SSOAR

Open Access Repository
www.ssoar.info

Persönliche Codes 'reloaded'

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. (2008). Persönliche Codes 'reloaded'. *Methoden, Daten, Analysen (mda)*, *2*(1), 59-70. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-126543

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 „reloaded“

Personal Codes “reloaded”

Andreas Pöge

Zusammenfassung
Bei Längsschnittuntersuchungen mit sensiblen Fragestellungen werden die Fragebogenzuordnungen zwischen den einzelnen Erhebungswellen aus Datenschutzgründen oftmals mit persönlichen und selbstgenerierten Codes vorgenommen – so auch in dem DFG-Projekt „Kriminalität in der modernen Stadt“. Die Ergebnisse mit dieser Zuordnungsmethode, insbesondere bezüglich der Probleme bei der Durchführung, der Ausschöpfungsquote und der Verzerrung der resultierenden Paneldaten, waren Gegenstand der Ausführungen in dem Artikel „Persönliche Codes bei Längsschnittstudien: Ein Erfahrungsbericht“ (Pöge 2005b), der sich auf die Münsteraner Teilstudie bezog. Mittlerweile wurde die Studie auf den Erhebungsort Duisburg ausgedehnt, wobei die Erkenntnisse aus Münster zu Modifkationen des Verfahrens führten. Hier sollen die teils sehr deutlichen Verbesserungen der Datenqualität aufgezeigt werden, die damit in Duisburg erreicht werden konnten.

Abstract
In panel studies the assignments of questionnaires are often carried out between the waves using personal codes - also in the DFG project “Juvenile delinquency in modern town”. The results with this allocation method, in particular with regard to the problems by the realisation, the exhaustion rate and the bias of the resultant panel data, have been published in the article “personal codes with profile studies: A report of experience” (Pöge 2005b) which refers to the study in Münster. Meanwhile the study was expanded to Duisburg and the findings from Münster led to modifications of the procedure. Here the very clear improvements in the exhaustion rate and smaller bias in Duisburg shall be described.

1 Einleitung

Eine grundlegende Problematik bei Panelstudien stellt die Zuordnung der Fragebögen aus den einzelnen Erhebungswellen dar, die von denselben Personen ausgefüllt wurden. Ist aus Gründen des sensiblen Befragungsgegenstandes (beispielsweise Krankheitsverläufe, kriminelle Verhaltensweisen etc.) oder des besonderen Persönlichkeitsschutzes der Befragten (Studien mit minderjährigen Teilnehmerinnen und -nehmern) besonderes Augenmerk auf Datenschutzbelange zu richten und daher die Erhebung und Speicherung beispielsweise von Adressen, Versicherungs- oder Matrikelnummern unerwünscht, bietet sich ein Verfahren mit persönlichen und selbstgenerierten Codes an (Schnell/Bachteler et al. 2006: 129). Die einzelnen Codefragen sollten dabei zeitstabile Merkmale mit persönlichem Bezug zu den einzelnen Respondenten aufweisen – aus diesem Grund wird oftmals auf Buchstaben des eigenen Namens oder dem Verwandter, auf Geburtsdatumsziffern oder andere persönliche Merkmale zurückgegriffen (Grube/Morgan et al. 1989: 159). Die Länge der Codes umfasst dabei meist sechs bis zehn Zeichen bzw. Ziffern (Schnell/Bachteler et al. 2006: 129). Problematisch an solch einem Verfahren sind Fehler bei der Erzeugung der Codes, da dann die Fragebögen nur noch unter erschwerten Bedingungen, nämlich mithilfe von fehlertoleranten Verfahren einander zugeordnet werden können und somit teils erheblicher Datenverlust droht.

Die folgenden Ausführungen stellen eine Fortschreibung des in der ZA-Information 56 (S. 50–69) veröffentlichten Artikels „Persönliche Codes bei Längsschnittstudien: Ein Erfahrungsbericht" dar. Wie dort vorgestellt, wurden bzw. werden im Zuge der Panelstudie *Kriminalität in der modernen Stadt*[1] mithilfe von solchen persönlichen und selbstgenerierten Codes, die über Codeblätter abgefragt wurden, Paneldatensätze erstellt. In der Münsteraner Teilstudie kam es im Hinblick auf die Fragebogenzuordnung der einzelnen Jahre zu Problemen, die zusammen mit der Problemlösestrategie und den Ergebnissen im genannten Artikel dargestellt wurden. Parallel wurde ab dem Jahr 2002 begonnen, in *Duisburg* einen Paneldatensatz aufzubauen, wobei die Erfahrungen mit den Münsteraner Codes berücksichtigt wurden. Auf Grundlage des Duisburger Vier-Wellen-Datensatzes 2002 bis 2005 sollen die Modifikationen im Codeverfahren aufgezeigt werden, die zu einer deutlich besseren Datenqualität in den Duisburger Paneldaten führen.

1 Projektleitung: Prof. Dr. Klaus Boers, Institut für Kriminalwissenschaften, WWU Münster und Prof. Dr. Jost Reinecke, Fakultät für Soziologie, Universität Bielefeld.

2 Schülerbefragungen in Duisburg

Bei den im Jahr 2002 begonnenen Untersuchungen in Duisburg wurden Schülerinnen und Schüler der siebten Jahrgangsstufe befragt, die eine Sonder-, Haupt-, Real-, Gesamtschule oder ein Gymnasium besuchten.[2] In den Wellen *zwei* bis *drei* wurden die Probanden der nun achten und neunten Jahrgangsstufe wiederbefragt, in der *vierten* Welle zusätzlich zu den Jugendlichen der nunmehr zehnten Klassen auch diejenigen, die nicht versetzt worden waren, also im Jahr 2005 die neunte Jahrgangsstufe wiederholten.

Die in den vier Querschnittsbefragungen gewonnenen Daten zeigen – nach den durchgeführten Plausibilitätskontrollen – als Stichprobengrößen 3.243 bis 3.411 verwertbare Fälle. Vergleicht man diese realisierten Fallzahlen mit den offiziellen Belegungszahlen der an den Befragungen *teilnehmenden Schulen* aus den jeweiligen Schulstatistiken, so ergeben sich sehr gute Rücklaufquoten zwischen 85 und 92%[3].

3 Gestaltung des Codeblattes

Um den für die Fragebogenzuordnung zwischen den Wellen erforderlichen Code zu erzeugen, sollte jede befragte Schülerin bzw. jeder Schüler in jedem Jahr ein Codeblatt mit verschiedenen Codefragen ausfüllen, so dass bei einem stabilen Antwortverhalten die Codeblätter und die über eine eindeutige Nummer mit ihnen verknüpften Fragebögen einander zugeordnet werden könnten. Als grundlegende Problembereiche waren in den früheren Münsteraner Befragungen die Probleme der Identifizierung und der Reproduktion aufgetreten (Pöge 2005a: 7ff.; Pöge 2005b). Das heißt, dass zum einen nicht alle in einem Querschnitt aufgetretenen Codes eindeutig waren und zum anderen, dass sich ein nicht unerheblicher Teil der Codes, die zwischen den Jahren hätten gleich sein müssen, in der Praxis unter-

2 Die Befragungen fanden als schriftliche Fragebogeninterviews im Klassenverband statt. Sie wurden von Interviewerinnen und Interviewern begleitet, die den Jugendlichen zu Beginn der doppelstündigen Befragungsphase Ausfüllanweisungen und Informationen zur Studie, zu Datenschutzfragen und zum Ablauf gaben.

3 Zur Dokumentation der Duisburger Querschnittsstudien der Jahre 2002 bis 2005 siehe Motzke/Brondies (2004); Brondies (2004); Hilfert (2005); Kunadt (2006), zu der Panelstudie siehe Pöge (2007). Die Befragungen werden seit dem Jahr 2005 im jährlichen Rhythmus weiter fortgeführt, wobei (zunächst nur teilweise) auf eine postalische Befragungsform umgestellt wurde, da die Jugendlichen nicht mehr alle eine allgemeinbildende Schule besuchen.

schied. Schon im letzten Befragungsjahr in Münster erfolgte daher eine grundlegende Überarbeitung des Codeblattes, die für die Duisburger Studie ab der zweiten Welle übernommen und weitergeführt wurde. Es wurde insbesondere versucht, die Schwierigkeit der Codefragen noch weiter herabzusetzen und Layoutmängel zu beseitigen.

Bei den Befragungen in Duisburg wurde zunächst wie in Münster ein fünfstelliger Code verwendet, bei dem allerdings aufgrund der Erfahrungen zum Teil bereits einfachere Fragen mit größerer Antwortvarianz zum Einsatz kamen.[4] So mussten zum Beispiel nicht mehr der letzte Buchstabe der Haarfarbe des Vaters, sondern der erste Buchstabe des Vornamens des Vaters angegeben werden.[5] Im Jahr 2003 wurde auf Grundlage der bisher gewonnenen Erkenntnisse der Duisburger Code um eine Frage erweitert und das Layout des gesamten Codeblattes überarbeitet. Im Zuge dieser Umstellung musste kein handschriftliches Ausfüllen mehr erfolgen, sondern alle Antwortvorgaben wurden zum Ankreuzen aufgeführt. Ab dem Befragungsjahr 2003 wurden außerdem zusätzlich Fragen nach einer Befragungsteilnahme im Vorjahr, einem Schulwechsel sowie einer eventuellen Nichtversetzung im vergangenen Jahr gestellt. Da im letzten Erhebungsjahr 2005 auch die nicht versetzten Schülerinnen und Schüler befragt wurden, musste auf dem Codeblatt eine Zusatzfrage nach der entsprechenden Jahrgangsstufe gestellt werden (siehe Abbildung 1). In Ergänzung zu diesen acht bzw. neun Fragen standen für die Fragebogenzuordnungen das Geschlecht der Befragten sowie die Schule, die mit einer Kennnummer erhoben wurde, zur Verfügung. Zusätzlich zu den Hinweisen auf dem Codeblatt selber wurden von den jeweiligen Interviewerinnen und Interviewern Ausfüllanweisungen gegeben. Zu nennenswerten Problemen oder Schwierigkeiten kam es beim Ausfüllen der Blätter nicht.

Zum Problembereich der Identifizierung ist mit den abgeänderten Fragen und deren größerer Antwortvarianz sowie der zusätzlichen Codestelle festzuhalten, dass in Duisburg ab dem Jahr 2003 fast alle Codes eindeutig sind. In Münster waren in allen vier Erhebungsjahren nur bei rund drei Vierteln der Schülerinnen

4 Probleme mit bewussten Falschangaben oder unbewussten Fehlern bei der Ausfüllung des Codeblattes tauchen selbstverständlich auch in Duisburg auf. Allerdings zeigten sich zumindest die unbewussten Fehler durch die leichtere Beantwortbarkeit der Fragen als positiv beeinflussbar.

5 Durch die wesentlich größere Datengrundlage in Duisburg war es möglich, wie ursprünglich auch in Münster geplant, oftmals die *ersten* Buchstaben der Antworten auf die einzelnen Codefragen notieren zu lassen ohne eine Verletzung der Anonymität durch eventuell mögliche Identifizierung einzelner Personen befürchten zu müssen. Das Notieren dieser jeweils ersten Buchstaben scheint in der Tat eine wesentliche Erleichterung für die ausfüllenden Personen zu sein.

und Schülern eindeutige Codes aufgetreten – bei rund einem Viertel der Befragten wiesen unterschiedliche Personen einen gleichen Code auf und waren somit ohne Hinzunahme weiterer Informationen nicht eindeutig zu identifzieren. In Duisburg hingegen haben bei dem modifzierten fünfstelligen Code aus dem Jahr 2002 immerhin rund 80%, bei dem sechststelligen Code, der in allen weiteren Befragungsjahren Anwendung fand, fast alle Befragten einen eindeutigen Code (siehe Tabelle 1). Insofern kann das Problem der Identifzierung durch die geschilderten Codemodifizierungen als gelöst betrachtet werden.

Abbildung 1 Das Codeblatt in Duisburg 2005

Bitte kreuze bei jeder der sechs Fragen immer nur ein Feld an!
Wenn du eine der Fragen überhaupt nicht beantworten kannst, kreuze bitte kein Feld an!

Hier nun die sechs Fragen zur Erstellung deines persönlichen Codes:

1	Bitte kreuze den **ersten** Buchstaben des Vornamens deines **Vaters** (oder einer Person, die für dich einem Vater am nächsten kommt) an. (z. B. Anton, Bernd, Hans-Peter usw.) a b c d e f g h i j k l m n o p q r s t u v w x y z ä ö ü ß
2	Bitte kreuze den **ersten** Buchstaben des Vornamens deiner **Mutter** (oder einer Person, die für dich einer Mutter am nächsten kommt) an. (z. B. Anna, Beate, Jutta, Maria, usw.) a b c d e f g h i j k l m n o p q r s t u v w x y z ä ö ü ß
3	Bitte kreuze den **ersten** Buchstaben deines **Vornamens** an (z. B. Michael, Thomas, Ute usw.) a b c d e f g h i j k l m n o p q r s t u v w x y z ä ö ü ß
4	Bitte kreuze den **Tag** deines **Geburtsdatums** an (z. B. Geburtstag am 7. Januar = 7, am 12. Mai = 12, am 31. Oktober = 31) 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
5	Bitte kreuze den **letzten** Buchstaben deiner natürlichen **Haarfarbe** an. (z. B. braun, Glatze, schwarz, usw.) a b c d e f g h i j k l m n o p q r s t u v w x y z ä ö ü ß
6	Bitte kreuze den **letzten** Buchstaben deiner **Augenfarbe** an. (z. B. braun, grün, grau, usw.) a b c d e f g h i j k l m n o p q r s t u v w x y z ä ö ü ß

Hast du im letzten Jahr an der Befragung teilgenommen?	☐ ja	☐ nein
Hast du im letzten Jahr die Schule gewechselt?	☐ ja	☐ nein
Bist du im letzten Jahr sitzen geblieben?	☐ ja	☐ nein
In welcher Klasse bist du?	⑨	⑩

Tabelle 1 Häufigkeiten der Codes

	2002		2003		2004		2005	
Häufigkeit	Anzahl Codes	Gesamt	Anzahl Codes	Gesamt	Anzahl Codes	Gesamt	Anzahl Codes	Gesamt
1	3064	3064	3352	3352	3350	3350	3358	3358
2	324	648	32	64	40	80	56	112
3	18	54	-	-	-	-	-	-
4	-	-	-	-	-	-	-	-
5	5	25	-	-	-	-	-	-
Summe[a]	3411	3791	3384	3384	3390	3430	3414	3470

[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.*

Das Problem der Reproduktion, dass also Befragte in den einzelnen Wellen einen nicht übereinstimmenden Code angeben, existiert allerdings auch in Duisburg noch in einem nicht zu vernachlässigendem Maße, konnte aber gleichwohl reduziert werden. In Münster fanden sich zwischen jeweils zwei Erhebungszeitpunkten lediglich rund 500 bis 600 Zuordnungen mit exakt übereinstimmendem Code, was bei allen realisierten Zuordnungen einen Prozentsatz von 40 bis 50% ausmachte (Pöge 2005a: 7f; Pöge 2005b: 66). In Duisburg hingegen liegt die Quote recht stabil bei rund 60% aller gefundenen Zuordnungen (siehe exemplarisch für die Zuordnungen zwischen 2003 und 2004 Tabelle 2). Die Reproduktion des Codes hat also mithilfe des verbesserten Layouts und der vereinfachten Fragen deutlich besser funktioniert, obwohl eine Codestelle mehr fehlerfrei angegeben werden musste als in Münster.

4 Das angewendete Zuordnungsverfahren

Um die Fragebögen der einzelnen Erhebungsjahre über die Codes einander zuzuordnen, wurde auch in Duisburg das im genannten Artikel für Münster vorgestellte *fehlertolerante Verfahren mit manuellem Handschriftenvergleich* angewendet. Es bestand zunächst aus drei, später aus vier Schritten: In einem *ersten Schritt* wurden maschinell alle exakt übereinstimmenden Codes aus zwei Erhebungswellen heraus-

gefunden.[6] Die zusammengehörigen Fragebögen und Codeblätter wurden daraufhin einer manuellen Handschriftenkontrolle[7] unterzogen, wobei die offensichtlich nicht passenden Zuordnungen aussortiert wurden. Die passenden Fragebogennummern wurden daraufhin 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. Seit dem Jahr 2003 war es mit dem eingesetzten erweiterten Code möglich, in einem *vierten Schritt* drei Fehler zuzulassen. Auch in diesen Schritten wurden als Validierung der Zuordnungen Handschriftenvergleiche durchgeführt, die offensichtlich nicht passenden Zuordnungen verworfen und vor der Durchführung des nächsten Schrittes die erfolgreich zugeordneten Nummern aus dem Datensatz entfernt. Um die im weiteren Verlauf vorgestellten, realisierten Fallzahlen erreichen zu können, waren pro Datensatz rund 4.000 manuelle Überprüfungen per Handschriftenvergleich nötig (exemplarisch siehe hierzu Tabelle 2).

Tabelle 2 Zugelassene Fehler, Anzahl der handschriftlichen Kontrollen und deren Ergebnisse (vor Plausibilitätskontrollen) 2003/2004

Fehler	passt	passt nicht	fehlt[a]	Kontrollen
0	1528	42	16	1586
1	735	175	7	917
2	292	853	5	1150
3	67	119	1	187
Summe	2622	1189	29	3840

[a] *In der Spalte „fehlt“ sind die Anzahlen der unauffindbaren Fragebögen ausgewiesen.*

Es stellte sich heraus, dass die Fragebogenpärchen auf Grundlage der Codezuordnung ohne Fehler meist tatsächlich von derselben Person ausgefüllt wurden. Bei den Zuordnungen mit zugelassenen Fehlern ist der Grad der falsch positiven Über-

6 Das Heraussuchen der Zuordnungsvorschläge wurde mithilfe des Office-Programms Access und SQL-Skripten durchgeführt.

7 Hierbei wurden vor allem die offenen Fragen im Fragebogen mit den handschriftlichen Eintragungen verwendet.

einstimmungen deutlich höher. Hier kommt wiederum zum Tragen, dass unter Auslassung einer oder mehrerer Codestellen der Code uneindeutiger wird. So passt dann ein Fragebogen mit reduziertem Code vermeintlich zu mehreren Fragebögen der jeweils anderen Welle. Dieser Umstand lässt noch längere Codes ratsam erscheinen.

Unter Berücksichtigung der gültigen Fälle – derjenigen also, die auch den Plausibilitätskontrollen der einzelnen Querschnitte standhielten, liegen die mit dem vorgestellten Zuordnungsverfahren erreichten Fallzahlen zwischen 2.472 und 2.640 für die Zwei-Wellen-, bei 2.012 und 2.186 für die Drei-Wellen-Panels und bei 1.715 für das Vier-Wellen-Panel (Tabelle 3).[8]

Tabelle 3 Zuordnungsgüte der Paneldaten in Duisburg und Münster im Vergleich (N_e bedeutet erwartetes N; N_b tatsächlich beobachtetes N)

(a) Münster

	N_e	N_b	Quote (%)
$P_{t1,2}$	1683	1271	75,5
$P_{t2,3}$	1710	1373	80,3
$P_{t3,4}$	1705	1406	82,2
$P_{t1,2,3}$	1502	997	66,4
$P_{t2,3,4}$	1497	1075	71,8
$P_{t1,2,3,4}$	1316	813	61,8

(b) Duisburg

	N_e	N_b	Quote (%)
$P_{t1,2}$	3075	2472	80,4
$P_{t2,3}$	3010	2596	86,2
$P_{t3,4^*}$	3106	2640	85,0
$P_{t1,2,3}$	2626	2012	76,6
$P_{t2,3,4^*}$	2755	2186	79,3
$P_{t1,2,3,4^*}$	2403	1715	71,4

Die Duisburger Paneldatensätze sind somit weitaus umfangreicher als die Münsteraner (siehe ebenfalls Tabelle 3), was vor allem auf die größere Querschnittsdatengrundlage zurückzuführen ist – allerdings leistet auch das deutlich besser funktionierende Zuordnungsverfahren seinen Beitrag zu den guten Ergebnissen. Die in

8 Zum vierten Zeitpunkt (2005) wurden in Duisburg zusätzlich zum zehnten Jahrgang auch Schülerinnen und Schüler der neunten Jahrgangsstufe befragt (Wiederholer). Die hochgestellten Sternchen bezeichnen diejenigen Datensätze, bei denen, um einen Vergleich mit den Münsteraner Daten zu ermöglichen, zum vierten Zeitpunkt *nur Jugendliche der zehnten Jahrgangsstufe* enthalten sind.

Duisburg mit dem beschriebenen Verfahren erreichten Zuordnungsquoten[9] für die Zwei-Wellen-Panels aus jeweils zwei aufeinanderfolgenden Zeitpunkten liegen zwischen 80 und 86%. Die Quote ist dabei zwischen den Jahren 2002 (t_1) und 2003 (t_2) deutlich niedriger als zwischen den späteren Zeitpunkten. Auch die absolute Fallzahl steigt im Laufe der Jahre an und ist beim Panel 2004/2005 am höchsten. Dies mag zum einen am steigenden Alter der Befragten und der möglicherweise damit verbundenen höheren Fähigkeit, die Codefragen richtig zu 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 77 und 79%, die des Vier-Wellen-Panels erwartungsgemäß niedriger bei 71%. Im Vergleich zu den Quoten aus der Münsteraner Panelstudie zeigen sich damit deutlich bessere Ergebnisse (siehe Tabelle 3). Es zeigt sich also, dass durch den verbesserten Code ein deutlich besseres Ergebnis bei der Ausschöpfung erzielt werden konnte.

Auch die Duisburger Paneldaten wurden auf Verzerrungen geprüft, mit dem Ergebnis, dass die Anzahl der Fehler bei Beantwortung des Codeblattes mit dem Geschlecht und der Schulbildung der Befragten korreliert. Während die Verzerrungen auf Ebene des gesamten Paneldatensatzes im Vergleich mit den Querschnittsdaten und der Schulstatistik nur leicht sind, lassen sich bei der Aufschlüsselung nach Fehlern deutliche Verschiebungen erkennen: 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. Vermutlich spielen kognitive Fähigkeiten – bei aller Einfachheit der Fragen – ebenso eine Rolle wie Konzentrationsfähigkeit und -wille. Im Vergleich zu den Verzerrungen in den Münsteraner Daten ist auf Ebene des Gesamtpanels festzustellen, dass der verbesserte Code hier zu deutlich geringer verzerrten Daten führt. Zwar unterscheiden sich die Abweichungen in der Verteilung nach Geschlecht kaum voneinander, die Verbesserung im Bereich der Verteilung nach Schulform ist jedoch erheblich.

9 Hier wird dieselbe Art der Quotenberechnung zugrundegelegt, wie für die Münsteraner Daten vorgestellt (Pöge 2005a: 62ff.).

5 Empfehlungen für die Praxis

Resümiert man nun nach acht Jahren die Erfahrungen, die in unserem Forschungsprojekt mit der Zuordnung von Fragebögen gemacht wurden, so ist festzuhalten, dass nach anfänglichen Schwierigkeiten ein durchaus gangbarer und erfolgreicher Weg gefunden wurde, die Paneldatensätze zu erstellen.

Aus vorwiegend datenschutzpsychologischen Gründen wurde ein Zuordnungsverfahren über persönliche Codes gewählt, damit gar nicht erst Bedenken im Hinblick auf die Anonymität bei den Befragten auftreten sollten. Dies erschien insbesondere wegen des jungen Alters der Befragungsteilnehmerinnen und -teilnehmer, welches sie besonders schützenswert macht, und des sehr sensiblen Datenmaterials (unter anderem Angabe von verübten Straftaten) geboten. Das Alternativverfahren eines Adresspanels mit Einwilligung zur Adressspeicherung kam daher zunächst nicht in Frage. Allerdings zeigen vergleichbare Studien auch mit diesem Verfahren keine deutlich besseren Ausschöpfungsquoten. So tritt erfahrungsgemäß teils schon bei der ersten Adresserhebung ein Ausfall von ca. 50% auf (exemplarisch hierzu siehe Böttger/Ehret et al. 2003: 35ff.), der im Laufe solcher Adresspanels deutlich größer wird. Mithilfe von selbstgenerierten Codes können oftmals bessere Quoten erreicht werden, wobei hier den Erhebungsabständen eine bedeutende Rolle zukommt. So liegen die Zuordnungsquoten bei monatlichem Befragungsintervall deutlich höher als bei jährlichem (siehe Kearney/Hopkins et al. 1984; Grube/Morgan et al. 1989).

In den beiden hier veröffentlichten Artikeln wurde vorgestellt, mit welchen Problemen in der Praxis mit selbstgenerierten persönlichen Codes zu rechnen ist. Im Vergleich der Erhebungsorte Münster und Duisburg wurde aufgezeigt, wie es möglich ist, durch Codemodifikationen deutliche Verbesserungen in den Zuordnungsquoten zu erreichen. Hierzu kann man zusammenfassend festhalten, dass bei der Planung eines solchen Zuordnungsverfahrens zuallererst die Variationsmöglichkeit des Codes geprüft werden muss. Es muss gewährleistet sein, dass jede Person mit hinreichender Wahrscheinlichkeit einen eindeutigen Code aufweist. In diese Überlegungen muss außerdem einbezogen werden, dass sich möglicherweise später einzelne Codestellen als fehlerhaft erweisen und bei der späteren Zuordnung nicht zur Verfügung stehen. Der Code sollte auch in diesem Fall noch eindeutig sein. Als Daumenregel stellte sich in unserer Untersuchung heraus, dass der Code auch unter Auslassung zweier Stellen eindeutig sein sollte. Als Stellschrauben für die Anforderungen an die Eindeutigkeit sind zum einen die Länge des Codes bzw. die Anzahl der Codestellen und zum anderen die Art der einzelnen

Codefragen im Hinblick auf deren Antwortvariation zu nennen. Es wurde hier gezeigt, wie abgeschätzt werden kann, ob der Code die genannten Anforderungen erfüllt (Pöge 2005a: 63; Pöge 2005b: 60).

Der zweite Punkt, den es zu beachten gilt, ist die Einfachheit der Codefragen. Wie hier aufgezeigt werden konnte, hat diese einen deutlichen Effekt auf die zu erzielenden Zuordnungsquoten. Zwar zeigen sich Unterschiede zwischen Geschlecht und Schulform der Befragten in dem Sinne, dass Mädchen sowie Jugendliche mit höherer Schulbildung weniger Probleme mit dem korrekten Beantworten der Fragen haben, aber um es auf eine kurze Formel zu bringen: Je einfacher die Codefragen sind, desto besser. Dies gilt dabei sowohl für den Inhalt der Fragen als auch für den zu notierenden Teil der Antwort. So fiel es den Befragten leichter, den ersten Buchstaben einer Antwort zu notieren als den letzten.

Neben diesen beiden grundlegenden Bereichen spielte in unserer Untersuchung auch das Layout der auszufüllenden Codeblätter eine Rolle. Es erwies sich als sehr hilfreich, handschriftliches Ausfüllen zu vermeiden und alle Antwortmöglichkeiten (Buchstaben und Zahlen) zum Ankreuzen aufzuführen. Hierdurch wurde die spätere Datenerfassung wesentlich erleichtert und deutlich weniger fehleranfällig.

In Bezug auf das Zuordnungsverfahren, welches mit den selbstgenerierten Codes arbeitet, ist zu sagen, dass es fehlerhafte Zuordnungen ermöglichen sollte, wenn nicht gar muss, um hinreichend gute Ergebnisse zu liefern. In beiden Erhebungen wäre die Ausschöpfungsquote mit fehlerfreien Codes unbefriedigend gewesen. In der Praxis stellte sich mit den vorgestellten Codes in unseren Stichproben die Zulassung von zwei Fehlern als guter Richtwert heraus. Mit dem erweiterten Code in Duisburg war auch die Zulassung von drei Fehlern möglich, dies erwies sich durch den relativ geringen Ertrag jedoch nur noch bedingt als sinnvoll.

In unserer Untersuchung wurde die Validierung der elektronisch gefundenen Fragebogenpärchen mithilfe von manuellen Handschriftenvergleichen durchgeführt, wobei in Duisburg pro Welle ein Aufwand von rund 4.000 Kontrollen aufgewendet wurde. Hierbei wurden mit erheblichem finanziellen Aufwand studentische Hilfskräfte eingesetzt. Schnell/Bachteler et al. haben ein fehlertolerantes Record-Linkage-Verfahren vorgestellt, welches das Heraussuchen der fehlerhaften Zuordnungsvorschläge automatisieren kann. Somit könnte möglicherweise auf die semi-automatische Vorgehensweise mit SQL-Skripten verzichtet werden. Auch sie schlagen bei ihrer Methode allerdings längere und komplexere Codes vor als bislang üblich (Schnell/Bachteler et al. 2006).

Berücksichtigt man die aufgeführten Punkte und setzt bei Paneluntersuchungen durchdachte, wohlkonzeptionierte Codes ein, können – wie vorgestellt –

durchaus beachtliche Zuordnungsquoten erreicht werden. Besonders in Studien, in denen selbstgenerierte persönliche Codes ratsam erscheinen, können sie sich trotz aller Probleme als probates Mittel der Fragebogenzuordnung erweisen.

Literatur

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

Brondies, M., 2004: Methodendokumentation der kriminologischen Schülerbefragung in Duisburg 2003. Münster und Trier: Schriftenreihe Jugendkriminalität in der modernen Stadt – Methoden (6).

Grube, J. W., M. Morgan und K. A. Kearney, 1989: Using Self-Generated Identification Codes to Match Questionnaires in Panel Studies of Adolescent Substant Use. Addictive Behaviors 14: 159-171.

Hilfert, N., 2005: Methodendokumentation der kriminologischen Schülerbefragung in Duisburg 2004. Münster und Bielefeld: Schriftenreihe Jugendkriminalität in der modernen Stadt – Methoden (10).

Kearney, K. A., R. H. Hopkins und A. L. Mauss et al., 1984: Self-Generated Identification Codes for Anonymous Collection of Longitudinal Questionnaire Data. The Public Opinion Quarterly 48: 370-378.

Kunadt, S., 2006: Methodendokumentation der kriminologischen Schülerbefragung in Duisburg 2005. Münster und Bielefeld: Schriftenreihe Jugendkriminalität in der modernen Stadt – Methoden (11).

Motzke, K. und M. Brondies, 2004: Methodendokumentation der kriminologischen Schülerbefragung in Duisburg 2002. Münster und Trier. Schriftenreihe Jugendkriminalität in der modernen Stadt – Methoden (5).

Pöge, A., 2005a: Methodendokumentation der kriminologischen Schülerbefragung in Münster 2000-2003 (Vier-Wellen-Panel). Münster und Trier: Schriftenreihe Jugendkriminalität in der modernen Stadt – Methoden (9).

Pöge, A., 2005b: Persönliche Codes bei Längsschnittstudien. Ein Erfahrungsbericht. ZA-Information 56: 50–69.

Pöge, A., 2007: Methodendokumentation der kriminologischen Schülerbefragung in Duisburg 2002 bis 2005 (Vier-Wellen-Panel). Münster und Bielefeld: Schriftenreihe Jugendkriminalität in der modernen Stadt – Methoden (13).

Schnell, R., T. Bachteler und J. Reiher, 2006: Die Anwendung statistischer Record-Linkage-Methoden auf selbstgenerierte Codes bei Längsschnitterhebungen. ZA-Information 59: 128-142.

Korrespondenzadresse: Andreas Pöge
Universität Bielefeld
Fakultät für Soziologie
Postfach 10 01 31
D-33501 Bielefeld
Andreas.Poege@uni-bielefeld.de