

## Leitprojekt MED<sup>2</sup>ICIN

---

# Virtuelle Kohorten als entwicklungsbegleitendes Tool für den Digitalen Zwilling

### Virtuelle Kohorten:

---

- stellen wertvolle Trainingsdatensets in der digitalen Gesundheitsforschung dar
- überbrücken die Zeiträume der Akquisition realer medizinischer Daten
- sind wertvoller Baustein in der medizinischen IT-Prototypentwicklung

### Hintergrund

Die Nutzung von Gesundheitsdaten in Deutschland unterliegt vielfältigen ethischen bzw. juristischen Einschränkungen aus Gründen des Datenschutzes, der Eigentumsrechte oder der Regulatorik und Gesetzgebung. Der Zugang zu »humanen Daten« sowohl einzelner Patientinnen und Patienten als auch vollständiger Studien bedarf daher vielfacher Freigabeprozesse, wie z.B. positiver Ethikvoten oder Risikofolgeabschätzungen und dem Abschluss entsprechender Datenzugangs- oder Datennutzungsverträge. Regulatorischer und zeitlicher Aufwand der Genehmigungsvorgänge zur Datenakquisition schränken technologische Entwicklungen in der Medizininformatik und ihren

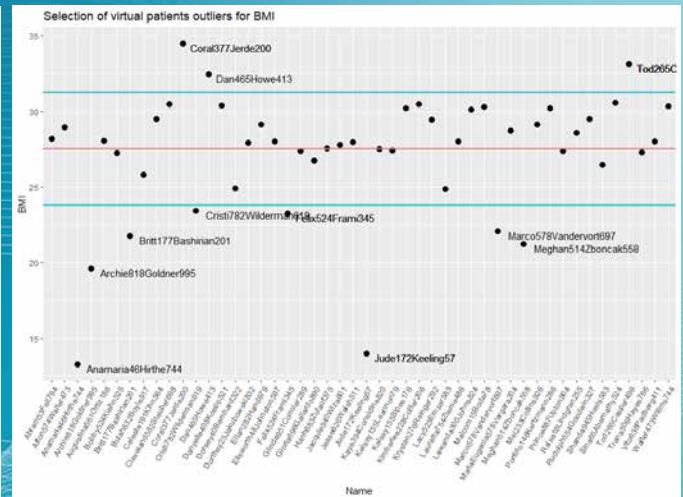
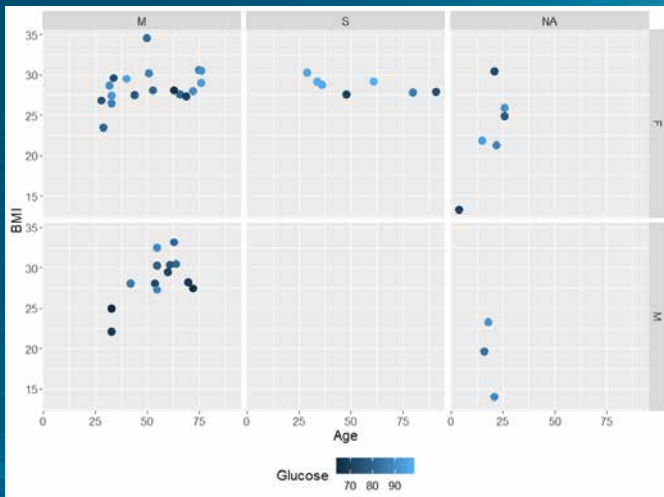
Einsatz wie im Falle des Digitalen Zwillings ein. Zuverlässige und qualitativ hochwertige synthetische Patientendaten könnten daher eine Lösung darstellen, die es ermöglicht vorerst auch ohne »Realdaten« Entwicklungsvorhaben voranzutreiben.

### Synthetische Daten als »Workaround«

Der Anspruch besteht darin, Datensätze mit medizinischen Informationen zu generieren, die in Inhalt und Format Abbild realer klinischer Daten sind sowie auch in statistischer Hinsicht die Kohorte valide beschreiben, nicht aber in Zusammenhang mit realen Personen stehen und daher die Nutzung aus rechtlicher Sicht deutlich vereinfachen. Die Erstellung und Verwendung virtueller Kohorten kann dann ein wertvoller Baustein für die Entwicklung neuer medizinischer Analysen und Modelle sein. Das ITMP verfügt über ein Portfolio von proprietären Modellen und hat mit selbst entwickelten und Open-Source Tools eine Bibliothek von virtuellen Kohorten aufgesetzt.

### Der Nutzen synthetischer Patientendaten

Die Verwendung synthetischer Daten stößt auf immer größeres Interesse in Forschungs- und Entwicklungsvorhaben, in denen einerseits medizinische Daten benötigt werden, es aber andererseits schwierig ist, diese zu erheben, Zugriff auf bestehende



Daten zu erhalten oder umfangreiche Datensätze für eine spezielle Fragestellung zu finden. Die Datenwissenschaft hat in den letzten Jahrzehnten daher quasi ein eigenes Forschungsfeld über synthetische (virtuelle) Daten geschaffen. Klinische Daten stellen dabei keine Ausnahme dar. Die inhärent komplexe Natur klinischer Daten und ihre nahezu unüberschaubare zeitabhängige Dimension zwingt die Forschenden dazu, Tools zu entwickeln, mit denen virtuelle Populationen (virtuelle Kohorten [VC]) generiert werden können und mittels derer Aspekte der modernen Datenwissenschaft getestet werden konnten:

- Datenprovenienz
- Extraktion und/oder Erzeugung von Metadaten
- Datenkuration
- Datenaggregation
- Deskriptive Analytik
- Prädiktive Analytik (Modellierung)
- Inferenz

### Virtuelle Kohorten in MED<sup>2</sup>ICIN

Um die medizinische und ökonomische Modellierung in Erwartung der realen Daten zu unterstützen, stellt eine virtuelle Kohorte für den Digitalen Zwillingen in MED<sup>2</sup>ICIN ein interessantes entwicklungsbegleitendes Tool dar. Aus diesem Grund wurde am ITMP das Open-Source-Tool SYNTHEA (<https://github.com/synthetichealth/synthea>) verwendet, um beispielhaft eine virtuelle Kolorektalkarzinom-Kohorte mit 5.000 Datensätzen zu erzeugen. Es ist von Vorteil, dass SYNTHEA Populationen generiert (nicht nur Erkrankte), aus der dann Betroffene für eine Kohorte ausgewählt werden können. Außerdem basiert SYNTHEA auf einem longitudinalen Datensatz, der für die jeweilige Krankheit angepasst werden kann. Das Herzstück des Programms sind Krankheitsmodule, die bei der Generierung der Population berücksichtigt werden. Das jeweilige Krankheitsmodul gibt die Zeitpunkte, den Zustand der Erkrankten und die Erfolgsrate der Behandlung über den gesamten zeitlichen Verlauf vor. Es bietet verschiedene

Pfade basierend auf Entscheidungen (entweder Behandlungs- oder Diagnoseereignisse oder auftreten anderer Krankheiten aus anderen Modulen) oder Gruppierungen/Klassifikationen (z.B. Alter, Geschlecht, Rauchen, Schlaf, Sport). Die gewählten Pfade beeinflussen schlussendlich die Parameterausgabe.

### Das Fraunhofer ITMP im Leitprojekt MED<sup>2</sup>ICIN

Das Fraunhofer ITMP positioniert sich entlang der pharmazeutischen Wertschöpfungskette von der Target-Identifizierung über das Wirkstoffscreening, die translationale präklinische Validierung bis hin zu klinischen Prüfungen, im 4D-Prinzip. Ein Schwerpunkt liegt dabei auf der Untersuchung immun-mediierter Erkrankungen. Durch jahrelange Erfahrung und enge Zusammenarbeit mit Klinikpersonal bietet das ITMP ein hochspezialisiertes Spektrum an Krankheitsmodellen, Formulierung von Therapiepfaden und Bioanalytik. Klinische Fragestellung und Prüfungen werden bei uns in mit einem breiten Netzwerk von Industriepartnern erarbeitet und nach dem Standard »Quality by Design« durchgeführt.

#### Fraunhofer-Institut für Translationale Medizin und Pharmakologie ITMP

##### Ansprechpartner

Andrea Zaliani  
 Innovationsbereich Medical Data Science  
[Andrea.Zaliani@itmp.fraunhofer.de](mailto:Andrea.Zaliani@itmp.fraunhofer.de)

Schnackenburgallee 114  
 22525 Hamburg



[s.fhg.de/med2icin](https://s.fhg.de/med2icin)