

Kurzfassung des Forschungsberichts 2009 und 2010

Prof. Dr. Rainer Schnell
Lehrstuhl für empirische Sozialforschung

6. Dezember 2010

Vorbemerkung

Der vorliegende Bericht ist weder in Hinsicht auf die durchgeführten Projekte noch in Hinsicht auf die Publikationen vollständig. Es werden nur diejenigen Forschungsaktivitäten benannt, die sich unmittelbar auf die Anforderungen des Forschungsberichts der Universität beziehen. Einzelheiten finden sich auf der Homepage der Arbeitsgruppe: www.uni-due.de/methods.

1 Zusammensetzung der Arbeitsgruppe

Die Arbeitsgruppe bestand aus den folgenden wissenschaftlichen Mitarbeitern:

- Tobias Bachteler
- Tobias Gramlich
- Marcel Noack
- Jörg Reiher
- Volker Stubbra
- Sabine von Thenen

2 Forschungsprojekte der Arbeitsgruppe

Derzeit laufen in der Arbeitsgruppe folgende Projekte:

- Safelink (DFG): Entwicklung kryptographischer Record-Linkage-Verfahren
- QDDS-3 (DFG): Entwicklung eines Fragebogendokumentationssystems
- Panelcodes (DFG): Verbesserung selbst-generierter Panelidentifikatoren
- Memory-Error (DFG): Untersuchung des Effekts von Erinnerungsfehlern auf Ereignisanalysen

- Nonresponse (DFG): Untersuchung des Nonresponsebias des PASS-Panels zur Hartz-IV-Evaluation anhand administrativer Daten

Bereits genehmigt, aber noch nicht angelaufen sind die Projekte

- TCS (BMBF): Untersuchung neuer Stichprobenpläne für seltene Populationen
- Record-Linkage-Zentrum (DFG): Einrichtung eines bundesweiten Zentrums für Record-Linkage Anwendungen

Derzeit in der Begutachtung befindet sich das Projekt

- Geocodes (DFG): Abstandserhaltende Verschlüsselungsverfahren für Geoinformationen

Abgeschlossen wurde das Projekt

- Arbeitsmarktbeteiligung Behinderter (im Auftrag des IAB)

3 Herausragende Ergebnisse der durchgeführten Projekte

3.1 DFG-Projekt SAFELINK

Im Rahmen des Projekts wurde ein neues Record-Linkage-Verfahren entwickelt, das trotz starker Verschlüsselung Fehler in den zur Zusammenführung verwendeten Namen erlaubt. Die zentrale Idee dabei ist es, q-Gramme der Identifikatoren mit kryptographischen Hashfunktionen in Bloom-Filtern zu speichern (Schnell et al. 2009). Testergebnisse anhand simulierter und echter Datenbestände zeigen, dass das Verfahren ähnlich gute Ergebnisse erbringt wie unverschlüsselte Identifikatoren. Die Ergebnisse des Projekts wurden unter anderem auf dem International Methodology Symposium 2010 (Statistics Canada, Ottawa) vorgestellt.

3.1.1 Kooperationen im Rahmen des Projektes

- Krebsregister Bremen
- Bremer Institut für Präventionsforschung und Sozialmedizin (Bips)
- Krebsregister Rheinland-Pfalz
- Institut für Medizinische Biometrie, Epidemiologie und Informatik (IMBEI), Universität Mainz
- Institut für Arbeitsmarkt- und Berufsforschung (IAB)
- Forschungsdatenzentrum der Bundesagentur für Arbeit im IAB
- Robert-Koch-Institut Berlin
- Swiss National Cohort

3.1.2 Ausgewählte Publikationen im Rahmen des Projektes

Schnell, R., Bachteler, T. & Reiher, J. (2009): Privacy-preserving record linkage using Bloom filters; in: BMC Medical Informatics and Decision Making 9 (41).

Schnell, R., Bachteler, T. & Reiher, J. (2009): Entwicklung einer neuen fehlertoleranten Methode bei der Verknüpfung von personenbezogenen Datenbanken unter Gewährleistung des Datenschutzes; in: Methoden - Daten - Analysen 3 (2) 203-217.

Schnell, R. (2009): Record-linkage from a technical point of view, Working Paper No. 124, RatSWD Working Paper Series, Rat für Sozial- und Wirtschaftsdaten.

Giersiepen, K., Bachteler, T., Gramlich, T., Reiher, J., Schubert, B., Novopashenny, I. & Schnell, R. (2010): Zur Leistungsfähigkeit des Record-Linkage zwischen epidemiologischen Krebsregistern und dem Mammographie-Screening; in: Bundesgesundheitsblatt 53 (7) 740-747.

3.2 DFG Projekt PANELCODES: Stability of self-generated panel codes

In Panelstudien zu sensitiven Themen werden zur Zusammenführung von Daten mehrerer Panel-Wellen sehr häufig Codes verwendet, die von den Befragten selbst generiert werden. Die Literatur zu solchen selbst-generierten Panelcodes berichtet von einem beträchtlichen Verlust von Befragten über die Zeit, da die generierten Codes instabil sind. Die Arbeitsgruppe konnte zeigen, dass sich die Verluste stark reduzieren lassen, wenn wesentlich mehr Komponenten als bisher üblich und die Codes mit einer Distanzfunktion für Zeichenketten abgeglichen werden. In Simulationen und Feldexperimenten zeigte sich das neu entwickelte Verfahren den bisher üblichen Methoden überlegen (Schnell et al. 2010).

3.2.1 Kooperationen im Rahmen des Projektes

- Universität Bielefeld, Lehrstuhl Methoden der empirischen Sozialforschung (Prof. Dr. Jost Reinecke)

3.2.2 Publikationen im Rahmen des Projekts

Schnell, R., Bachteler, T. & Reiher, J. (2010): Improving the use of self-generated identification codes; in: Evaluation Review 34 (5) 391-418.

3.3 IAB Projekt Erwerbsbeteiligung von Menschen mit Behinderung

Im Rahmen dieses für das Institut für Arbeitsmarkt- und Berufsforschung (IAB) der Bundesanstalt für Arbeit durchgeführten Projekts wurde die erste vollständige Liste der Datenquellen erstellt, die in der Bundesrepublik über die Erwerbstätigkeit von Behinderten vorhanden sind. Das Ziel des Projekts bestand in der Entwicklung eines Stichprobenplans, der für die Durchführung einer Untersuchung der Erwerbstätigkeit von Behinderten erforderlich ist. Eine entsprechende Untersuchung auf der Basis eines solchen Stichprobenplans wird zukünftig die Grundlage des Behindertenberichts der Bundesregierung bilden müssen. Der Abschlussbericht wurde als Arbeitspapier des Rates für Sozial- und Wirtschaftsdaten veröffentlicht (Schnell/Stubbra 2010).

3.3.1 Publikationen im Rahmen des Projekts

Schnell, R. & Stubbra, V. (2010): Datengrundlagen zur Erwerbsbeteiligung von Menschen mit Behinderung in der Bundesrepublik, Working Paper No. 148, RatSWD Working Paper Series, Rat für Sozial- und Wirtschaftsdaten.

3.3.2 Kooperationen im Rahmen des Projektes

- Institut für Arbeitsmarkt- und Berufsforschung (IAB)

4 Publikationen außerhalb der Projekte

Gramlich, T., Bachteler, T., Schimpl-Neimanns, B. & Schnell, R. (2009): Panelerhebungen der amtlichen Statistik als Datenquellen für die Sozialwissenschaften, Working Paper No. 132, RatSWD Working Paper Series, Rat für Sozial- und Wirtschaftsdaten.

Eifler, S., Thume, D. & Schnell, R. (2009): Unterschiede zwischen subjektiven und objektiven Messungen von Zeichen öffentlicher Unordnung („Signs of Incivility“); in: Weichbold, M., Bacher, J. & Wolf, C. (eds.): Umfrageforschung: Herausforderungen und Grenzen (Österreichische Zeitschrift für Soziologie, Sonderheft No. 9), Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 415-441.

Schnell, R. (2009): Biologische Variablen in sozialwissenschaftlichen Surveys, Working Paper No. 107, RatSWD Working Paper Series, Rat für Sozial- und Wirtschaftsdaten.

Schnell, R. (2009): Biometrische Daten: 8. Wissenschaftliche Tagung; in: König, C., Stahl, M. & Wiegand, E. (eds.): Nicht-reaktive Erhebungsverfahren, Bonn: pp. 45-60.

Schnell, R. (2010): Biological variables in social surveys, Working Paper No. 138, RatSWD Working Paper Series, Rat für Sozial- und Wirtschaftsdaten.

Dye, L., Mansfield, M., Lasikiewicz, N., Mahawish, L., Schnell, R., Talbot, D., Chauhan, H., Croden, F. & Lawton, C. (2010): Correspondence of continuous interstitial glucose measurement against arterialised and capillary glucose following an oral glucose tolerance test in healthy volunteers; in: British Journal of Nutrition 103 (1) 134-140.

Gramlich, T., Bachteler, T., Schimpl-Neimanns, B. & Schnell, R. (2010): Panelerhebungen der amtlichen Statistik als Datenquellen für die Sozialwissenschaften; in: AStA Wirtschafts- und Sozialstatistisches Archiv 4 (3) 153-183.

5 Wichtigste Publikationen

Schnell, R., Bachteler, T. & Reiher, J. (2009): Privacy-preserving record linkage using Bloom filters; in: BMC Medical Informatics and Decision Making 9 (41).

Schnell, R., Bachteler, T. & Reiher, J. (2010): Improving the use of self-generated identification codes; in: Evaluation Review 34 (5) 391-418.

6 Forschungsperspektive der Arbeitsgruppe

Die Arbeitsgruppe wird weiter an der Verbesserung von Stichprobenplänen, dem Nonresponse-Problem und der Weiterentwicklung von Record-Linkage-Techniken arbeiten. Mit dem bereits von der DFG genehmigten Record-Linkage-Zentrum wird den Sozial- und Humanwissenschaften eine nachhaltige Infrastruktureinrichtung angeboten, die durch Serviceleistungen und Forschungsarbeiten die Zahl und Qualität von Record-Linkage-Anwendungen in den Fachwissenschaften steigern soll. Ein Beispiel dafür ist das bereits bei der DFG beantragte Projekt Geocodes, in dem ein von der Arbeitsgruppe entwickeltes Verfahren untersucht wird. Das Verfahren erlaubt die Berechnung von Entfernungen von Geokoordinaten eines Scientific-Use-Files bei Wahrung der Anonymität der Objekte (Geoprivacy). Die Arbeitsgruppe wird sich bemühen, die Sicherheit des Verfahrens nachzuweisen und es im Rahmen des Record-Linkage-Zentrums den inhaltlich arbeitenden Kollegen zur Verfügung zu stellen.