



# Das Data Dictionary der Nationalen Kohorte

Stefan Ostrzinski

TMF-Workshop Metadatenmanagement  
Berlin, 18.12.2012

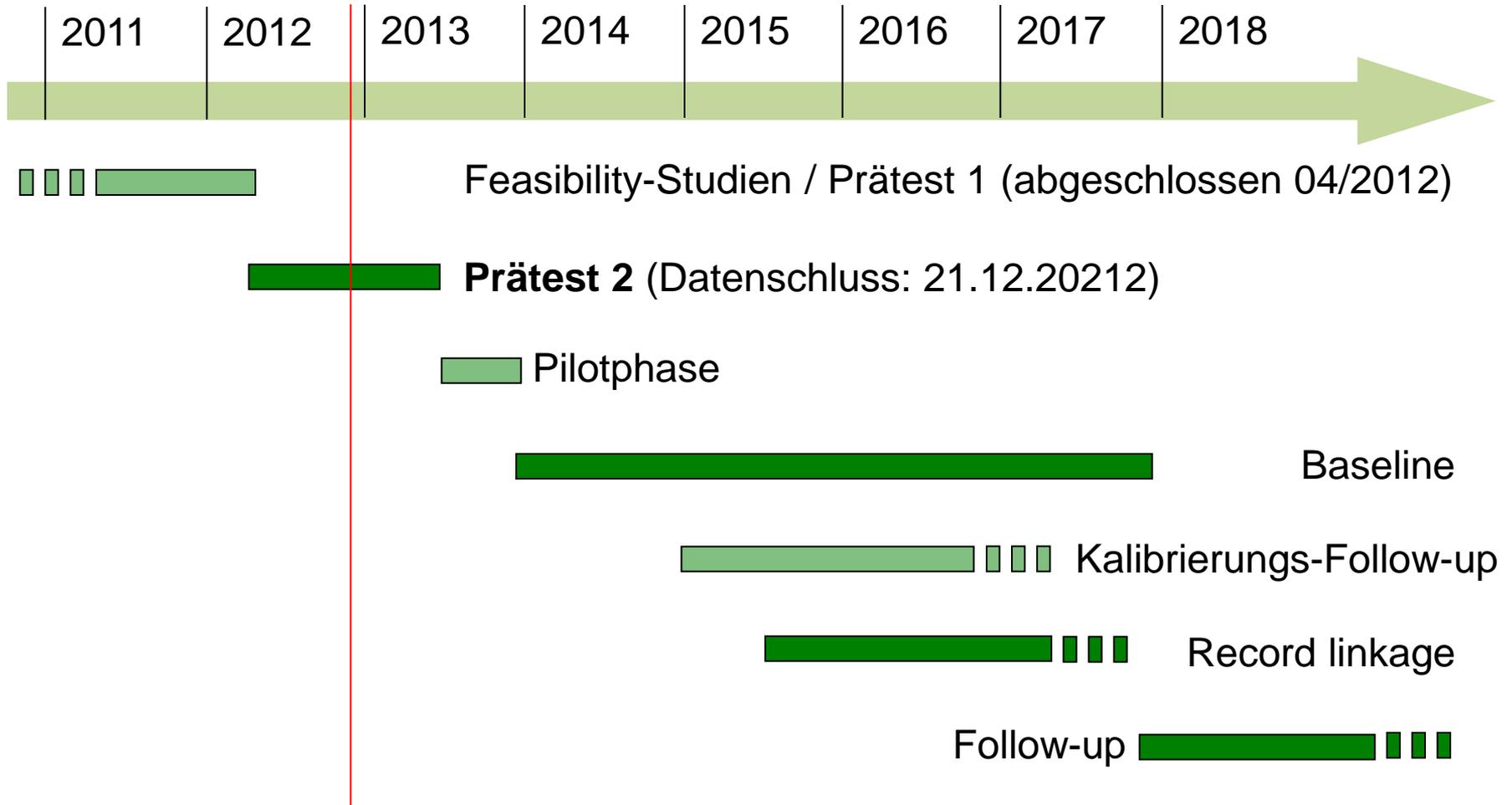
# Studiendesign



- Bevölkerungsbasierte prospektive Kohortenstudie
- Altersbereich 20-69 Jahre zur Baseline-Untersuchung (m/w)
- Repräsentativer Ansatz (Meldeamtsstichprobe)
- 8 regionale Cluster, 18 Studienzentren
- Level 1 (Basisprogramm)  
ca. 2,5 h, N = 200.000
- Level 2 (intensiviertes Programm)  
ca. 4 h, N = 40.000
- Level 3 (zusätzliche Fragestellungen)  
N variabel
- MRT-Programm, N≈30.000
- Baseline-Untersuchung 4 Jahre  
Kalibrierung nach 1 Jahr  
Nachbeobachtung alle 2-3 Jahre  
Follow-Up-Untersuchung 4 Jahre



# Zeitplan



# Datenmanagement: Herausforderungen



- **Dezentrale Organisation**  
Viele Beteiligte mit unterschiedlichen Voraussetzungen und Interessen
  - **Lange Projektlaufzeit**  
Daten gewinnen im Lauf der Jahrzehnte an wissenschaftlichem Wert
  - **Dezentrale Erhebung vs. zentrale Verarbeitung**  
Effiziente Integration und detaillierte Regelung der Nutzung erforderlich
  - **Dezentrale Erhebung – Qualitätssicherung**  
Vergleichbarkeit über alle Studienzentren und über den Zeitverlauf
  - **Ethik / Datenschutz**  
Ethische und rechtliche Anforderungen „erschweren“ IT-Unterstützung
- ⇒ **Standardisierung, Zentralisierung, Automation**

# Data Dictionary als zentrales Element



- Definition / Design der Studie
- Erstellung der Erhebungswerkzeuge
  - Interview, Untersuchungen, Selbstausfüller-Fragebögen
  - Dokumentation Durchführung und Ergebnisse, Nachdokumentation
- Datenerhebung
  - Plausibilitäts-, Konsistenz- und Vollständigkeitsprüfungen
  - Warnhinweise, Fehlermeldungen
- Datenbereinigung, prozessbegleitende Qualitätssicherung
- Datenweitergabe für wissenschaftliche Auswertung
  - Beantragung, Übergabe und Rückführung
- Wissenschaftliche Analysen/Datenauswertung

Questions 1, 2

# Data Dictionary: Allgemeine Anforderungen



- Multizentrisch, Quer- und Längsschnitt
- Vollständigkeit
  - Alle benötigten Informationen abgebildet
- „Normalisierung“
  - Abhängigkeiten in der Struktur, nicht in den Daten (keine impliziten Constraints)
- Versionierung
  - Zuordnung zwischen Studiendaten und Versionsstand des Data Dictionary
- Nutzung von „Vorarbeiten“
  - SHIP, SFB-TRR19, Transferstelle FVCM, GANI\_MED





# Mindestinhalte Data Dictionary

- Untersuchung:
  - Name, Label, Beschreibung, Projektzuordnung
- Variable:
  - Name, Label, Beschreibung, Reihenfolge
  - Skalenniveau, Datentyp, Art des Eingabeelements
  - Optional: Belegungsliste, Einheit, Überschrift, Gliederung, Hilfetext
- Belegung:
  - Liste der erlaubten Werte, Flags
- Validierungs- und Sprungregeln

# Validierungs- und Sprungregeln



- “weiche” und “harte” Validierungsregeln  
→ Warnung / Fehlermeldung + Ablehnung
- “Sprungregeln”  
→ “Ein- / Ausblenden” von Variablen
- Berechnungsregeln für transformierte Variablen?
- Gleiche Beschreibungsform für alle Arten
- derzeit realisiert als JavaScript-Code
- Input: Variablen, wenige speziell ermittelte Werte  
(Alter und Geschlecht, Projektzuordnung)

# JavaScript-Definition von Regeln



- Vorteile
  - (fast) jedes denkbare Regelkonstrukt realisierbar
  - funktioniert ;-)
- Nachteile
  - Programmierkenntnisse erforderlich
  - „Freiheiten“ von JavaScript (z. B. „NULL vs. 0-Problem“, Abhängigkeiten von verwendeter JavaScript-Engine)
- Alternativen
  - „Metastruktur“ aus „UND“/„ODER“/„NICHT“-Verknüpfungen

# Fazit zum Thema Regeln



- Wir haben eine funktionierende Lösung mit Vor- und Nachteilen.
- „Die optimale Lösung“ gibt es nicht.
- Problem immer: „Möglichkeiten“ vs. „Sicherheit“
- Für jedes Projekt abzuwägen:
  - Wer definiert die Anforderungen?
  - Wer programmiert?
  - Was ist abzubilden?

# Erhebungsinstrumente



Abmelden Benutzer: fitzerk@Neubrandenburg UntZ Proband: 2810009992@Pretest 2

Passwort ändern Startseite

- Intro
- Blutdruck
- Anthropometrie
- Handkraft
- Start Interview
- SES
- MAN\_HKE
- MAN\_KRE
- MAN\_SWE
- MAN\_SDE
- MAN\_RME
- MAN\_LUN
- MAN\_ALL
- MAN\_MDE
- MAN\_GLF
- MAN\_HTE
- MAN\_NBE
- MAN\_NME**
- MAN\_AOF
- MAN\_INF
- MAN\_UNF
- MAN\_OP
- Frauen
- Ende
- Blutentnahme
- Blutverarbeitung Basis
- NeuroKog Config
- Zufriedenheit
- Abschlussgespräch
- Geruchstestung
- Touchscreen Intro
- Knochenbrüche
- Männer
- Touchscreen: NeuroKog

**MAN\_NME** MNME\_AN\_G1

Neurologische und psychische Erkrankungen

Wie würden Sie Ihr Gedächtnis zurzeit beurteilen, würden Sie sagen, es ist...

Weiter Zurück Zur nächsten

Bitte auswählen...  
Bitte auswählen...  
Sehr gut  
Gut  
Mittelmäßig  
Schlecht  
Sehr schlecht  
[Weiß nicht]  
[Keine Angabe]

Notiz anlegen Sprung

- Webbasierte Erhebungsanwendung
- Touchscreen für Selbstausfüller-Fragebögen



- Dynamische Formularerstellung (Datenbank)
- Geräteintegration
- Zentrale Studiendatenbank

# Formularerstellung



⇒ Papierformular

⇒ “Formale Beschreibung” (Data Dictionary)

⇒ Datenbank

⇒ Webanwendung

⇒ Webformular

“Untersuchung”: Label

**Sozioökonomischer Status und soziodemographische Faktoren**

SES01

Geschlecht

INT: nicht fragen

1

Männlich

2

Weiblich

“Untersuchung”: Beschreibung

*Ich möchte Ihnen zunächst einige Fragen zu Ihrer Person stellen*

SES02

Wann sind Sie geboren?  
(TT.MM.JJJJ)

INT: Antwort des Probanden laut  
wiederholen

|\_|\_|\_| · |\_|\_| · |\_|\_|\_|\_|  
TT MM JJ

**Nationalität/ Ethnizität/ Muttersprache**

SES03a

Wo sind Sie geboren?

*Bitte verwenden Sie die heutige  
Staatsbezeichnung!*

1

In Deutschland (in den heutigen Grenzen)

2

In einem anderen Land und zwar:

\_\_\_\_\_ Geburtsland

SES03b

In welcher Stadt sind Sie geboren?

Variablenblock

Abmelden "Untersuchung": Label

Neubrandenburg UntZ Proband: 1000001@Feasibility

### Fragebogen SES

"Untersuchung": Name

#### Nationalität/ Ethnizität/ Muttersprache

Sozioökonomischer Status und soziodemographische Faktoren

#### Wo sind sie geboren ?

Bitte verwenden Sie die heutige Staatsbezeichnung!

Deutschland (in den heutigen Grenzen)

ja  nein

"Untersuchung": Beschreibung

Bitte auswählen...

In diesem Land:

[Weiter](#) [Zurück](#) [Zur nächsten leeren Variablen](#) [Notiz anlegen](#) [Sprung](#)

- Blutdruck
- SES
- SES DAIMON
- MAN a AAA DAIMON
- MAN a HKE DAIMON
- MAN b SWF DAIMON
- MAN c SDF DAIMON
- MAN d NME DAIMON
- MAN e LU DAIMON
- MAN f ALL DAIMON
- MAN g MDE DAIMON
- MAN h GLE DAIMON
- MAN i HTE DAIMON
- MAN j NBE DAIMON
- MAN k NME DAIMON
- MAN l AOE DAIMON
- MAN m INF DAIMON
- MAN o SBH DAIMON
- MAN p UNF DAIMON
- FRAUEN DAIMON
- IPAQ DAIMON
- END DAIMON
- Anthro DAIMON
- Blutdruck DAIMON

Variablenblock

Screenshot der Webanwendung



## Sozioökonomischer Status und soziodemographische Faktoren

<b>SES01</b>	Geschlecht	<input type="checkbox"/> 1	Männlich
<b>Variablenname (oder Teil davon)</b>		<input type="checkbox"/> 2	Weiblich
INT: nicht fragen			

*Ich möchte Ihnen zunächst einige Fragen zu Ihrer Person stellen.*

Block 1: Header

<b>SES02</b>	Wann sind Sie geboren? (TT.MM.JJJJ)
INT: Antwort des Probanden laut wiederholen	

TT	MM	JJ
----	----	----

### Nationalität/ Ethnizität/ Muttersprache

<b>SES03a</b>	Wo sind Sie geboren? <i>Bitte verwenden Sie die heutige Staatsbezeichnung!</i>
---------------	---

<input type="checkbox"/> 1	In Deutschland (in den heutigen Grenzen)
<input type="checkbox"/> 2	In einem anderen Land und zwar:
_____ Geburtsland	

Block 2 oder Variable

Je "Kästchen" eine Variable

Block 2: Hilfe

Block 2: Header

Abmelden

Benutzer: osi@Neubrandenburg UntZ

Block 1: Header

Passwort ändern  
Probandenauswahl

### Fragebogen SES

#### Nationalität/ Ethnizität/ Muttersprache

Sozioökonomischer Status und soziodemographische Faktoren

- SES DAIMON
- MAN a AAA DAIMON
- MAN a HKE DAIMON
- MAN b SWE DAIMON
- MAN c SDE DAIMON
- MAN d RME DAIMON
- MAN e LUN DAIMON
- MAN f ALL DAIMON
- MAN g MDE DAIMON
- MAN h GLE DAIMON
- MAN i HTE DAIMON
- MAN j NBE DAIMON
- MAN k NME DAIMON
- MAN l AOE DAIMON
- MAN m INF DAIMON
- MAN o SPH DAIMON
- MAN p INF DAIMON
- FRAUEN DAIMON
- IPAQ DAIMON
- END DAIMON

#### Wo sind sie geboren ?

Bitte verwenden Sie die heutige Staatsbezeichnung!

Deutschland (in den heutigen Grenzen)

ja  nein

In einem anderen Land und zwar:

In diesem Land:

Definierte Wahlmöglichkeiten

Je "Kästchen" eine Variable

Weiter Zurück Zur nächsten leeren Variablen Notiz anlegen Sprung

Block 2: Hilfe

Block 2: Header

Block 2 oder Variable



# Formularelemente

- Radio-Button  
(horizontal / vertikal)
- Select-Liste
- Datum / Uhrzeit
- Textfeld  
(einzeilig / mehrzeilig)
- Checkbox



## Grundsätze:

- Freitexte (Textfelder) und Checkboxes sollten vermieden werden.
- Es sollten keine Default-Werte vorgegeben werden, ggf. Pseudo-Default-Werte, die nicht erfolgreich validiert werden können.

Hintergrund: Es ist sonst nicht möglich, hinterher zu erkennen, ob ein Feld absichtlich oder versehentlich nicht gefüllt wurde.



Passwort ändern

Startseite

- ✓ Intro
- ✓ NeuroKog Config
- [Kontaktmöglichkeiten](#)
- [Blutdruck](#)
- [Anthropometrie](#)
- [Handkraft](#)
- [Start Interview](#)
- SES
- MAN\_HKE
- MAN\_KRE
- MAN\_SWE
- MAN\_SDE
- MAN\_RME
- MAN\_LUN
- MAN\_ALL
- MAN\_MDE
- MAN\_GLE
- MAN\_HTE
- MAN\_NBE
- MAN\_NME
- MAN\_AOE
- MAN\_INF
- MAN\_UNF
- MAN\_OP
- Frauen
- Ende
- [Blutentnahme](#)
- [Blutverarbeitung Basis](#)
- [Laborwerte](#)
- [Zeiterfassung TS-Core](#)
- ✓ [Touchscreen-Intro](#)
- ✓ [Tabak](#)
- [Alkohol](#)
- [Frühkindliche Faktoren](#)
- [Gewichtsgeschichte](#)
- Phänotyp
- Screening
- Gesundh.bez. Lebensqual.
- WHOQOL BREF
- ✓ [Neurokognitive Tests](#)
- [Neuropsychologische Tests](#)
- [Zufriedenheit](#)
- [Abschlussgespräch](#)
- [Geruchstestung](#)
- [Blutverarbeitung C1](#)
- [Speicheluntersuchung](#)
- ✓ [Urinuntersuchung](#)
- [Stuhluntersuchung](#)
- [Nasenabstrich](#)

## Geruchstestung

### Geruchstestung mit Sniffin Sticks - Screening 12

Untersuchernummer

Beginn



TT.MM.JJJJ hh:mm

Ich führe jetzt einen Test zur Prüfung Ihres Riechvermögens durch. Dafür halte ich Ihnen gleich mehrfach einen Stift unter die Nase. Sie versuchen dann den jeweiligen Geruch zu erkennen und die richtige Antwort aus vier vorgegebenen Möglichkeiten auszuwählen.

Lassen Sie mich aber zunächst fragen:

Haben Sie oder hatten Sie in den letzten 6 Wochen einen Schnupfen?

- ja  nein  [keine Angabe]

“Missing“-Werte

Wie frei ist Ihre Atmung durch die Nase im Moment?

**INT:** Bitte nennen Sie mir auf dieser Skala (Skala zeigen) von 1-10 das Kästchen, das Ihre Nasenatmung am ehesten beschreibt. 1 bedeutet die Nase ist völlig frei und 10 bedeutet die Nase ist völlig verstopft?

Hilfetexte und Anweisungen wie in SOP definiert

- 1 - völlig frei  2  3  4  5  6  7  8  9  10 - völlig verstopft

Weiter

Zurück

Zur nächsten leeren Variablen

Notiz anlegen

Sprung

Für diesen Probanden durchzuführende Untersuchungen

SOPs online verfügbar

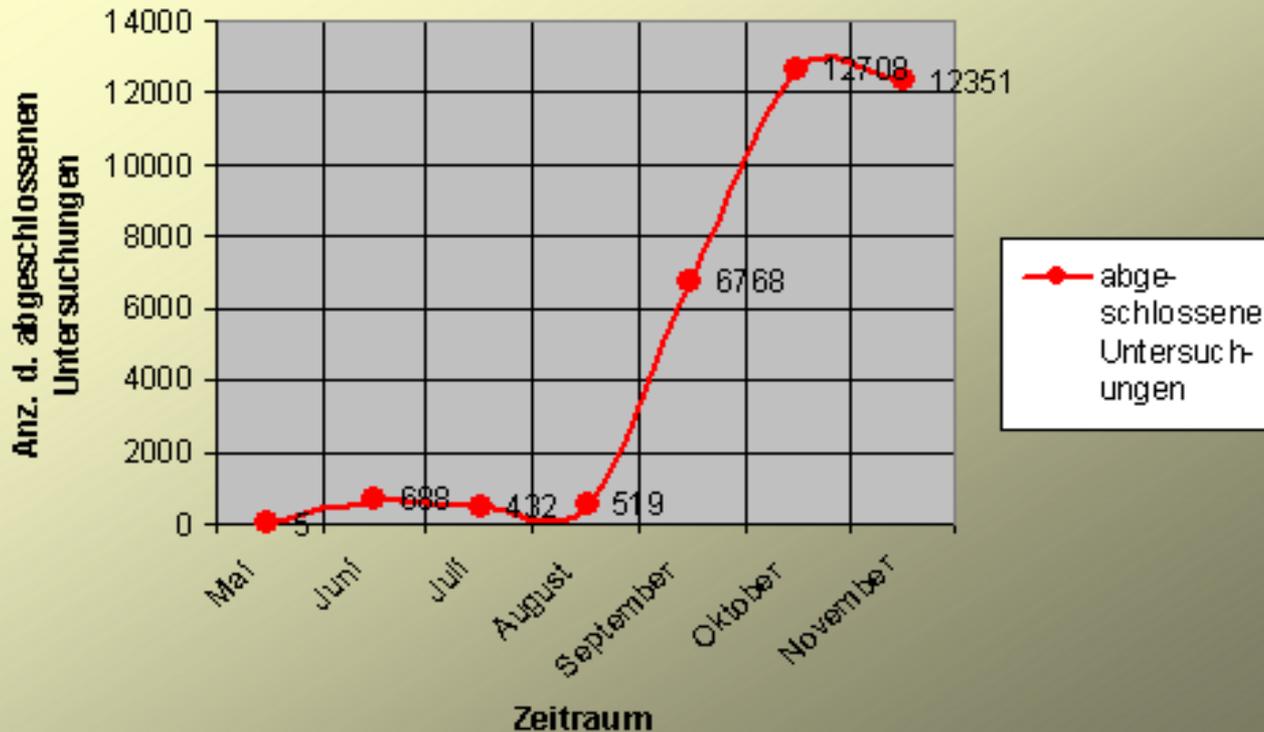
Ablaufdaten ...

# Datenmanagement Nationale Kohorte

## Zahlen & Fakten Prätest 2



### Pretest 2 - abgeschlossene Untersuchungen



Untersuchungen  
(Formulare):

64

Variablen:

2.461

Regeln (Validierung  
und Sprünge):

1.263

Untersucher:

306

bisher untersuchte  
Probanden:

5313

**Σ 33471 abgeschlossene Untersuchungen für den Pretest 2**  
(30.05.-23.11.2012)

Stand: 26.11.2012

# Data Dictionary:

## Weitere realisierte Use cases



- DD-Export
  - CSV
- Datenweitergabe
  - CSV
  - SAS-Dataset + Formatkatalog

# Data Dictionary der Nationalen Kohorte



Danke für Ihre Aufmerksamkeit.



Dipl.-Math. Stefan Ostrzinski

Universitätsmedizin Greifswald

ICM-VC - Zentrales Datenmanagement

E-Mail: [stefan.ostrzinski@uni-greifswald.de](mailto:stefan.ostrzinski@uni-greifswald.de)

Tel.: +49 3834 86 7545