



Minimierung von Redundanzen und
Inkonsistenzen im Datenmonitoring
einer multizentrischen
internationalen epidemiologischen
Studie



This project has received funding from the European Union's
Seventh Framework Programme for research, technological
development and demonstration under grant agreement No. 266044



Gliederung

1. Hintergrund
2. Wie betreiben wir Datenmanagement?
3. Wieso überhaupt Redundanzen und Inkonsistenzen?
4. Wie bekämpfen wir Redundanzen und Inkonsistenzen?
5. Was ist das Datenmanagement-Access-Tool?
6. Ergebnisse



1. Hintergrund

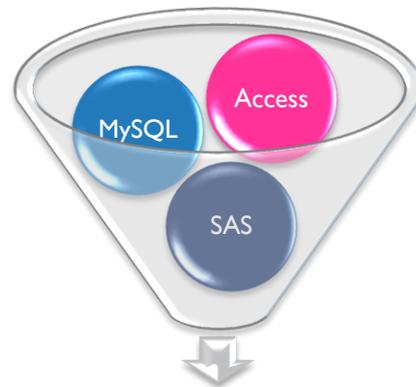


This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement No. 266044

1. Was ist I.Family?

- Follow-Up der Idefics-Studie
 - Einflüsse auf Ernährung, Lebensstil und Gesundheit europäischer Kinder, Jugendlicher und ihrer Eltern
 - Studienstart 2012
 - Survey Start April 2013
 - > 17000 Teilnehmer (Kinder und Familien) aus 8 Ländern
 - Italien, Estland, Zypern, Belgien, Schweden, Deutschland, Ungarn, Spanien
- Instrumente u. a.
 - Bewegungsmessungen
 - Anthropometrie
 - Neuropsychologische Tests
 - etc.

2. Wie betreiben wir Datenmanagement?



Datenmanagement

1. Warum Access?

- Nahezu flächendeckender Einsatz
 - Gehört zum erweiterten Kreis vom MS-Office-Paket
 - Falls nicht vorhanden, einfach zu besorgen
- Verhältnismäßig einfache Bedienung
 - Datenbank mit integrierter Entwicklungsumgebung (IDE)
 - Personal kann schnell angelernt werden



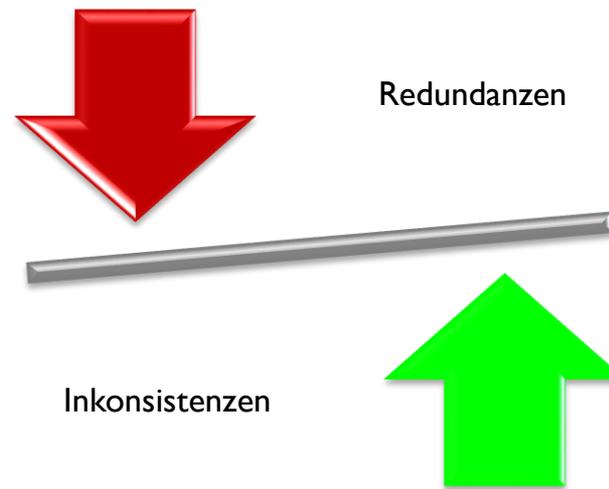
2. Datenbankenarten

- Eingabedatenbanken
 - Liegen auf einem Zentralen Daten Server
 - Zugriff für Studienzentren
 - Erst- und Zweiteingabe-Datenbank
 - MySQL-Tabellen als Backend
 - Weiterverarbeitung in SAS
- Lokale Datenbanken
 - Eingabe der Instrumente durch die Studienzentren
 - Standardisierte Upload-Routine
 - Mitgeliefertes Tool
 - One-Button-Upload
 - Daten in gezippter Form
 - Weiterverarbeitung in SAS

3. Wie funktionieren Pooling-Datenbanken?

- Zusammenführung der Daten aller Länder
 - Instrumentenweise
 - Gelieferte Daten vorerst in Import-Tabellen
 - Auswahlabfragen überprüfen auf bereits vorhandene Daten
 - Löschanfrage löscht bereits vorhandene Daten
 - Zusammenführung der Daten eines Instruments

3. Wieso überhaupt Redundanzen und Inkonsistenzen?



1. Kompromiss zwischen

- Datenbankprogrammierern
 - So restriktiv wie möglich
 - Wenig Freiheit bei Eingabe
- Eingabekräften
 - So einfach wie möglich
 - Beste Bedienbarkeit
- Probanden
 - So schnell wie möglich
 - Möglichst kurz und nicht langatmig



2. Extern erstellte Instrumente

- Eventuell wenig Einfluss auf deren Erstellung
 - Aufbau der Tabellen
 - Vergabe von Primärschlüsseln
- Möglicherweise keine Plausibilitätsprüfungen auf
 - IDs
 - Werte

3. Nach der Datenaufbereitung

- SAS-Programme
 - Änderungen von Daten
 - Mögliches Entstehen von Duplikaten

4. Wie bekämpfen wir Redundanzen und Inkonsistenzen?



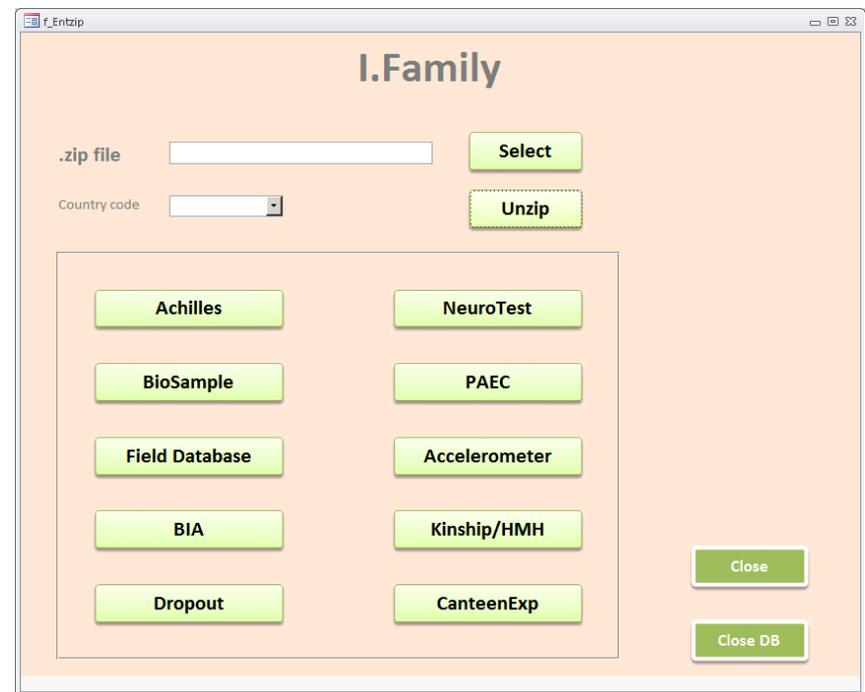
1. Vor Import

- Access-Tool analysiert eingelesene Daten
 - Abfragen geben fehlerhafte IDs als Message-Box aus
- Dokumentation fehlerhafter IDs
- Nachfragen fehlerhafter IDs bei den Studienzentren

2. Korrektur fehlerhafter IDs

- In den Pooling-Datenbanken
 - Regelfall bei lokalen Daten
- In MySQL-Datenbanken
 - Regelfall bei zentralen Daten
- Im SAS-Aufbereitungsprogramm
 - Ausnahmefälle

5. Was ist das Datenmanagement-Access-Tool?



1. Wer hat's erstellt?

- Marina Bleskina
 - (ehem.) Studentin
 - Hochschule Hannover
 - Studiengang Medizinisches Informationsmanagement
 - Bachelorarbeit

2. Wer nutzt es?

- Fachgruppe IT, Datenmanagement und Medizinische Dokumentation
 - Datenaufbereitung der I-Family-Studie

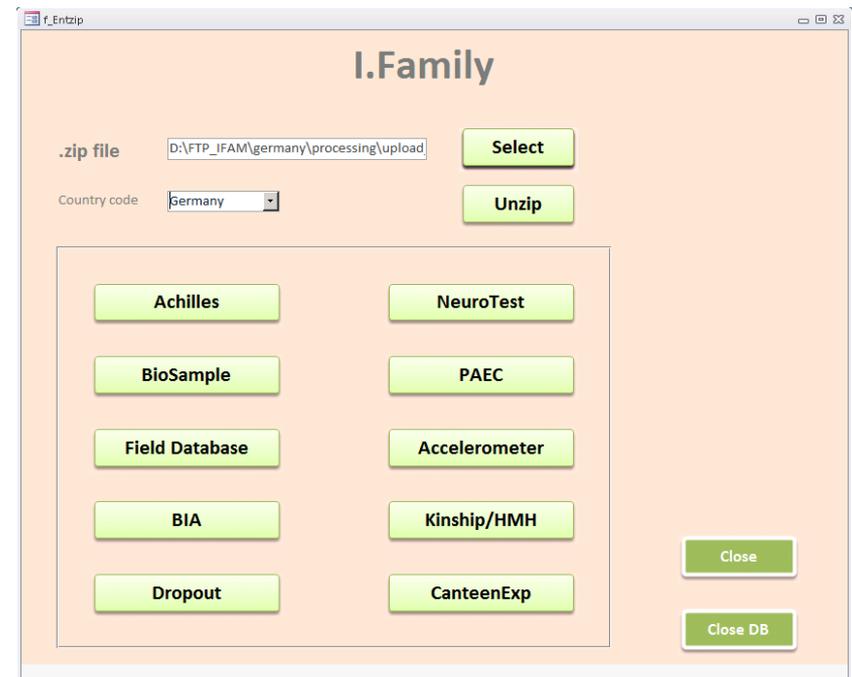
3. Wofür ist es gedacht?

- Hilfe zur Dokumentation der Uploads
 - Auszählung der Datensätze durch VBA
 - Schneller und sicherer als manuelles Öffnen jedes einzelnen Instruments und Auszählen der Datensätze
- Aufspüren fehlerhafter IDs
 - Identifizierung durch VBA
 - Fehlermeldungen als Message-Boxen
 - VBA For-Schleife



4. Wie wird es angewandt?

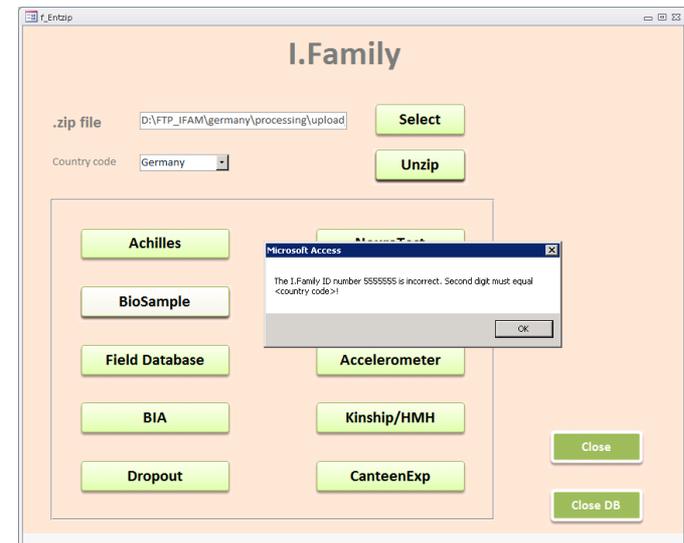
- Vorgehen
 - Zip-File auswählen
 - Land auswählen
 - Zip-File entpacken
 - Auswahl des Instruments
 - Dokumentation der Anzahlen
 - Dokumentation der fehlerhaften IDs
 - Ggf. Kontaktaufnahme mit den Studienzentren





5. Wie hilft es bei der Minimierung von Redundanzen und Inkonsistenzen?

- Tool überprüft beim Öffnen die enthaltenen IDs
 - Zu kurz
 - Zu lang
 - Stimmt der Ländercode
 - Jedem Land ein Ländercode zugeordnet
 - Bereits ausgewählter Ländercode wird mit ID verglichen
 - IDs mit falschem Ländercode würden einem anderen Land zugeordnet werden, ohne bemerkt zu werden



6. Was hilft das?

- Dokumentation
 - Notieren fehlerhafter IDs
 - Bei Studienzentren nachfragen
- Verbesserter Arbeitsfluss
 - Bei regelmäßiger Überprüfung
 - Anfragen bei Studienzentren vor nächstem Datenstand
 - Bearbeitung der Uploads anderer Länder möglich
 - Bei Antwort der Studienzentren
 - Datenkorrektur vor dem eigentlichen Import möglich
- Vorteile
 - Daten kommen „sauberer“ in die Datenbanken
 - Erhöhte initiale Datenqualität
 - Frühzeitige Korrekturen ersparen Zeit

6. Was sind die Ergebnisse?



1. Unter Verwendung des Tools

- Das Tool lohnt sich, weil es
 - Zeit spart
 - Qualität steigert
 - das Auszählen der Datensätze vereinfacht
 - die Überprüfung und Korrektur der erfassten Daten beschleunigt
 - nachfolgende Arbeitsschritte vereinfacht
 - Poolen der Daten
 - Weiterverarbeitung in SAS

Noch Fragen



Kontakt: zaja@bips.uni-bremen.de

XING: Marcus Zaja



This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement No. 266044



**Vielen Dank für Ihre
Aufmerksamkeit!**



This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement No. 266044