

## Kapitel 11

# Datenwissenschaften in der Personalisierten Medizin

Ziel der Personalisierten Medizin ist es, die Wahrscheinlichkeit des individuellen Therapieerfolgs zu erhöhen, das Risiko von unerwünschten Nebenwirkungen zu reduzieren und frühere und genauere Diagnosen und Präventionsmassnahmen zu ermöglichen. Der Schlüssel zu diesen Verbesserungen ist eine hochauflösende Stratifizierung (d. h. eine Einteilung in homogenere Untergruppen) von Individuen. Während in der Vergangenheit Therapieindikationen oft sehr breit gefasst waren («one drug for all»), versucht die Personalisierte Medizin kleinere und spezifischere Patientengruppen zu identifizieren, die dann in gleicher Weise behandelt werden, bis hin zur 1-Personen-Gruppe, also der individualisierten Therapie. Dies gilt für die somatische genauso wie für die psychiatrische Medizin.

Die Stratifizierung basiert auf Biomarkern, also Grössen, die für jede Patientin messbar und hinreichend reproduzierbar sind und die die ursprünglich heterogene Patientengruppe in homogenere Untergruppen separieren. Viele Biomarker stammen aus molekularen Messungen, wie z. B. Genomik, Transkriptomik, Proteomik, Metabolomik usw. und aus bildgebenden Verfahren wie z. B. MRT, CT usw. Daneben gewinnen auch Daten zum Lifestyle, z. B. erfasst durch Wearables, an Gewicht. Von besonderer Bedeutung ist auch die Krankengeschichte der Patientin, die idealerweise in ihrer elektronischen Patientenakte zusammengefasst und zugänglich ist.

Wie findet man nun Untergruppen von Patienten mit ähnlichem Therapieansprechen? Dazu müssen in der Fülle der oben erwähnten Daten Biomarker entdeckt werden, d. h. solche Messungen, die eine Einteilung in homogenere Untergruppen ermöglichen. Ein solcher Biomarker ist z. B. HER2. Brustkrebs-Patientinnen, deren Tumorzellen diesen Rezeptor überexprimieren, können mit Trastuzumab (Herceptin®) behandelt werden, da der Wirkstoff (ein monoklonaler Antikörper) spezifisch gegen diesen Rezeptor wirkt. Wenn der Tumor aber nicht HER2 überexprimiert, dann hat Trastuzumab auch keinen positiven Effekt und kann sogar Schaden verursachen. Die FDA zählt zurzeit etwa 160 solcher pharmakogenetischer Biomarker, die für die Stratifizierung anhand des Therapieansprechens zugelassen sind.

## 11.1. Biomarkersignaturen

Ein einzelner Biomarker ist jedoch oft nicht ausreichend, um eine gute Stratifizierung zu erreichen. Vielen Erkrankungen liegen komplexe molekulare Veränderungen zugrunde, die sich nur durch die Kombination mehrerer Biomarker erfassen lassen. Ein Beispiel einer solchen Biomarkersignatur ist der 2007 von der FDA zugelassene MammaPrint™. Dieser prognostische Brustkrebstest besteht aus 70 Genen, die gemeinsam die Entstehung von Metastasen und den Erfolg einer Chemotherapie vorhersagen. Die Stratifizierung in eine Gruppe niedrigen und eine Gruppe hohen Risikos erfolgt anhand eines Scores, der definiert ist als die gewichtete Summe der Expressionswerte der 70 Gene. Zur vollständigen Beschreibung der Signatur gehören also neben den 70 Genen auch 70 Gewichte und ein Schwellenwert, der die Niedrigrisiko- von der Hochrisikogruppe separiert.

Komplexe Biomarkersignaturen können nur noch mithilfe von Computerprogrammen dargestellt und angewandt werden, da sie eine Vielzahl von Parametern beinhalten und verknüpfen. Methoden der Bioinformatik und der künstlichen Intelligenz erlauben es, Algorithmen und Programme zu entwickeln, die Vorhersagen anhand sehr grosser Datenmengen, z. B. vollständiger molekularer Profile, machen. Dieser Ansatz findet z. B. in molekularen Tumorboards Anwendung. Genomweite DNA- und RNA-Profile werden bereits heute routinemässig für die Diagnostik und Therapieauswahl in der Onkologie herangezogen. Um diese Daten zu nutzen, bedarf es zunächst einer sorgfältigen statistischen Analyse, da Omics-Daten generell mit erheblichen Amplifizierungs- und Sequenzierfehlern behaftet sind. Die normalisierten Profile werden dann integriert und gemeinsam hinsichtlich Therapieoptionen interpretiert. Das Ergebnis wird meist mittels eines Reports von einem Experten in klinischer Bioinformatik an das Tumorboard kommuniziert.

## 11.2. Ableitung prädiktiver Modelle

Mithilfe der Bioinformatik und der Datenwissenschaften lassen sich aus den grossen Datenmengen, die jeden einzelnen Patienten sehr genau charakterisieren, prädiktive Modelle ableiten. Zu diesem Zweck kommen Methoden der Künstlichen Intelligenz, insbesondere des maschinellen Lernens, des Data Minings und der Statistik zum Einsatz. Die Entdeckung komplexer Biomarkersignaturen und -modelle ist aus einer Reihe von Gründen eine sehr schwierige statistische und rechnerische Aufgabe. Zum einen zeichnen sich die zugrunde liegenden Daten bedingt durch Messungenauigkeiten fast immer durch eine

hohe Unvollständigkeit und Fehlerhaftigkeit aus. Zum anderen gibt es in typischen Datensätzen viel mehr beobachtete Werte je Patient als Patienten insgesamt. Während Big Data im Allgemeinen schlicht für grosse Datenmengen steht, ist es in der Medizin meist die hohe Dimension der Daten (z.B. genomweite Messungen, Bilder bestehend aus vielen Pixeln, lange Zeitreihen von Wearables), die das Erkennen von Mustern erschwert. So gibt es z.B. mehr als  $10^{200}$  Möglichkeiten, 70 aus 20 000 Genen auszuwählen.

Maschinelle Lernverfahren sind Methoden, die in solch riesigen Datenmengen ein oder einige wenige besonders vielversprechende Modelle finden. Beim sogenannten überwachten maschinellen Lernen werden funktionale Zusammenhänge zwischen den Prädiktoren, d.h. den beschreibenden Merkmalen, und einer Zielvariablen, also der vorherzusagenden Grösse, gelernt. Relevante Zielvariablen können Klassifizierungen, z.B. in niedriges oder hohes Risiko, kontinuierliche Werte, z.B. die Krankheitsschwere, oder Ereigniszeiten, wie etwa die Überlebenszeit, sein. Der MammaPrint™ ist demnach ein lineares Modell, das Tumore anhand von 70 Genexpressionswerten binär klassifiziert. Beim unüberwachten maschinellen Lernen ist dagegen keine Zielvariable definiert. Beim Clustering wird z.B. versucht, die Beobachtungen anhand ihrer Merkmale so zu gruppieren, dass homogenere Teilgruppen entstehen. Beide Klassen maschineller Lernverfahren finden häufig Anwendung in der Medizin zur Identifizierung von Patientenuntergruppen (Stratifizierung) und zur Konstruktion prädiktiver Modelle des Therapieansprechens (Vorhersage).

### 11.3. Validierung

Wenn statistische und maschinelle Lernmodelle für Diagnostik, Prognostik und die Vorhersage von Therapieerfolg benutzt werden, dann kommt der Korrektheit und Genauigkeit der Vorhersagen grosse Bedeutung zu. Um die Güte der Vorhersage oder Stratifizierung abzuschätzen, reicht es nicht aus, diese auf der Basis desselben Datensatzes zu erheben, der für die Konstruktion des Modells verwandt wurde. Es ist bekannt, dass dieser Schätzer überoptimistisch ist. Mit anderen Worten, er wird einen kleineren Fehler suggerieren, als tatsächlich auf neuen, zuvor ungesehenen Daten zu erwarten ist. Eine solide Validierung von klinischen Modellen umfasst deshalb folgende Schritte: (1) eine interne Validierung auf dem ursprünglichen Datensatz mittels Kreuzvalidierung, bei der sukzessive ein Teil der Daten zum Schätzen der Vorhersagegenauigkeit zurückgehalten und das Modell auf den verbleibenden Daten gelernt wird; (2) eine externe Validierung in einer unabhängigen Kohorte, um den möglichen Bias bei der Zusammenstellung

der ursprünglichen Entdeckungskohorte zu eliminieren; und (3) Validierung in einer prospektiven klinischen Studie im Vergleich zum «Standard of Care». Der gesamte Prozess ist sehr zeit- und kostenintensiv, und es überrascht daher nicht, dass es bisher nur wenige klinisch validierte Modelle gibt.

#### 11.4. Infrastrukturelle Voraussetzungen

Bioinformatik und Datenwissenschaften können nur dann für die Medizin nützliche Modelle generieren, wenn die zugrunde liegenden Daten in der nötigen Quantität und Qualität zur Verfügung stehen. Im Hinblick darauf bedarf es national und international einer Harmonisierung der unterschiedlichen Datentypen und Informationssysteme. Grosser Handlungsbedarf besteht auch bei der Datenorganisation, d. h. beim Erfassen und Aufbereiten von Patientendaten, von biologischen Basisdaten (einschliesslich Biodatenbanken) und Omics-Daten sowie bei strukturierten Prozessen. Diese sollen eine nationale Interoperabilität von verschlüsselten Daten aus den Spitälern und Forschungsstrukturen gewährleisten, und zwar unter Beachtung von gesetzlichen Vorgaben, optimaler Datensicherheit und Datenschutzbedingungen.

In der Schweiz haben diese Herausforderungen den Bund dazu veranlasst, die SAMW 2016 mit dem Aufbau des «Swiss Personalized Health Network» (SPHN) zu beauftragen. Mit dem SPHN wird die Infrastruktur aufgebaut, um die vielfältigen gesundheitsbezogenen Daten für die Forschung und Innovation nutzbar zu machen.

Für den Zeitraum 2017–2020 liegt der Fokus beim Aufbau einer national koordinierten Dateninfrastruktur. Damit können lokale und regionale Informationssysteme harmonisiert und die Dateninteroperabilität gewährleistet werden. Ein weiterer Schwerpunkt liegt bei den klinischen Datenmanagementsystemen, die den sicheren Austausch von Patientendaten (z. B. Krankheitsphänotypen) ermöglichen sollen. Langfristig ist auch die Integration von öffentlichen Gesundheitsdaten, Kohortenstudien und Daten von gesunden Bürgern denkbar.

## 11.5. Grenzen der Datenwissenschaften

Auch wenn bereits heute sehr mächtige datenwissenschaftliche Werkzeuge zur Verfügung stehen, auf denen ganze Wirtschaftszweige basieren, so bleiben speziell für die biomedizinischen Datenwissenschaften noch viele Fragen offen. Zum Beispiel ist in medizinischen Anwendungen neben der Güte der Vorhersage oder Stratifizierung auch die Interpretierbarkeit der Modelle von grossem Interesse, da sie die Akzeptanz von modellbasierten Vorhersagen erhöht. Die Methoden des maschinellen Lernens sind zwar gut geeignet, komplexe Muster in grossen Datenmengen zu finden und möglichst genaue Vorhersagen zu machen, aber sie können keine mechanistischen oder kausalen Zusammenhänge aufklären. Die Verfahren basieren meist auf der Entdeckung von Korrelationen, aber Korrelationen implizieren im Allgemeinen keine Kausalität. Auf der Suche nach Biomarkersignaturen führt dies z. B. häufig dazu, dass eine Vielzahl von verschiedenen Signaturen die gleiche optimale Vorhersagegüte erreichen. Nach Entdeckung des MammaPrint™ wurden beispielsweise viele weitere 70-Gen-Signaturen von gleicher Qualität gefunden. Die Modelle, die maschinelle Lernverfahren finden, sollten daher stets als Hypothesen angesehen werden, die anschliessend mit anderen wissenschaftlichen Methoden (z. B. Randomized Controlled Trials, Longitudinal Observational Studies) überprüft werden müssen.

Die Personalisierte Medizin ist ohne die Bioinformatik und die Datenwissenschaften nicht denkbar. Grosse Genom-, Bild- und Lifestyle-Datensätze sowie Fortschritte in der Künstlichen Intelligenz geben Anlass zur Hoffnung, dass auf diese Weise auch medizinischer Fortschritt langfristig und nachhaltig generiert werden kann, der zu besseren Therapien und niedrigeren Kosten führt. Dennoch ist die Liste der Erfolgsgeschichten bisher überschaubar. Es ist wahrscheinlich, dass dies zumindest zum Teil eine Konsequenz der molekularen Komplexität vieler Krankheiten ist und auch damit zusammenhängt, dass selbst moderne Hochdurchsatzverfahren von diesen Systemen nur sehr beschränkte Beobachtungen erlauben.



Schweizerische Akademie der Medizinischen Wissenschaften  
Académie Suisse des Sciences Médicales  
Accademia Svizzera delle Scienze Mediche  
Swiss Academy of Medical Sciences

### Herausgeberin

Schweizerische Akademie der Medizinischen Wissenschaften  
Haus der Akademien, Laupenstrasse 7, CH-3001 Bern  
mail@samw.ch, www.samw.ch

### Gestaltung

Howald Fosco Biberstein, Basel

### Übersetzung

Apostroph, Bern

### Umschlagbild

adobestock – joyt; istock – teekid

Deutsche und französische Version (pdf) auf [samw.ch/grundlagen-personalisierte-medizin](http://samw.ch/grundlagen-personalisierte-medizin)



Copyright: ©2019 Schweizerische Akademie der Medizinischen Wissenschaften  
Dies ist eine Open-Access-Publikation, lizenziert unter «Creative Commons Attribution» (<http://creativecommons.org/licenses/by/4.0>). Der Inhalt dieser Publikation darf uneingeschränkt und in allen Formen genutzt, geteilt und wiedergegeben werden, solange der Urheber und die Quelle angemessen angegeben werden.

### Zitiervorschlag:

Schweizerische Akademie der Medizinischen Wissenschaften (2019)  
Personalisierte Medizin. Grundlagen für die interprofessionelle Aus-, Weiter- und Fortbildung von Gesundheitsfachleuten.  
Swiss Academies Communications 14 (6).

ISSN (online): 2297-1807

DOI: <http://doi.org/10.5281/zenodo.3265194>



### SDGs: Die internationalen Nachhaltigkeitsziele der UNO

Mit dieser Publikation leistet die Schweizerische Akademie der Medizinischen Wissenschaften einen Beitrag zu SDG 3:  
«Ein gesundes Leben für alle Menschen jeden Alters gewährleisten und ihr Wohlergehen fördern»

[sustainabledevelopment.un.org](http://sustainabledevelopment.un.org)  
[www.eda.admin.ch/agenda2030](http://www.eda.admin.ch/agenda2030) → agenda 2030 →  
→ 17 Ziele für nachhaltige Entwicklung