



Datennachnutzung in der Psychologie: Bielefelder Zwillingsdaten

Rainer Riemann

Universität Bielefeld

Fakultät für Psychologie und Sportwissenschaft

Abteilung Psychologie

Überblick

Beschreibung unserer Datensätze

Probleme der Datenverwaltung

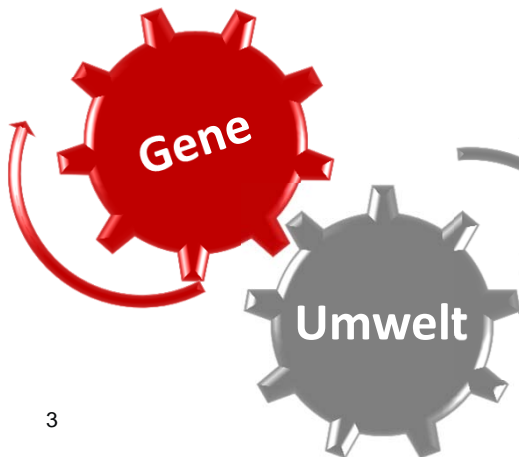
„I have a dream“
oder Wünsche an eine Datenverwaltung

Datensätze der verhaltensgenetischen Forschung in Bielefeld

Seit etwa 17 Jahren werden in der Arbeitsgruppe „Differentielle Psychologie“ Zwillingsdaten gesammelt. Initiatoren der Zwillingsforschung in Bielefeld, waren mein „Vorgänger“ Prof. Dr. A. Angleitner und Prof. Dr. Jan Strelau (Warschau).

Es sind in den vergangenen Jahren umfangreiche Datensätze gesammelt worden, die prinzipiell als „public use files“ für weitere Analysen dem Fachpublikum zugänglich gemacht werden sollen / könnten.

Schwerpunkte der Datensammlung bilden Persönlichkeits- und Temperamentstests, die mittels Fragebogen erhoben wurden. Diese werden durch Daten aus Leistungstests und Verhaltensaufzeichnungen ergänzt.



Nachdem die Fragen nach genetischen und Umwelteinflüssen weitgehend abgearbeitet wurden, stehen zunehmend komplexere Fragestellungen im Fokus der Forschung:
z.B. welche Einflussgrößen liegen der Stabilität und Veränderung von Persönlichkeitsmerkmalen zugrunde?

Komplexität der Daten

- Viele Variablen, da Zwillinge vergleichsweise schwer zu rekrutieren sind.
- Unterschiedliche Datenquellen: Selbstberichtsdaten (Fragebogen), Testdaten, Bekanntenbeurteilungen.
- Unterschiedliche „Form“ und Anordnung derselben Daten.



➡ Datensätze werden schnell sehr unübersichtlich und sind von Außenstehenden nur nach längerer Einarbeitung zu nutzen.

Beispiel

- Für einen Teil unserer Probanden haben wir Verhaltensbeobachtungen in kontrollierten Situationen durchgeführt.
- Das Verhalten in jeder dieser 15 Situationen wurde von jeweils 4 Beurteilern für jeden Zwilling (2 x 300) in Bezug auf 30 Adjektive eingeschätzt (≈ 1080000 Datenwerte)



Beispiel

- Trotz der für psychologische Forschung bereits recht großen Anzahl von Messwerten pro Individuum, bereiten diese zunächst kein Problem.
- Von Interesse sind für viele Fragestellungen ohnehin Zusammenfassungen dieser Datenpunkte.

Zusammenfassungen:

- Pro Zwilling über jede Situation, über bestimmte Adjektive über alle Beurteiler
- ... pro Situation über Adjektive und Beurteiler
- ... pro Adjektiv über Situationen und Beurteiler
- ... pro Beurteiler über Adjektive und Situationen
-

Beispiel: Besonderheit von Zwillingsdaten

Daten werden in unterschiedlichen Strukturen benötigt:

- a) Fallweise (jedes Zwillingsgeschwister ist ein Fall)
- b) Paarweise
- c) „Double entry“

Verschieden Transformationen der Daten werden benötigt:

- a) Rohdaten
- b) um Alters- und Geschlechtseffekte bereinigte Daten
- c) ipsativierte Daten
- d) ...

{Aggregate} x {Transformationen} x {Datenstrukturen}

Beispiel: Besonderheit von Zwillingdaten

Bisherige Lösung: eher adhoc, durch die „Verwertung“
getrieben.

Verwendung „wachsender“ Datendateien mit mitwachsenden
Dokumentationen.

Nachteil: Schlechte Nutzung durch „Dritte“.
Fehleranfälligkeit, aufwändige Kontrolle

{Aggregate} x {Transformationen} x {Datenstrukturen}

Entwicklungen, die über das klassische Zwillingsdesign hinausgehen

Bereits erwähnt:
Zunehmend
umfassendere
Datenerhebungen

No	Action	Measure				Duration
		OM	UBO	ES	RAT	
1	Taking blood samples					60 min.
2	Welcoming the twins, first objective measures (incl. actometers, photograph)	T	P			15 min.
3	Way to the experimental rooms		P			5 min.
4	In the waiting-room: List of emotional states 1: self-rating				P	3 min.
5	'Please introduce yourself' (presentation)			P		5 min.
6	Psychometric intelligence (APM, short) parallel: (EXP: short-list)	I	P			25 min.
7	Big Five Short-List: self-rating				P	5 min.
8	'4-picture-story': structure and tell			P		5 min.
9	TAT-pictures			P		15 min.
10	List of emotional states 2: self-rating				P	3 min.
11	PRF-Achievement Scale: self-rating				P	5 min.
12	Experimental temperament measures at the computer	T	P			15 min.
13	Break					5 min.
14	Self Monitoring Scale: self-rating				P	10 min.
15	Elementary Cognitive Tasks	I	P			20 min.
16	List of emotional states 3: self-rating				P	3 min.
17	Waiting-room situation with confederate		P			10 min.
18	'Please tell us a joke'			P		5 min.
19	Telephone-role play (situation 1)			P		5 min.
20	Telephone-role play (situation 2)			P		5 min.
21	"Get to know each other" (conversation with confederate)			P		10 min.
22	Memory-test (incl. moderate frustration)			P		5 min.
23	Logical reasoning (incl. moderate frustration & possibility of cheating)	I	P	P		5 min.
24	List of emotional states 4: self-rating				P	3 min.
25	Lunchbreak					55 min.
26	Joint twin-interview		P			30 min.
27	Back into the experimental rooms		P			5 min.
28	Encyclopaedia task (creativity)			P		5 min.
29	Rigging up a paper-tower			P		5 min.
30	'Headlines' (presentation task)			P		10 min.
31	Questionnaire on parenting				P	10 min.
32	'Divergent brick' (creativity)			P		5 min.
33	'Sing a song'			P		5 min.
34	List of emotional states 5: self-rating				P	3 min.
35	Break					10 min.
36	Separate twin interview					15 min.
37	Psychometric intelligence: LPS [parallel: EXP: Big Five short-list 2]	I	P			25 min.
38	Reactivity to stress experiment		P			30 min.
39	Return to the main office					5 min.
40	Rating task 'headlines'				P	10 min.
41	Questions and information, organisational matters, dismissal		P			35 min.
42	EXP: NEO-FFI rating on the respective participant				P	15 min.

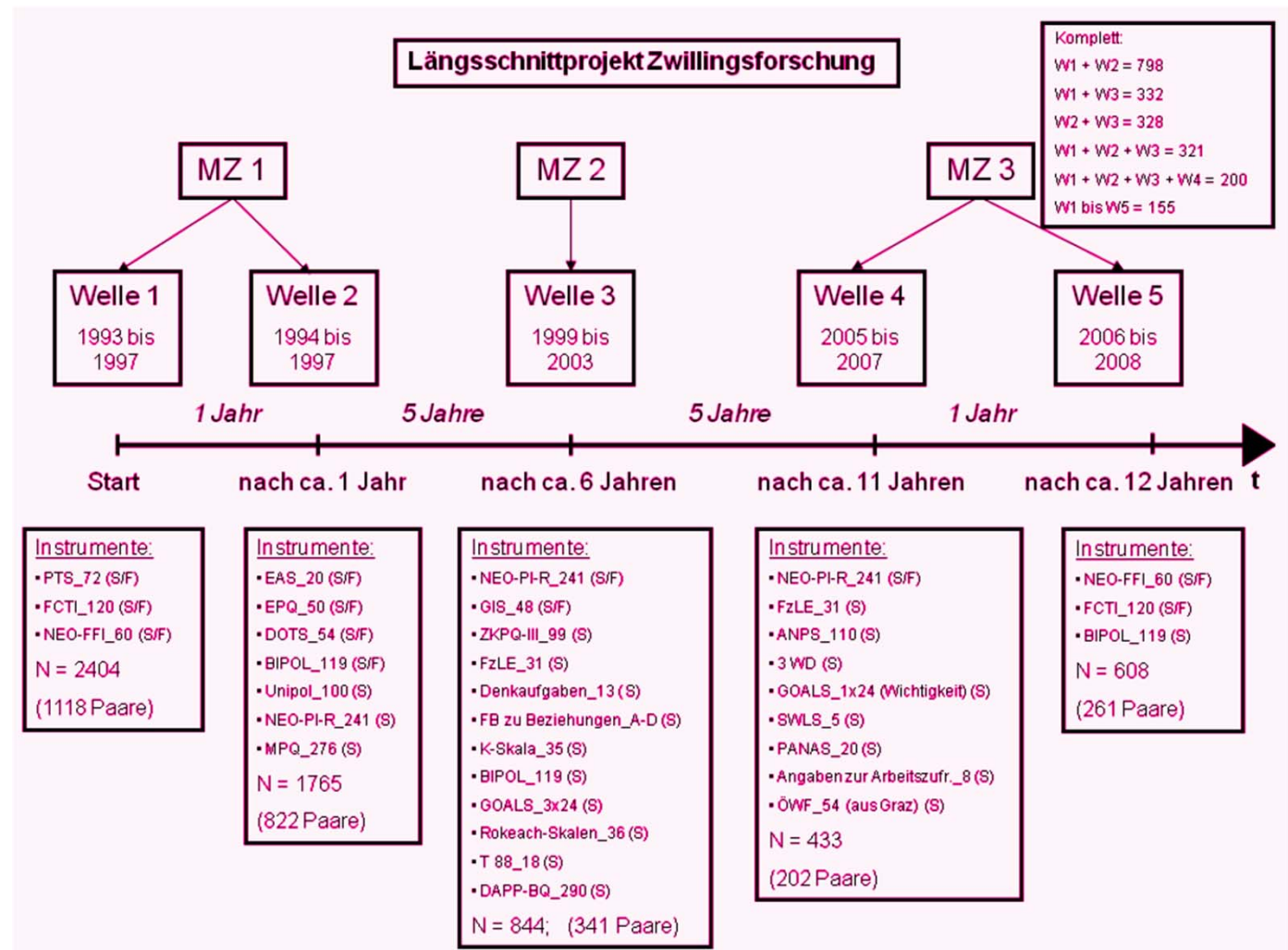
Entwicklungen, die über das klassische Zwillingdesign hinausgehen

Integration nicht numerischen Materials
(z.B. Videoaufzeichnungen)

Gegenwärtig haben wir umfangreiches Videomaterial digitalisiert, das ursprünglich auf VHS Videokassetten vorlag . Die Nutzung ist sehr aufwändig, da die Sequenzen nun in einer fest Ordnung auf Stapeln von DVD in einer festen Ordnung vorliegen und für weitere Analysen jeweils neu aufbereitet werden müssen.

Entwicklungen, die über das klassische Zwillingsdesign hinausgehen

Ausdehnung über die Zeit
(Längsschnitt)



Entwicklungen, die über das klassische Zwillingsdesign hinausgehen

Ausdehnung auf die Zwillingsfamilie: Eltern, Geschwister, Kinder, weitere Verwandte.

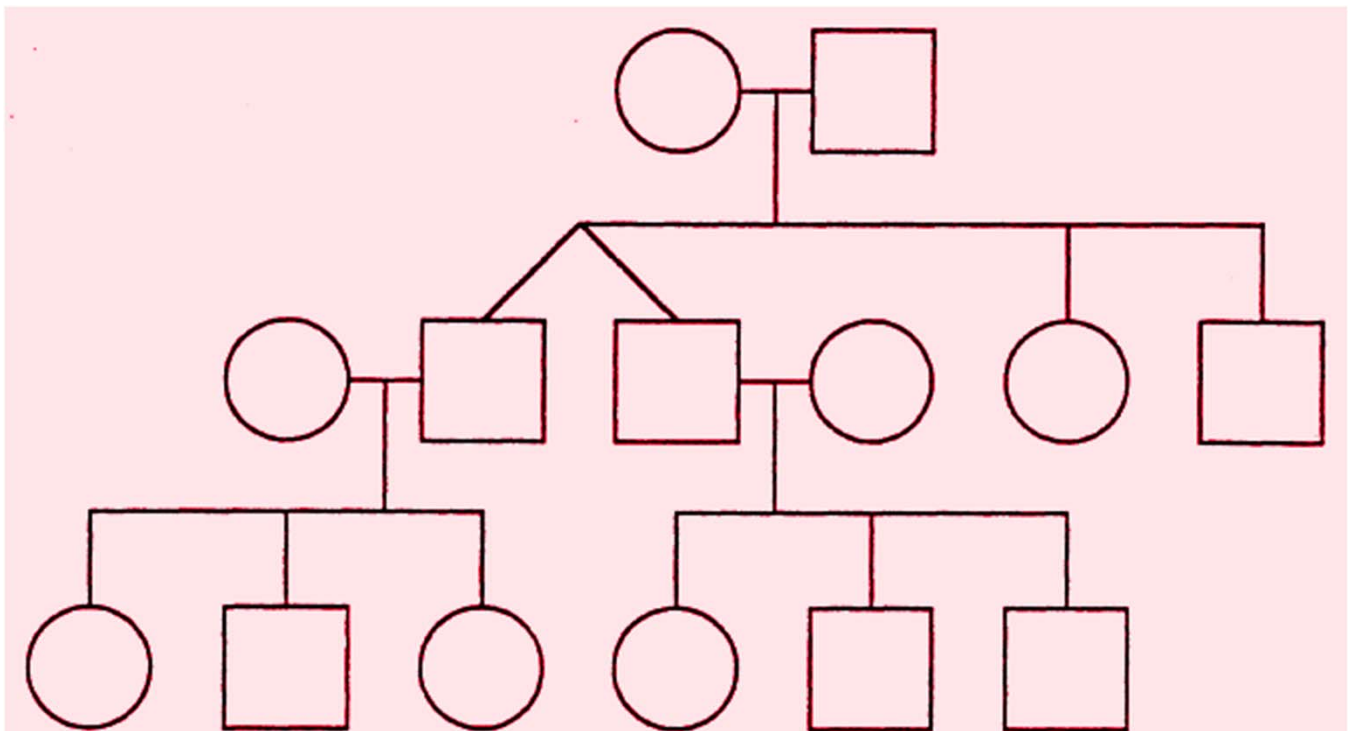


FIGURE 11-1. *Idealized family structure for the Virginia 30,000 study. Standard conventions are used for drawing pedigrees: Circles denote women and squares denote men. Horizontal bars between circles and squares denote spouses. Offspring from a mating are indicated by vertical lines connected by horizontal bars. Twins are connected by diagonal lines.*

Fazit

Es liegen komplexe Datensätze vor, deren Speicherung und Nutzung kein technisches Problem darstellt (Ausnahme vielleicht Videomaterial).

Auch in der psychologischen Forschung wird der Wert solcher Datensätze erkannt.

- Qualitätssicherung
- Datenhandling

Einzelne Arbeitsgruppen können solche Datensätze kaum in angemessener Zeit ausschöpfen.

Die Komplexität erschwert die Nachnutzung.

Fazit

In der psychologischen Forschung gibt es gelegentlich die Schwierigkeit, dass sensible personenbezogene Daten erhoben wurden.

Hier müssen Vorgehensweisen entwickelt werden, diese Daten hinreichend zu anonymisieren.
Schon bei der Planung müssen Möglichkeiten und Grenzen der Weitergabe von Daten geprüft werden (Ethikgutachten).

Es wird Fälle geben, in denen eine Nachnutzung von Daten nicht vertretbar ist.

Wie könnte ein Lösung aussehen

Bisher: Speicherung mit Hilfe von Statistikpaketen oder als Textdateien. Auf diese Weise werden Daten in der Regel auch zur Nachnutzung angeboten.

Wir benötigen spezielle Werkzeuge, die die Verwaltung, Qualitätskontrolle, Selektion und Transformation (Aufbereitung) von Daten unterstützen.

Überspitzt:

Wir haben für das Sammeln von Daten einige 100000€ ausgegeben.

Wir arbeiten hart für einen neuen Datensatz einige Mio. € ausgeben zu können,

..... um diese Daten dann auf Lochkarten zu stanzen.

Wie könnte ein Lösung aussehen

Deklaratives versus prozedurales Gedächtnis:

Es wäre sehr unhandlich zu versuchen, von Beginn an alle wahrscheinlich benötigten Transformationen von Daten vorzunehmen und abzuspeichern. Ein Werkzeug sollte in der Lage sein, solche Transformationen durchzuführen, wenn die entsprechenden Daten benötigt werden.

- Aggregation
- Transformation
- Formate
- Datenstrukturen

Wie könnte ein Lösung aussehen

Hierarchischer Zugriff auf die Daten, der die Datenstruktur abbildet.

Psychologische Datensätze weisen in der Regel eine hierarische Binnenstruktur der Art auf:

- Messzeitpunkt
- Befragte Person (z.B. Selbst- versus Bekanntenbeurteilung, Mutter)
- Instrument (Test, Fragebogen)
- Skala
- Transformation (Rohwerte, bereinigte Werte)
- Items (Fragen, Testaufgaben)

Wie könnte ein Lösung aussehen

Organisation von Fällen

Gerade in längsschnittlichen Datensätzen benötigen wir einfache Verfahren zur Auswahl von Fällen (z.B. Vergleich der Werte von Probanden, die nur zum ersten Messzeitpunkt teilgenommen haben mit denen, die zu allen Zeitpunkten teilgenommen haben).

Fehlende Werte

Routinen zur Behandlung von fehlenden Werten sollten in ein Datenverwaltungswerkzeug integriert sein.

Wie könnte ein Lösung aussehen

Datensätze verändern sich mit jeder Analyse.

Ein Werkzeug zur Datenverwaltung muss Mechanismen enthalten, derartige Veränderungen nach festgelegten Qualitätsstandards in den Datensatz einzupflegen.

Dies ist besonders für public use Files von Bedeutung.
Beispiel: Entwicklung von Kontrollskalen für etablierte Fragebogen.

Wie könnte ein Lösung aussehen

Manager versus Programmierer

Wir sollten hier an einem Paradigmenwechsel arbeiten. „Data handling“ soll im Regelfall nicht Aufgabe eines „Programmierer-Experten für Datensatz x“ sein, sondern durch den „Wissenschaftler-Nutzer“ selbst in kurzer Zeit zu bewerkstelligen sein.

Hierfür ist - über das Gesagte hinaus – eine enge Integration von Datendokumentation und Datenverwaltung erforderlich.

Wie könnte ein Lösung aussehen

Langfristig ist zu fordern, dass der Datenaufbereitung, Datenverwaltung und -nutzung durch die „scientific community“ auch in der Psychologie mehr Gewicht beigemessen wird.

Entsprechende Anforderungen sind bereits Bestandteil von DFG-Förderprogrammen, dies sollte weiter ausgebaut werden.

Ich danke für Ihre Aufmerksamkeit und Geduld und freue mich auf Ihre kritischen Fragen!