

# DIFUTURE Use Case Parkinson's Disease

Studien Outcomes / Datenlieferung / Lessons  
learned

**Benjamin Sailer, Christoph Kiefer**  
21.11.23

© UNIVERSITÄTSKLINIKUM TÜBINGEN.

EBERHARD KARLS  
**UNIVERSITÄT  
TÜBINGEN**



**Universitätsklinikum  
Tübingen**

# Data Integration for Future Medicine (DIFUTURE)

- “consortium **Data Integration for Future Medicine (DIFUTURE)** will establish Data Integration Centers (DICs) at university medical centers”
- “approach is disease-oriented and use-case driven, i.e. following the needs of physicians and researchers and aiming at measurable benefits for our patients”
- “DIFUTURE will implement several disease-oriented use cases in a step-wise manner”
  - “first use case focuses on the development of an infrastructure for personalized optimal treatment of **Multiple Sclerosis (MS)**”
  - “use case on Parkinson’s Disease aims at the [...] infrastructure supporting research on early diagnosis of the different subtypes of **Parkinson’s disease (PD)**”
- “The early use cases will serve as blueprints, which show that our infrastructure is able to support a wide variety of application scenarios”



# LOC-PD (1)

- “Longitudinal cohort study of patients with Parkinson’s disease (LOC-PD) in early and late disease stages”
- Study objectives:
  - Two-fold:
    - “evaluate the predictive value of a single or a combination of markers [...] after 36 months in early PD”  
-> LOC-EARLY
    - “define predictors for a positive treatment outcome after deep brain stimulation in later PD stages”  
-> LOC-DBS
  - Non-clinical objective:
    - “To evaluate the IT-infrastructure of the DIFUTURE consortium in order to assess the principal feasibility of data capture and sharing techniques developed”



# LOC-PD (2) - Timeline

	Tasks	2018		2019				2020				2021				2022				2023				2024					
Study Phase		quarter (proposed trial duration)																											
		Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4		
Planning	Writing of study protocol																												
	Ethic approval letter																												
Execution	Screening and inclusion period																												
	12 months follow-up visit (LOC-EARLY 12)																												
	24 months follow-up visit (LOC-EARLY 24)																												
	23 months follow-up visit (LOC-EARLY 36)																												
Analysis	Interim data analysis																												
	Final data analysis																												
	preparation of manuscript																												

End of funding



# Datenerhebung

- Klin. Daten
  - Demographische Daten
  - Symptome
  - Untersuchungen
    - Hoehn & Yahr
    - UMSARS
    - ...
  - Fragebögen
    - MoCA
    - BDI-II
    - UPDRS
    - ...
- Einwilligung



# Datenerhebung

- Klin. Daten
  - Demographische Daten
  - Symptome
  - Untersuchungen
    - Hoehn & Yahr
    - UMSARS
    - ...
  - Fragebögen
    - MoCA
    - BDI-II
    - UPDRS
    - ...
- Einwilligung
- Laboruntersuchungen
  - Blut (Plasma/Serum)
  - Stuhl
  - ...
- Bildgebung
  - Zerebrales MRT
- Medikation
- Sonstiges
  - für Tiefe Hirnstimulation
    - Stimulationsparameter
    - Medikationspumpeneinstellungen
  - Bioproben



# Datenerhebung - Systeme (1)

- Klin. Daten

- Demographische Daten
- Symptome
- Untersuchungen
  - Hoehn & Yahr
  - UMSARS
  - ...
- Fragebögen
  - MoCA
  - BDI-II
  - UPDRS
  - ...

SAP

- Einwilligung

- Laboruntersuchungen

- Blut (Plasma/Serum)
- Stuhl
- ...

Swisslab

- Bildgebung

- Zerebrales MRT

RIS

- Medikation

- Sonstiges

REDCap

- für Tiefe Hirnstimulation
  - Stimulationsparameter
  - Medikationspumpen-einstellungen
- Bioproben

SAP

Studien-  
system DIS



# Datenerhebung - Systeme (2)

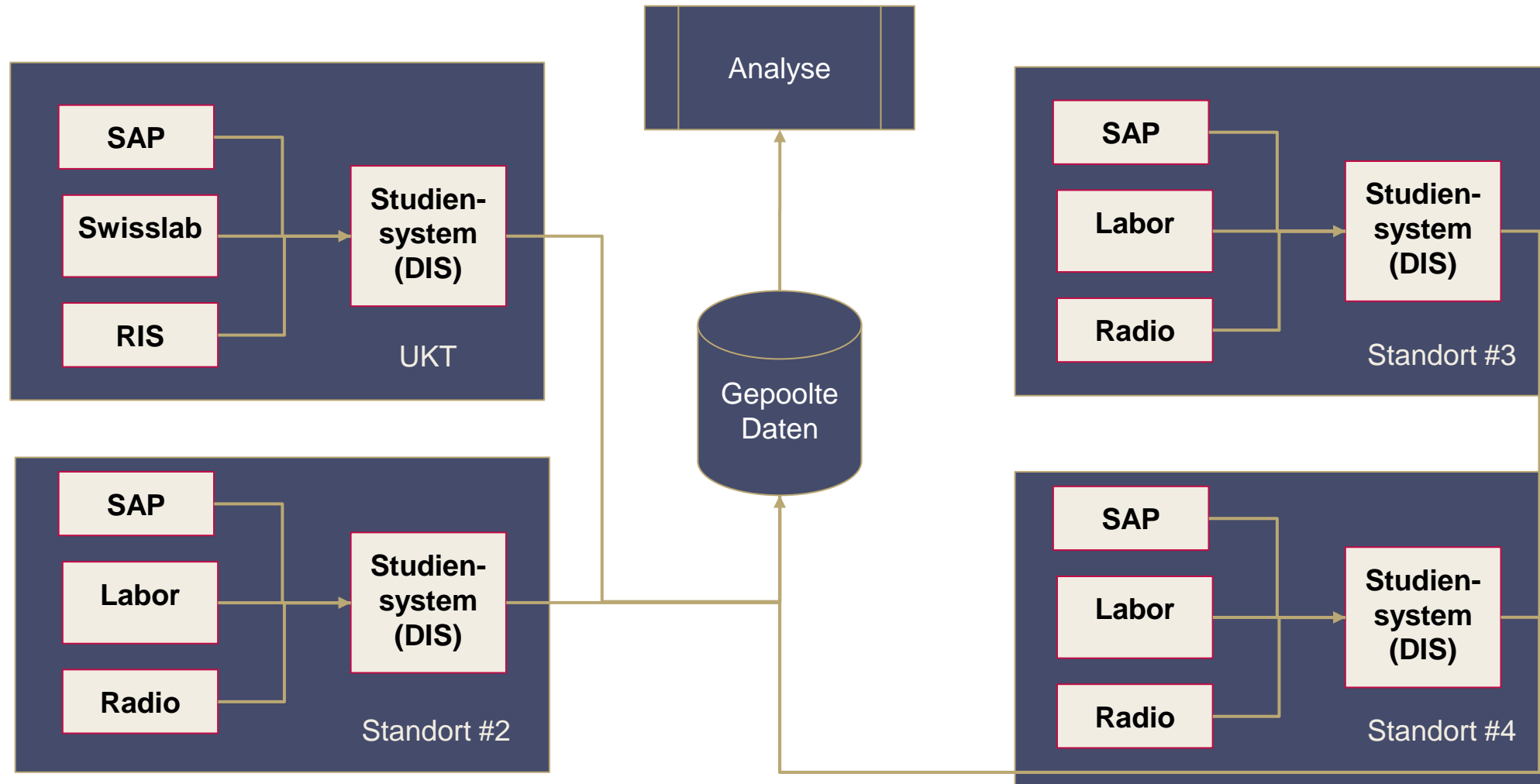
- SAP i.s.h.med.:
  - Klinisches Informationssystem am UKT
  - Daten werden händisch in **P**arametrierbare **M**ed. **D**okumentation (PMD) eingetragen
  - Eine Visite = ein “Haupt”-PMD + mehrere “Unter”-PMDs
- Swisslab:
  - Laborinformationssystem am UKT
  - Befunde enthalten Patienten-ID und Fallnummer (eine Visite = ein Fall)
- RIS:
  - **R**adiologie**i**nformationssystem (RIS)
  - Befundung (semi-)strukturiert über vorgefertigte Textbausteine





# IT-Infrastruktur

DIS: über alle Standorte harmonisiert  
RIS: Radiologieinformationssystem



# Datenbereitstellung - Umsetzung

- SAP i.s.h.med. (klin. Daten):
  - Täglicher Export aus dem SAP von ausgewählten PMDs (als XML)
  - Verarbeitung auf dezidiertem VM
  - Übergabe an Studiensysteme
- Swisslab (Labor):
  - Tägliche Queries an Swisslab Datenbank basierend auf Fall-ID + Analytodelist
  - Zuordnung im Studiensystem via Fall-ID
- RIS (Bildgebung):
  - Manueller Prozess (in Arbeit)
- REDCap (Medikation):
  - Tägliche Abfragen von Studiensystem für Studienpatienten (full-dump)
  - Zuordnung über Patienten-ID und Visitendatum



# Datenqualität (1)

- Versch. Lösungen für die Sicherstellung der Datenqualität
  - Eingabeprüfungen in Primärsystemen
  - Diverse Python und SQL Skripte/Queries für spezifische Teilbereiche (Labordaten, PMDs, etc.)
  - (Semi-)Automatische Datenqualitätsberichte
    - Idee: monatliche Reports für Kliniker als Alternative zum Monitoring
    - RMarkdown Skript, was auf Export der Studiendaten ausgeführt wird
    - Ergebnis: eine HTML Datei, die Probleme in den Daten darstellt (inkl. Feldposition im Primärsystem)



# Datenqualität (2)

Auszug aus automatischem  
Datenqualitätsbericht mit Testdaten

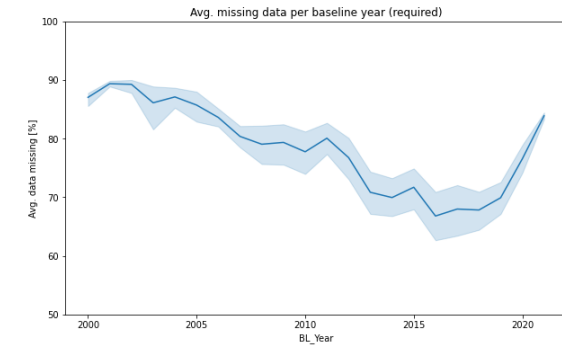
IshId	ExtCaselId	visit_name	Quelle	Dokument	Reiter	Sektion	Untersektion	Label	value	error
70230900	490482119	Visit M0 - Baseline	PMD	NEBPDAMB	Allgemein	Diagnose		Dementielles Syndrom		missing value (mandatory)
70230900	490482119	Visit M0 - Baseline	PMD	NEBPDAMB	Allgemein	Diagnose- und Manifestationsdatum		Datum der Erstmanifestation		missing value (mandatory)
70230900	490482119	Visit M0 - Baseline	PMD	NEBPDAMB	Allgemein	Familienanamnese		Familienanamnese erhoben		missing value (mandatory)

PatientID	IshId	ExtCaselId	visit_type	visit_name	Quelle	Dokument	Reiter	Sektion	Untersektion	Gewicht	Größe	error
7W28Q74RO	70230900	44871200	LEM24	Visit M24 - FU	PMD	NEBPDAMB	Allgemein	Allgemein		185.232	162	Weight is higher than height.
7W28Q74RO	70230900	169461222	LEM36	Visit M36 - FU	PMD	NEBPDAMB	Allgemein	Allgemein		208.516	185	Weight is higher than height.



# Retrospektive Datenanalyse (1) - verwendete Datensätze

- Retrospektive-Daten in Tübingen
  - RedCap-Export
  - Versorgungsdaten mit weniger guter Datenqualität
- PPMI-Daten
  - Allg. sehr gute Datenqualität
  - Vorhandene Daten:
    - Labor-Daten (CSF, DNA, RNA, Plasma etc.)
    - Imaging-Daten (MRI, DTI, SPECT etc.)
    - Klinische-Daten (MDS-UPDRS, MoCA, GDS-15, Hoehn & Yahr etc.)
    - Genetische-Daten (WGS, Exome Sequencing, Gene Sequencing etc.)
- Auch Imaging-Daten werden für Analyse verwendet:
  - MRI 3D-T1: commercial segmentation tool (AIRAscore structure 2.0.1)
  - 218 Patienten mit 33 Hirnregionen



PARKINSON'S  
PROGRESSION  
MARKERS  
INITIATIVE

Play a Part in Parkinson's Research

**AIRAm**ed  
artificial intelligence in radiology



# Retrospektive Datenanalyse (2) - Mapping & Endpunkte

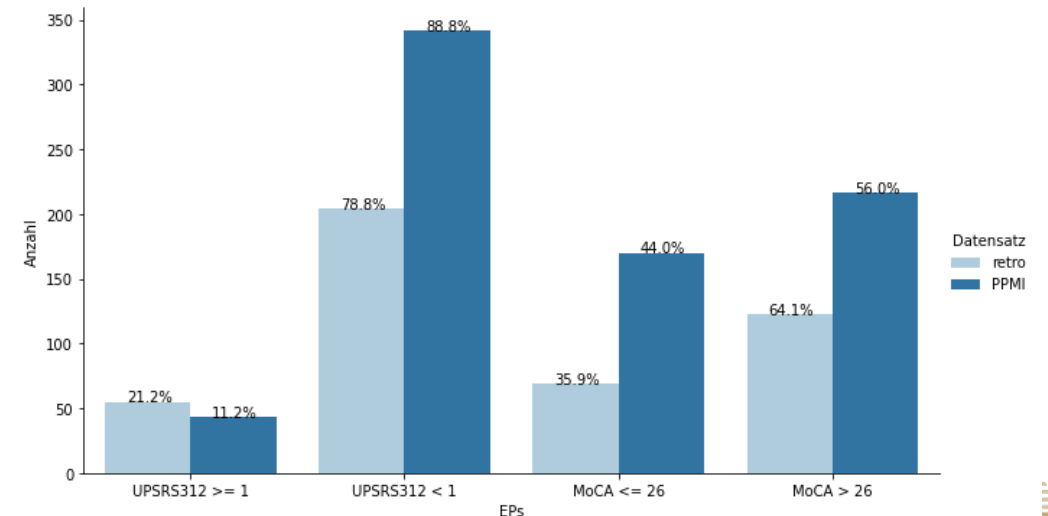
- Mapping:

- Daten müssen vor Verwendung auf gemeinsamen Datensatz gemappt werden
- PPMI: 388 PD Patienten
- Retro-Tübingen: 967 PD Patienten  
→ Insg. 1355 PD Patienten
- PPMI: 1492 Visiten
- Retro-Tübingen: 2213 Visiten  
→ Insg. 3705 Visiten
- PPMI: 192 Felder
- Retro-Tübingen: 341 Felder

- Endpunkte (M36):

- MCI (Mild Cognitive Impairment):
  - MoCA  $\leq 26$
- PI (Postural Instability):
  - UPDRS 3.12  $\geq 1$

Data & Assessments for LOC-EARLY and LOC-DBS study patients		
Label	type	options / comments
Demographic data including family history, education, ethnicity		
Variable	Description	
APPRDX	Appropriate Diagnosis	
EVENT_ID	Visit Label	
PATNO	Patient number	
SITE	Blinded Center Number	
ab_asyn	Ratio of CSF A-beta 1-42 to CSF Alpha-synuclein	
abeta	CSF A-beta 1-42 (2016 assay)	
age	Age at Baseline	
age_cat	Categorical Age at Baseline	
agediag	Age at PD Diagnosis	
ageonset	Age at Symptom Onset	
ai_caudate	Asymmetry Index (Caudate)	
ai_putamen	Asymmetry Index (Putamen)	
ai_striatum	Asymmetry Index (Striatum)	
APOE	APOE Genotype	



# Retrospektive Datenanalyse (3) - geplante Publikation (in progress)

**Ziele der Analyse in Zusammenarbeit mit Klinikern erarbeitet:**

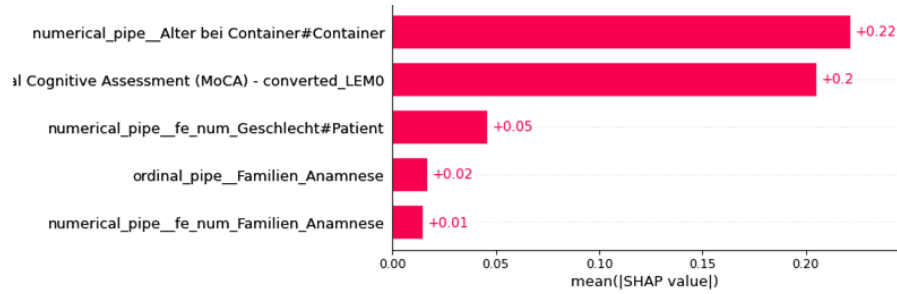
Prädiktion von klinischem Endpunkt “Mild Cognitive Impairment” (MCI) und “Postural Instability” (PI) im Vergleich von..

1. Minimal/core clinical dataset:  
PPMI (exploratory cohort) vs. TuePaC (validation cohort)
2. Minimal/core clinical data vs. extended dataset (excluding imaging):  
PPMI (exploratory cohort) vs. TuePaC (validation cohort)
3. PPMI cohort (only the subset of patients with imaging data):  
extended dataset (excluding imaging) vs. extended dataset (including imaging)



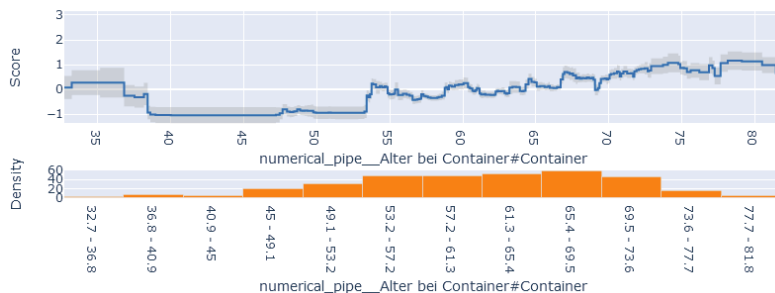
# Retrospektive Datenanalyse (4) - ausgewählte Ergebnisse

- Exemplarisch - Prädiktion von MCI:
  - BL-Daten für EBM-Modell (minimal)
    - AUC-Wert: 0,66
    - Train: PPMI, Test: TuePaC
- Shapley-values bei Prädiktion von MCI:

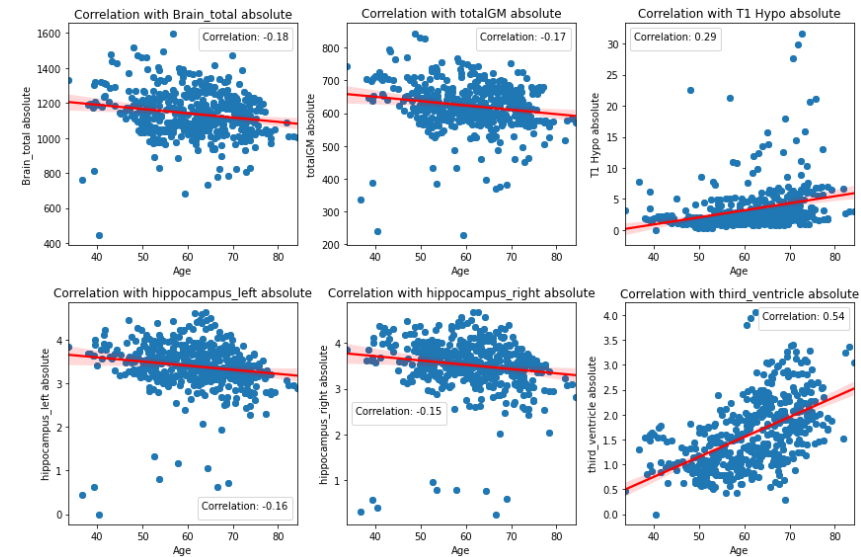


- EBM-spezifische Darstellung von “Alter”:

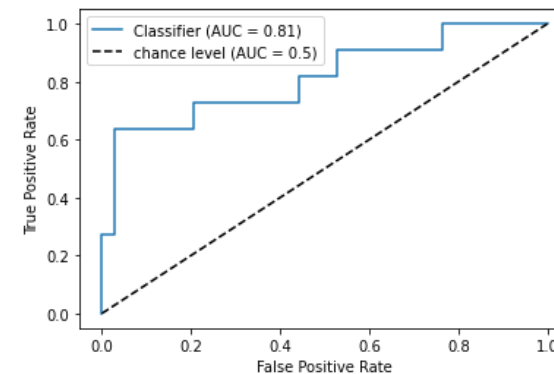
Term: numerical\_pipe\_Alter bei Container#Container (continuous)



- Imaging (Atrophie in untersch. Hirnregionen):



- AUC-Kurve (mit Imaging-Daten):





# Studienergebnisse aus klin. Sicht

- 4 Standorte beteiligt (MRI/TUM, KUM/LMU, UKT, UKU)
- Patientenrekrutierung:
  - LOC-EARLY: 182 (geplant: 400)
  - LOC-DBS: 26 (geplant: 114)
- Studienstart während Corona -> massive Auswirkungen auf Rekrutierung
- Finanzierung während Studie ausgelaufen
- Studienergebnisse: keine mit Studiendaten, Analyse mit retrospektiven Daten: Ende '23
- Weitere Projekte:
  - Bilddatenanalyse (geplant)



# Studienergebnisse aus IT-Sicht

- Insgesamt:
  - PMDs für Datenerhebung entwickelt + integriert -> langfristige Nutzung möglich
    - Generierung von Textblöcken aus PMDs für Arztbriefschreibung
  - Erfahrungen mit klin. Primärsystemen gesammelt
  - Kontakte zu Klinikern + anderen beteiligten Standorten
- LOC-PD:
  - (studienspezifische) Datenqualitätsberichte
  - (Code-)Template für Analyseprojekte



# Probleme

- Organisatorisch:
  - Kommunikation teils schwierig, viele Stakeholder (Kliniker <-> IT, standortintern, standortübergreifend) & hohe Fluktuation
  - COVID-19:
    - Patientenrekrutierung eingebrochen
    - Kapazitäten gebunden (GECCO/NUM CODEX)
- Technisch:
  - PMD-Entwicklung und -Integration sehr aufwändig und langwierig
  - Viele manuelle Schritte (PMD Freigabe, Eintragen von Fall-IDs, Einwilligung)  
-> hohe Fehleranfälligkeit für Übertragung ins Studiensystem
  - Wenig Übersicht über Daten / teilweise kein/später Datenzugriff
  - Datenqualität: keine (frühe) standortübergreifende Lösung

# Lessons learned

- Regelmäßiger Austausch wichtig
- Frühzeitiges + ausführliches Monitoring der ETL Schnittstellen
- Früher Zugriff auf Daten wichtig:
  - Frühzeitige Korrekturen bei Problemen
  - Regelmäßige Datenqualitätsberichte
- Prozesse, Umsetzung, etc. haben signifikant (!) mehr Zeit benötigt als antizipiert



# Ausblick

- Bereitstellung der LOC-PD Bilddaten noch in Arbeit
- Finalisierung Analyse: Retrospektive PD Daten + PPMI Daten
- Perspektivisch (?):
  - Ausweiten der Datenqualitätsberichte als generische Lösung für andere Projekte





# Fragen?

**Benjamin Sailer, Christoph Kiefer**  
21.11.23

© UNIVERSITÄTSKLINIKUM TÜBINGEN.

EBERHARD KARLS  
**UNIVERSITÄT  
TÜBINGEN**



**Universitätsklinikum  
Tübingen**