prognose*“). Die Stabilität hängt im Wesentlichen von den äußeren Rahmenbedingungen ab. Diese kann weder der Statistiker noch der Nutznießer einer Prognose beeinflussen oder vorhersehen. Sich ändernde Rahmenbedingungen sind häufig der Grund für Fehlprognosen.

Das Erstellen und das Aktualisieren eines Prognosemodells ist also keine routinemäßige Angelegenheit. Neben Erfahrung im Verarbeiten umfangreichen Datenmaterials und der Kenntnis komplexer statistischer Analysetechniken sind dazu fachliches Spezialwissen und die intensive Zusammenarbeit zwischen Statistikern und Vertretern des jeweiligen Fachgebiets (z. B. Mediziner, Klimaforscher oder Wirtschaftswissenschaftler) vonnöten. Insbesondere Langzeitprognosen erfordern kontinuierliche Beobachtungen der Entwicklungen sowie die Bereitschaft auf beiden Seiten, die zugrunde liegenden Annahmen und die Validität der Prognose kritisch zu hinterfragen, auf sich ändernde Rahmenbedingungen adäquat zu reagieren, die Prognose – falls erforderlich – zu revidieren und Handlungsmöglichkeiten aufzuzeigen.

## 3 Anwendungsbeispiele

### 3.1 Prognosen in der Medizin

„Wie lautet meine Prognose?“ oder „Wie stehen meine Chancen?“ Derlei Fragen stellt ein Patient, wenn seine Krankheit lebensgefährlich ist oder wenn mit gravierenden Beeinträchtigungen zu rechnen ist. Konkret handelt es sich dabei um eine Vorhersage bezüglich eines individuellen Krankheitsverlaufs. Diese kann sich auf das Überleben, das Auftreten eines Rezidivs, die Wiederherstellung der Arbeitsfähigkeit des Patienten oder die Funktionsweise eines Organs beziehen.

Für den behandelnden Arzt mag eine ehrliche Antwort heikel sein. Diese hat schließlich weitreichende Konsequenzen: Eine gute Prognose vermittelt Hoffnung und

Zuversicht, während eine schlechte oder gar eine infauste Prognose die Betroffenen mit einer bitteren Realität konfrontiert. Andererseits ist klar, dass eine solche Prognose nicht absolut verlässlich ist, da auf den Heilungsprozess mitunter Einflüsse wirken, die ein Arzt im Einzelfall weder vorhersehen noch beeinflussen kann und die eventuell sogar seiner Erfahrung widersprechen.

Bereits in der hippokratischen Medizin war die Kunst der Prognose hochgeschätzt [9]. Sie ließ erkennen, dass der Arzt über eine hohe Expertise verfügte; außerdem diente sie dazu, eine geeignete Therapie zu planen. Schließlich dämpfte sie überzogene Erwartungen, so dass der Arzt bei einem ungünstigen Krankheitsverlauf nicht belangt werden konnte. Diese Argumente sind im Wesentlichen auch heute noch aktuell.

Die Güte einer solchen Prognose lässt sich anhand der in Abschnitt 2.2 dargelegten Kriterien beurteilen. Zur Nicht-Trivialität: Kein Arzt sollte einem Patienten eine nichtsagende Prognose wie „Alle Menschen müssen sterben“ zukommen lassen. Damit würde er sich seiner Verantwortung entziehen. – Das Kriterium der Objektivität betrifft die Rahmenbedingungen, auf denen die Prognose basiert. Zunächst spielt die individuelle Erfahrung des behandelnden Arztes eine wesentliche Rolle. Wenn er Hunderte Patienten mit einem ähnlichen Krankheitsbild behandelt hat, darf angenommen werden, dass er intuitiv den zu erwartenden Endzustand eines Individualfalls einzuschätzen vermag. Eine weitere Basis stellen Prognosestudien dar, in denen der Einfluss demografischer und prognostischer Faktoren sowie der Einfluss einer speziellen Therapie untersucht werden und die dann im Idealfall den zu erwartenden Nutzen einer Therapie objektiv quantifizieren. Dies setzt freilich voraus, dass der Arzt über den aktuellen Stand der Forschung umfassend informiert ist und dass die Ergebnisse der Studien auf den individuellen Patienten übertragbar sind. – Das dritte Kriterium (die Validität) fordert, dass der Patient ehrlich informiert wird. Wenn er wissen möchte, wie lange er noch zu leben hat, sollte der Arzt nicht auf Ersatzparameter (z. B. Laborwerte, Dauer des Krankenhausaufenthalts) ausweichen, die für den Patienten von untergeordneter Bedeutung sind.

Das härteste Kriterium ist die Überprüfbarkeit des Wahrheitsgehalts der Prognose. Man wird es dem Arzt verzeihen, wenn sich seine Prognose im Nachhinein als zu pessimistisch erweist, weil die Krankheit wider Erwarten einen günstigeren Verlauf genommen hat als ursprünglich angenommen. Im umgekehrten Fall – wenn der Endzustand des Patienten schlechter ist als prognostiziert – wird man eher geneigt sein, die Kompetenz des Arztes in Zweifel zu ziehen. Dabei ist zu bedenken: Eine Prognose darf nicht mit der gewünschten Zukunft verwechselt werden. Kein Arzt kann mit Sicherheit einen individuellen Krankheitsverlauf vorhersagen; aus den Ergebnissen einer Prognosestudie lassen sich keine konkreten Hinweise bezüglich eines Einzelfalls ableiten. Dies sollte dem Patienten klar kommuniziert werden.

Bei mittel- oder längerfristigen Prognosen erscheint es darüber hinaus sinnvoll und notwendig, die zugrunde liegenden Annahmen hin und wieder zu überprüfen. Eine Änderung der Prognose kann beispielsweise erforderlich sein, wenn bei einem Patienten eine seltene, nicht erwartete Nebenwirkung einer Therapie auftritt oder wenn sich aus unvorhersehbaren Gründen dessen Allgemeinzustand verschlechtert. Das kann dazu führen, dass die angewandte Therapie abgesetzt, modifiziert oder durch eine andere Therapie ersetzt wird. Änderungen können auch durch mangelnde Compliance des Patienten bedingt sein: Sei es, dass er kurative Maßnahmen ablehnt oder dass er nicht motiviert ist, den Anordnungen seiner Ärzte Folge zu leisten. Dies führt in der Regel zu einer Verschlechterung der Prognose oder dazu, dass der erwartete Endzustand (z. B. Tod, Organversagen) früher eintritt. Andererseits ist es nicht ausgeschlossen, dass der gesundheitliche Zustand eines Patienten sich wider Erwarten verbessert, was ebenfalls eine Korrektur der ursprünglichen Prognose erfordert. Auch Fortschritte in der Forschung – etwa eine neu entwickelte Therapie – können bewirken, dass Heilung möglich ist, die kurze Zeit zuvor ausgeschlossen schien.

In jedem Fall können unvorhersehbare Ereignisse den Krankheitsverlauf positiv oder negativ beeinflussen und zu einer überraschenden Wende führen – auch wenn dies eher selten geschieht. Dessen sollten sich sowohl die Patienten als auch die behandelnden Ärzte bewusst sein.

## 3.2 Wahlprognosen

Obwohl in den letzten Jahren teilweise eine eher geringe Wahlbeteiligung zu beobachten war, ist das Interesse der Bürger eines Landes oder einer Kommune an künftig zu erwartenden Wahlergebnissen hoch. Um Wahlverhalten zu erforschen und Prognosen zu erstellen, können Aggregatdaten- oder Individualdatenanalysen verwendet werden. Aggregatdaten beinhalten amtliche Ergebnisse von Stimmbezirken. Aufgrund von Studien aus der Wahlforschung ist bekannt, welche sozialen, wirtschaftlichen und strukturellen Faktoren das Wahlverhalten beeinflussen. Diese Faktoren können zusammen mit den Ergebnissen von vorangegangenen Wahlen als Prädiktoren in einem multiplen Regressionsmodell verwendet werden, um die Ergebnisse einer bevorstehenden Wahl zu prognostizieren. Außerdem lassen sich mit Regressionsanalysen Wählerwanderungen untersuchen. Der Vorteil von Aggregatdatenanalysen liegt darin, dass man auf vorhandenes Datenmaterial zurückgreifen kann, und dass der Einfluss der einzelnen Indikatoren quantifizierbar ist. Nachteilig ist allerdings, dass eine darauf basierende Prognose nur dann valide Ergebnisse liefert, wenn sich das Wahlverhalten der wahlberechtigten Bürger nicht grundlegend ändert.

Andere Prognosen basieren auf Individualdaten, die man durch Befragungen bei einem Teil der wahlberechtigten Bürger erhält. Dazu zählen das politische Barometer, das jede Woche von der ARD ausgestrahlt wird, ebenso wie Umfragen, die vor politischen Wahlen durchgeführt werden. Dabei muss darauf geachtet werden, dass die untersuchte Stichprobe repräsentativ für die Gesamtheit aller Wahlberechtigten ist, und dass diese Stichprobe hinreichend groß ist. Die erforderliche Größe richtet sich danach, wie präzise die Ergebnisse sein sollten. Wenn beispielsweise prognostiziert werden soll, ob eine Partei bei einer Bundestagswahl an der 5%-Hürde scheitert, ist eine wesentlich größere Stichprobe erforderlich als bei einer Umfrage vor einer Bürgermeisterwahl mit zwei Kandidaten und einem Favoriten. Im ersten Fall kann eine Abweichung des Umfrageergebnisses vom tatsächlichen Wahlausgang in Höhe von 0,1% zu einer gravierenden Fehlprognose führen, während es im zweiten Fall auf ein paar Prozentpunkte mehr oder weniger nicht ankommt.

Die Prognosen der Umfragen, die am Wahltag durchgeführt werden, liegen in der Regel nahe an den offiziellen Wahlergebnissen. Das spricht für die Qualität der Prognoseverfahren. Bei diesen Umfragen arbeiten die Meinungsforscher unter idealen Bedingungen. Den Teilnehmern wird direkt nach der Stimmabgabe eine Frage gestellt, die einfach zu beantworten ist und kaum durch Erinnerungslücken beeinträchtigt sein dürfte: „Welche Partei haben Sie gewählt?“ Mit der zusätzlichen Frage „Welcher Partei haben Sie bei der letzten Wahl Ihre Stimme gegeben?“ lassen sich Informationen zu Wählerwanderungen gewinnen.

Diese Umfragen bieten mehrere Vorteile: Da die Befragten ihre Antwort anonym abgeben, besteht kein Anlass, falsche Angaben zu machen. Außerdem ist der Stichprobenumfang in aller Regel sehr hoch, was dazu führt, dass die Schätzungen das endgültige Wahlergebnis präzise widerspiegeln. So kann oft schon kurz nach Schließung der Wahllokale eine zuverlässige Prognose gegeben werden.

Wahlprognosen, die längere Zeit vor einer Wahl erstellt werden, genießen einen weniger guten Ruf. Bei telefonischen Umfragen, bei der nur eine vergleichsweise geringe Zahl von wahlberechtigten Personen befragt wird, stellen sich die Rahmenbedingungen komplett anders dar. Abgesehen von der unpräzisen Schätzung aufgrund der geringen Fallzahl können diverse Fehlerquellen dazu beitragen, dass die Umfrageergebnisse das endgültige Ergebnis nicht korrekt widerspiegeln. Diese Verzerrungen sind im Wesentlichen durch fehlende oder falsche Angaben bedingt. Besondere Vorsicht ist bei Online-Umfragen geboten. Hier ist kaum anzunehmen, dass die Stichprobe repräsentativ ist, da nur eine sehr spezielle Klientel erreicht wird.

Es gibt vielerlei Gründe für fehlende Antworten. Manche Zeitgenossen verweigern die Teilnahme – sei es, weil sie darin eine Verletzung des Wahlheimnisses sehen oder weil sie nicht zugeben möchten, einer extremen Partei ihre Stimme zu geben. Wenn der überwiegende Teil dieser *Non-Responder* für eine bestimmte Partei sympathisiert,

führt dies zu einem Bias. Seriöse Umfrageinstitute bemühen sich, solche Verzerrungen mit einer statistischen Nachbearbeitung zu korrigieren, indem sie die Stimmanteile der einzelnen Parteien unterschiedlich gewichten.

Fehler aufgrund von falschen Angaben sind wesentlich schwerer zu erkennen und zu beheben. Diese entstehen beispielsweise, wenn Befragte ihre Meinung kurzfristig ändern. Sie versichern, eine bestimmte Partei zu wählen – und gehen dann nicht zur Wahl oder votieren anders. Andere behaupten, nicht zu wählen – und tun es am Wahltag doch. Wiederum andere geben absichtlich falsche Antworten, um die Anrufer zu foppen. Manche Wähler sind unehrlich, weil sie fürchten, dass ihre Meinung negativ bewertet wird (das Phänomen der „sozialen Erwünschtheit“). Ferner ist denkbar, dass Wähler aufgrund von Umfrageergebnissen ihr Verhalten ändern, etwa um eine Partei zu unterstützen, die an der 5 %-Hürde zu scheitern droht, oder weil sie glauben, dass das Wahlergebnis bereits feststeht. Umfrageergebnisse können auch auf die politischen Akteure Einfluss ausüben und dazu führen, dass sie in der Endphase des Wahlkampfes besonders aktiv werden oder dass sie resignieren. Schließlich können unvorhergesehene Ereignisse kurz vor einer Wahl dazu führen, dass Wähler ihre ursprüngliche Entscheidung revidieren oder politisch Verantwortliche ihre Strategie ändern.

Aus diesen Gründen erwiesen sich manche Prognosen in der Vergangenheit als falsch. Es sei an die Wahl des US-Präsidenten am 8. November 2016 erinnert. Lange Zeit sah Hillary Clinton wie die sichere Gewinnerin aus, letzten Endes wurde jedoch Donald Trump zum Präsidenten gewählt. Wie konnte es dazu kommen? Tatsächlich erhielt Clinton mehr Stimmen als Trump (65,4 im Vergleich zu 62,8 Millionen). Insofern lag die Wahlprognose nicht völlig daneben. Dass dennoch Trump Präsident wurde, liegt zum einen am US-amerikanischen Wahlsystem. In Staaten, bei denen die Stimmanteile der Kandidaten nahe bei 50 % lagen, waren präzise Prognosen extrem schwierig. So waren im Bundesstaat Florida 48,1 % für Clinton und 47,5 % für Trump prognostiziert worden; am Ende erhielten die Kandidaten 47,7 % bzw. 49,1 %. Die prognostizierten Werte wichen von den tatsächlich erreichten zwar nur wenig ab. Trotzdem hatte dies massive Auswirkungen: Alle 29 Wahlmänner dieses Bundesstaates votierten für Trump. Andere Ursachen sind psychologisch bedingt: Viele Trump-Wähler bekannten sich nicht zu ihrer Entscheidung und gaben bei Umfragen keine oder eine falsche Antwort, was zu einer Verzerrung zugunsten Clintons führte. – Ein weiterer Grund liegt in der falschen Interpretation von Umfrageergebnissen: Aus einer Prognose, die Clintons Sieg mit einer Wahrscheinlichkeit von 70 % prophezeite, leiteten einige Wähler und Medien die Botschaft ab, Clinton werde zweifellos die nächste Präsidentin werden. Manche Medien verbreiteten die Nachricht, dass 70 % der Wahlmänner für Clinton (und nur 30 % für Clinton) votieren würden. Auch wenn dies barer Unsinn ist: Manche Wahlhelfer der Demokratischen Partei glaubten angesichts dieser Zahlen, dass weitere Anstrengungen für ihre Kandidatin unnötig wären, und viele US-Bürger gaben sich der Illusion hin,

dass die Wahl entschieden und ihre Stimmabgabe nicht mehr erforderlich wäre. Ein fataler Irrglaube, wie sich bald erweisen sollte. Das Umfrageergebnis zeigte lediglich, dass der Ausgang der Wahl offen war.

Ein anderes Beispiel stellt die Bundestagswahl 2002 dar. Lange Zeit sah die CDU/CSU mit ihrem Kanzlerkandidaten Edmund Stoiber wie der sichere Gewinner aus; sogar kurz nach Schließung der Wahllokale wurde von manchen Medien ein Sieg der Unionsparteien prognostiziert. Am Ende kam es zu einer Fortsetzung der roten-grünen Koalition unter Bundeskanzler Gerhard Schröder. Wie konnte dies passieren? Wenige Wochen vor der Wahl kam es nach extremen Regenfällen zu einer Flutkatastrophe, von der weite Teile Deutschlands betroffen waren. Dieses Ereignis und das Krisenmanagement der damaligen Regierung führten offensichtlich dazu, dass viele Wähler ihre Meinung kurzfristig änderten und der rot-grünen Regierung zu einem knappen Sieg verhalfen. Die falsche Prognose kurz nach Schließung der Wahllokale könnte darauf zurückzuführen sein, dass sie auf einer nicht-repräsentativen Stichprobe basierte, in der bayrische Stimmbezirke überproportional häufig vertreten waren.

Überraschende Wahlergebnisse bedeuten also nicht zwingend, dass das Umfrageverfahren nicht valide war. Sie sind eher darauf zurückzuführen, dass die gewonnenen Informationen unzuverlässig sind. Darin zeigt sich der Unterschied zwischen medizinischen Prognosen und Wahlumfragen. Der Zustand eines Patienten kann vom behandelnden Arzt objektiv beurteilt werden, nicht jedoch das Wahlverhalten nach einer Umfrage.

### 3.3 Demografische Prognosen

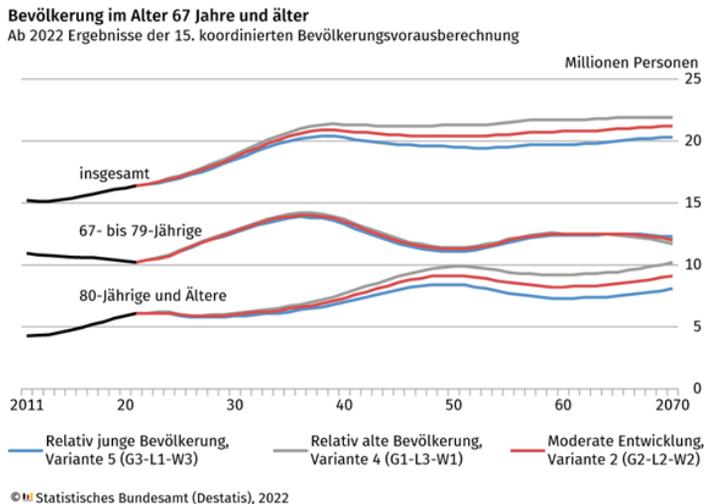
Demografie ist eine wissenschaftliche Disziplin, die sich mit Populationen und ihren Strukturen befasst. Zur Beschreibung einer Population und deren Änderungen über die Zeit sind drei Prozesse relevant: Geburtenverhalten, Alterung und Migration. Demografische Prognosen betreffen die zukünftige Entwicklung einer Population. Gebräuchliche Kenngrößen sind die Geburtenrate (Anzahl der Lebendgeborenen pro Jahr bezogen auf 1000 Einwohner), die Fertilitätsrate (Anzahl der Kinder, die eine Frau durchschnittlich zur Welt bringt), die durchschnittliche Lebenserwartung bei der Geburt und altersspezifische Mortalitätsraten. Mittels dieser Werte lassen sich Populationen untereinander vergleichen, zeitliche Änderungen beschreiben und Prognosen erstellen. Die dafür erforderlichen Daten werden rekrutiert aus Registern, Befragungen oder mittels Volkszählungen.

Als einer der Begründer der Demografie gilt der Kurzwarenhändler **John Graunt** (1620–1674), der bereits im 17. Jahrhundert basierend auf Londoner Geburts- und Sterberegistern Gesetzmäßigkeiten bezüglich der Bevölkerungsentwicklung herleitete.

**Johann Peter Süßmilch** erstellte 1741 das bahnbrechende Werk der deutschen Bevölkerungsstatistik mit dem Titel „Die göttliche Ordnung in den Änderungen des menschlichen Geschlechts“. Aufgrund seiner Berechnungen war es erstmals möglich, unter Annahme bestimmter Rahmenbedingungen Prognosen bezüglich der künftigen Entwicklung einer Bevölkerung zu erstellen.

Das öffentliche Interesse an Demografie hat in den letzten Jahren zugenommen. In den meisten Ländern Europas ist ein Altern der Gesellschaft zu beobachten. Die Lebenserwartung steigt, während die Fertilitätsrate bei weniger als zwei Kindern pro Frau liegt. Dies bereitet vielen Menschen Sorge: Ältere fürchten um ihre Versorgung, während jüngere Angst vor einer ungewissen Zukunft verspüren. Es ist Aufgabe der Politik, rechtzeitig Lösungen zu erarbeiten, um unliebsamen Entwicklungen entgegenzuwirken und die Ressourcen gerecht unter den Generationen und – global gesehen – unter den Völkern der Welt zu verteilen.

Demografen sehen sich bei der Erstellung ihrer Prognosen mit großen Herausforderungen konfrontiert. Um Bevölkerungsentwicklungen auf nationaler und internationaler Ebene zu prognostizieren, müssen sie realistische und nachvollziehbare Annahmen bezüglich des künftigen Geburtenverhaltens, der zu erwartenden Sterblichkeit und der Zu- und Abwanderung treffen. Dies ist aufgrund der instabilen politischen, ökonomischen und sozialen Verhältnisse weltweit und aufgrund der zahlreichen Einflussfaktoren und deren komplexen Wechselwirkungen problematisch. Es kommt hinzu, dass in der Demografie in aller Regel Langzeitprognosen erstellt werden, die per se mit einer großen Unsicherheit assoziiert sind. Deshalb werden häufig mehrere Szenarien simuliert, um künftige Entwicklungen zu prognostizieren (siehe Abbildung 2).



**Abbildung 2:** Mehrere Varianten der Bevölkerungsentwicklung in Deutschland bis 2070

Dies bedeutet jedoch keineswegs, dass demoskopische Prognosen generell unzuverlässig oder gar sinnlos sind. Wenn die zugrunde liegenden Annahmen auf realen Fakten basieren, die keinen gravierenden zeitlichen Schwankungen unterliegen, stellen diese Prognosen ein unverzichtbares Mittel zur Planung dar. So lässt sich beispielsweise aufgrund der Anzahl von Kleinkindern, die in einer Region leben, und der Altersstruktur der aktuell unterrichtenden Lehrer durchaus zuverlässig abschätzen, wie hoch der Bedarf an Schulgebäuden und Lehrpersonal in wenigen Jahren sein wird.

Dieses Beispiel zeigt: Demographische Prognosen sind unerlässlich für politische Entscheidungen und zukunftsweisende Planungen. Diese wiederum sind erforderlich, um eine auch auf längere Sicht gut funktionierende Infrastruktur zu gewährleisten und um auf unerwartete Änderungen adäquat reagieren zu können. Dies betrifft neben dem Bildungswesen die Wirtschaft, den Arbeitsmarkt, das Verkehrswesen, den Wohnungsbau, den Rentenmarkt, die medizinische Versorgung, das Versicherungswesen und vieles mehr.

Vorsicht ist jedoch geboten vor allzu weit in der Zukunft liegenden Horrorszenarien. Wenn etwa Politiker prophezeien, dass aufgrund der Bevölkerungsentwicklung in Deutschland in einigen Jahrzehnten Arbeitskräftemangel herrschen würde, die Renten nicht mehr finanzierbar wären und die Sozialsysteme zusammenbrechen würden, darf man das getrost als Kaffeesatzleserei abtun. Um die Wertigkeit von Prognosen über 50 Jahren zu beurteilen, lohnt sich ein Blick zurück in das Jahr 1974. Niemand hätte damals die Auflösung des Ostblocks, eine weltweite Pandemie oder einen Krieg in der Ukraine für möglich gehalten; keiner konnte sich eine Welt mit Smartphones und Internet vorstellen. Nutzbringender als höchst zweifelhafte Langzeitprognosen sind Überlegungen, welche Änderungen im nächsten Jahr anstehen und wie diese zu bewältigen sind.

## 4 Diskussion

### 4.1 Der Umgang mit Wahrscheinlichkeiten

Bezüglich des Umgangs mit Wahrscheinlichkeiten hat der Mensch eine paradoxe Natur. Einerseits sind viele Zeitgenossen davon überzeugt, dass sie zumindest hin und wieder in der Zukunft liegende Ereignisse aufgrund persönlicher Erfahrungen vorhersagen oder die Wahrscheinlichkeit für deren Eintreten intuitiv abschätzen können. Andererseits fehlt den meisten Menschen die Fähigkeit, vorgegebene Wahrscheinlichkeiten adäquat zu interpretieren.

Laut der 1979 erschienenen „Deutsche Risikostudie Kernkraftwerke“ liegt das geschätzte Risiko für eine Kernschmelze bei 1 zu 10.000 pro Betriebsjahr. Nach Sicherheitsanalysen, die zu Beginn des 21. Jahrhunderts erstellt wurden, ist die

Wahrscheinlichkeit einer Kernschmelze für das Kernkraftwerk Fukushima bei 1 zu einer Million pro Jahr noch wesentlich geringer. Die Nuklearkatastrophen in Tschernobyl am 2. Mai 1986 und in Fukushima am 11. März 2011 mit jeweils verheerenden Auswirkungen scheinen diesen Prognosen zu widersprechen. Waren die Berechnungen falsch?

Diese Wahrscheinlichkeiten suggerieren, dass keiner der Menschen, der heute die Erde bewohnt, eine solche Katastrophe erleben wird. Dabei ist zu beachten: Bei derzeit mehr als mehr als 400 Reaktorblöcken weltweit droht rein rechnerisch alle 25 Jahre ein unbeherrschbarer Reaktorunfall. Die rohe Wahrscheinlichkeit besagt zudem nichts darüber, wann sich ein solcher Unfall ereignen wird.

Vertreter zahlreicher Fachrichtungen wie Klimaforscher, Versicherungsmathematiker oder Demoskopien, Journalisten, Meinungsforscher und Hobby-Prognostiker agieren gerne mit Zahlen. Argumente aufgrund objektiven Datenmaterials sind nachvollziehbar; darauf basierende Prognosen scheinen glaubhaft und seriös zu sein. Gänzlich unstatthaft ist dies jedoch in Situationen, bei denen alleine der Zufall ausschlaggebend ist. Dazu zählen Vorhersagen, die die nächste Ziehung der Lottozahlen betreffen, wie etwa „Die 11 wird mit hoher Wahrscheinlichkeit gezogen werden“. Derlei unseriöse Vorhersagen gründen sich auf dem irrigen Glauben, dass bei endlich vielen Lottospielen die Häufigkeiten der Ziehungen für jede Zahl identisch sind und dass deshalb – wenn die 11 in den letzten Spielen seltener als alle anderen Zahlen gezogen wurde – die Wahrscheinlichkeit für die 11 beim nächsten Spiel besonders hoch ist. In Wirklichkeit sind die Ziehungen unabhängig voneinander: Die Wahrscheinlichkeit, dass eine bestimmte Zahl gezogen wird, beträgt bei jeder Ziehung für jede Zahl konstant  $6/49$ ; die Wahrscheinlichkeit für einen 6er im Lotto liegt bei 1 zu 14 Millionen (jeweils ohne Berücksichtigung der Zusatzzahl).

Auch bei sportlichen Ereignissen werden zuweilen Prognosen höchst zweifelhafter Seriosität gewagt. Traditionell werden im Sport diverse Parameter ermittelt. So werden beispielsweise bei jedem Fußballspiel von jedem einzelnen Spieler die Anzahl von Toren, Torschüssen, Frei-, Eck- und Strafstoßen und Fouls sowie die gelaufenen Meter und die Durchschnittsgeschwindigkeit erfasst. Bei Wettbewerben werden die Einsätze pro Spieler sowie die Zahl von Siegen, Niederlagen und Unentschieden jedes Teams gezählt. Mit diesem umfangreichen Datenmaterial lassen sich vielfältige Statistiken erstellen. Für Prognosezwecke sind diese jedoch untauglich. Häufig wird zwar argumentiert, dass eine Mannschaft, die bei Duellen in der Vergangenheit häufig als Sieger hervorgegangen ist, mit hoher Wahrscheinlichkeit auch beim nächsten Duell siegen wird. Diese „Wahrscheinlichkeit“ lässt sich sogar anhand des Anteils der gewonnenen Spiele quantifizieren. Manche Fans zeigen sich siegesgewiss, wenn diese Wahrscheinlichkeit höher als 50 % ist (so wie einst viele US-Bürger, die ein Umfrageergebnis von 70 % für Clinton als sicheren Sieg missdeuteten). Die Erfahrung lehrt jedoch, dass auf solche Vorhersagen kein Verlass ist. Über Sieg oder Niederlage entscheiden neben dem

Können und dem Engagement der Spieler diverse äußere Einflüsse und Unvorhersehbarkeiten, die sich unter dem Stichwort „Zufall“ zusammenfassen lassen. Aussagen wie „Die Statistik spricht für Deutschland“ sind daher eher als Wunschvorstellungen denn als ernst zu nehmende Prognosen aufzufassen.

## 4.2 Der Umgang mit Fehlprognosen

Zahlreiche vor langer Zeit erstellte Prognosen sorgen heute – nachdem einige Jahrzehnte vergangen sind – für Erstaunen, Erheiterung oder auch Mitleid ob der Naivität unserer ehemals lebenden Vorfahren. Dazu einige Beispiele: „Der Bauch, die Brust und das Gehirn werden dem Chirurgen für immer verschlossen bleiben.“ So äußerte sich **Sir John Eric Erichsen** (1818–1896), Leibarzt der Queen Victoria. **Thomas John Watson** (1874–1956), Vorstandsvorsitzender von IBM, prognostizierte im Jahr 1943, dass es keinen großen Markt für Computer geben werde. Mitte des 19. Jahrhunderts belegten Hochrechnungen, dass die Straßen von New York spätestens 1910 meterhoch mit Pferdemist bedeckt und damit unpassierbar sein würden. „Röntgenstrahlen werden sich als Schwindel erweisen“ unkte 1883 der Physiker **William Thomson Lord Kelvin** (1824–1907). „Die weltweite Nachfrage nach Kraftfahrzeugen wird eine Million nicht überschreiten, allein schon aus Mangel an verfügbaren Chauffeuren“ mutmaßte der Erfinder des Automobils **Gottlieb Daimler** (1834–1900).

Beispiele für Fehlprognosen finden sich auch in der neueren Vergangenheit. In den 1980er Jahren flößte der Begriff „Waldsterben“ vielen Menschen Angst ein. „Auf den Wetterbericht ist kein Verlass“: Dieser Satz erfreut sich nach wie vor breiter Zustimmung. Vorhersagen bezüglich des Verlaufs der Covid-19-Pandemie konnten häufig nicht bestätigt werden. Weitere Beispiele über fatale Irrungen und überraschende Wendungen finden sich in [5].

Es gilt zu bedenken, dass diese Prognosen zu ihrer Zeit plausibel schienen. Gegen Ende des 19. Jahrhunderts schienen chirurgische Eingriffe am Thorax oder am Hirn undenkbar zu sein. Erichsen konnte nicht ahnen, welche rasanten Fortschritte die medizinische Forschung machen würde. Zu Watsons Lebzeiten waren Computer extrem teuer; ihre Bedienung erforderte hohen technischen Sachverstand. Dass Computer einst den Alltag fast aller Menschen prägen würde, war unter diesen Voraussetzungen nicht vorstellbar. Die Vorhersage bezüglich des Pferdemistes wurde mit Wahrscheinlichkeitsrechnung hergeleitet. Die Eigenschaften der Röntgenstrahlung (die gleichzeitig Partikel- und Wellencharakter zeigte) verwirrte viele Naturwissenschaftler, nicht nur Kelvin. Daimlers Prophezeiung war nach damaligen Erkenntnissen (als sich kaum jemand ein Auto leisten oder es bedienen konnte) glaubwürdig.

Diese Beispiele zeigen: Zukunftserwartungen können sich vom tatsächlichen Fortgang der Geschichte unterscheiden. Ein Prognosemodell ist eben nur so gut wie die zugrunde liegenden Annahmen. Wenn diese sich aufgrund technischer, sozialer oder demografischer Entwicklungen oder aufgrund von wissenschaftlichen Erkenntnissen ändern, verliert die Prognose ihre Gültigkeit. Das bedeutet nicht, dass sie schlecht war; es zeigt vielmehr, dass sie nur im zeitlichen Kontext zu bewerten ist.

In Szenarien, die maßgeblich von menschlichen Verhaltensweisen bestimmt sind, beobachtet man häufig das sogenannte Prognoseparadox. Dieses entsteht, wenn eine Prognose zu Verhaltensänderungen führt. Diese sollen verhindern, dass das prognostizierte Ereignis eintritt. Damit ist die Prognose bereits zum Zeitpunkt ihrer Veröffentlichung falsch. In Abschnitt 3.2 wurde ausgeführt, dass das Ergebnis von Umfragen vor einer Wahl häufig dazu beiträgt, dass bedingt durch das Wählerverhalten das Wahlergebnis anders ausfällt. Das in den 1980er Jahren prognostizierte Waldsterben hat dazu geführt, dass seitens der Politik, der Forstwirtschaft und der Industrie Anstrengungen unternommen wurden, um genau dies zu verhindern. Besonders problematisch erwiesen sich Prognosen bezüglich der Covid-19-Pandemie. Zahlreiche Quellen der Unsicherheit (beispielsweise bezüglich des Infektionsmechanismus, der schwer zu schätzenden Basisreproduktionszahl, der Wettereinflüsse und zahlreicher anderer Faktoren) erschwerten das Erstellen einer validen Prognose. Schließlich führten politische Maßnahmen und menschliches Verhalten dazu, dass sich die Rahmenbedingungen und damit auch die Prognose änderten. Es wäre unfair, dies allein deren Entwicklern anzulasten.

Noch eine Anmerkung zu Langzeitprognosen: Prognosen, die sich auf einen Zeitraum beziehen, der in weiterer Ferne liegt, sind freilich mit extrem hohen Unsicherheiten assoziiert. Man könnte geneigt sein, diesen Prognosen ihren Sinn abzuspochen, weil sie letzten Endes nicht überprüfbar sind. Dies wäre jedoch zu kurz gedacht. Als Beispiel sei der prognostizierte Klimawandel erwähnt. Aufgrund umfangreicher Datenmengen und komplexer Algorithmen lassen sich zwar mittlerweile recht präzise Kurzzeitprognosen für das Wetter in den nächsten Tagen erstellen. Wie sieht es jedoch mit der Validität von Langzeitprognosen für das Klima aus, die sich über einen Zeitraum von Jahrzehnten oder gar Jahrhunderten erstrecken? Freilich sind genaue Vorhersagen unmöglich. Es ist jedoch bekannt, dass die Entwicklung des Klimas wesentlich von Treibhausgasemissionen bestimmt wird. Vor diesem Hintergrund erscheint es sinnvoll, unterschiedliche Szenarien zu entwickeln, um weitere mögliche Entwicklungen zu prognostizieren. Auf diese Weise lassen sich mögliche Konsequenzen aufzeigen, was wiederum dazu führen mag, dass Gegenmaßnahmen ergriffen werden, um das schlimmste aller vorstellbaren Szenarien zu verhindern. Insofern können Langzeitprognosen vieles in Gang setzen, auch wenn die Prognose selbst höchst ungewiss ist.

## 5 Schlussfolgerungen

Viele Entscheidungen beruhen auf Prognosen – sei es in der Politik, in der Ökonomie, der Medizin oder im privaten Umfeld. Demzufolge wird unser aller Leben in hohem Maße direkt oder indirekt von Prognosen bestimmt, auch wenn wir uns dessen nicht permanent bewusst sind. Insofern sind Prognosen wichtig, da sich aus ihnen konkrete Handlungsempfehlungen ableiten lassen. Ansonsten wäre das Überangebot der auf uns einströmenden Informationen kaum zu bewältigen.

Freilich ist man geneigt, prognostischen Methoden aus früheren Zeiten wie das Befragen eines Orakels ihren Sinn abzuspochen. Dabei sollte man jedoch nicht vergessen, dass die so erstellten Prognosen den Menschen in der Antike oder im Mittelalter plausibel erschienen, dass sie Orientierungshilfen boten und dass darauf basierend weitreichende politische oder kulturelle Entscheidungen getroffen wurden. Insofern gibt es durchaus Parallelen zwischen den obskuren Methoden, die in früheren Zeiten angewandt wurden, und den heute bevorzugten mathematisch-statistischen Techniken. Es liegt in der Natur der Sache, dass eine Prognose unsicher ist und dass sich deren Qualität nicht unmittelbar nach deren Erstellung, sondern erst zu einem späteren Zeitpunkt gänzlich beurteilen lässt. Insbesondere Langzeitprognosen und Prognosen, die von menschlichen Verhaltensweisen und politischen Entscheidungen beeinflussbar sind, erweisen sich häufig als falsch. Der große Mathematiker Laplace irrte sich, als er postulierte, dass eines fernen Tages die Zukunft vollständig vorhersehbar wäre. Manche Menschen neigen sogleich zu hämischen Kommentaren, wenn sich herausstellt, wenn sich eine Prognose als Fehlprognose erweist. Dabei sollte man jedoch bedenken, dass jede Prognose auf einem Modell basiert, das die Wirklichkeit nur vereinfacht darstellt. Manche Prognosen (die sich später als falsch herausstellten) haben sogar viel Gutes bewirkt, weil sie problematische Entwicklungen erst aufgezeigt und rechtzeitig Gegensteuerungen bewirkt haben.

Die Schwächen der Prognose sprechen also nicht grundsätzlich gegen ihren Einsatz. Prognosen helfen uns nicht nur, rationale Entscheidungen zu treffen; sie können auch für künftige Prozesse und Probleme sensibilisieren. Die im Titel dieses Beitrags aufgeworfene Frage „Top oder Flop?“ ist demzufolge eindeutig mit „Top“ zu beantworten.

## Literatur

Auer B., Rottmann H.: *Statistik und Ökonometrie für Wirtschaftswissenschaftler*. Kapitel 6: Prognose mit geschätzten

Regressionsmodellen. Springer Fachmedien Wiesbaden (2015)

- Bacher J., Müller W., Ruderstorfer S.:** Statistische Prognoseverfahren für die Sozialwissenschaften. In: Bachleitner R et al. (Hrsg.): *Empirische Prognoseverfahren in den Sozialwissenschaften, Zukunft und Forschung*. Springer Fachmedien Wiesbaden (2016)
- Gelaschwili S.:** Einführung in die Statistische Modellierung und Prognose. In: *Statistische Diskussionsbeiträge* Nr. 26. Universität Postdam (2007)
- Labisch A.:** Prognosen in den Wissenschaften – Einleitung. *Acta Historica Leopoldina* 79, 7–13 (2021)
- Radkau J.:** *Geschichte der Zukunft. Prognosen, Visionen, Irrungen in Deutschland von 1945 bis heute*. Hanser-Verlag 2017
- Shumway RH., Stoffer DS:** *Time Series Analysis and Its Applications: With R Examples*. Springer-Verlag, 4. Auflage (2017)
- Weiß C.:** Regressionsanalysen. In: *Basiswissen Medizinische Statistik*, Springer Verlag, 7. Auflage 2019
- Weiß C.:** Entwicklung der Medizinischen Statistik in Deutschland. Der lange Weg dahin. *GMDS Med Inform Biom Epidemiol* (2005)
- Wildner M.:** Diagnose, Prognose, Prädiktion. *Gesundheitswesen* 76: 185–186 (2014)

## Über die Autorin

**Christel Weiß** ist Professorin für Biomathematik und Epidemiologie an der Medizinischen Fakultät Mannheim der Universität Heidelberg. In ihren Verantwortungsbereich fallen Lehrveranstaltungen für Studierende der Medizin und Masterkurs-Absolventen, Seminare sowie die Beratung von Ärzten, wissenschaftlichen Mitarbeitern und Doktoranden bei der Planung und Durchführung von klinischen und epidemiologischen Studien. Frau Weiß ist Autorin des Lehrbuchs „Basiswissen Medizinische Statistik“ (erschieden im Springer-Verlag, demnächst 8. Auflage), des Ratgebers „Promotion. Die medizinische Doktorarbeit – von der Themensuche bis zur Dissertation“ (zusammen mit Prof. Dr. Axel Bauer, erschienen im Thieme-Verlag, 4. Auflage) sowie Autorin oder Koautorin zahlreicher Papers und Buchbeiträge.

## Korrespondenzadresse

Prof. Dr. Christel Weiß  
 Medizinische Fakultät Mannheim der Universität Heidelberg  
 Abteilung für Medizinische Statistik und Biomathematik, Theodor-Kutzer-Ufer 1  
 68167 Mannheim, Germany  
[christel.weiss@medma.uni-heidelberg.de](mailto:christel.weiss@medma.uni-heidelberg.de)  
<https://www.umm.uni-heidelberg.de/cpd/digitale-gesundheit/biomedizinische-informatik/medizinische-statistik-biomathematik-und-informationsverarbeitung>