SSOAR

Open Access Repository
www.ssoar.info

Persönliche Codes bei Längsschnittstudien: ein Erfahrungsbericht

Pöge, Andreas

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:
Pöge, A. (2005). Persönliche Codes bei Längsschnittstudien: ein Erfahrungsbericht. *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, 56, 50-69. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-198538

Nutzungsbedingungen:
Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.
Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:
This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.
By using this particular document, you accept the above-stated conditions of use.

gesis Leibniz-Institut für Sozialwissenschaften

Mitglied der Leibniz-Gemeinschaft

Persönliche Codes bei Längsschnittstudien: Ein Erfahrungsbericht

von Andreas Pöge [1]

Zusammenfassung

Anhand der Münsteraner Längsschnittdaten des DFG-Projektes „Jugendkriminalität in der modernen Stadt“ *wird die Fragebogenzuordnung zwischen den Erhebungswellen mit Hilfe eines Verfahrens über persönliche Codes dargestellt und dessen Auswirkungen geschildert. Solch ein Verfahren ist im Hinblick auf Fehleranfälligkeit und Uneindeutigkeit der Codes und einer Verzerrung der daraus resultierenden Paneldaten nicht unproblematisch. Trotz dieser Probleme ist bei der gewählten Erhebungsmethode, Schülerinterviews im Klassenverband, und der besonderen Berücksichtigung datenschutzrelevanter Belange das genannte Verfahren alternativlos.*

Abstract

On the basis of longitudinal data taken from the DFG project “Juvenile delinquency in modern town” *the assignment of questionnaires between the survey waves by means of personal codes is described and the effects are analysed. Such a method is problematic because the codes tend to be ambiguous, prone to errors and the resulting panel data is biased. Despite these problems there is no alternative to the chosen method of assignment with regard to the survey method (interviews with pupils in their classrooms) and in consideration of data protection.*

1 ***Andreas Pöge***, M.A., Universität Trier, ist Wissenschaftlicher Mitarbeiter des DFG-Projektes „Jugendkriminalität in der modernen Stadt“. Universität Trier, FB IV – Soziologie, Universitätsring 15, 54286 Trier und Westfälische Wilhelms-Universität Münster, Institut für Kriminalwissenschaften, Abteilung IV – Kriminologie, Bispinghof 24/25, 48143 Münster.
E-Mail: poge@uni-muenster.de.

1 Einleitung

Längsschnitt- bzw. Paneluntersuchungen in der Umfrageforschung bringen zwangsläufig das Problem der Verbindung von Daten aus mehreren Erhebungswellen mit sich. Oftmals werden hierbei die Umfrageinformationen mit schon bekannten oder gleichzeitig erhobenen Adressdaten oder anderen datenschutzrechtlich sensiblen Informationen zusammengespielt, um über diese persönlichen Daten die gewünschten Zuordnungen über die Zeit gewährleisten zu können. Anonymität kann bei einem solchen Verfahren meist - mit oftmals negativen Auswirkungen auf die Rücklaufquoten - nicht gewährleistet werden. Im Zuge der nachfolgend näher vorgestellten DFG-geförderten Panelstudie *„Jugendkriminalität in der modernen Stadt“*[2] wurden Jugendliche im Alter von 13 bis 17 Jahren zu Opferwerdung, selbstberichteter Delinquenz sowie verschiedensten Bereichen des täglichen Lebens befragt. Die Befragungen fanden als schriftliche Fragebogeninterviews im Klassenverband in einjährigem Rhythmus in den Städten Münster und Duisburg statt. Angesichts des geringen Alters der Probanden sollten datenschutzrelevante Belange besonders sorgfältig beachtet werden. Möglicherweise wären nicht alle Personen in der Lage, ihre eigenen Interessen zu erkennen und zu vertreten. Außerdem wollte man besondere datenschutz-psychologische Vorsicht walten lassen, um die Stichprobenrückläufe (auch im Hinblick auf die nötige Zustimmung der Erziehungsberechtigten) nicht zu gefährden. Mit der *Landesbeauftragten für Datenschutz und Informationsfreiheit Nordrhein-Westfalen* wurde deshalb ein Verfahren vereinbart, mit welchem unter Zuhilfenahme von persönlichen Codes ein Zusammenspielen der Daten aus den einzelnen Erhebungswellen, bei Anonymität der Probanden, ermöglicht werden sollte. Dieser Bericht stellt das geplante Zuordnungsverfahren vor und dokumentiert dabei aufgetretene Probleme und deren Lösungsstrategien.

2 Das Projekt „Jugendkriminalität in der modernen Stadt“

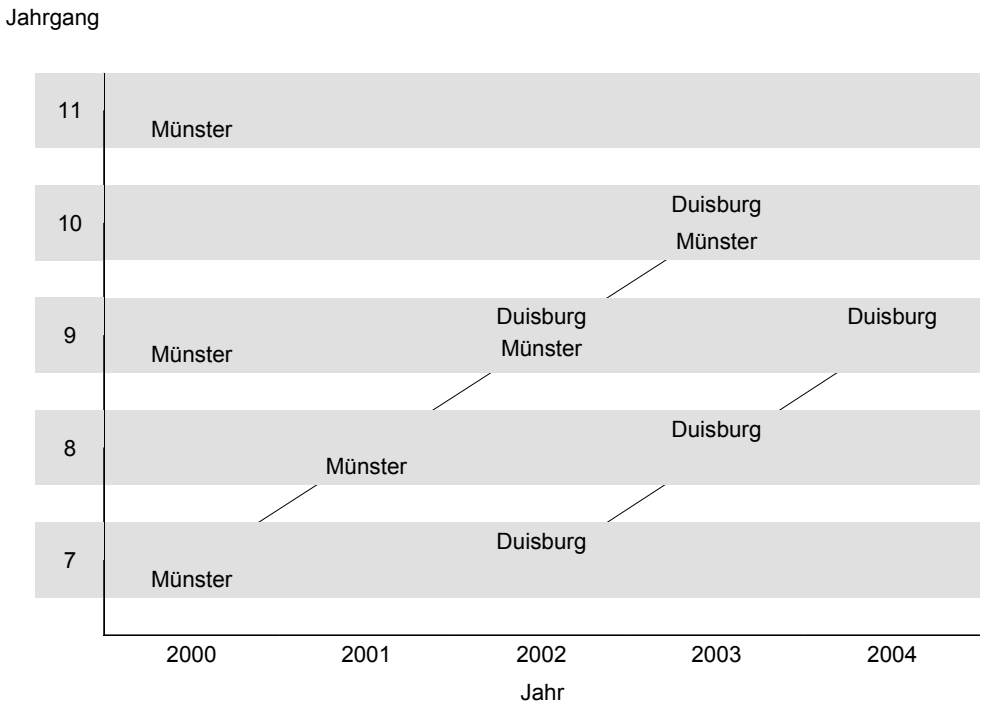
Die Ziele des DFG-Projektes *„Jugendkriminalität in der modernen Stadt“* liegen unter anderem in der Untersuchung der Entstehung und Entwicklung delinquenter bzw. devianter Handlungsstile bei Jugendlichen und deren sozialer Kontrolle (und zwar formeller Art durch Polizei und Justiz und informeller Art durch Schule und Familie).[3]

---

2 Projektleitung: Prof. Dr. ***Klaus Boers***, Institut für Kriminalwissenschaften, WWU Münster und Prof. Dr. ***Jost Reinecke***, Fakultät für Soziologie, Universität Bielefeld, DFG-Förderzeichen: Bo 1234/6-1, 6-2 und Re 832/4-1, 4-2.

3 Zu ersten Ergebnissen siehe ***Boers, Reinecke, Wittenberg*** und ***Motzke*** (2002).

Abbildung 1 Erhebungsdesign 2000-2004



Untersucht werden unter anderem die dynamischen Zusammenhänge zwischen Sozialstruktur und Delinquenz, sowie die Art der Moderierung von Selektionsprozessen formeller Kontrollinstanzen über die Jugendphase hinweg. Insbesondere wird hierbei den Fragen nachgegangen, ob und inwieweit Delinquenz episodaler Teil der Dynamik jugendlicher Milieus[4] ist oder ob und inwieweit sich Delinquenz in der institutionellen Eigendynamik selektiver Kontrollprozesse zu kriminellen Karrieren verdichtet.

Der methodische Zugang zu diesen Problemkomplexen erforderte ein für diesen theoretischen und inhaltlichen Kontext in Deutschland zum ersten Mal geplantes kohortenspezifisches Längsschnittdesign, das in zwei westdeutschen Städten unterschiedlicher Größe und Sozialstruktur, Münster und Duisburg, umgesetzt wurde. Die durch wiederholte Schülerbefragungen erhobenen Längsschnittinformationen (Paneldaten) sollen sowohl die Analyse von Ursache-Wirkungsbeziehungen zwischen den zentralen Untersuchungsvariablen (einschließlich möglicher Rückkopplungsprozesse) als auch die Analyse von Stabilität und Veränderung der individuellen Kriminalitätsraten ermöglichen.

4 Zu ersten Ergebnisse siehe ***Boers*** und ***Pöge*** (2003).

Seit dem Jahr 2000 wurden in einjährigen Abständen an Münsteraner[5], seit dem Jahr 2002 auch an Duisburger Schulen Fragebogeninterviews durchgeführt. Entsprechend der methodischen Konzeption des Forschungsprojektes als Panelstudie, wurden dabei dieselben Personen wiederholt befragt. Der genaue Erhebungsplan kann Abbildung 1 entnommen werden.[6]

2.1 Durchführung der Befragungen

Aus Gründen des Umfangs und der bisher zur Verfügung stehenden Daten (erste Erhebung Münster im Jahr 2000, erste Erhebung Duisburg im Jahr 2002) wird im Folgenden nur auf die Münsteraner Daten eingegangen.

In der ersten Münsteraner Schülerbefragung im Jahr 2000 wurde eine Vollerhebung der Schülerinnen und Schüler der Jahrgangsstufe 7 angestrebt. Ergänzt wurde die Studie durch Stichproben aus den Klassen des 9. und 11. Jahrgangs sowie den Eingangsklassen an den Münsteraner Berufskollegs.[7] Die zweite Schülerbefragung im Jahr 2001 strebte eine Wiederbefragung der Schülerinnen und Schüler aus der ersten Studie an. Alle Schülerinnen und Schüler der Jahrgangsstufe 8, die Schulen besuchten, die an der ersten Erhebung teilgenommen hatten, sollten erneut wiederbefragt werden.[8] Der geschilderten Erhebungssystematik folgend sollten in der dritten und vierten Schülerbefragung im Jahr 2002[9] und 2003[10] alle Schülerinnen und Schüler der Jahrgangsstufe 9 bzw. 10, die eine Schule besuchten, die an den ersten beiden Erhebungen teilgenommen hatte, wiederbefragt werden. Der eingesetzte Fragebogen zu Opferwerdung, selbstberichteter Delinquenz, Ethnizität[11], Erziehungsstilen[12], Konfliktverhalten, Kriminalitätseinstellungen[13] sowie zu Lebens-, Freizeit- und Konsumstilen war für eine schriftliche, anonyme Befragung der Schülerinnen und Schüler konzipiert. So weit wie möglich wurden Fragen verwendet, die bereits in anderen Jugendstudien eingesetzt wurden. Die Fragebögen der

5 Zu ersten Ergebnissen siehe ***Boers*** und ***Kurz*** (2000).

6 Zur Dokumentation der Münsteraner Panel-Studie 2000-2003 vgl. ***Pöge*** (2005b).

7 Zur Dokumentation der Münsteraner Studie aus dem Jahr 2000 vgl. ***Motzke*** und ***Wittenberg*** (2004).

8 Zur Dokumentation der Münsteraner Studie aus dem Jahr 2001 vgl. ***Wittenberg*** (2004a).

9 Zur Dokumentation der Studie aus dem Jahr 2002 vgl. ***Wittenberg*** (2004b).

10 Zur Dokumentation der Studie aus dem Jahr 2003 vgl. ***Wittenberg*** (2004c).

11 Zu ersten Ergebnissen siehe ***Pöge*** (2005a).

12 Zu ersten Ergebnissen siehe ***Raithel*** (2002).

13 Zu ersten Ergebnissen siehe ***Pöge*** (2002).

einzelnen Jahre sind aufgrund des Panelansatzes der Studie weitgehend identisch.[14] Vor den jeweiligen Haupterhebungen wurden die Eltern und Schüler über die Untersuchungsziele informiert und auf die Freiwilligkeit der Teilnahme hingewiesen. Die schriftlichen Befragungen fanden im Klassenverband statt, der zeitliche Rahmen erstreckte sich über eine Schuldoppelstunde. In wenigen Einzelfällen kam es jedoch vor, dass die Beantwortung der Fragen mehr Zeit in Anspruch nahm, insbesondere wenn sprachliche Schwierigkeiten die Beantwortung beeinträchtigten.

2.2 Datengrundlage der Querschnitte

Einige grundlegende Informationen der Münsteraner Querschnitts-Datengrundlage können Tabelle 1 entnommen werden. Die Stichprobengröße der einzelnen Querschnitte liegt zwischen 1.819 und 1.949 Personen. Grundlage hierfür sind Schülerinnen und Schüler, die Schulen besuchten, welche den Befragungen in jedem Zeitpunkt zugestimmt hatten. Unberücksichtigt blieben Fälle, deren Fragebögen Plausibilitätskontrollen nicht standhielten. Um den Rücklauf bestimmen zu können, wurde die Schulstatistik herangezogen, wiederum Bezug nehmend auf die Schulen, die in jedem Befragungszeitpunkt an der Umfrage teilnahmen. Die sich hieraus ergebenden Quoten liegen, wie in Tabelle 1 ersichtlich, zwischen 86 und 88 Prozent, sind also als sehr gut zu bezeichnen.

Tabelle 1 Zusammenfassung der Befragungen in Münster 2000-2003

	Münster 2000 (t_1)	Münster 2001 (t_2)	Münster 2002 (t_3)	Münster 2003 (t_4)
Jahrgang	7	8	9	10
Schulstatistik (N)	2.105	2.181	2.251	2.077
Stichprobe (N)	1.850	1.915	1.947	1.819
Rücklaufquote (%)	88	88	86	88
Geschlecht (Stichprobe)				
Anteil Mädchen (%)	49	50	51	50
Anteil Jungen (%)	51	50	49	50
Alter (Stichprobe)				
Durchschnittsalter in Jahren	13	14	15	16

14 Muster der verwendeten Fragebögen können über den Autor bezogen werden.

3 Geplante Fragebogenzuordung mit persönlichen Codes

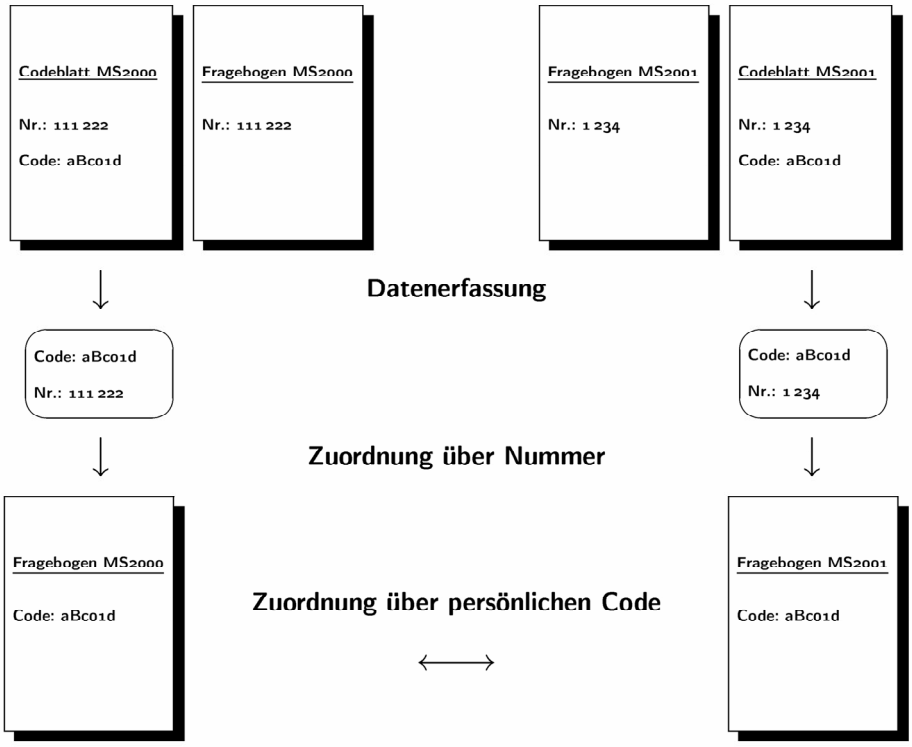
Um eine Zuordnung der Fragebögen aus den verschiedenen Erhebungswellen zu ermöglichen, wurden über Codeblätter abgefragte Codes verwendet. In jeder Erhebungswelle wurden solche Codeblätter den Schülerinnen und Schülern zum Ausfüllen vorgelegt, so dass bei einem stabilen Antwortverhalten der Schüler, die Codeblätter, die von ein und derselben Person in den verschiedenen Wellen ausgefüllt wurden, denselben Code aufweisen mussten (siehe Abbildung 2). Im Einzelnen waren die geplanten Schritte wie folgt: Während der Befragung im Klassenverband füllen die Schülerinnen und Schüler zunächst die Codeblätter und die Fragebögen sorgfältig aus. Auf einem Codeblatt und einem Fragebogen, die dieselbe Person ausfüllt, befindet sich dieselbe Nummer. Die Codeblätter und Fragebögen werden getrennt voneinander in Briefumschläge gesteckt, verschlossen und abgegeben. Codeblätter und Fragebögen gehen getrennt voneinander in die Datenerfassung, das einzige Verbindeglied stellt die Nummer auf Codeblatt und Fragebogen dar. Daraufhin werden Codeblätter und die Original-Fragebögen vernichtet. Die persönlichen Codes können dann mit den Fragebogendaten anhand der Codeblatt- bzw. Fragebogennummer zusammengespielt werden. Dieses Procedere sollte in jeder Erhebungswelle wiederholt werden, wodurch die Fragebogendaten über den Code einander zugeordnet werden könnten, ohne dass jemals ursprüngliche Informationen aus den Codeblatt-Fragen mit Informationen aus den Fragebögen zusammengeführt werden müssten. Dieses Verfahren sollte somit die völlige Anonymität der befragten Personen gewährleisten. Die persönlichen Fragen, die auf den Codeblättern gestellt wurden, betrafen die Privatsphäre der Befragten und deren Familien und mussten mit dem Datenschutz abgestimmt werden.

Während der ersten Befragung im Jahr 2000 sollte auf den Codeblättern[15] hierbei Folgendes handschriftlich notiert werden:

1. Der letzte Buchstabe der natürlichen Haarfarbe des Vaters
2. Der erste Buchstabe des Vornamens der Oma mütterlicherseits
3. Der letzte Buchstabe der eigenen natürlichen Haarfarbe
4. Die beiden Tagesziffern des eigenen Geburtstages
5. Der letzte Buchstabe der eigenen Augenfarbe

15 Muster der verwendeten Codeblätter können über den Autor bezogen werden.

Abbildung 2 Geplantes Zuordnungsverfahren in Münster 2000/2001



Im Jahr 2001 wurde das Codeblatt um die Fragen nach Klassenwechsel und Schulwechsel ergänzt, hierbei konnte eine Antwortvorgabe angekreuzt werden. Für die Befragung in 2002 wurde das Codeblatt um die Frage nach einer Befragungsteilnahme im vorigen Jahr erweitert. In 2003 wurde das Layout grundlegend geändert, so dass kein handschriftliches Ausfüllen mehr erfolgen musste, sondern alle Antwortvorgaben zum Ankreuzen aufgeführt wurden. Außerdem wurde eine weitere Frage aufgenommen, nämlich die nach dem ersten Buchstaben des eigenen Vornamens.[16] Darüber hinaus wurde die Frage nach einem Klassenwechsel zum besseren Verständnis umformuliert.

Das geschilderte Verfahren mit diesen Fragen konnte nur unter zwei Grundvoraussetzungen funktionieren. Als erste Grundvoraussetzung mussten die Befragungsteilnehmerinnen und -teilnehmer in den unterschiedlichen Wellen *exakt* denselben Code aufweisen, da die Zuordnung über den gesamten Code erfolgen sollte. Dies bedeutete insbesondere, dass die Codefragen auf den Codeblättern der Befragungsjahre von den Probanden *exakt* gleich beantwortet werden mussten. Ergäben sich

16 Diese Frage blieb jedoch ohne Bedeutung, da in Münster auf eine weitere Befragungswelle verzichtet wurde.

Differenzen im Antwortmuster ein und derselben Person bei den Codefragen der unterschiedlichen Jahre, könnte eine Zuordnung in der geplanten Art und Weise (über den kompletten Code) nicht mehr stattfinden! Als zweite Grundvoraussetzung musste gewährleistet sein, dass jede Person einen eindeutigen Code aufwies. Hätten mehrere Personen denselben Code, könnte eine Zuordnung in der geplanten Art und Weise (ohne Hinzunahme weiterer Informationen) nicht mehr stattfinden!

3.1 Durchführung des Verfahrens

Bei der technischen Durchführung der Zuordnungen der Fragebögen aus den einzelnen Wellen traten Probleme auf, die die Grundvoraussetzungen des Verfahrens verletzten und Änderungen notwendig machten. Eine Analyse der Daten, die über die ausgefüllten Codeblätter zur Verfügung stehen, lässt in Bezug auf die gebildeten Codes erkennen, dass in jeder Welle nicht jede Schülerin und jeder Schüler einen eindeutigen (d. h. nur einmal vorkommenden) Code aufweist. Wie in Tabelle 2 ersichtlich, treten in jeder der vier Erhebungswellen jeweils nur bei ungefähr drei Vierteln der Probanden eindeutige Codes auf (z. B. gibt es in 2000 nur 1.497 einmal vorkommende Codes). Bei einem Viertel der Befragten weisen unterschiedliche Personen einen gleichen Code auf (in 2001 trat sogar ein Code sechs Mal auf). Diese Personen sind über den Code ohne Hinzunahme weiterer Informationen nicht eindeutig zu identifizieren. Im Folgenden soll dieses Problem als Problem der *Identifizierung* bezeichnet werden. Wählt man aus jeder Welle nur die *eindeutigen* Codes aus und ordnet zwischen den Wellen diese Codes einander zu, so finden sich zwischen jeweils zwei Erhebungszeitpunkten lediglich rund 600 Zuordnungen mit *exakt* gleichem Code (für t_3/t_4 siehe Tabelle 5). Man kann also davon ausgehen, dass gleiche Personen in unterschiedlichen Wellen nicht denselben Code aufweisen. Die Hauptursache hierfür liegt darin, dass die Schülerinnen und Schüler den Code nicht exakt in jeder Welle reproduzieren konnten. Dieses Problem soll im Folgenden als Problem der *Reproduktion* bezeichnet werden. Die Grundprobleme des Verfahrens sind also:

- Mehrfaches Auftreten gleicher Codes innerhalb der Wellen (*Identifizierung*)
- Zu geringes Auftreten gleicher Codes zwischen den Wellen (*Reproduktion*)

Die oben aufgeführten zwei Grundvoraussetzungen des geplanten Verfahrens sind offensichtlich (und schwerwiegend) verletzt. Nachfolgend werden die Ursachen für die aufgetretenen Probleme analysiert und die schließlich angewendete Problemlösestrategie erläutert.

Tabelle 2 Häufigkeiten der Codes

	2000 (t_1)		2001 (t_2)		2002 (t_3)		2003 (t_4)	
Häufigkeit	Anz. Codes	Gesamt	Anz. Codes	Gesamt	Anz. Codes	Gesamt	Anz. Codes	Gesamt
1	1.497	1.497	1.538	1.538	1.533	1.533	1.449	1.449
2	208	416	161	322	176	352	151	302
3	28	84	21	63	24	72	21	63
4	3	12	1	4	6	24	4	16
5	1	5	-	-	1	5	1	5
6	-	-	1	6	-	-	-	-
Summe[a]	1.737	2.014	1.722	1.933	1.740	1.986	1.626	1.835

a Die Anzahlen für die Codeblätter können von den Stichprobengrößen abweichen. Zum einen wurden die Codeblätter von Schulen, die nicht an allen Befragungszeitpunkten teilnahmen, zum anderen Codeblätter deren Fragebögen durch die Plausibilitätskontrollen fielen, verwendet.

3.2 Probleme der Zuordnungsmethode

Die Ursachen für die aufgetretenen Zuordnungsprobleme der Identifizierung und Reproduktion sind unterschiedlicher Art und überlappen sich teilweise. Zusammenfassend lassen sich folgende Hauptursachen ausmachen:

- Datenschutzbedingte Vorgaben, die das Ausfüllen der Codeblätter durch komplizierte Fragestellungen erschwerten
- Zu hohe Schwierigkeit der Codefragen, die eine Reproduzierbarkeit erschwerte
- Ungünstige Auswahl der Fragen bzw. Antwort-Buchstaben, die eine eindeutige Identifizierung der Codes verhinderte
- Mangelnde Reproduktion der Codes durch die Schüler aufgrund von Flüchtigkeitsfehlern oder beabsichtigten Falschangaben beim Ausfüllen der Codeblätter
- Layoutmängel und dadurch unleserliches Ausfüllen der Codeblätter

Datenschutzvorgaben

Durch die nötige Abstimmung mit dem Datenschutz musste bei der Fragengestaltung Rücksicht genommen werden. Dies resultierte in einem Kompromiss dergestalt, dass nicht bei den Antworten auf jede Frage (wie ursprünglich geplant) die ersten Buchstaben notiert werden konnten, sondern zum Teil die jeweils letzten. Die Begründung hierfür lag in der Befürchtung, dass, falls die im Klassenverband ausgefüllten Codeblätter in die Hände von Personen mit genügend Information über die Schülerinnen und Schüler fielen, diese zu identifizieren wären. Dann hätten die datenschutzrelevanten Informationen aus den Fragebögen im ungünstigsten Fall von unbefugten Personen den einzelnen Schülerinnen und Schülern zugeordnet werden können und diesen hieraus Nachteile entstehen können. Der Wechsel von „erster Buchstabe…“ und „letzter Buchstabe…“ scheint jedoch möglicherweise für die ausfüllenden Personen zu unübersichtlich gewesen zu sein.

Zu hohe Schwierigkeit der Codefragen

Da ein erheblicher Teil der Schülerinnen und Schüler nicht in der Lage war, das Codeblatt in beiden Jahren richtig oder zumindest gleich auszufüllen, muss die Frage gestellt werden, ob die Codefragen zu schwierig waren. Möglicherweise waren nicht alle Schülerinnen und Schüler der vier befragten Jahrgangsstufen in der Lage, die oben genannten Fragen richtig zu beantworten oder sich bei der Beantwortung genügend zu konzentrieren. Zwar waren einige Fragen schwieriger zu beantworten, z. B. Haarfarbe des Vaters oder Vorname der Oma mütterlicherseits, falls persönliche Gründe die Beantwortung schwierig oder unmöglich machten (z. B. beim Fehlen einer solchen Bezugsperson). Jedoch waren die Interviewerinnen und Interviewer diesbezüglichen angewiesen worden, die problematischen Fälle vor dem Ausfüllen anzusprechen bzw. in diesen Fällen Antwortvorgaben zu nennen.

Ungünstige Auswahl und Anzahl an Fragen für die Identifizierung

Neben der Tatsache, dass Codes, die in beiden Jahren gleich sein müssten, unterschiedlich sind, ist das Auftreten von Mehrfachcodes das gewichtigste Problem. Offensichtlich eignet sich der Code, so wie er konzipiert ist, nicht dazu, die ca. 2.000 Umfrageteilnehmerinnen und -teilnehmer eindeutig zu bestimmen. Dieser Code setzt sich, wie oben erläutert, aus Buchstaben bzw. Zahlen zusammen, die aus den Antworten der fünf Codefragen hervorgehen. Die Anzahl der Kombinationsmöglichkeiten, die sich aus diesen Zahlen und Buchstaben ergeben, ist nicht zahlreich genug, als dass mit genügender Wahrscheinlichkeit jede Schülerin und jeder

Schüler einen einzigartigen Code erhält.[17] Um eine Identifizierung der Codeblätter bzw. Fragebögen schon in *einer* Welle zu ermöglichen, müssen zusätzliche Informationen hinzugezogen werden. Eine Lösung für dieses Problem kann nur in der Modifizierung der Fragen bestehen, so dass eine höhere Anzahl von Antwortbuchstaben bzw. -zahlen erreicht wird, oder in der Erhöhung der Anzahl der Codevariablen bzw. -fragen selbst. Eine Änderung der Fragen hätte den großen Nachteil, dass der Code über die Zeit dann nicht mehr derselbe sein kann, also Bögen aus späteren Wellen nicht mehr über den Code zugeordnet werden könnten. Eine Erhöhung der Anzahl der Codevariablen hat den Nachteil, dass dadurch natürlich auch die Fehleranfälligkeit des Codes weiter steigen und somit das Problem der Reproduktion eher verschärft würde.

Beabsichtigte Falschangaben und Flüchtigkeitsfehler

Bei der Beantwortung der einzelnen Fragen auf dem Codeblatt können selbstverständlich sowohl beabsichtigte wie unbeabsichtigte Fehler auftreten, so dass der Code ein und desselben Schülers oder derselben Schülerin aus zwei Jahren unterschiedlich sein kann, obwohl er gleich sein müsste. Grundsätzlich kann dieses Problem bei jeder Frage auftreten, absichtliche Falschangaben können nicht vermieden werden, jedoch sollten leicht und eindeutig zu beantwortende Fragen unbeabsichtigte Fehler vermeiden helfen.

Layoutmängel und dadurch unleserliches Ausfüllen der Codeblätter

Ein nicht vom Forschungsteam direkt beeinflussbares Problem, welches in diesem Ausmaß nicht zu erwarten war, ist das zum Teil unleserliche Ausfüllen der Codeblätter. Dies führte dazu, dass in einer Reihe von Fällen die Codes nicht eindeutig zu entziffern waren, somit nicht oder nur eingeschränkt für die EDV zur Verfügung standen und deshalb zu fehlerhaften bzw. keinen Zuordnungen führten. Eine Lösung dieses Problems ist für die Datensätze 2000, 2001 und 2002 nicht möglich, jedoch

17 Allgemein kann hierzu folgendes Problem formuliert werden: Bei einem Zufallsexperiment mit „Zurücklegen“ wird aus einer Menge der Größe *N* von unterschiedlichen Codes eine bestimmte Anzahl *m* von Codes gezogen. Die Frage ist nun, wie groß die Codemenge sein muss, damit die Wahrscheinlichkeit *P(N,m)*, dass kein Code mehrfach gezogen wird, mindestens einen bestimmten Wert, z. B. 0,95, übersteigt. Eine allgemeine, näherungsweise Lösung lautet:

$$N(m) \approx 9{,}74x^2 - 9{,}4x + 1{,}1. \qquad (1)$$

In unserem Fall, bei ca. 2.000 zu identifizierenden Personen lautet die Lösung näherungsweise 39 Millionen. Die Codemenge *N* ergibt sich aus der Multiplikation der Merkmalsausprägungen der einzelnen Codestellen. Im vorliegenden Fall reicht die Anzahl der Ausprägungen nicht aus, um die erforderliche Größe der Codemenge *N* zu erreichen (Zur Herleitung vgl. ***Pöge*** 2005b, S. 63.)

wurde das Codeblattlayout dergestalt geändert, dass in der Erhebungswelle 2003 kein handschriftliches Ausfüllen mehr nötig war, sondern nur noch Antwortvorgaben angekreuzt werden mussten. Ein solches Layout mit vorgegebenen Antwortkästchen, die nur angekreuzt werden müssen, erwies sich als sehr viel vorteilhafter. Zwar verschlechterte sich dadurch die Übersichtlichkeit des Codeblattes leicht, jedoch wiegt der Vorteil der besseren Lesbarkeit dieses Problem wieder auf.

4 Zuordnungsverfahren mit manuellem Handschriftenvergleich

Wie oben aufgezeigt, hätte das ursprünglich geplante Verfahren nur zu einer Zuordnung von rund 600 Personen geführt (für t_3/t_4 siehe Tabelle 5). Beide Grundvoraussetzungen des Verfahrens sind nicht ausreichend gewährleistet, um eine Zuordnung eines befriedigend großen Teils der Probanden zu ermöglichen. Um weitere Zuordnungen realisieren zu können, musste das ursprüngliche Zuordnungsverfahren abgeändert werden. Das Problem der Identifizierung erforderte die Hinzunahme weiterer Informationen zum Code, das Problem der Reproduktion erforderte das Zuordnen von Personen über „fehlerhafte“ Codes. Nur so konnten weitere Zuordnungen gefunden werden. Um Fehler im Code zulassen zu können, musste auf eine Zuordnung über die Codes als *komplette* Zeichen-Zahlenkette verzichtet werden. Bei der Identifizierung sowohl der einzelnen Codes, als auch bei deren Zuordnung zwischen den Wellen wurden darüber hinaus die Schulinformation[18] und das Geschlecht der Probanden hinzugenommen. Das schließlich angewendete, *fehlertolerante Verfahren mit manuellem Handschriftenvergleich* bestand aus drei Schritten:

In einem *ersten Schritt* wurden maschinell alle exakt übereinstimmenden Codes aus zwei Erhebungswellen herausgefunden. Die zusammengehörigen Fragebögen und Codeblätter wurden daraufhin einer manuellen Handschriftenkontrolle unterzogen, wobei die offensichtlich nicht passenden Zuordnungen aussortiert wurden. Die passenden Fragebogennummern wurden aus dem Datensatz genommen, so dass sie für die nachfolgenden Schritte nicht mehr zur Verfügung standen. Im *zweiten Schritt* wurde nach Codeübereinstimmungen unter Zulassung eines Fehlers und im *dritten Schritt* unter Zulassung von zwei Fehlern gesucht und die zugehörigen Bogennummern herausgeschrieben. Auch in diesen beiden Schritten wurden als Validierung der Zuordnungen Handschriftenvergleiche durchgeführt, die offensichtlich nicht passenden Zuordnungen verworfen und vor der Durchführung des nächsten Schrittes

18 Die Hinzunahme der Schulinformation war beim ursprünglichen Verfahren für die Identifizierung einkalkuliert worden; in diesen Ausführungen wurde auf eine Darstellung bewusst verzichtet, um die beschränkte Leistungsfähigkeit des Codes stärker zu betonen.

die passenden Nummern aus dem Datensatz entfernt. Das geschilderte Verfahren ist codegeleitet und hierarchisch. Zum einen ist das erste Kriterium für eine Zuordnung nach wie vor der Code, es wurden nur offensichtliche Fehlzuordnungen ausgesondert, zum anderen wird durch die schrittweise Durchführung gewährleistet, dass weniger Fehler im Code bei der Zuordnung bevorzugt werden.

4.1 Realisierte Paneldatensätze Münster 2000-2003

Mit Hilfe des oben beschriebenen Zuordnungsverfahrens mit manuellem Handschriftenvergleich konnten die in Tabelle 3 aufgeführten Datensatzgrößen erreicht werden. So war es möglich, in den Daten aus jeweils zwei aufeinander folgenden Erhebungszeitpunkten zwischen ca. 1.270 und ca. 1.400 Zuordnungen zu finden. Für drei aufeinander folgende Erhebungspunkte ergaben sich 997 bzw. 1.075 und für alle vier Zeitpunkte 813 Zuordnungen.

Tabelle 3 Paneldatensätze Münster 2000-2003

Datensatz	*N*	Beschreibung
$P_{t1,t2}$	1.271	Zwei-Wellen-Panel Münster 2000/2001
$P_{t2,t3}$	1.373	Zwei-Wellen-Panel Münster 2001/2002
$P_{t3,t4}$	1.406	Zwei-Wellen-Panel Münster 2002/2003
$P_{t1,t2,t3}$	997	Drei-Wellen-Panel Münster 2000/2001/2002
$P_{t2,t3,t4}$	1.075	Drei-Wellen-Panel Münster 2001/2002/2003
$P_{t1,t2,t3,t4}$	813	Vier-Wellen-Panel Münster 2000/2001/2002/2003

4.2 Bewertung der Ausschöpfung

Will man eine Bewertung der Ausschöpfungsquote anhand der realisierten Paneldatensätze vornehmen, so muss zunächst eine Unterscheidung zwischen *Zuordnungsverfahren* und *gesamter Erhebungsmethode* getroffen werden. Da zur eigentlichen Grundgesamtheit der Münsteraner Schulbefragung alle Schülerinnen und Schüler gehören, die zu den Befragungszeitpunkten auf eine Münsteraner Schule gingen, könnte die Bewertung der gesamten Erhebungsmethode anhand der Schülerzahlen der Schulstatistik erfolgen. Dieser Fragestellung soll hier jedoch nicht weiter nachgegangen werden, vielmehr soll das Zuordnungsverfahren begutachtet werden. Es soll die Frage im Vordergrund stehen, wie viele der Fragebögen, die hätten zugeordnet werden können, mit dem oben vorgestellten, tatsächlich angewendeten hierarchischen Zuordnungsverfahren tatsächlich gefunden werden konnten.

Die Frage nach der Anzahl derjenigen Schülerinnen und Schüler, deren Bögen zwischen den Befragungswellen hätten zugeordnet werden können, ist dabei schwieriger zu beantworten, als dies auf den ersten Blick erscheint.[19] Zunächst beschränkt sich auf Schulebene die Grundgesamtheit auf die Jugendlichen der befragten Jahrgänge, die eine Schule besuchten, die zu allen Zeitpunkten der Befragung zustimmten. Diese Anzahl kann relativ genau den jeweiligen Schulstatistiken entnommen werden (siehe Tabelle 1). Auch die Schulstatistiken selbst sind meist nicht exakt, da die ausgewiesenen Belegungszahlen zu den *Statistikstichtagen* durch zwischenzeitliche Ab- und Zugänge der Schulen von den tatsächlichen Belegungen an den *Befragungsstichtagen* abweichen können; diese Abweichungen scheinen jedoch vernachlässigbar klein. Darüber hinaus problematisch sind Veränderungen der Schulpopulation über den *gesamten Befragungszeitraum* (vier Jahre) im Sinne von Zu- und Wegzügen nach bzw. aus Münster, Schulwechsler in bzw. von Schulen, die nicht an den Befragungen teilnahmen, Sitzenbleiber etc. Durch die Wahl der Erhebungsmethode (Fragebogeninterviews im Klassenverband) ist die erreichbare Probandenzahl auf diejenigen beschränkt, die an den jeweiligen Befragungstagen anwesend waren. Da die Fragebogenzuordnungen über die oben geschilderte Methode unter Zuhilfenahme von Codeblättern vorgenommen wurden, reduziert sich die Zahl weiter auf diejenigen Befragten, die ihr Codeblatt verwertbar ausgefüllt hatten.[20] Da als einzige Referenzdaten die aggregierten Schulstatistiken der vier Befragungsjahre zur Verfügung stehen und diese keinen befriedigenden Aufschluss über die oben genannten Probleme geben können, kann die Zahl der Probanden, die *tatsächlich mehrfach befragt wurden*, also hätten zugeordnet werden können, nur geschätzt werden. Um die zu erwartende Zahl (N_e) der möglichen Zuordnungen

19 Dazu folgendes Beispiel: Zum Zeitpunkt t_1 gingen 2.105 Schülerinnen und Schüler auf die von uns befragten Schulen, 1.850 von ihnen konnten von uns (verwertbar) befragt werden, 255 nicht. Zum Zeitpunkt t_2 waren es laut Schulstatistik 2.181 Probanden, von denen wir 1.915 befragen konnten, 266 nicht. Wie viele der Probanden aus t_1 haben wir zu t_2 wiederbefragt bzw. wie viele der Fragebögen aus t_1 können Bögen aus t_2 theoretisch zugeordnet werden? Nimmt man an, dass die Schulpopulation (relativ) stabil blieb, kann man zwei Extrempunkte formulieren. Zum einen könnten alle 1.850 wiederbefragt worden sein, die 255 Ausfälle (t_1) sind in den 266 Ausfällen (t_2) enthalten. Zum anderen könnten alle 255 Ausfälle (t_1) zum zweiten Zeitpunkt wiederbefragt worden sein, alle 266 Ausfälle (t_2) könnten allerdings im ersten Zeitpunkt schon befragt worden sein, so dass nur noch 1.329 Personen zu beiden Zeitpunkten befragt wurden.
Die realisierten Zuordnungen von 1.271 ergeben somit Quoten von 96 Prozent im besten oder 68 Prozent im schlechtesten Fall. Beides ist unwahrscheinlich.

20 Lässt man diese Betrachtungen außer Acht, so sind noch 813 Jugendliche von den ursprünglichen 1.850 im Vier-Wellen-Panel enthalten. Dies sind Jugendliche, die im beobachteten Zeitraum dreimal versetzt wurden, nie umgezogen sind, nie auf eine Schule außerhalb der Stichprobe umgeschult wurden, nie an den Befragungstagen gefehlt haben und bei jeder Befragung einen hinreichenden Code geliefert haben. Bei einem sehr konservativen Standpunkt kann hieraus mit einer gewissen Berechtigung eine Panelmortalität von 56 Prozent abgeleitet werden.

zwischen den jeweiligen Zeitpunkten schätzen zu können, seien folgende Annahmen zugelassen:

1. In der größten offiziellen Schulpopulation der Befragungszeitpunkte sind alle Schülerinnen und Schüler der anderen Zeitpunkte enthalten
2. Die Differenzen zwischen den Stichproben und den jeweiligen Schulstatistiken beruhen auf zufälligen, stichprobenneutralen Ausfällen

Annahme 1 ist notwendig, um eine Referenzgröße über die Anzahl aller Probanden zu bestimmen, die überhaupt zu irgendeinem Zeitpunkt hätten teilnehmen können. Problematisch ist die Annahme deshalb, da Veränderungen der Schulpopulation stattgefunden haben, die auf Aggregatebene nicht mehr erkennbar sind. Es wird vernachlässigt, dass zu jedem Zeitpunkt ein gewisser Teil an Schülerinnen und Schülern nicht in der größten Population enthalten waren. Dies führt zu einer Unterschätzung der Zuordnungsquoten.

Annahme 2 ist notwendig, um die Zahl der Ausfälle (Querschnitts-Rückläufe) einzubeziehen. Sie ist problematisch, da in der Realität die Befragungsteilnahme vermutlich nicht zufällig, sondern von bestimmten Faktoren abhängig ist. So haben erste Analysen gezeigt, dass der Rücklauf in den Querschnitten unter anderem von Geschlecht und Schulform abhängt, es kann also vermutet werden, dass auch der Anteil der Personen, die zu mehreren Zeitpunkten nicht teilnahmen, nicht zufällig verteilt ist, sondern von denselben Faktoren abhängt. Dies führt tendenziell zu einer Überschätzung der Zuordnungsquoten.

Lässt man die o. g. Annahmen zu, kann formuliert werden: Die zu erwartende Anzahl der Befragungsteilnehmer (N_e) zu bestimmten Zeitpunkten (t_i) ergibt sich aus der größten Schulpopulation ($\max(N(S_{ti}))$ multipliziert mit dem Produkt der Quotienten aus den jeweiligen Anzahlen der realisierten Querschnittsstichproben ($N(Q_{ti})$) und den Schulpopulationen nach Schulstatistik ($N(S_{ti})$).[21]

$$N_e = \max_{i=1,\ldots,k}(N(S_{ti})) * \prod_{i=1}^{k} \frac{(N(Q_{ti}))}{(N(S_{ti}))} \qquad (2)$$

21 Anders ausgedrückt entspricht dies dem Produkt der größten Ausgangsstichprobe und den multiplizierten Rücklaufquoten der beteiligten Querschnitte.

Aus den zu erwartenden Anzahlen (N_e) und den tatsächlich beobachteten bzw. realisierten Anzahlen der Zuordnungen (N_b) lassen sich nun die in Tabelle 4 angeführten Quoten berechnen. Sie dürften, wie oben angedeutet, die „wahre" Zuordnungsquote eher unterschätzen, da auf aggregierter Ebene der Schulstatistiken nicht sichtbare Veränderungen (Sitzenbleiber, Umzüge etc., die sich pro Jahr (aber nicht im Panel) ausgleichen), nicht berücksichtigt werden. Dieser Anteil liegt schätzungsweise bei mindestens zehn Prozent. Eine Überschätzung der Quoten durch nicht berücksichtigte, systematische Ausfälle erscheint im Vergleich dazu weniger bedeutsam.

Tabelle 4 Zuordnungsgüte der Paneldaten Münster 2000-2003

	N_e	N_b	Zuordnungsquote (%)
$P_{t1,t2}$	1.683	1.271	76
$P_{t2,t3}$	1.710	1.373	80
$P_{t3,t4}$	1.705	1.406	82
$P_{t1,t2,t3}$	1.502	997	66
$P_{t2,t3,t4}$	1.497	1.075	72
$P_{t1,t2,t3,t4}$	1.316	813	62

N_e bedeutet erwartetes N nach Formel 2

N_b bedeutet beobachtetes N

Um die Zuordnungsquoten der realisierten Paneldatensätze zu bewerten, können Quoten, die in anderen Panelstudien erreicht wurden, herangezogen werden. Ein übliches Verfahren bei Paneluntersuchungen stellt die Zuordnung der Fragebögen über erhobene Adressen der Schülerinnen und Schüler dar. Hierbei tritt erfahrungsgemäß schon bei der *Adresserhebung* ein Ausfall von ca. 50 Prozent auf, der im Laufe solcher Addresspanels deutlich größer wird.[22]

Die Zuordnungsquoten für die Zwei-Wellen-Panels aus jeweils zwei aufeinander folgenden Zeitpunkten liegen zwischen 76 und 82 Prozent. Es ist ersichtlich, dass die Quoten im Verlauf der Befragungen besser wurden. So wurde die schlechteste Zuordnungsquote zwischen den Jahren 2000 (t_1) und 2001 (t_2), die beste zwischen den Jahren 2002 (t_3) und 2003 (t_4) verwirklicht; auch die absolute Fallzahl stieg entsprechend an. Dies mag zum einen am steigenden Alter der Befragten und der möglicherweise damit verbundenen höheren Fähigkeit, die Codefragen richtig zu

22 Exemplarisch hierzu siehe ***Böttger*** et al. (2003, S. 35 ff.) mit einem Ausfall von 42 Prozent zwischen t_1 und t_2 verglichen mit hier 24 Prozent (realistische, korrigierte Schätzung) bzw. 31 Prozent (konservativste Schätzung) Ausfall zwischen t_1 und t_2.

beantworten bzw. sich zu konzentrieren liegen. Zum anderen mag auch ein positiver Gewöhnungseffekt bei den Probanden eine Rolle spielen. Die Quoten der beiden lückenlosen Drei-Wellen-Panels liegen bei 66 und 72 Prozent, die des Vier-Wellen-Panels erwartungsgemäß niedriger bei 62 Prozent. Die realisierten Quoten liegen damit deutlich über denen von Adresspanels.

4.3 Verzerrung der Paneldaten

Am Beispiel des Paneldatensatzes $P_{t3,t4}$ (Münster 2002/2003) soll erläutert werden, welche Auswirkungen das angewendete Zuordnungsverfahren in Bezug auf die Repräsentativität der Daten hat. In Tabelle 5 finden sich die Paneldaten aufgeschlüsselt nach den zugelassenen Fehlern in der Codezuordnung. Zum Vergleich sind dort die Münsteraner Querschnittsdaten aus 2003 und die Daten der Schulstatistik desselben Jahres aufgeführt.

Tabelle 5 Verzerrung der Paneldaten Münster 2002/2003

	Codezuordnung $P_{t3,t4}$			Gesamt	Gesamt	Gesamt
	0 Fehler	1 Fehler	2 Fehler	$P_{t3,t4}$	$Q_{,t4}$	S_{t4}
N	637	523	246	1.406	1.819	2.077
Geschlecht						
Mädchen (%)	59	49	41	52	50	49
Jungen (%)	41	51	59	48	50	51
Schulform						
Hauptschüler (%)	16	21	27	19	23	23
Realschüler (%)	33	33	33	33	32	31
Gymnasiasten (%)	50	44	34	45	41	39
Sonderschüler (%)	2	3	5	3	4	6

P bedeutet Panel

Q bedeutet Querschnitt

S bedeutet Schulstatistik (nur befragte Schulen)

Es wird deutlich, dass zwar der größte Teil (N = 637) der Schülerinnen und Schüler bei der Beantwortung des Codeblattes in 2002 und 2003 keinen Fehler machte, der Anteil derjenigen, die einen Fehler (N = 523) machten jedoch kaum kleiner ist. Der Anteil der Probanden, die zwei Fehler (N = 246) machten ist zwar deutlich geringer, jedoch ebenfalls noch erheblich. Bei dem ursprünglich geplanten Zuordnungsverfahren mit Hilfe des gesamten Codes, hätte man nur auf die 637 Zuordnungen ohne Fehler zurückgreifen können. Ein Verzicht auf die Zuordnungen mit einem oder zwei Fehlern erscheint daher unvernünftig. Vielmehr rechtfertigt der Ertrag der gefundenen fehlerhaften Zuordnungen den oben geschilderten Aufwand eines Zuordnungsverfahrens mit manuellem Handschriftenvergleich.

Aus Tabelle 5 wird darüber hinaus ersichtlich, dass die Anzahl der Fehler bei der Beantwortung des Codeblattes mit Geschlecht und Schulbildung der Befragten korrelieren. Während die Verzerrungen auf Ebene des Gesamt-Paneldatensatzes (der Zuordnungen bis zu zwei Fehler beinhaltet) nur leicht sind, lassen sich bei der Aufschlüsselung der Zuordnungen nach Fehlern deutliche Verschiebungen erkennen. Augenscheinlich war es den weiblichen Befragungsteilnehmern eher möglich, in zwei Jahren fehlerfrei zu antworten, der Anteil der Mädchen ist bei den Zuordnungen mit keinem Fehler deutlich größer als derjenige der Jungen. Auch die Schulbildung hat einen (erwartungsgemäßen) Effekt. Je höher die Schulbildung, desto weniger Fehler wurden bei der Beantwortung gemacht. Über die Ursachen können an dieser Stelle nur Mutmaßungen angestellt werden. Es ist zu vermuten, dass kognitive Fähigkeiten - bei aller Einfachheit der Fragen - ebenso eine Rolle spielen, wie Konzentrationsfähigkeit und -wille.

Auch unter dem Gesichtspunkt der Verzerrung der Daten erscheint die Beschränkung nur auf fehlerfreie Zuordnungen nicht wünschenswert. Muss der Code zwischen den Jahren komplett übereinstimmen, so sind die zu erwartenden Datensätze deutlich verzerrt. Erst ein fehlertolerantes Zuordnungsverfahren unter Zulassung von (bis zu zwei) Fehlern liefert hier Daten, deren Verzerrung in Bezug auf Geschlecht und Schulbildung deutlich geringer ist.

5 Fazit

Resümiert man die Erfahrungen, die in unserer Studie mit einem Zuordnungsverfahren über persönliche Codes gemacht wurden, so muss man zu dem Ergebnis kommen, dass sich das ursprünglich geplante Verfahren in der Praxis nur teilweise bewährt hat, jedoch mit Hilfe der geschilderten Problemlösestrategie als Ergebnis bessere Zuordnungsquoten realisiert werden konnten als in vergleichbaren Adresspanels. Die Grundvoraussetzungen des Verfahrens erwiesen sich zunächst als unzureichend erfüllt, so dass wegen der zu geringen Zuordnungsquote und einer starken Verzerrung der Daten eine Abweichung vom ursprünglich geplanten Verfahren notwendig war. Die Voraussetzung, dass jeder Proband einen eindeutigen Code aufweisen muss, war in unserer Studie nicht gegeben. Zu einem großen Teil lag dies an Kompromissen (Art der Fragen, Fragestellung), die während der Verhandlungen mit dem Datenschutz eingegangen werden mussten. Für zukünftige Forschungsvorhaben erscheint diese Voraussetzung allerdings durch die Wahl geeigneter Codes bzw. Codefragen realisierbar.

Als sehr viel schwieriger zu realisieren, dürfte sich die zweite Voraussetzung des Verfahrens - exakt übereinstimmende Codes in den einzelnen Erhebungswellen - erweisen. In unserer Studie zeigte sich, dass ein sehr großer Teil der befragten Jugendlichen nicht in der Lage war, die Codefragen in den unterschiedlichen Jahren gleich zu beantworten. Bei einem Zuordnungsverfahren über den kompletten Code wäre dann eine angemessene Zuordnungsquote nicht mehr zu erreichen gewesen. Für die Phase der Zuordnungsfindungen mussten deshalb „fehlerhafte“ Codes zugelassen werden, um die Fragebögen, die von denselben Personen in unterschiedlichen Wellen ausgefüllt wurden, einander zuordnen zu können.

Die Schwierigkeitsstufe der gestellten Fragen war unseres Erachtens nach Schülerinnen und Schüler der befragten Altersstufen zumutbar. Es zeigte sich darüber hinaus, dass die Fehlerhaftigkeit des Codes bei fortschreitendem Alter der Befragten nicht wesentlich geringer wurde. Es steht also eher zu vermuten, dass generell ein nicht zu vernachlässigender Teil von fehlerhaften bzw. differierenden Antworten zu erwarten ist, stellt man Jugendlichen in mehreren Jahren fünf oder mehr Fragen. Ein Zuordnungsverfahren über Codes, welcher auf Einzelfragen beruht, muss die Möglichkeit eines fehlertoleranten Zusammenspiels von Daten bieten.

Literatur

Boers, K. und ***Kurz, P.*** (2000). Schule, Familie, Einstellungen, Lebensstile, delinquentes und abweichendes Verhalten. Erste Ergebnisse der Münsteraner Schülerbefragung 2000. Eigendruck. Münster.

Boers, K. und ***Pöge, A.*** (2003). Wertorientierungen und Jugenddelinquenz. In: ***S. Lamnek*** und ***M. Boatca*** (Hrsg.), Geschlecht, Gewalt, Gesellschaft (S. 246-269). Opladen: Leske + Budrich.

Boers, K., Reinecke, J., Wittenberg, J. und ***Motzke, K.*** (2002). Wertorientierungen, Freizeitstile und Jugenddelinquenz. Neue Kriminalpolitik, 14(4), 141-146.

Böttger, A., Ehret, B., Othold, F., Prein, G., Schumann, K. F. und ***Seus, L.*** (2003). Methoden der Untersuchung. In: ***K. F. Schumann*** (Hrsg.), Delinquenz im Lebensverlauf. Bremer Längsschnittstudie zum Übergang von der Schule in den Beruf bei ehemaligen Hauptschülern (Bd. 2, S. 35-66). Weinheim, München: Juventa.

Motzke, K. und ***Wittenberg, J.*** (2004). Methodendokumentation der kriminologischen Schülerbefragung in Münster 2000 (Schriftenreihe „Jugendkriminalität in der modernen Stadt - Methoden“ Nr. 1). Münster, Trier.

Pöge, A. (2002). Der Zusammenhang von Lebensmilieu, Einstellungen und deviantem Verhalten bei Jugendlichen. Münster. (unveröffentlichte Magisterarbeit)

Pöge, A. (2005a). Ethnicity and self-reported delinquency. How to define ethnicity? In: ***Queloz***, ***R. Brossard***, ***R.F. Bütikofer***, ***B. Meyer-Bisch*** und ***D. Pittet*** (Hrsg.), Délinquance des jeunes et justice des mineurs. Les défis des migrations et de la pluralité ethnique. Youth Crime and Juvenile Justice. The challenge of migration and ethnic diversity. Berne, Bruylant, Bruxelles: Editions Staempfli. (in Druck)

Pöge, A. (2005b). Methodendokumentation der kriminologischen Schülerbefragung in Münster 2000-2003 (Vier-Wellen-Panel) (Schriftenreihe „Jugendkriminalität in der modernen Stadt - Methoden“ Nr. 9). Münster, Trier.

Raithel, J. (2002). Jugendkriminalität und elterliches Erziehungsverhalten. Neue Kriminalpolitik, 14(2), 62-65.

Wittenberg, J. (2004a). Methodendokumentation der kriminologischen Schülerbefragung in Münster 2001 (Schriftenreihe „Jugendkriminalität in der modernen Stadt - Methoden“ Nr. 2). Münster, Trier.

Wittenberg, J. (2004b). Methodendokumentation der kriminologischen Schülerbefragung in Münster 2002 (Schriftenreihe „Jugendkriminalität in der modernen Stadt - Methoden“ Nr. 4). Münster, Trier.

Wittenberg, J. (2004c). Methodendokumentation der kriminologischen Schülerbefragung in Münster 2003 (Schriftenreihe „Jugendkriminalität in der modernen Stadt - Methoden“ Nr. 7). Münster, Trier.