

- Mehlkop, Guido und Rolf Becker (2004), Soziale Schichtung und Delinquenz. Eine empirische Anwendung einer Rational-Choice-Ansatzes mit Hilfe von Querschnittsdaten des ALLBUS 1990 und 2000, *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 1, 56, 95–126.
- Merton, Rober King (1968), *Social Theory of Social Structure*, New York: The Free Press.
- Metje, Matthias, (1994), *Wählerschaft und Sozialstruktur im Generationenwechsel. Eine Generationenanalyse des Wahlverhaltens bei Bundestagswahlen*, Wiesbaden: Deutscher Universitätsverlag.
- Mooser, Joseph (1997), Die «Geistige Landesverteidigung» in den 1930er Jahren, *Schweizerische Zeitschrift für Geschichte*, 4, 47, 685–708.
- Müller, Walter (1998), Erwartete und unerwartete Folgen der Bildungsexpansion, in: Jürgen Friedrichs, M. Rainer Lepsius und Karl Ulrich Mayer, Eds., *Die Diagnosefähigkeit der Soziologie*, Sonderheft der Kölner Zeitschrift für Soziologie und Sozialpsychologie, 38, Opladen: Westdeutscher Verlag, 81–112.
- Opp, Karl-Dieter (1974), *Abweichendes Verhalten und Gesellschaftsstruktur*, Darmstadt und Neuwied: Luchterhand.
- Ormel, Johan; Siegwart Lindenberg, Nardin Steverink und Lois M. Verbrugge (1999), Subjective Well-Being and Social Production Functions, *Social Indicators Research*, 46, 61–90.
- Oswald, Hans (2000), East German adolescents' attitudes towards West German democracy, in: Peter M. Daly, Hans Walter Frischkopf, Trudis Goldsmith-Reber und Horst Richter, Eds., *Images of Germany*, New York: Peter Lang, 121–134.
- Picht, Georg (1964), *Die deutsche Bildungskatastrophe. Analyse und Dokumentation*, Freiburg: Walter.
- Plum, Wolfgang (1982), Kohortenanalyse von Umfragedaten. Zur Identifizierung möglicher Einflussfaktoren politischen Verhaltens und politischer Einstellungen, *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 3, 34, 509–532.
- Rokeach, Milton (1972), *Beliefs, Attitudes and Values. A Theory of Organization and Change*, San Francisco: Jossey-Bass.
- Runciman, Walter G. (1966), *Relative Deprivation and Social Justice*, London: Routledge & Kegan Paul.
- Ryder, Norman B. (1965), The cohort as a concept in the study of social change, *American Sociological Review*, 30, 843–861.
- Sacchi, Stefan (1992), Postmaterialismus in der Schweiz von 1972 bis 1990, *Schweizerische Zeitschrift für Soziologie*, 1, 18, 87–117.
- Shavit, Yossi und Walter Müller, Eds., (1998), *From School to Work. A Comparative Study of Educational Qualifications and Occupational Destinations*, Oxford: Clarendon Press.
- Solga, Heike und Sandra Wagner (2001), Paradoxie der Bildungsexpansion. Die doppelte Benachteiligung von Hauptschülern, *Zeitschrift für Erziehungswissenschaft*, 1, 4, 107–127.
- Theodorson, George A. und Achilles G. Theodorson (1969), *A Modern Dictionary of Sociology*, New York: Cromwell.
- Tuma, Nancy B. und Michael T. Hannan, (1984), *Social Dynamics. Models and Methods*, Orlando: Academic Press.
- Volken, Thomas (2002), *Elemente des Vertrauens*, Berlin: Peter Lang.
- Wagner, Michael (2001), *Kohortenstudien in Deutschland. Expertise für die Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik*, Köln: Universität zu Köln.
- Weber, Max ([1921] 1972), *Wirtschaft und Gesellschaft*, Tübingen: J.C.B. Mohr (Paul Siebeck).

Überlegungen zum Berner Stichprobenplan¹

Ben Jann*

1 Einleitung

Wie auch in anderen Ländern ohne zentrales Einwohnerregister stellt in der Schweiz die Ziehung einer repräsentativen Bevölkerungsstichprobe die empirische Sozialforschung vor schwerwiegende Probleme. Ohne Urliste der Mitglieder der zu untersuchenden Population lässt sich naturgemäss nur schwer eine unverzerrte Zufallsstichprobe aus derselben ziehen. Zudem ist es, wenn man über ein ansprechendes und praktikables Stichprobenverfahren verfügt, auch nicht immer ganz einfach festzustellen, ob das Verfahren tatsächlich unverzerrte Resultate liefert. Die Erstellung von qualitativ hochwertigen Stichproben kann also sehr zeitaufwändig sein und unter Umständen erhebliche Kosten verursachen. Somit ist wohl nicht zuletzt aufgrund beschränkter Forschungsbudgets damit zu rechnen, dass ein grosser Teil der empirischen Umfrageforschung in der Schweiz auf zumindest zweifelhaften Stichproben beruht.

Die einfachste und wahrscheinlich am häufigsten angewandte Stichprobenmethode in der Schweiz ist die Ziehung einer Haushaltsstichprobe aus dem Telefonverzeichnis (vgl. etwa Jann, 2001; für Deutschland Schnell, 1997), die in einem zweiten Schritt nach der Erfassung der Haushaltsstrukturen durch Ziehung von einzelnen Zielpersonen in den Haushalten in eine Personenstichprobe überführt wird. Die Probleme dieser Methode sind bekannt: Erstens gibt es einen gewichtigen Anteil der Bevölkerung, der auf diesem Weg nicht erreicht werden kann, weil kein oder nur ein nicht registrierter Telefonanschluss vorliegt (zu etwas älteren Schätzungen dieses Anteils siehe zum Beispiel die Studien von Schmutz und Grau, 1998 und 2000; Experten gehen heute von einem Anteil nicht über das Telefonverzeichnis erreichbarer Personen von 10 bis 15 Prozent aus). Zweitens handelt es sich bei der Ziehung aus dem Telefonregister wie angesprochen zumeist um ein zweistufiges Verfahren, bei dem innerhalb der Haushalte auf der zweiten Stufe eine Personenauswahl getroffen wird. Man kann sich hierbei leicht zusätzliche Verzerrungen einhandeln, wenn die Auswahl der Zielperson im Haushalt zum Beispiel nach dem Schwedenschlüssel oder mit der Geburtstagsmethode durch die Haushalte selbst vorgenommen wird.

* Ben Jann, ETH Zürich, Soziologie, SEW E 28, Scheuchzerstrasse 70, CH-8092 Zürich, E-Mail: jann@soz.gess.ethz.ch.

1 In Gedenken an Herbert Iff († 18. April 1998). Für wertvolle Hinweise bedanke ich mich bei Ruth Meyer Schweizer, Stefan Sacchi und einem anonymen Gutachter.

(dies gilt allerdings nicht für CATI-Umfragen, bei denen die Auswahl der Zielperson normalerweise per Computer nach telefonischer Erfassung der Haushaltsstruktur erfolgt). Zudem führt die Zweistufigkeit des Verfahrens meistens zu einer Stichprobe mit nicht identischen Auswahlwahrscheinlichkeiten, was tendenziell zu geringerer Stichprobeneffizienz führt und die Unannehmlichkeiten einer Gewichtung bei der Datenanalyse nach sich zieht. Ein drittes schwerwiegendes Problem von Telefonstichproben betrifft das Vorhandensein von Mehrfacheinträgen im Telefonregister beziehungsweise von unterschiedlichen Einträgen, die auf die gleichen Haushalte verweisen, und den damit verbundenen Schwierigkeiten, die Auswahlwahrscheinlichkeiten einzelner Personen genau zu bestimmen.

Alternativen zu den Stichproben aus dem Telefonregister gibt es wenige, so dass zum Beispiel auch das Bundesamt für Statistik für Grossbefragungen wie etwa die jährlich durchgeführte Schweizerische Arbeitskräfteerhebung (SAKE) darauf zurückgreift. Auch stecken Random-Digit-Dialing-Verfahren (RDD), mit denen man zumindest die nicht eingetragenen Haushalte erreichen könnte, in der Schweiz noch in den Kinderschuhen. Abklärungen durch das Bundesamt für Statistik sind im Gange und erste Gehversuche werden zum Beispiel vom LINK-Institut unternommen. Zudem scheint die BIK Aschpurwis + Behrens GmbH aus Hamburg neuerdings RDD-Stichproben für die Schweiz anzubieten, die sich an das Verfahren von Gabler und Häder anlehnen (1997; vgl. auch den Übersichtsartikel von Häder und Glemser, 2006). Ein Problem ist aber natürlich auch bei RDD-Verfahren, dass ein Teil der Bevölkerung so nicht erreichbar ist (zur Problematik des Ausschlusses von ganzen Bevölkerungsteilen aus allgemeinen Bevölkerungsumfragen siehe etwa Schnell, 1991) und der Schritt vom Telefonanschluss zur effektiv zu interviewenden Zielperson steinig sein kann.

Telefonstichproben (und RDD-Stichproben, wenn ein entsprechendes Verfahren einsatzbereit ist) mögen also zwar verhältnismässig einfach und preiswert zu erlangen sein, sie sind aber aus den angesprochen Gründen methodisch nicht wirklich befriedigend. Eine bessere Datenquelle zur Ziehung von qualitativ hochwertigen Stichproben wären da ohne Zweifel die Einwohnerregister der Gemeinden. Die Register der Gemeinden, obwohl zurzeit noch in recht heterogener Form bezüglich der Art der Datenerfassung und der enthaltenen Information vorliegend, geben sicherlich das aktuellste und vollständigste Abbild der Bevölkerung der Schweiz.² Nicht abgedeckt werden einzig einige Randgruppen wie illegal eingewanderte Personen, die sich der Registrierung durch die Behörden bewusst entziehen. Geringfügige Unschärfen ergeben sich zudem aufgrund von Mobilitätsprozessen. Insgesamt ist aber an den Einwohnerregisterdaten eigentlich nur problematisch, dass die Daten

2 Auf Bundesebene gibt es konkrete Bestrebungen zur Harmonisierung der Einwohnerregister insbesondere auch zum Zwecke bevölkerungstatistischer Erhebungen (siehe z. B. Bundesamt für Statistik, 2004). So hat das Parlament am 23. Juni 2006 das Registerharmonisierungsgesetz (RHG) verabschiedet (teilweises Inkrafttreten des RHG am 1. November 2006) und es ist vorgesehen, die Volkszählung 2010 zumindest teilweise als Registererhebung durchzuführen.

nicht in zentralisierter Form zugänglich sind. Entschliesst man sich, eine Stichprobe mit Hilfe der Register zu ziehen, muss man in einem ersten Schritt also ein Verfahren zur Auswahl von Gemeinden anwenden. In einem zweiten Schritt wird dann mit den Gemeindeverwaltungen Kontakt aufgenommen, um die Zielpersonen der Stichprobe zu bestimmen.³

Ein Gemeinde-Auswahlverfahren kann sein, mit Hilfe des vom Bundesamt für Statistik geführten Gemeindeverzeichnisses, das unter anderem Angaben zur Anzahl der Einwohner in den Gemeinden enthält, eine künstliche Liste der Schweizer Bevölkerung zu erstellen, und aus dieser Liste mit einem Zufallsalgorithmus eine einfache Wahrscheinlichkeitsauswahl zu ziehen. Gemeinden, die mit mindestens einem Treffer in der fiktiven Stichprobe enthalten sind, werden dann zwecks Ziehung einer entsprechenden Anzahl realer Zielpersonen aus deren Einwohnerregister angeschrieben. Der Nachteil dieses Verfahrens ist, dass es zu einer Stichprobe mit sehr feiner geographischer Granularität führt. Das heisst, in einer so gezogenen Stichprobe sind sehr viele Gemeinden enthalten und in einem grossen Teil dieser Gemeinden muss vielleicht nur gerade je ein Interview durchgeführt werden. Eine Folge davon ist, dass die Adressbeschaffung sehr zeitaufwändig und auch kostspielig wird. Zudem entstehen etwa bei Face-to-Face-Interviews hohe Reisespesen.

Eine Lösung zur Verminderung der geographischen Streuung und somit der Reduktion der Kosten ist die Ziehung einer Klumpenstichprobe, bei der zum Beispiel Gemeinden mit einem PPS-Verfahren (*Probability Proportional to Size*) gezogen werden, und dann pro gezogene Gemeinde eine fixe Anzahl Interviews durchgeführt wird.⁴ Die Anzahl in der Stichprobe enthaltenen Gemeinden kann so im Vergleich zu einer einfachen Wahrscheinlichkeitsauswahl stark reduziert werden. Allerdings leidet aber auch die statistische Effizienz, das heisst, die Schätzung von Populationsparametern auf Grundlage solcher Klumpenstichproben ist unter Umständen mit einer deutlich erhöhten Unsicherheit behaftet.

Vor dem Hintergrund, dass eine einfache Wahrscheinlichkeitsauswahl zwar gute statistische Eigenschaften aufweist, aber nur aufwändig zu erlangen ist, und umgekehrt bei einer kostengünstigen Klumpenstichprobe die statistische Effizienz zu wünschen übrig lässt, haben Fritschi, Meyer und Schweizer (1976) einen Vorschlag für eine «geteilte» Zufallsstichprobe ausgearbeitet. Das Verfahren hat sich

3 Zu erwähnen ist hier, dass das Bundesamt für Statistik im Rahmen des neuen Systems der Haushalts- und Personenstatistiken (SHAPE) und im Zuge der Harmonisierung der Personenregister versucht, unter anderem auf Grundlage der Einwohnerregister eine zentrale Datenbank aufzubauen, die die Ziehung von qualitativ hochwertigen Stichproben aus der Schweizer Bevölkerung erlauben soll (Bundesamt für Statistik, 2007, 8). Vor 2010 ist mit dieser Möglichkeit aber nicht zu rechnen und der Zugang dürfte wohl eher restriktiv gehandhabt werden (nur für Forschungsprojekte von «nationalem Interesse»).

4 Beziehungsweise in grossen Gemeinden unter Umständen ein Vielfaches dieser fixen Anzahl. Bei sehr schiefen Verteilungen wie der Gemeindegrössenverteilung macht nur ein PPS-Verfahren *mit Zurücklegen* Sinn (einzelne Gemeinden hätten in einem Verfahren ohne Zurücklegen Auswahlwahrscheinlichkeit grösser eins), so dass eine Gemeinde mehrmals in die PPS-Stichprobe gelangen kann.

in Anlehnung an die Herkunft der Autoren als der «Berner Stichprobenplan» in der Literatur niedergeschlagen. Zwar wurde der Stichprobenplan in der Schweiz für eine Reihe von Studien verwendet (die prominenteste Arbeit, bei der auf das Verfahren zurückgegriffen wurde, ist die Armutsstudie von Leu et al., 1997; vgl. weiter z. B. Meyer et al., 1982, oder Wydler et al., 1996). Da es sich beim Berner Stichprobenplan aber um einen viel versprechenden Ansatz handelt, dessen Vorteile gegenüber den üblichen Telefonregisterstichproben meiner Meinung nach bisher zu wenig genutzt werden, möchte ich nachfolgend einige Überlegungen zu diesem Verfahren präsentieren.

Im nächsten Abschnitt wird das von Fritschi et al. (1976) vorgeschlagene Verfahren genauer beschrieben. Es wird unter anderem gezeigt, dass das Verfahren zu leicht verzerrten Stichproben führt, und es wird ein korrigiertes Verfahren entwickelt, das diese Verzerrungen ausgleicht. Der dritte Abschnitt befasst sich mit einer nahe liegenden Vereinfachung des Verfahrens und es werden die Eigenschaften des Berner Stichprobenplans anhand von Simulationsergebnissen illustriert. Der vierte Abschnitt fasst die Ergebnisse zusammen.

2 Das «geteilte» Stichprobenverfahren von Fritschi, Meyer und Schweizer

Der Ausgangspunkt des Klumpungsverfahrens nach Fritschi et al. (1976) ist die Überlegung, dass eine Klumpung der Stichprobe für relativ grosse Gemeinden wenig Sinn macht, weil auf diese Gemeinden auch bei einer einfachen Wahrscheinlichkeitsauswahl eine hinreichend grosse Anzahl Stichprobenmitglieder entfällt. Durch die Klumpung würde man in diesem Teil der Population unnötigerweise eine Menge Präzision verschenken. Das Problem sind vielmehr die vielen kleinen Gemeinden, in denen bei einer einfachen Wahrscheinlichkeitsauswahl jeweils nur eine kleine Anzahl Interviews durchgeführt werden muss. Fritschi et al. (1976) schlagen deshalb das folgende Verfahren zur Erstellung einer «geteilten» Zufallsstichprobe vor, bei dem eine zuvor festgelegte Mindestzahl von k Stichprobenmitgliedern pro Gemeinde nicht unterschritten wird: In einem ersten Schritt ziehe man ausgehend von einem Gemeindeverzeichnis mit Angaben zu den Einwohnerzahlen (bzw. der Anzahl Stimm- und Wahlberechtigten im Anwendungsfall von Fritschi et al.) eine (hypothetische) Stichprobe aus der durch das Verzeichnis definierten Population. Die Ziehung erfolgt nach den Regeln einer einfachen Wahrscheinlichkeitsauswahl mit konstanter Auswahlwahrscheinlichkeit $p = n/N$, wobei n dem gewünschten Umfang der Stichprobe und N der Populationsgrösse entspricht. Für jede Gemeinde wird dann die Anzahl «Treffer», also die Anzahl hypothetischer Stichprobenmitglieder, die auf diese Gemeinde entfallen sind, gezählt. Gemeinden, die die gewünschte Mindesttrefferzahl k erreichen, werden mit ihren Treffern für die Endstichprobe vorgemerkt. Aus der Population in den verbleibenden Gemeinden, also den

Gemeinden mit weniger als k oder gar keinen Treffern, ziehe man dann in einem zweiten Schritt eine Stichprobe mit reduzierter Trefferwahrscheinlichkeit $p' = p/k$. Gemeinden mit mindestens einem Treffer in diesem zweiten Schritt gelangen ebenfalls in die Endstichprobe, und zwar mit je k Stichprobenmitgliedern pro Treffer. Die Endstichprobe setzt sich somit zusammen aus je einer Zielperson pro Treffer in den Gemeinden mit k oder mehr Treffern im ersten Schritt und jeweils k Zielpersonen pro Treffer im zweiten Schritt. Die Bestimmung der effektiven Befragungspersonen erfolgt schliesslich in der Praxis am besten über die Einwohner- oder Stimmregister der in der Stichprobe enthaltenen Gemeinden. Als Alternative wäre auch ein Random-Route-Verfahren denkbar.

Der Ansatz von Fritschi et al. (1976) ist einleuchtend: Zuerst eine einfache Wahrscheinlichkeitsauswahl durchführen, dann denjenigen Teil der Stichprobe, in dem nur wenige Treffer pro Gemeinde vorliegen, durch eine Klumpenstichprobe ersetzen. Anders als von Fritschi et al. vermeintlich bewiesen, handelt es sich aber nicht um eine Methode, bei der die *a priori* Auswahlwahrscheinlichkeiten in allen Gemeinden gleich sind, und die Verteilungen von mit der Gemeindegrösse zusammenhängenden Merkmalen werden verzerrt wiedergegeben. Dies soll nun dargelegt werden.

Sei P_{ij} die Wahrscheinlichkeit, dass Person i aus Gemeinde j in die Stichprobe gelangt. Bei einer einfachen Wahrscheinlichkeitsauswahl ist P_{ij} für alle Personen aus der Grundgesamtheit gleich, also

$$P_{ij} = \frac{n}{N} \quad \text{für alle } i, j$$

wenn n der Stichprobenumfang und N die Populationsgrösse ist. Stichproben, die diese Eigenschaft besitzen, werden manchmal als EPSEM-Stichproben bezeichnet (*Equal Probability of Selection Method*, Babbie, 1979, 330). Zwar ist es möglich, auch aus Stichproben mit nicht identischen Auswahlwahrscheinlichkeiten auf die Grundgesamtheit zu schliessen, sofern die Auswahlwahrscheinlichkeiten bekannt sind. Im Allgemeinen sind jedoch EPSEM-Stichproben aus theoretischen (mehr statistische Effizienz) sowie praktischen Gründen (es werden weniger komplexe Schätzer benötigt) vorzuziehen (eine Ausnahme mögen je nach Eigenschaften des Untersuchungsgegenstandes disproportional geschichtete Stichproben bilden). Im vorliegenden Fall eines gemeindebasierten Stichprobenverfahrens sollte insbesondere darauf geachtet werden, dass die Auswahlwahrscheinlichkeiten nicht mit der Gemeindegrösse zusammenhängen (oder aber, dass der Zusammenhang genau bekannt ist und mit entsprechenden Gewichten bei der Datenanalyse ausgeglichen werden kann).

Wie erläutert wird beim zweistufigen Berner Stichprobenplan in einem ersten Schritt eine einfache Wahrscheinlichkeitsauswahl mit Auswahlwahrscheinlichkeit $p = n/N$ durchgeführt. Es wird sodann die Anzahl «Treffer», also die Anzahl ausgewähl-

ter Personen pro Gemeinde gezählt und in einem zweiten Schritt unter den Gemeinden, die eine kritische Anzahl Treffer k nicht erreicht haben, eine Klumpenstichprobe gezogen. Die Stichprobe setzt sich schliesslich zusammen aus den im ersten Schritt gewählten Personen in den Gemeinden mit mehr als k Treffern und den Klumpen zu je k Personen aus dem zweiten Schritt. Sei X eine Zufallsvariable der Anzahl Treffer pro Gemeinde im ersten Schritt und Y eine von X unabhängige Zufallsvariable der Anzahl Klumpen pro Gemeinde im zweiten Schritt. Die Auswahlwahrscheinlichkeit von Person i aus Gemeinde j kann dann geschrieben werden als

$$P_{ij} = P(X_j \geq k) \frac{E(X_j | X_j \geq k)}{N_j} + P(X_j < k) \frac{E(Y_j | X_j < k) \cdot k}{N_j} \quad (1)$$

wobei $P(X_j \geq k) = 1 - P(X_j < k)$ der Wahrscheinlichkeit entspricht, dass in Gemeinde j im ersten Schritt k oder mehr Treffer erreicht werden. N_j entspricht der Anzahl Personen in Gemeinde j und $E(\cdot)$ symbolisiert den Erwartungswert.

Wird mit Zurücklegen gezogen, können X und Y als binomialverteilte Zufallsvariablen aufgefasst werden, also

$$X_j \sim B(N_j, p) \text{ und } Y_j \sim B(N_j, p')$$

mit $p = n/N$ und $p' = p/k$ als den Trefferwahrscheinlichkeiten. Es gilt insbesondere

$$E(X_j) = p \cdot N_j \text{ und } E(Y_j) = p' \cdot N_j$$

Aus der Unabhängigkeit von X und Y folgt zudem $E(Y_j | X_j < k) = E(Y_j)$. Wenn der Berner Stichprobenplan nun EPSEM-Charakter haben soll, dann muss $P_{ij} = n/N$ gelten (für alle i und j). Durch Einsetzen in (1) erhält man nach einigen Umformungen jedoch

$$P_{ij} = \frac{n}{N} + P(X_j \geq k) \left(\frac{E(X_j | X_j \geq k)}{N_j} - \frac{n}{N} \right) \quad (2)$$

Da der zweite Summand in Gleichung (2) von N_j abhängt und nur in Spezialfällen gleich null ist, wird die Bedingung konstanter Auswahlwahrscheinlichkeiten verletzt.

Es ist somit zu erwarten, dass der Berner Stichprobenplan nach Fritschi et al., der im ersten wie auch im zweiten Schritt jeweils konstante Auswahlwahrscheinlichkeiten vorsieht, zu verzerrten Ergebnissen führt. Die Verzerrung kann indes korrigiert werden, indem zum Beispiel im zweiten Schritt variable, wie folgt aus

Formel (1) ableitbare Auswahlwahrscheinlichkeiten eingesetzt werden (zur numerischen Anwendung von Gleichung 3 siehe Anhang):⁵

$$p'_j = \frac{\frac{n}{N} - P(X_j \geq k) \frac{E(X_j | X_j \geq k)}{N_j}}{k \cdot [1 - P(X_j \geq k)]} \quad (3)$$

Die Eigenschaften der Verzerrung im unkorrigierten Berner Stichprobenplan können mit Hilfe einer Simulation veranschaulicht werden. Als Grundlage für die Simulation verwende ich das Gemeindeverzeichnis 2002, das Angaben zu den Einwohnerzahlen aller politischen Gemeinden enthält (Bundesamt für Statistik, 2002). Gezogen wird jeweils eine Stichprobe im Umfang von $n = 3750$ Personen, die minimale Anzahl Zielpersonen pro Gemeinde ist auf $k = 10$ festgelegt (dies entspricht bezüglich der Anzahl in die Stichprobe gelangender Gemeinden in etwa den im Anwendungsbeispiel von Fritschi et al. verwendeten Parametern; Fritschi et al. legten ihrem Verfahren die kleinere Population der Wahlberechtigten zu Grunde). Tabelle 1 zeigt die Simulationsergebnisse über 10'000 Runden. Angegeben ist einerseits die faktische Gemeindegrössenverteilung in der Population und andererseits die aus den Stichproben resultierende gemittelte Gemeindegrössenverteilung für verschiedene Ziehungsverfahren. Man erkennt, dass bei der ursprünglichen Version des Berner Stichprobenplans die meisten Gemeindegrössenkategorien leicht unterrepräsentiert sind. Zum Beispiel leben 12.06 Prozent der Bevölkerung in Gemeinden mit 2000 bis 3499 Einwohnern, in den Stichproben auf Grundlage des Berner Stichprobenplans sind es im Schnitt aber nur 11.71 Prozent. Eine Kategorie jedoch, Gemeindegrösse 10'000 bis 49'999, ist deutlich übervertreten. Ähnlich wie bei einer einfachen Wahrscheinlichkeitsauswahl (letzte Spalte) treten diese Verzerrungen beim «korrigierten» Berner Stichprobenplan, bei dem die Auswahlwahrscheinlichkeiten im zweiten Schritt nach Formel (3) bestimmt wurden, jedoch nicht auf.

Ein eindrückliches Bild der Natur der Verzerrung im unkorrigierten Berner Stichprobenplan liefert Abbildung 1. Dargestellt ist die relative Verteilung der gemittelten Gemeindegrössenverteilung in den Stichproben und der Gemeindegrössenverteilung in der Population (bzw. genauer: für jede Gemeinde ist das Verhältnis ihres durchschnittlichen Anteils an der Stichprobe und ihres tatsächlichen Anteils an der Gesamtbevölkerung abgebildet). Man sieht sehr deutlich, dass Personen aus Gemeinden mit ca. 10'000 bis 30'000 Einwohnern zu häufig in die Stichprobe

5 Ein alternatives Korrekturverfahren mit einem ähnlichen Effekt bestünde darin, die Auswahlwahrscheinlichkeiten im ersten Schritt variabel zu gestalten, dafür aber im zweiten Schritt konstant zu halten. Eine weitere Korrekturstrategie wäre, die Stichprobe mit dem unkorrigierten Verfahren zu ziehen, jedoch bei der Datenanalyse entsprechende Design-Gewichte einzusetzen.

Tabelle 1: Gemeindegrößenverteilung im Berner Stichprobenplan

Gemeindegrösse	Bevölkerungsanteil	Stichprobenverteilungen (10'000 Replikationen)		
		Berner Stichprobenplan (unkorrigiert)	Berner Stichprobenplan (korrigiert)	einfache Zufallsauswahl
1 – 249	0.98	0.94	0.98	0.98
250 – 499	2.47	2.39	2.48	2.47
500 – 999	5.56	5.36	5.58	5.57
1000 – 1999	10.49	10.18	10.48	10.48
2000 – 3499	12.06	11.71	12.08	12.07
3500 – 4999	9.89	9.60	9.87	9.90
5000 – 9999	16.79	16.49	16.77	16.80
10000 – 49999	25.67	27.25	25.66	25.65
50000 und mehr	16.09	16.08	16.10	16.08
Total	100.00	100.00	100.00	100.00

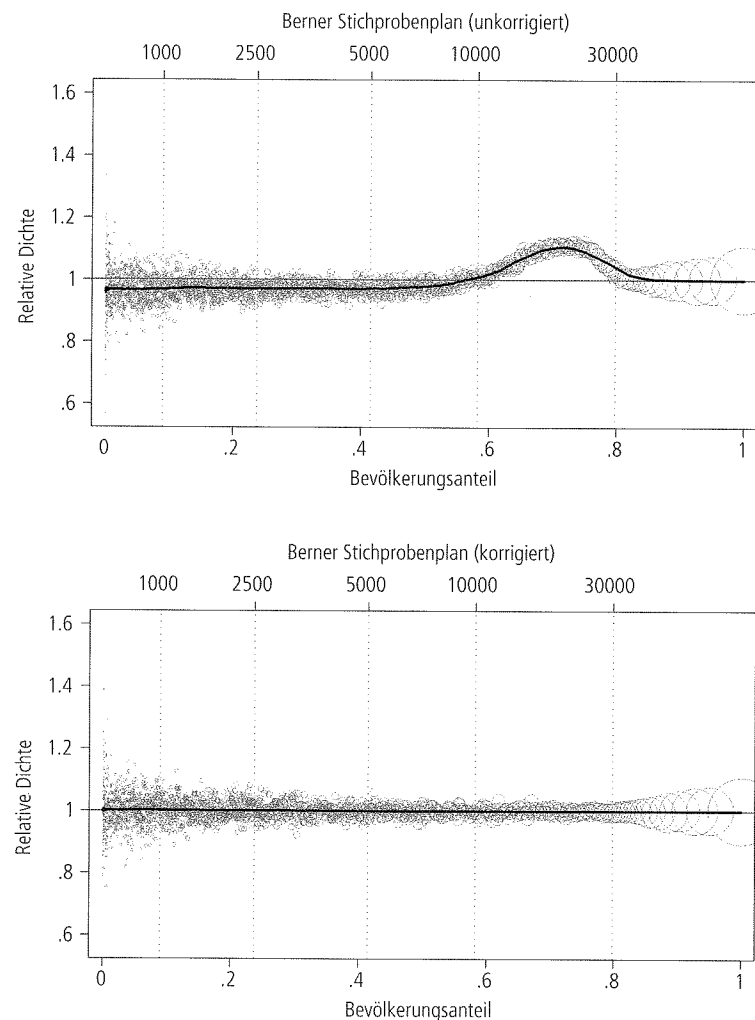
Quelle: Simulationsergebnisse (10'000 Replikationen) auf Grundlage des Gemeindeverzeichnis 2002 (BFS, 2002, 2876 Gemeinden mit insgesamt 7'241'468 Einwohnern) mit Stichprobengrösse 3750 und Klumpengrösse 10.

gelangen. Natürlich sind dies gerade diejenigen Gemeinden, die an der «Grenze» zur Klumpenstichprobe liegen. Das sind Gemeinden mit einer erwarteten Trefferzahl nahe der Mindestklumpengrösse k , die also je nach Ausgang der einfachen Wahrscheinlichkeitsauswahl im ersten Schritt manchmal in das Klumpenauswahlverfahren des zweiten Schritts gelangen und manchmal nicht. Beim korrigierten Berner Stichprobenplan mit Auswahlwahrscheinlichkeiten nach Formel (3) wird die Verzerrung, wie man erkennen kann, erfolgreich geglättet.

Anzumerken ist, dass die hier recht bedeutend erscheinende Verzerrung in der von Fritschi et al. (1976) gewählten Umsetzung ihres Stichprobenverfahrens deutlich geringer war. Fritschi et al. verwendeten wohl aus praktischen Gründen einen Ziehungsalgorithmus, bei dem ausgehend von einem Zufallsstartwert die nach Gemeinden strukturierte Population mit fixen Auswahlritten durchwandert wird (sog. systematisches Auswahlverfahren). Dieses Verfahren führt zu einer geringeren Variation der Anzahl Treffer in einer Gemeinde mit gegebener Gemeindegrösse als ein nicht systematisches Verfahren, was die Verzerrung der Gemeindegrößenverteilung stark reduziert.⁶

⁶ Der von Fritschi et al. (1976) angewendete Algorithmus kann als Verfahren angesehen werden, bei dem quasi nach Gemeindegrösse geschichtet gezogen wird. Eine systematische Schichtung nach Gemeindegrösse kann durchaus wünschenswert sein. Allerdings sollte man dann gleich ein «echtes» Schichtungsverfahren mit einfacheren theoretischen Eigenschaften verwenden (vgl. auch Abschnitt 4).

Abbildung 1: Vergleich der Gemeindeverteilung des Berner Stichprobenplans (gemittelt über 10'000 Ziehungen) mit der Gemeindeverteilung in der Grundgesamtheit



Legende: Dargestellt ist für jede Gemeinde das Verhältnis zwischen ihrem durchschnittlichen Anteil in der Stichprobe (über 10'000 Ziehungen) und ihrem Anteil an der Gesamtbevölkerung. Ein Wert von eins bedeutet, dass die Mitglieder der Gemeinde im Durchschnitt genau ihrem Anteil an der Gesamtbevölkerung entsprechend repräsentiert werden (zum Konzept der relativen Verteilung siehe Handcock und Morris, 1999). Die Gemeinden sind ihrer Grösse nach auf der proportional zum kumulierten Bevölkerungsanteil skalierten Abszisse geordnet, zudem richtet sich die Grösse der Symbole nach der Anzahl Einwohner in den entsprechenden Gemeinden. Da in einer endlichen Simulation die Werte vor allem für kleine Gemeinden beträchtlich streuen, wurde zur Verdeutlichung des Trends eine geglättete Kurve eingezeichnet (Sasieni, 1998).

3 Ein vereinfachtes «geteiltes» Verfahren

Zwar ist es, wie oben dargelegt, möglich, das Verfahren von Fritschi et al. (1976) so zu modifizieren, dass unverzerrte Stichproben erzeugt werden. Alternativ könnten auch nachträglich Gewichte berechnet werden, um der Verzerrung bei der Datenanalyse entgegenzuwirken. Ich möchte hier aber zeigen, dass beide Möglichkeiten wenig praktische Relevanz besitzen, da es ein Verfahren zur Ziehung einer «geteilten» Stichprobe gibt, das in vergleichbarer Art zu einer Reduktion der Anzahl Gemeinden in der Stichprobe führt, jedoch einfacher umzusetzen ist und erst noch etwas bessere statistische Eigenschaften aufweist. Die Komplikationen des Verfahrens von Fritschi et al. rühren daher, dass die Gemeinden *ex post*, also erst *nach* der ersten Ziehung einer einfachen Wahrscheinlichkeitsauswahl aufgrund der empirisch realisierten Anzahl Treffer in zwei Gruppen eingeteilt werden. Eine offensichtliche Strategie zur Vermeidung dieser Probleme ist, die Gruppenzuteilung ganz einfach *ex ante* aufgrund der theoretisch erwarteten Anzahl Treffer vorzunehmen. Ich schlage also ein Verfahren vor, bei dem – gegeben die Trefferwahrscheinlichkeit $p = n/N$ – die Gemeinden zuerst in eine Gruppe mit erwarteter Trefferzahl gleich k oder grösser, also mit $E(X_j) = p \cdot N_j \geq k$, und eine Gruppe mit erwarteter Trefferzahl kleiner k , also mit $E(X_j) < k$, aufgeteilt werden. In einem zweiten Schritt wird dann aus der in Gruppe 1 mit den grossen Gemeinden enthaltenen Population eine einfache Wahrscheinlichkeitsauswahl gezogen und aus Gruppe 2 mit den kleinen Gemeinden eine Klumpenstichprobe mit je k Zielpersonen pro Klumpen. Der angestrebte Stichprobenumfang n wird dabei im Verhältnis der durch die beiden Gruppen abgedeckten Bevölkerungsanteile aufgeteilt. Das heisst, der zu realisierende Stichprobenanteil in Gruppe 1 ist

$$n_1 = \frac{n}{N} \cdot \sum_{E(X_j) \geq k} N_j = \sum_{E(X_j) \geq k} E(X_j)$$

und die Anzahl in Gruppe 2 zu ziehender Klumpen beträgt

$$\frac{n_2}{k} = \frac{n}{k \cdot N} \cdot \sum_{E(X_j) < k} N_j = \frac{1}{k} \cdot \sum_{E(X_j) < k} E(X_j)$$

wobei bei nicht ganzzahligen Ergebnissen für n_1 und n_2/k in der Praxis Zufallsrundungen eingesetzt werden können. Es folgt unmittelbar, dass es sich bei der beschriebenen *ex ante* geteilten Zufallsstichprobe um ein Verfahren mit identischen Auswahlwahrscheinlichkeiten handelt, da

$$P_{j|E(X_j) \geq k} = \frac{n_1}{\sum_{E(X_j) \geq k} N_j} = P_{j|E(X_j) < k} = \frac{k \cdot \frac{n_2}{k}}{\sum_{E(X_j) < k} N_j} = \frac{n}{N}$$

Die praktische Durchführung der Ziehung einer *ex ante* geteilten Zufallsstichprobe ist bedeutend einfacher als die Ziehung einer Stichprobe nach dem korrigierten Verfahren von Fritschi et al. (1976), da keine komplizierten Ergebnisse wie Formel (3) benötigt werden. Das Verfahren hat aber noch einen anderen Vorteil: Durch die vorgängige Aufteilung der Population entsteht ein Schichtungseffekt, der die Effizienz der Stichprobe für Merkmale, die einen Zusammenhang zur Gemeindegrösse aufweisen, im Vergleich zum Stichprobenplan von Fritschi et al. zusätzlich erhöht.

Tabelle 2 zeigt die Ergebnisse einer Simulation, bei der das hier vorgeschlagene *ex ante* geteilte Stichprobenverfahren mit einer einfachen Wahrscheinlichkeitsauswahl, einer Klumpenstichprobe und dem *ex post* geteilten Stichprobenverfahren (dem korrigierten Berner Stichprobenplan nach Formel 3) verglichen wird. Grundlage der Simulation ist wiederum das Gemeindeverzeichnis 2002 (Bundesamt für Statistik, 2002). Gezogen werden Stichproben im Umfang von $n = 1000$. Die Klumpengrösse in den geklumpten Stichproben beträgt $k = 8$.⁷ Die Simulationsergebnisse in Tabelle 2 beziehen sich auf die Stichprobenschätzer der Erwartungswerte von verschiedenen Merkmalen mit unterschiedlichem linearen Zusammenhang zur Gemeindegrösse (r) und mit unterschiedlicher Intra-Klassen-Korrelation in den Gemeinden (ρ ; die Intra-Klassen-Korrelation ist ein Mass für die interne Homogenität in den Gemeinden). Sämtliche Merkmale sind standardisiert, das heisst, sie haben in der Population einen Mittelwert von 0 und eine Standardabweichung von 1. Dargestellt sind für jedes Stichprobenverfahren die Durchschnitte und Standardabweichungen der Mittelwertsschätzer über 10'000 Ziehungen.

Im Fuss von Tabelle 2 finden sich Angaben zur durchschnittlichen Anzahl PSU (*Primary Sampling Units*, d. h. Anzahl Klumpen)⁸ und der durchschnittlichen Anzahl Gemeinden in den Stichproben. Man erkennt sehr schön, dass alle Klumpungsverfahren in vergleichbarer Weise zu einer Reduktion der Anzahl Gemeinden auf etwas mehr als 100 führen – im Vergleich zu den knapp 600 Gemeinden, die bei einer einfachen Wahrscheinlichkeitsauswahl in die Stichprobe gelangen. Verbunden mit der Verringerung der Anzahl Gemeinden ist naturgemäss eine Reduktion der Anzahl PSU, wobei aber diese Reduktion beim Berner Stichprobenplan (275 bzw. 266 PSU) deutlich geringer ausfällt als bei der reinen Klumpenstichprobe (125 PSU). Gerade in dieser geringeren Reduktion der Anzahl PSU bei gleichzeitig vergleichbarer Reduktion der Anzahl Gemeinden liegt die Stärke des Berner Stichprobenplans gegenüber der reinen Klumpenstichprobe, wobei dieser Vorteil allerdings

7 Man beachte, dass mit diesen Simulationsparametern nur gerade acht der rund 3000 Gemeinden in den ungeklumpten Zweig der *ex ante* geteilten Stichprobe gelangen (es sind dies Zürich, Genf, Basel, Bern, Lausanne, Winterthur, St. Gallen und Luzern).

8 Bei der einfachen Wahrscheinlichkeitsauswahl sind die PSU identisch mit den Zielpersonen, das heisst, eine Stichprobe von Umfang n enthält n Klumpen mit jeweils einem Mitglied. Die reine Klumpenstichprobe enthält n/k Klumpen mit je k Mitgliedern. Bei den geteilten Verfahren enthalten die Stichproben eine Mischung aus Ein-Personen- und k -Personen-Klumpen.

nur unter bestimmten, noch zu identifizierenden Bedingungen tatsächlich zu einer verbesserten Stichprobeneffizienz führt.

An den Resultaten im Hauptteil von Tabelle 2 erkennt man, dass mit allen Stichprobenverfahren die Populationsmittelwerte der vier Merkmale erwartungstreu geschätzt werden. Die Verfahren unterscheiden sich aber deutlich hinsichtlich der Streuung der Mittelwertsschätzer über die einzelnen Stichproben. Zum Beispiel ist bei der Klumpenstichprobe der Standardfehler des Mittelwertsschätzers für Merkmal X1, welches perfekt mit der Gemeindegrösse korreliert (X1 ist eine lineare Transformation der Gemeindegrösse), fast dreimal so gross wie bei der einfachen Wahrscheinlichkeitsauswahl. Der Design-Effekt in der Simulation, also das Verhältnis der Varianzen, beträgt in diesem Fall 8.12, was in etwa dem nach der Formel von Kish (1965, 162) erwarteten Design-Effekt für Klumpenstichproben von $1 + \rho(k - 1) = 8$ entspricht (vgl. Tabelle 3). Beim korrigierten Berner Stichprobenplan ist der Standardfehler für Merkmal X1 jedoch nur geringfügig grösser als bei der einfachen Zufallsauswahl und beim *ex ante* geteilten Verfahren ist er sogar bedeutend kleiner (der Design-Effekt beträgt 0.36). Letzteres liegt am angesprochenen Schichtungseffekt durch die vorgängige Teilung der Stichprobe in zwei Gemeindegrössengruppen, aus denen dann separate Stichproben gezogen werden.

Bei Merkmal X2, das mittelstark mit der Gemeindegrösse korreliert, aber über keine über das durch den linearen Zusammenhang bedingte Ausmass hinausgehende Intra-Klassen-Korrelation verfügt,⁹ schneidet der Berner Stichprobenplan ebenfalls sehr gut ab. Der Design-Effekt ist für die *ex ante* geteilte Stichprobe immer noch Varianz vermindern und beträgt 0.83 (gegenüber 2.79 für die Klumpenstichprobe; Tabelle 3, Spalte 2). Wird nun die Intra-Klassen-Korrelation aber systematisch erhöht, verlieren die geteilten Stichprobenverfahren zunehmend an Boden. Merkmal X3 weist wie X2 einen linearen Zusammenhang von 0.5 auf, die Intra-Klassen-Korrelation wurde aber künstlich von den minimal 0.25 auf ebenfalls 0.5 angehoben. Die Streuung ist nun in der *ex ante* geteilten Stichprobe deutlich höher als in der einfachen Wahrscheinlichkeitsauswahl, der Design-Effekt ist aber mit 2.40 immer noch nur etwa halb so gross wie bei der reinen Klumpenstichprobe.

Die Vorteile des Berner Stichprobenplans verschwinden schliesslich fast vollständig, wenn die interne Homogenität der Gemeinden auf hohem Niveau beibehalten, der lineare Zusammenhang zur Gemeindegrösse aber eliminiert wird. Merkmal X4 ist unabhängig von der Gemeindegrösse an sich, weist aber eine relativ starke interne Homogenität in den Gemeinden auf. Das heisst, Personen aus der gleichen Gemeinde sind sich bezüglich X4 – unabhängig von der Gemeindegrösse – im Durchschnitt bedeutend ähnlicher als Personen aus unterschiedlichen Gemeinden. Die Eigenschaften des Berner Stichprobenplans verschlechtern sich hier verglichen mit der einfachen Wahrscheinlichkeitsauswahl deutlich. Der Design-Effekt erreicht

⁹ Aus dem linearen Zusammenhang zwischen einem Merkmal X und der Gemeindegrösse folgt ein Mindestmass an interner Homogenität in den Gemeinden von $\rho^{\min} = r^2$.

Tabelle 2: Mittelwerte und Standardfehler (in Klammern) von Erwartungswertsschätzern verschiedener Stichprobenverfahren

	Erwartungswert	Simulationsergebnisse (10'000 Replikationen)			
		einfache Zufallsauswahl	Klumpenstichprobe	Berner Stichprobenplan	
				ex post geteilte Stichprobe	ex ante geteilte Stichprobe
X1 ($r = 1.00, \rho = 1.00$)	0.000	0.000 (0.032)	0.001 (0.090)	0.000 (0.033)	0.000 (0.019)
X2 ($r = 0.50, \rho = 0.50$)	0.000	0.000 (0.032)	0.000 (0.053)	0.000 (0.032)	0.000 (0.029)
X3 ($r = 0.50, \rho = 0.50$)	0.000	0.000 (0.032)	0.000 (0.067)	0.000 (0.051)	-0.001 (0.049)
X4 ($r = 0.00, \rho = 0.50$)	0.000	0.000 (0.032)	-0.001 (0.067)	0.000 (0.064)	-0.001 (0.065)
X5 ($r = 0.00, \rho = 0.50^*$)	0.000	0.000 (0.032)	-0.001 (0.067)	0.000 (0.033)	0.000 (0.032)
X6 ($r = 0.00, \rho = 0.00$)	0.000	0.000 (0.032)	0.000 (0.032)	0.000 (0.032)	0.000 (0.031)
X7 ($r = 0.10, \rho = 0.07$)	0.000	0.000 (0.032)	0.000 (0.039)	0.000 (0.037)	-0.001 (0.037)
Anzahl PSU		1000.0	125.0	275.1	265.8
Anzahl Gemeinden		568.8	104.7	105.9	105.9

Quelle: Simulationsergebnisse (10'000 Replikationen) auf Grundlage des Gemeindeverzeichnisses 2002 (BFS, 2002, 2876 Gemeinden mit insgesamt 7'241'468 Einwohnern) mit Stichprobengrösse 1000 und Klumpengrösse 8. Bei den Variablen X1 bis X7 handelt es sich um künstlich für die Population erzeugte Merkmale (r : Korrelation mit der Gemeindegrösse; ρ : Intra-Klassen-Korrelation, * mit $\rho = 1$ in den grossen und $\rho = 0$ in den kleinen Gemeinden).

nun fast einen Wert, wie er für die reine Klumpenstichprobe festgestellt werden kann, und liegt – etwas überraschend – deutlich über dem theoretisch erwarteten Design-Effekt für eine reine Klumpenstichprobe mit einer dem Wert für die geteilte Stichprobe entsprechenden Anzahl von 266 Klumpen bzw. PSU (vgl. die letzte Zeile von Tabelle 3). Das heisst, dass sich die grössere Anzahl PSU im geteilten Verfahren bei Merkmal X4 nicht voll ausspielen kann. Daraus lässt sich schliessen, dass es auf die genaue Struktur ankommt, in der sich die Intra-Klassen-Korrelation eines Merkmals präsentiert. Bei den bisherigen Merkmalen wurde von homoskedastischer Varianz innerhalb der Gemeinden ausgegangen. Die Intra-Klassen-Korrelation kann nun aber wiederum selbst eine Funktion der Gemeindegrösse sein, was sich auf die Effizienz des Berner Stichprobenplans auszuwirken scheint. Merkmal X5 repräsentiert den Extremfall in dem in den grossen Gemeinden mit $E(X_j) \geq k$ perfekte interne Homogenität herrscht, die kleinen Gemeinden mit $E(X_j) < k$ jedoch eine Intra-Klassen-Korrelation von $\rho = 0$ aufweisen. Die Varianzen in den beiden

Teilen der Population wurden zudem so gewählt, dass sich über alle Gemeinden hinweg eine Intra-Klassen-Korrelation von 0.5 ergibt. An den Simulationsresultaten in den Tabellen 2 und 3 erkennt man, dass in dieser Situation, von der man auf den ersten Blick annehmen könnte, sie unterscheidet sich kaum von der Situation für Merkmal X4, das *ex ante* geteilte Stichprobenverfahren wiederum sehr gut abschneidet und einen neutralen Design-Effekt von 1 erreicht. An was liegt das? Die Begründung ist klar: In den kleinen Gemeinden ist die geteilte Stichprobe eine Klumpenstichprobe und bei einer Intra-Klassen-Korrelation von 0 ist die Klumpenstichprobe gleich effizient wie eine einfache Wahrscheinlichkeitsauswahl (vgl. unten). In den grossen Gemeinden entspricht die geteilte Stichprobe einer einfachen Wahrscheinlichkeitsauswahl, die bei einer Intra-Klassen-Korrelation von 1 einen maximalen Effizienzvorteil gegenüber der Klumpenstichprobe aufweist. Die *ex ante* geteilte Stichprobe *musst* in dieser Situation folglich einen Design-Effekt von 1 haben.¹⁰ Umgekehrt lässt sich auch ableiten, unter welchen Bedingungen die geteilte Stichprobe genau gleich ineffizient ist wie eine reine Klumpenstichprobe. Dies ist der Fall, wenn in den kleinen Gemeinden perfekte interne Homogenität herrscht und die grossen Gemeinden perfekte Heterogenität aufweisen. Gegeben die Korrelation zwischen einem Merkmal und der Gemeindegrösse ist null, liegt der Berner Stichprobenplan bezüglich der statistischen Effizienz also je nach Struktur der internen Homogenität in den Gemeinden zwischen den beiden Polen der einfachen Wahrscheinlichkeitsauswahl und der Klumpenstichprobe. Bei Korrelation zwischen dem Merkmal und der Gemeindegrösse (bzw. genauer: bei positiver Intra-Klassen-Korrelation in den beiden Gemeindegrössenschichten) kommt beim *ex ante* geteilten Stichprobenverfahren zusätzlich der Schichtungseffekt zum Tragen, so dass für einzelne Merkmale unter Umständen eine grössere statistische Effizienz als in der einfachen Wahrscheinlichkeitsauswahl erreicht werden kann.

Merkmal X6 illustriert weiterhin den Fall, in dem mit allen Stichprobenverfahren die gleichen Ergebnisse erzielt werden. Wenn keine Intra-Klassen-Korrelation in den Gemeinden besteht, das heisst, wenn sich Personen aus der gleichen Gemeinde bezüglich eines Merkmals nicht ähnlicher sind als Personen aus unterschiedlichen Gemeinden, dann spielt es für die Effizienz der Stichprobe keine Rolle, ob die Stichprobe nach Gemeinden geklumpt ist oder nicht. Selbst in der reinen Klumpenstichprobe ist in diesem (in der Realität wohl eher selten anzutreffenden) Fall also der Standardfehler nicht grösser als in der einfachen Wahrscheinlichkeitsauswahl.

Bei den Merkmalen X1 bis X6 ging es darum, die Eigenschaften der verschiedenen Stichprobenpläne unter theoretisch interessanten Varianten von Parameterkonstellationen zu evaluieren. Natürlich wird man aber in der Empirie kaum z. B. Intra-Klassen-Korrelationen von 0.5 oder höher finden (nach Kalton, 1983, 31, liegt

Tabelle 3: Simulierte und theoretische Design-Effekte (Verhältnis der Varianzen)

	X1 (r = 1.00, p = 1.00)	X2 (r = 0.50, p = 0.25)	X3 (r = 0.50, p = 0.50)	X4 (r = 0.00, p = 0.50)	X5 (r = 0.00, p = 0.50 ^a)	X6 (r = 0.00, p = 0.00)	X7 (r = 0.10, p = 0.07)
Klumpenstichprobe vs. einfache Zufallsauswahl	8.12	2.79	4.55	4.55	4.55	1.01	1.50
<i>ex ante</i> geteilte Stichprobe vs. einfache Zufallsauswahl	0.36	0.83	2.40	4.21	1.00	0.99	1.36
<i>ex ante</i> geteilte Stichprobe vs. Klumpenstichprobe	0.04	0.30	0.53	0.93	0.22	0.97	0.90
<i>ex ante</i> geteilte Stichprobe vs. <i>ex post</i> geteilte Stichprobe	0.33	0.79	0.92	1.01	0.92	0.98	0.98
Theoretische Design-Effekte für Klumpenstichproben							
– mit 125 Klumpen	8.00	2.75	4.50	4.50	4.50	1.00	1.49
– mit 266 Klumpen	3.76	1.69	2.38	2.38	2.38	1.00	1.19

Quelle: Beruhend auf den Zahlen in Tabelle 2.

r: Korrelation mit der Gemeindegrösse; p: Intra-Klassen-Korrelation. ^a mit $p = 1$ in den grossen und $p = 0$ in den kleinen Gemeinden.

p i. d. R. im Bereich 0–0.15, wobei es aber je nach Thema Ausnahmen geben kann). Merkmal X7 repräsentiert einen realistischen moderaten Fall mit einem linearen Zusammenhang zur Gemeindegrösse von 0.1 und einer Intra-Klassen-Korrelation von 0.07. Diese Zahlen entsprechen den Werten für das Merkmal Nationalität (Schweizer vs. Ausländer) in der Volkszählung 2000 (Quelle: Bundesamt für Statistik, harmonisierte Personen-Daten der eidgenössischen Volkszählungen 1970–2000, eigene Berechnungen). Auch hier zeigt sich der Berner Stichprobenplan gegenüber der reinen Klumpenstichprobe vorteilhaft, die Unterschiede sind aber nicht so dramatisch wie in den Fällen mit hohen Korrelationen. Der Design-Effekt für die reine Klumpenstichprobe beträgt 1.50, für die *ex ante* geteilte Stichprobe 1.36. Mit anderen Worten: Knapp 30% des Effizienzverlustes durch die Klumpung können in diesem Fall mit dem Berner Stichprobenverfahren wieder gutgemacht werden.

Tabelle 3 mit den Design-Effekten enthält schliesslich noch den Vergleich zwischen dem *ex ante* geteilten Stichprobenverfahren und dem *ex post* geteilten Verfahren, bei dem die Auswahlwahrscheinlichkeiten nach Formel (3) bestimmt wurden. Man erkennt, dass aufgrund des Schichtungseffekts die *ex ante* geteilte Stichprobe in fast allen Situationen zumindest geringfügig besser abschneidet als die *ex post* geteilte Stichprobe und nie eine grössere Varianz aufweist. Die Wahl des einfacher umzusetzenden *ex ante* geteilten Stichprobenverfahrens ist somit die «dominante» Strategie: man kann sich gegenüber dem *ex post* geteilten Verfahren nur verbessern, nicht aber verschlechtern. Ähnliches gilt für den Vergleich der *ex ante* geteilten Stichprobe mit der reinen Klumpenstichprobe (Tabelle 3, Zeile 3).

10 Die Tatsache, dass die *ex ante* geteilte Stichprobe geschichtet ist, spielt hier keine Rolle, da X4 unkorreliert ist mit der Gemeindegrösse. Das heisst, der Mittelwert von X4 ist in beiden Schichten identisch und es kann somit kein Schichtungseffekt entstehen.

4 Zusammenfassung und Diskussion

Die in der Schweiz am häufigsten eingesetzte Methode zur Erstellung einer Bevölkerungsstichprobe – die Ziehung aus dem Telefonregister – ist vor allem aufgrund mangelnder Abdeckung der Grundgesamtheit nur mit Vorbehalt für Studien, die wissenschaftliche Qualitätsstandards erfüllen sollen, in Betracht zu ziehen. Der Berner Stichprobenplan, bei dem die Stichprobe über die Einwohner- oder Stimmregister der Gemeinden bestimmt wird, kann hier eine Alternative sein. Zwar sind die Kosten und der Zeitaufwand der Adressbeschaffung höher als bei einer Telefonregisterstichprobe, die zweifelsohne stark verbesserte Qualität der resultierenden Ausgangsstichprobe lässt den Zusatzaufwand, der sich durch die Klumpung der Stichprobe auf wenige Gemeinden in überschaubarem Rahmen bewegt, aber durchaus als lohnenswert erscheinen. Ohnehin ist der Aufwand für die Stichprobenziehung eher klein, wenn er am Aufwand gemessen wird, der in der Regel für die Durchführung der Interviews nötig ist. Man spart sicherlich am falschen Ort, wenn man auf billige, qualitativ schlechte Stichprobenverfahren setzt, auch wenn durch die Kosteneinsparung ein paar zusätzliche Interviews realisiert werden können.¹¹

Die Eigenschaften des Berner Stichprobenplans wurden in den Abschnitten 2 und 3 erläutert. Es wurde erstens gezeigt, dass das ursprünglich von Fritschi et al. (1976) vorgeschlagene Verfahren zu leicht verzerrten Stichproben führt, was aber mit ein paar Modifikationen korrigiert werden kann. Zudem wurde mit dem *ex ante* geteilten Stichprobenverfahren ein alternativer Ansatz vorgeschlagen, der einfacher umzusetzen ist und erst noch etwas bessere statistische Eigenschaften aufweist als der ursprüngliche Stichprobenplan. Unabhängig davon, ob man jetzt das Originalverfahren verwendet oder das hier vorgeschlagene vereinfachte Verfahren, zeigt sich der Berner Stichprobenplan vor allem dann gegenüber der Klumpenstichprobe als vorteilhaft, wenn ein Zusammenhang zwischen dem Untersuchungsmerkmal und der Gemeindegrösse besteht. Bei Merkmalen ohne Zusammenhang zur Gemeindegrösse sind die Verhältnisse etwas komplizierter. Die genaue Struktur der Intra-Klassen-Korrelation ist dann von Bedeutung. Etwas überraschend erzielt der Berner Stichprobenplan besonders gute Ergebnisse, wenn die interne Homogenität in den grossen Gemeinden am stärksten ist. Dies erscheint etwas ungünstig, da man eher davon ausgehen würde, dass, falls überhaupt Unterschiede bestehen, für die meisten Merkmale die interne Homogenität in kleinen, ländlichen Gemeinden stärker ist als in den Städten. Trotz dieses Vorbehalts zeigen die Simulationen jedoch, dass der

11 Längerfristig wird es zumindest für Grossprojekte wahrscheinlich möglich sein, qualitativ hochwertige Stichproben direkt aus den harmonisierten, beim Bundesamt für Statistik zusammengezogenen Einwohnerregistern zu ziehen (vgl. Fussnote 3). Der Nutzen des Berner Stichprobenplans als Verfahren zur Erlangung von Stichproben mit verbesserter Abdeckung würde sich in diesem Fall erübrigen. Weiterhin relevant bliebe der Berner Stichprobenplan aber als effiziente Methode zur Reduzierung der Kosten von Face-to-Face- bzw. CAPI-Befragungen, bei denen die Interviewer zu den Befragungspersonen nach Hause fahren müssen.

Berner Stichprobenplan in der Regel besser abschneidet als ein einfaches Klumpungsverfahren und somit bei etwa gleich hohem Aufwand für die Beschaffung der Adressen der Zielpersonen insgesamt zu effizienteren Stichproben führt.

Auch die Datenanalyse gestaltet sich beim Berner Stichprobenplan kaum komplizierter als bei einer einfachen Klumpenstichprobe. Bei beiden Verfahren müssen Schätzer verwendet werden, die die Klumpenstruktur der Daten in Rechnung stellen, um die Varianzen der statistischen Koeffizienten nicht zu unterschätzen. Bei Stichproben auf Grundlage des *ex ante* geteilten Verfahrens ist zudem zu berücksichtigen, dass die Stichprobe aus zwei Schichten gezogen wurde, was die Varianzen im Allgemeinen verringert. Komplexe Schätzer, die Klumpungs- und Schichtungseffekte in angemessener Weise behandeln, stehen für eine Vielzahl statistischer Verfahren zur Verfügung (zur Analyse von komplexen Stichproben vgl. z. B. Lohr, 1999, oder Levy und Lemeshow, 2002) und sind mittlerweile in allgemeinen Statistik-Programmen recht gut unterstützt (vgl. z. B. StataCorp, 2005). Surveys, bei denen es weder Schichtungs- oder Klumpenstrukturen noch Gewichte zur Korrektur unterschiedlicher Auswahlwahrscheinlichkeiten zu berücksichtigen gilt, sind ohnehin recht selten. Die Datenanalyse dürfte somit beim Berner Stichprobenplan keine zusätzlichen Schwierigkeiten bereiten.

Bei allen Vorteilen des Verfahrens sind zwei kritische Aspekte des Berner Stichprobenplans jedoch noch zu thematisieren. Die hier präsentierten Simulationsergebnisse beruhen alle auf dem Gemeindeverzeichnis 2002 der Schweiz. Eine Eigenschaft der Schweizer Gemeinden ist, dass die Verteilung der Anzahl Personen, die in diesen Gemeinden leben, sehr schief ist. Das heisst, es gibt hunderte kleiner Gemeinden aber nur eine Hand voll grosse. Natürlich ist eine solche Datenstruktur eine Bedingung dafür, dass der Berner Stichprobenplan überhaupt zu anderen Resultaten führt als ein einfaches Klumpungsverfahren. Wird das geteilte Stichprobenverfahren auf Aggregateinheiten angewendet, die sich in ihrer Grösse nur wenig unterscheiden, erreicht unter Umständen keine einzige Einheit die notwendige Mindestzahl an Treffern für die Aufnahme in den ungeklumpten Zweig der Stichprobe. Eine schiefe Verteilung der Grössen der verwendeten Aggregateinheiten ist also eine Voraussetzung für eine sinnvolle Anwendung des Berner Stichprobenplans. Weiterhin benötigt eine Stichprobe nach dem Berner Verfahren neben der Stichprobengrösse n die Festlegung einer Mindestklumpengrösse k , was forschungswirtschaftliche Fragen aufwirft. Im Prinzip lässt sich k durch Gegenüberstellung von Kosten und statistischer Effizienz optimieren. Eine solche Optimierung ist aber nicht ganz trivial und hängt von einer Reihe von Annahmen ab. Den Ausgangspunkt für Überlegungen zur geeigneten Grösse von k können die Optimierungsformeln für klassische Klumpenstichproben bilden (vgl. z. B. Kalton, 1983, 36, oder Kish, 1965, Kap. 8).

Zum Abschluss bleibt darauf hinzuweisen, dass sich die hier vorgestellte Studie nur auf die Darlegung der Vorteile des Berner Verfahrens gegenüber einer klassischen Klumpenstichprobe bezieht. Wie sich der Berner Stichprobenplan im Vergleich zu

alternativen Verfahren wie z. B. einer geschichteten Klumpenstichprobe auf Grundlage von Gemeindetypologien und -größen (vgl. etwa Buchmann und Sacchi, 1997) bewährt, muss in einer Folgestudie geklärt werden. Eine solche Studie wäre auch von Interesse, weil sie möglicherweise Wege zur zusätzlichen Effizienzsteigerung durch die Kombination von Verfahren aufzeigen würde. Unter Umständen kann zum Beispiel die Stichprobeneffizienz erhöht werden, wenn dem Berner Stichprobenplan eine Schichtung nach Gemeindetypen vorgeschaltet wird.

5 Anhang

Die folgenden Ergebnisse werden zur Lösung von Gleichung (3) benötigt. Für eine binomialverteilte Zufallsvariable $X \sim B(n, p)$ ist die Wahrscheinlichkeit, k oder mehr Treffer zu erreichen, bekanntlich gegeben als

$$P(X \geq k) = \sum_{x=k}^n \binom{n}{x} p^x (1-p)^{n-x}$$

(vgl. z. B. Evans et al., 2000, 43–47). Funktionen zur automatischen Berechnung von $P(X \geq k)$ stehen in den meisten Statistikprogrammen zur Verfügung. Etwas aufwändiger ist die Berechnung von $E(X | X \geq k)$, also dem Erwartungswert einer binomialverteilten Zufallsvariablen, gegeben eine bestimmte Mindestzahl an Treffern wird erreicht. Der Erwartungswert einer diskreten Variablen entspricht dem Mittel der mit den Auftretenswahrscheinlichkeiten gewichteten Ausprägungen, im vorliegenden Fall also:

$$E(X | X \geq k) = \frac{k \cdot P(X = k) + (k+1) \cdot P(X = k+1) + \dots + n \cdot P(X = n)}{P(X \geq k)}$$

Dies lässt sich vereinfachen zu:

$$E(X | X \geq k) = k + \frac{P(X \geq k+1) + \dots + P(X = n)}{P(X \geq k)}$$

6 Literaturverzeichnis

- Babbie, Earl R. (1979), *The Practice of Social Research*, Second Edition, Belmont, CA: Wadsworth.
 Buchmann, Marlis und Stefan Sacchi (1997), *Berufsverlauf und Berufsidentität im sozio-technischen Wandel. Konzeption, Methodik und Repräsentativität einer retrospektiven Befragung der Geburtsjahrgänge 1949–51 und 1959–61*, ETH Zürich.

- Bundesamt für Statistik (2002), *Gemeinde- und Ortschaftenverzeichnis 2002*, Neuchâtel: BFS.
 Bundesamt für Statistik (2004), *Register im Dienst der Statistik. Einblick in das Grossprojekt Registerharmonisierung*, Neuchâtel: Bundesamt für Statistik.
 Bundesamt für Statistik (2007), *Das neue System der Haushalts- und Personenstatistiken. Ein Monitoring für die Zukunft*, Neuchâtel: Bundesamt für Statistik.
 Evans, Merran; Nicholas Hastings und Brian Peacock (2000), *Statistical Distributions*. Third Edition, New York: Wiley.
 Fritschi, Peter; Ruth Meyer und Willy Schweizer (1976), Ein neuer Stichprobenplan für ein gesamtschweizerisches Sample, *Schweizerische Zeitschrift für Soziologie*, 2, 3, 149–158.
 Gabler, Siegfried und Sabine Häder (1997), Überlegungen zu einem Stichprobendesign für Telefonumfragen in Deutschland, *ZUMA-Nachrichten*, 41, 7–19.
 Häder, Sabine und Axel Glemser (2006), Stichprobenziehung für Telefonumfragen in Deutschland, in: A. Diekmann (Hrsg), *Methoden der Sozialforschung, Sonderheft 44 der Kölner Zeitschrift für Soziologie und Sozialpsychologie*, Wiesbaden: VS Verlag, 148–171.
 Hancock, Mark S. und Martina Morris (1999), *Relative Distribution Methods in the Social Sciences*, New York: Springer.
 Jann, Ben (2001), *Stichprobenziehung aus TwixTel*, Universität Bern. Erhältlich unter <http://www.socio.ethz.ch/people/jann/wp/twixtel.pdf>.
 Kalton, Graham (1983), *Introduction to Survey Sampling*, Newbury Park: Sage.
 Kish, Leslie (1965), *Survey Sampling*, New York: Wiley.
 Leu, Robert E.; Stefan Burri und Tom Priester (1997), *Lebensqualität und Armut in der Schweiz*, Bern: Haupt.
 Levy, Paul S. und Stanley Lemeshow (1999), *Sampling of Populations: Methods and Applications*, 3rd ed., New York: Wiley.
 Lohr, Sharon L. (1999), *Sampling: Design and Analysis*, Pacific Grove: Duxbury Press.
 Meyer, Ruth; Karl Haltiner, Rudolf Hofer, Herbert Iff und Walter Rüegg (1982), *Fragen an die Zukunft. Die Bedeutung von Beruf, Bildung und Politik für die zwanzigjährigen Schweizerinnen und Schweizer*, Aarau: Sauerländer.
 Sasieni, Peter (1998), An adaptive variable span running line smoother, *Stata Technical Bulletin*, 41, 4–7.
 Schmutz, Susanna, und Peter Grau (1998), *Telefon-Anschlüsse der privaten Haushalte in der Schweiz. Situationsanalyse 1998*, Luzern: LINK Institut.
 Schmutz, Susanna, und Peter Grau (2000), *Sind die SchweizerInnen überhaupt noch zu erreichen? Telefonanschlüsse der privaten Haushalte in der Schweiz im Jahr 2000*, Luzern: LINK Institut.
 Schnell, Rainer (1991), Wer ist das Volk? Zur faktischen Grundgesamtheit bei «allgemeinen Bevölkerungsumfragen»: Undercoverage, Schwererreichbare und Nichtbefragbare. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 43, 1, 106–137.
 Schnell, Rainer (1997), Praktische Ziehung von Zufallsstichproben für Telefon-Surveys, *ZA-Informationen*, 40, 45–59.
 StataCorp (2005), *Stata Survey Data Reference Manual*. College Station, Texas: Stata Press.
 Wydler, Hans; Therese Walter, Achim Hattich, Rainer Hornung und Felix Gutzwiller (1996), *Die Gesundheit 20jähriger in der Schweiz: Ergebnisse der PRP 1993*, Aarau: Sauerländer.