Krebsforschungsdatenzentrum|onkoFDZ

KI-gestützte Evidenzgenerierung aus versorgungsnahen Daten Klinischer Krebsregister, GKV-Routinedaten, Klinikdaten und deren Linkage

Monika Klinkhammer-Schalke[2,3], Olaf Schoffer[1], Stefan Rolf Benz[4], Vinzenz Völkel[3], Judith Hansinger[3], Johannes Soff[4], Tobias Kussel[5], Kees Kleihues-van Tol[2], Bianca Franke[2], Jacqueline Witzmann[2], Melanie Börries[6], Jörg Janne Vehreschild[7], Christophe Kowalski[9], Michael Gerken[3], Fabian Baum[1], Daniel Maier[7], Fabienne Fox[7], Martin Lablans[5], Soo-Zin Kim-Wanner[8], Jochen Schmitt[1]

Einleitung

Das im Jahr 2021 verabschiedete Gesetz zur Zusammenführung von Krebsregisterdaten definiert die Rahmenbedingungen für eine bundesweite Synthese onkologischer Behandlungsdaten. Seither erarbeiten führende Organisationen und Vertretende der Krebsregistrierung gemeinsam eine Plattformlösung für eine bundesweite, anlassbezogene Datenzusammenführung und Verknüpfung von verschiedenen Datenquellen. Für dieses Konzept ist das Projekt Krebsforschungsdatenzentrum (onkoFDZ) von entscheidender Bedeutung: obwohl randomisierte kontrollierte Studien (RCTs) als der Goldstandard in der medizinischen Forschung gelten [1], sind sie nicht für alle Fragestellungen gleichermaßen geeignet. Deswegen sind Leitlinienprogramme bei spezifischen Fragestellungen lediglich limitiert fähig, Empfehlungen von hohem Evidenzgrad [2] zu formulieren.

Methodische Umsetzung

Methoden:

- anlassbezogene neuartige Linkage Methoden [4, 5, 6] komplementärer Datenbasen (KKR, GKV, CCC, DKTK, Oncobox Research)
- Brückenkopfmodell
- Datenlinkage mittels SMPC
- umfassende Datenharmonisierung
- Nutzung moderner Verfahren des maschinellen Lernens und der Statistik

Datenbasis:

- integrierte versorgungsnahe Daten

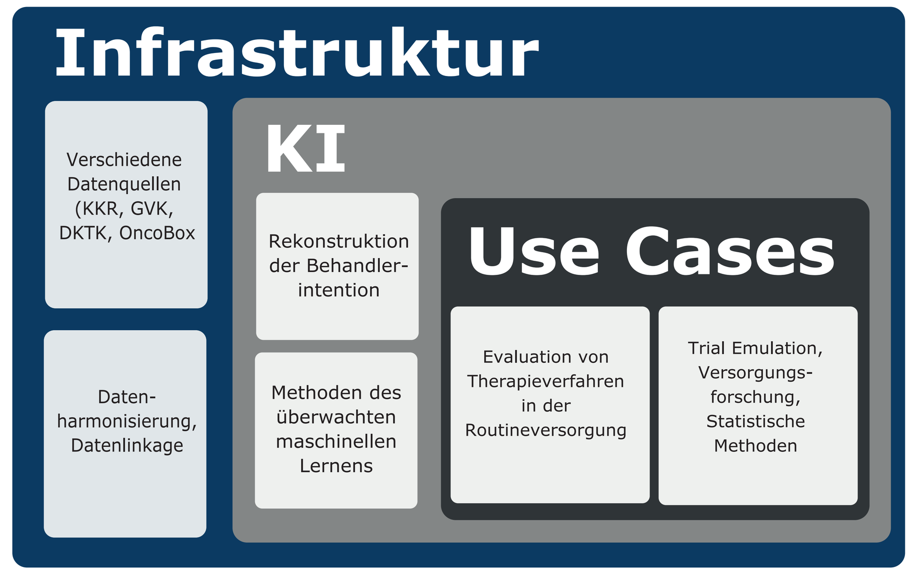

Infrastruktur
Verschiedene Datenquellen (KKR, GVK, DKTK, OncoBox
Daten-harmonisierung, Datenlinkage
KI
Rekonstruktion der Behandler-intention
Methoden des überwachten maschinellen Lernens
Use Cases
Evaluation von Therapieverfahren in der Routineversorgung
Trial Emulation, Versorgungs-forschung, Statistische Methoden

Abbildung: Übersicht Projektaufbau

Interdisziplinäre Einbeziehung von Fachexpertise:
Die klinischen, juristischen und methodischen Expertisen werden in diesem Projekt eng verzahnt.

Projektbeteiligte:

Kooperationspartner:

Anlassbezogenes Datenlinkage

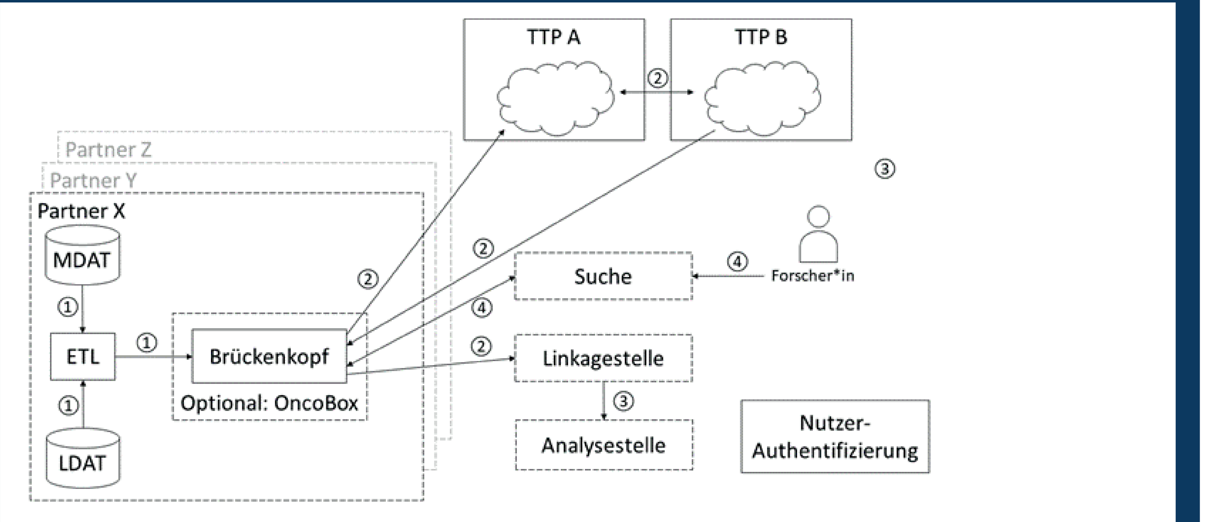
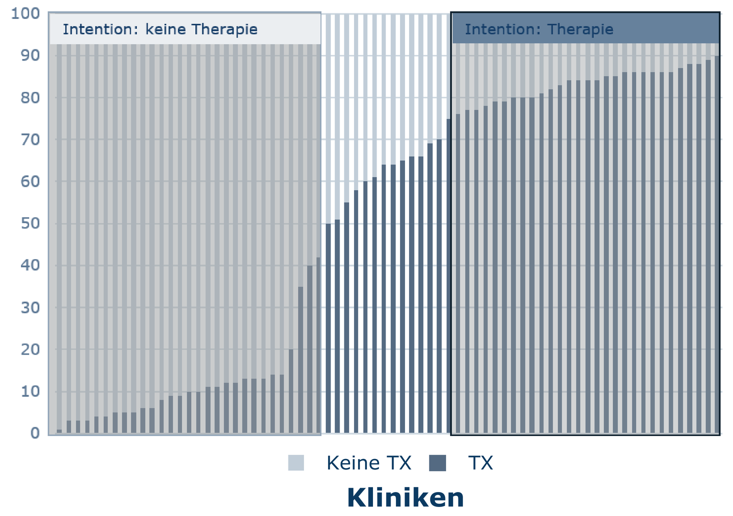

TTP A
TTP B
Partner Z
Partner Y
Partner X
MDAT
ETL
LDAT
Brückenkopf
Optional: OncoBox
Suche
Forscher*in
Linkagestelle
Analysestelle
Nutzer-Authentifizierung

Abbildung: Schematische Darstellung eines möglichen konzipierten Pseudonymisierungsablaufs (Brückenkopfmodell)

Ziele

Analysen umfassender, versorgungsnaher Daten können in solchen Fällen die Möglichkeit bieten, hochwertige Evidenz ergänzend zu RCTs zu generieren und Leitlinienlücken zu schließen. Daher liegt das übergeordnete Ziel im Projekt onkoFDZ in der Entwicklung einer IT-Infrastruktur, die eine sichere Zusammenführung der Daten aus unterschiedlichen Quellen realisieren lässt. Die entstandene Struktur soll anhand des Use Cases „Therapie des Darmkrebses" erprobt werden. Dabei werden klinische Fragestellungen fokussiert, zu denen die aktuelle Leitlinie Kolorektales Karzinom [3] keine oder keine eindeutige Empfehlung gibt.

Einführung des Prinzips der Behandlerpräferenz in die Registerforschung


Intention: keine Therapie
Intention: Therapie
0
10
20
30
40
50
60
70
80
90
100
Keine TX
TX
Kliniken

Behandlerintention wird aus den dokumentierten Behandlerentscheidungen rekonstruiert

Geplante Arbeitsschritte KI:

- Datenaufbereitung und Feature Engineering (Identifizierung von Einflussfaktoren)
- Simulation der Behandlerpräferenz mit Maschinellen Lern-Verfahren zur Minimierung des confounding by indication Problems → Ableitung von Propensity-Weights (Gewichtung von Einflussfaktoren)
- Datenauswertung mit Maschinellem Lernen und konventionellen statistischen Methoden → Abgleich der Ergebnisse

Datenauswertung für Use Case → Kolonkarzinom

- Trial Emulation unter Berücksichtigung der Propensity-Weights
- Analyse folgender (und weiterer) Beispielfragestellungen

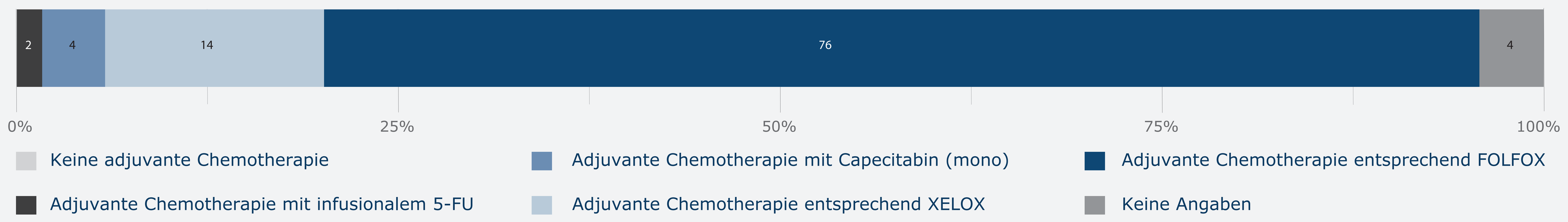
a. Adjuvante Chemotherapie > 75a beim Kolon-CA
b. Adjuvante Chemotherapie in UICC Stadium II
c. Adjuvante Chemotherapie Rektumkarzinom nach neoadjuvanter Vorbehandlung
d. Wahl der systemischen Therapie in Abhängigkeit von der molekularpathologischen Subgruppe und der Tumorlokalisation
e. Laparoskopische Operation und robotische Chirurgie beim Kolon- und Rektumkarzinom
f. Spättherapielinien/ Rekonstruktion von Therapielinien

Fallvignettenstudie Behandlerpräferenz an DKG-zertifizierten Zentren

314 befragte zertifizierte Zentren, Rücklaufquote 37% (N = 116)

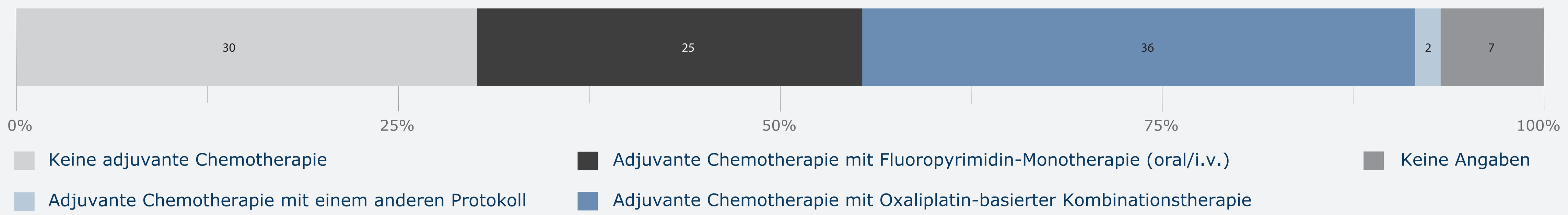
Adjuvante Chemotherapie Kolonkarzinom in UICC Stadium III

Kolonkarzinom der linken Flexur, postoperatives Stadium III (pT3, N2 (4/28) L0 V1 R0, G3), ECOG 0, Alter 60


2
4
14
76
4
0%
25%
50%
75%
100%
Keine adjuvante Chemotherapie
Adjuvante Chemotherapie mit Capecitabin (mono)
Adjuvante Chemotherapie entsprechend FOLFOX
Adjuvante Chemotherapie mit infusionalem 5-FU
Adjuvante Chemotherapie entsprechend XELOX
Keine Angaben

Adjuvante Chemotherapie Kolonkarzinom in UICC Stadium II

Kolonkarzinom UICC Stadium II pT4a N0 (0/25) L0 V0 Pn0 G2 R0, MSS, ECOG 0, Alter 68


30
25
36
2
7
0%
25%
50%
75%
100%
Keine adjuvante Chemotherapie
Adjuvante Chemotherapie mit Fluoropyrimidin-Monotherapie (oral/i.v.)
Keine Angaben
Adjuvante Chemotherapie mit einem anderen Protokoll
Adjuvante Chemotherapie mit Oxaliplatin-basierter Kombinationstherapie

Fazit:
Die Heterogenität ist überraschend, da sie keine Folge von unterschiedlichen Patientencharakteristika, sondern das Resultat unterschiedlicher ärztlicher Präferenzen, bzw. Folge einer unterschiedlichen Interpretation der Evidenz ist. Die gezeigten Unterschiede haben für die Analyse von versorgungsnahen Daten weitreichende Konsequenzen:

- Je höher die präferenz-basierte Heterogenität ist, desto höher kann auch die Validität bei Vergleichen eingeschätzt werden (hohe Zufallskomponente in der Therapiewahl).
- Eine hohe präferenz-basierte Heterogenität könnte in Zukunft die retrospektive Konstruktion von gut vergleichbaren Kollektiven für die Analyse von Effektivitätsuntersuchung ermöglichen. **Diese Möglichkeit wird derzeit im onkoFDZ-Projekt evaluiert.**

Verwertung

erwartete Ergebnisse:
In den einzelnen Arbeitsgruppen wurden die Konzepte und Studiendokumente unter Berücksichtigung datenschutzrechtlicher Anforderungen erarbeitet und befinden sich derzeit in einem finalen Abgleichsprozess. Diese sollen eine Grundlage zur Steuerung der Datenflüsse sowie einen Rahmen für die Datenauswertung bilden. Ein erfolgreicher IT-Infrastrukturaufbau ermöglicht die Nutzung einer umfassenden Datenbasis, die erforderlich ist, um mithilfe von ML-Verfahren mögliches Confounding zu kontrollieren und die primäre Behandlerintention auf Patientenebene zu approximieren.

Mehrwert für Wissenschaft und Forschung
Nachhaltigkeitsansatz:

- Implementierung als Toolkit
- open-source lizensiert und unter Creative Commons dokumentiert
- Konzeption eines Use&Access-Verfahrens für wissenschaftliche Nutzung
- Einbindung in das „Ökosystem" der verschiedenen Forschungsdatenzentren in Deutschland
- offene Einbindung in die wissenschaftliche Community

Diskussion:
Eine Zusammenführung der heterogenen Daten aus unterschiedlichen Quellen gestaltet sich in zweierlei Hinsicht herausfordernd:
(1) Sie unterliegt strengen datenschutzrechtlichen Anforderungen
(2) sie erfordert eine umfassende Harmonisierung.

Zusammenfassung

Die im Projekt Krebsforschungsdatenzentrum (onkoFDZ) entwickelte IT-Infrastruktur soll dazu beitragen, eine Basis für die Zusammenführung versorgungsnaher Daten unterschiedlicher Quellen (Klinische Krebsregister (KKR), zertifizierte Zentren der Deutschen Krebsgesellschaft, onkologische Spitzenzentren und gesetzliche Krankenkassen) zu schaffen. Es soll eine nachhaltige und permanente Infrastruktur implementiert werden, die der wissenschaftlichen Community zugänglich gemacht wird, um eine KI-basierte Generierung hochwertiger Evidenz aus versorgungsnahen Daten zu ermöglichen.

Literatur:

[1] IQWIG. Konzepte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V. Rapid Report A19-43, Vers. 1.1 vom 13.05.2020.

[2] Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. Plast Reconstr Surg. 2011 Jul; 128(1):305-310. doi: 10.1097/PRS.0b013e318219c171. PMID: 21701348; PMCID: PMC3124652.

[3] Leitlinienprogramm Onkologie (DKG, DKH, AWMF): S3-Leitlinie Kolorektales Karzinom, Kurzversion 2.1, 2019, AWMF Registrierungsnummer:021/007OL, http://www.leitlinienprogramm-onkologie.de/leitlinien/kolorektales-karzinom/ [abgerufen am: 10.05.2023]

[4] Stammler S, Kussel T, Schoppmann P, Stampe F, Tremper G, Katzenbeisser S, Hamacher K, Lablans M. Maizelliste SecureEpiLinker (MainSEL): privacy-preserving record linkage using secure multi-party computation. Bioinformatics 2020; 38(6):1657-1668. doi: 10.1093/bioinformatics/btaa764. PMID: 32871006 PMCID: PMC8896632

[5] Lablans M, Schmidt E, Ückert F. An Architecture for Translational Cancer Research As Exemplified by the German Cancer Consortium. JCO Clin Cancer Inform. 2018 Dec;2:1-8. doi: 10.1200/CCI.17.00062. PMID: 30652543

[6] Bobeth C, Tol KK, Rößler M, Bierbaum V, Gerken M, Günster C, Dröge P, Ruhnke T, Klinkhammer-Schalke M, Schmitt J, Schoffer O. Methodik und Zuordnungserfolg eines Linkage von Daten klinischer Krebsregister mit Abrechnungsdaten gesetzlicher Krankenkassen [Methodology and Attribution Success of a Data Linkage of Clinical Registry Data with Health Insurance Data]. Gesundheitswesen. 2023 Mar;85(S 02):S154-S161. German. doi: 10.1055/a-1984-0085. Epub 2023 Mar 20. PMID: 36940697. https://www.thieme-connect.com/products/ejournals/html/10.1055/a-1984-0085

Institutionen:

1. Zentrum für Evidenzbasierte Gesundheitsversorgung (ZEGV), Dresden
2. Arbeitsgemeinschaft Deutscher Tumorzentren e.V. (ADT), Berlin
3. Tumorzentrum – Zentrum für Qualitätssicherung und Versorgungsforschung an der Fakultät für Medizin der Universität Regensburg, Regensburg
4. Arbeitsgemeinschaft Deutscher Darmkrebszentren (addz e.V.), Hamburg
5. Deutsches Krebsforschungszentrum (DKFZ), Heidelberg
6. Albert-Ludwigs-Universität, Institut für Medizinische Bioinformation und Systemmedizin, Freiburg
7. Johann Wolfgang Goethe-Universität, Medizinische Klinik II, Frankfurt am Main
8. HLfGP, Abt. IV Hessisches Krebsregister, Frankfurt am Main
9. Deutsche Krebsgesellschaft (DKG)

gefördert durch:

das Bundesministerium für Gesundheit im Handlungsfeld „Digitalisierung", Forschungsschwerpunkt „Krebsregisterdaten zusammenführen und intelligent nutzen"

(**Förderkennzeichen: ZMI5-2522DAT14A-O**)

Projektvorstellung auf der Ministeriumshomepage: https://www.bundesgesundheitsministerium.de/ministerium/ressortforschung-1/handlungsfelder/forschungsschwerpunkte/krebsregisterdaten/onkofdz.html

Projektlaufzeit: 01.09.2022 bis 31.08.2025

Projektleitung: Jochen Schmitt, Monika Klinkhammer-Schalke

Projektkoordination: Olaf Schoffer, Bianca Franke

Gefördert durch:
Bundesministerium für Gesundheit
aufgrund eines Beschlusses des Deutschen Bundestages