

économie et statistiques

octobre 2006

14

**Die Umstellung des LFS (Labour Force Survey) auf RDD-CATI (Random Digit Dialing – Computer Assisted Telephone Interviewing)**

Prof. Dr. Rainer SCHNELL, Center for Quantitative Methods and Survey Research, Universität

Konstanz

Gutachten im Auftrag von STATEC

SAVOIR POUR AGIR

**statec**  
LUXEMBOURG

# Inhaltsverzeichnis

## Die Umstellung des LFS auf RDD-CATI

Enquête sur les Forces de Travail : Étude de faisabilité sur la collecte des données avec le système Random Digit Dialing – Computer Assisted Telephone Interviewing (RDD-CATI) – Résumé.....	3
<b>1. Gegenstand des Gutachtens .....</b>	<b>4</b>
<b>2. Rechtliche Grundlagen des LFS in Luxemburg.....</b>	<b>4</b>
<b>3. Beschreibung der bisherigen Praxis des LFS in Luxemburg .....</b>	<b>4</b>
3.1 Survey Design.....	4
3.1.1 Grundgesamtheit, Auswahlgesamtheit, Survey-Population .....	4
3.1.2 Schichtung .....	5
3.1.3 Stichprobenverfahren .....	5
3.1.4 Datenerhebung im Haushalt.....	5
3.1.5 Datenerhebungsmodus.....	5
3.2 Ergebnisse der Feldarbeit.....	5
3.2.1 Feldarbeit-Details.....	5
3.2.2 Nonresponse-Error.....	6
3.2.3 Andere Nonsampling-Errors.....	7
3.2.4 Sampling-Error.....	7
<b>4. Der LFS als RDD-CATI .....</b>	<b>9</b>
4.1 Probleme der Adaption des Erhebungsinstruments auf CATI-RDD.....	9
4.2 Sampling-Frame-Konstruktion für RDD.....	9
4.2.1 Verfügbarkeit der Telefonnummern.....	9
4.2.2 Aufbereitung des Sampling-Frames.....	9
4.2.3 Abschätzung der Zahl der privaten Anschlüsse .....	9
4.2.4 Abschätzung der Blockdichten .....	10
4.2.5 Abschätzung des Effekts nicht eingetragener privater Anschlüsse.....	12
4.3 Abschätzung des Coverage-Errors .....	13
4.4 Abschätzung des Sampling-Errors.....	14
4.5 Nonresponse .....	14
4.5.1 Ausmaß bei unterschiedlichen Feldprozeduren .....	15
4.5.2 Non-contacts und Callbacks.....	15
4.5.3 Anschreiben .....	15
4.5.4 Refusals und Refusal-Konversion.....	15
4.5.5 Diskussion der Möglichkeiten von Incentives.....	15
4.6 Implikationen des Auswahlgesetzes .....	16
4.7 Imputationen und deren Kosten.....	16
4.8 Diskussion der Möglichkeit von Record-Linkage und dessen Kosten .....	16
4.9 Veränderung der Datenerfassung oder des Data-Editing .....	17
4.10 Konsequenzen des RDD für die statistische Nutzung der Daten .....	17
4.10.1 Einfluss auf Gewichtungungsverfahren .....	17
4.10.2 Einfluss auf Hochrechnungsverfahren .....	17
4.10.3 Konsequenzen des RDD auf die Varianzschätzung .....	17
4.11 Pilotstudie.....	19
<b>5. Langfristige Perspektiven .....</b>	<b>20</b>
<b>6. Anhang: Konzeption und Durchführung des LFS in den Ländern Europas .....</b>	<b>21</b>
<b>7. Literatur .....</b>	<b>22</b>

## Enquête sur les Forces de Travail : Étude de faisabilité sur la collecte des données avec le système Random Digit Dialing – Computer Assisted Telephone Interviewing (RDD-CATI) – Résumé

L'Enquête sur les Forces de Travail est une enquête harmonisée au niveau de l'UE visant à collecter des informations sur la situation socio-économique de la population (emploi, chômage, départ à la retraite, éducation et formation). Jusqu'en 2002, c'était une enquête ponctuelle où la période de référence était une semaine au printemps. Avec le passage à l'enquête continue, toutes les semaines de l'année servent de période de référence et le STATEC doit transmettre, chaque trimestre, à EUROSTAT un fichier avec les données individuelles recueillies. Ce changement a entraîné de nouvelles contraintes quant aux délais à respecter au niveau des interviews auprès des ménages. Il a été décidé d'abandonner les interviews 'face à face' et de les remplacer par des interviews par téléphone. En supprimant les déplacements, souvent sans résultats (refus, absences), des enquêteurs auprès des ménages de l'échantillon, ce nouveau système devait introduire une plus grande souplesse dans le 'field-work' et garantir ainsi la réalisation d'un nombre suffisant d'interviews dans les délais requis.

L'enquête était confiée à une société de service privée disposant d'un 'Call Center' à laquelle le STATEC devait fournir les numéros de téléphone des ménages choisis. Comme auparavant, ces ménages étaient tirés dans un registre central de population, le Répertoire Général des Personnes Physiques géré par le Centre Informatique de l'Etat. Au STATEC de trouver les numéros de téléphone afférents. Cette tâche s'est révélée plus ardue et fastidieuse que prévu et pour quelque 40 % des ménages il n'a pas été possible de trouver ce numéro. Une alternative serait de tirer au hasard non plus des ménages avec nom et adresse mais des numéros de téléphone générés de façon aléatoire par un logiciel idoine. C'est cette alternative qui est examinée dans le papier du Professeur SCHNELL qui, par la même occasion, analyse la façon de procéder actuelle du 'Call Center'. Sa conclusion est qu'il est tout à fait possible de passer au RDD (Random Digit Dialing). Concernant les interviews menées jusqu'à présent, un certain nombre de déficiences sont relevées et des propositions d'amélioration sont présentées.

# Die Umstellung des LFS auf RDD-CATI

## 1. Gegenstand des Gutachtens

Gegenstand des Gutachtens ist die Prüfung der Möglichkeit der Durchführung des nationalen "Labour Force Surveys" (LFS) in Luxemburg als RDD-CATI-Erhebung.

## 2. Rechtliche Grundlagen des LFS in Luxemburg

Die maßgebliche Grundlage stellt die "Council Regulation No 577/98" der EU aus dem Jahre 1998 dar. Dort wird für alle Mitgliedstaaten die Durchführung des "Labour Force Surveys" geregelt. Die Vorgabe regelt folgende Punkte:

**Zeitlicher Abstand.** Der LFS soll kontinuierlich über das Jahr hinweg durchgeführt werden. Eine Auswertung soll vierteljährlich und jährlich erfolgen. Die Länder, die keinen kontinuierlichen Survey durchführen können, sollen eine jährliche Auswertung auf der Basis eines im Frühling stattfindenden Surveys erstellen.

**Grundgesamtheit.** Privatpersonen in Haushalten (L77/4).

**Erhebungsmethode.** Arbeitsstatus und Arbeitslosigkeit sollen über ein Interview mit der betroffenen Person bzw. als über ein Proxyinterview erhoben werden. Andere Variablen dürfen auch administrativen Quellen entnommen werden.

**Größe der Standardfehler.** Bei Staaten mit mehr als einer Million Einwohner soll für kontinuierliche Surveys der Standardfehler für die Schätzung der Quartalsänderung auf nationaler Ebene nicht größer als 2% der interessierenden Teilpopulation sein, wenn diese Teilpopulation 5% der Arbeitsbevölkerung ausmacht.

**Nonresponse-Verfahren.** Item-Nonresponse soll durch Imputationsverfahren korrigiert werden.

**Gewichtungsvariablen.** Neben Designgewichten werden als Gewichtungsvariablen Geschlecht, Alter (in 5-Jahres-Intervallen) und Ort (auf NUTS II Ebene) erwähnt.

**Erhebungsprogramm.** Demographische Angaben, Arbeitsstatus, Angaben zu Haupt- und Nebenberufstätigkeiten, Angaben zur Arbeitslosigkeit, Angaben zur Bildung und Angaben zum Einkommen (optional).

**Zeitpunkt der Datenlieferung.** Bei einem kontinuierlichen Survey sollten die Daten 12 Wochen nach dem Ende der Referenzperiode an die EU geliefert werden. Bei einem einmaligen jährlichen Survey beträgt die Zeit 9 Monate nach Ende der Referenzperiode.

Weiterhin sieht die "Council Regulation" vor, dass die Kommission dem Parlament und dem Rat ab dem Jahr 2000 alle drei Jahre einen Bericht über die Implementierung des "Labour Force Surveys" in allen Mitgliedstaaten vorlegen soll. Der Bericht soll Angaben über die Qualität des Surveys in jedem einzelnen Mitgliedsstaat machen. Entsprechend muss regelmäßig ein Bericht über die Qualität des LFS erstellt werden.

## 3. Beschreibung der bisherigen Praxis des LFS in Luxemburg

### 3.1 Survey Design

Der LFS wird seit 2003 als kontinuierlicher Survey über das Kalenderjahr hinweg durchgeführt. Die Stichprobe wird hierbei auf 52 Referenzwochen aufgeteilt. Die Interviews erfolgen drei Wochen lang nach der jeweiligen Referenzwoche. Die Ergebnisse werden aggregiert und als jährliches Ergebnis an die EU gemeldet. Die jeweiligen Statistiken werden nur auf nationaler Ebene ausgewiesen, da Luxemburg zu wenig Einwohner für eine kleinräumigere Ausweisung besitzt. Ein Viertel der Befragten im nächsten Kalenderjahr wieder befragt. Die Teilnahme am Survey ist freiwillig.

#### 3.1.1 Grundgesamtheit, Auswahlgesamtheit, Survey-Population

Die angestrebte Grundgesamtheit sind die Angehörigen privater Haushalte im Inland. Als "sampling unit" werden Haushalte verwendet. Die Ziehungsbasis stellt das RGPP ("Répertoire Général des Personnes Physiques") dar. In diesem zentralen

Register befinden sich die Angaben der Haushaltsvorstände privater Haushalte in Luxemburg. Aus diesem Register wird eine geschichtete Stichprobe gezogen. Für die gezogenen Haushalte werden in den Telefon-CDs Einträge recherchiert. Dabei konnten 2003 32% und 2004 40% der Haushalte nicht ermittelt werden. Die ermittelbaren Haushalte bilden die Survey-Population.

### 3.1.2 Schichtung

Bei der Stichprobe handelt es sich um eine proportional geschichtete Zufallsstichprobe. Geschichtet wird nach der regionalen Einteilung in Cantone sowie nach der Haushaltsgröße. Luxemburg besteht aus 13 Cantonen und die Haushaltsgröße wird in vier Klassen unterteilt 1, 2, 3, 4+. Damit ergeben sich 52 Schichten.

Die Schichten werden proportional zu ihren Anteilen in der Population besetzt, mit einem Minimum von 48 realisierten Haushaltsinterviews pro Schicht und Jahr. Da der Anteil Nonresponse ca. 60% der ausgewählten Haushalte beträgt, wurde die Minimumanzahl der Haushalte in der Bruttostichprobe für die kleinste Schicht auf 96 festgelegt.

### 3.1.3 Stichprobenverfahren

Für die ausgewählten Haushalte werden die Telefonnummern im öffentlichen Telefonbuch recherchiert. Ca. 40% der Bruttostichprobe können nicht gefunden werden. Haushaltsvorstände, deren Nummer nicht recherchiert werden kann, werden postalisch um ihre Telefonnummer zur Teilnahme an dem Survey gebeten. Ca. 9% der Befragten des LFS werden auf diese Weise rekrutiert. Haushalte, für die keine Telefonnummer recherchiert werden kann, gelten als Nonresponse.

### 3.1.4 Datenerhebung im Haushalt

Innerhalb eines gezogenen Haushaltes werden die demographischen Variablen für alle Haushaltsmitglieder erhoben. Die anderen Variablen des LFS werden nur für Haushaltsmitglieder im Alter zwischen 15 und 74 Jahren erhoben. Proxy-Interviews sind zulässig.

### 3.1.5 Datenerhebungsmodus

Die Interviews werden alle telefonisch als CATI durchgeführt. Die Interviewdauer liegt pro Person bei 10-15 Minuten.

## 3.2 Ergebnisse der Feldarbeit

Die folgenden Angaben beziehen sich auf den LFS aus dem Jahre 2004.<sup>1</sup>

		Prozent
<b>Bruttostichprobe</b>	<b>23 463</b>	<b>100.00</b>
Refusals		14.99
Non-contacts		11.13
Phones not found		40.49
<b>Nonresponse gesamt</b>		<b>66.61</b>
<b>Verbleibende Interviews</b>	<b>7 834</b>	<b>33.39</b>

Es lässt sich mit den uns vorliegenden Daten und Dokumentationen nicht eindeutig klären, für exakt wieviele Personen tatsächlich Daten erhoben wurden.<sup>2</sup>

### 3.2.1 Feldarbeit-Details

Die Daten der Kontaktprotokolle des LFS 2004 wurden einer umfangreichen Analyse unterzogen. Auffällig ist zunächst die besonders hohe Workload der Interviewer (Tabelle 1). Bemerkenswert ist die Tatsache, dass 42% der Interviews des LFS von ein und demselben Interviewer erhoben werden.

Diese hohe Zahl von Kontaktversuche pro Interviewer besitzt zunächst eine Reihe von Effekten für die Kontaktversuche selbst:

- 25% der Kontaktversuche werden nach 17:00 Uhr durchgeführt
- Lediglich 0,2% der Kontaktversuche werden vor 8:00 unternommen
- Bei erstmaliger Verweigerung wird kein weiterer Kontaktversuch mehr unternommen
- Zwischen 12:00–12:30 werden deutlich weniger Kontaktversuche durchgeführt als zu erwarten wäre.

<sup>1</sup> Im LFS 2003 wurde die Stichprobe nur zu einem Teil aus dem RGPP gezogen. Ein anderer Teil stammte aus dem Zensusdatensatz des Jahres 2001. Im Folgenden wird daher ausschließlich auf der Basis der Daten 2004 argumentiert.

<sup>2</sup> Das Problem der Dokumentation lässt sich an den Fallzahlen demonstrieren: Laut "Standard Quality Report 2004" (2004:17) wurden von den 7834 Haushaltsinterviews 7771 realisiert; die Ursache der Differenz ist unklar. In Berichten des CEPS ist sowohl für das Jahr 2004 als auch für 2003 von 8532 Haushalten in der Stichprobe die Rede (CEPS Rapport intermédiaire 2004: 1 und Tirage de l'Échantillon EFT 2003:1).

**Tabelle 1: Workload im LFS 2004**

Interviewer	Freq.	Percent	Cum.
1	3 312	42.22	42.22
2	2 523	32.16	74.38
3	1 143	14.57	88.95
4	315	4.02	92.96
5	246	3.14	96.10
6	257	3.28	99.38
7	49	0.62	100.00
<b>Total</b>	<b>7 845</b>	<b>100.00</b>	

Diese Zahlen deuten auf erhebliche Probleme im Call-Management und in der Allokation der Interviewer hin. Entsprechend den Einschränkungen in Art und Verteilung der Kontaktversuche wird das Ausmass und die Art des möglichen Nonresponse-Errors verständlich.

### 3.2.2 Nonresponse-Error

Die Nonresponse-Error Rate (ohne nicht ermittelte Haushalte) liegt bei 26.12%. 43% der Ausfälle sind dabei nicht-erreichte Haushalte, 57% entsprechend Verweigerungen. Bedeutsam werden diese Anteile bei der Betrachtung der einzelnen Ausfallsursachen (den Dispositionscodes in den Kontaktprotokollen).

Auffällig bei der Analyse der einzelnen Ausfallursachen sind zunächst die Ausfälle aufgrund von Sprachproblemen. Von den ursprünglich 1 274 kontaktierten Haushalten mit Sprachproblemen bleiben bei 795 letztlich ohne Erfolg (62.4%), das entspricht ca. 6% aller kontaktierten Telefonnummern. Sollten diese Zahlen korrekt sein, dann ist es nicht unwahrscheinlich, dass hinter diesen unbearbeiteten Nummern ein inhaltliches Problem steht.

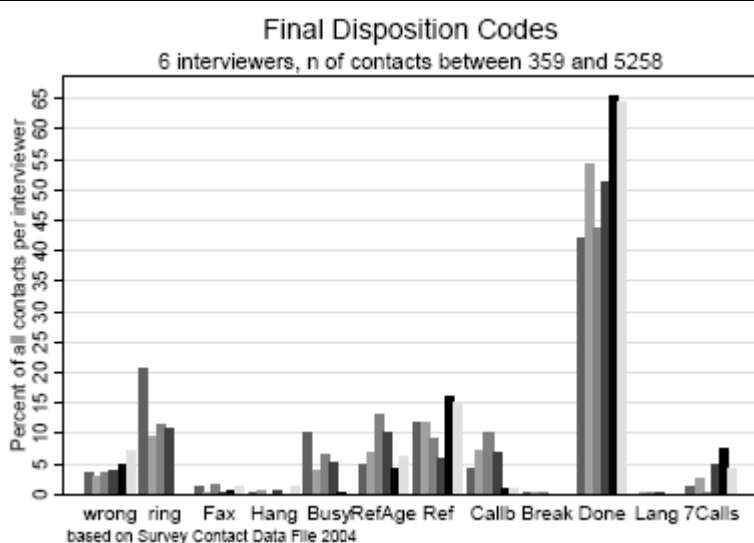
Sollte tatsächlich keine Kommunikation mit dem Haushalt auf luxemburgisch, französisch, portugiesisch oder deutsch möglich sein, dann wird auch eine entlohnte Beschäftigung eher schwierig. Dann wäre eine Ergänzung des Interviewerstabes um italienische und englischsprachige Interviewer sinnvoll. Allerdings entfielen sowohl 2003 als auch 2004 die meisten dieser Fälle auf Haushalte, in denen die Referenzperson aus Deutschland, Portugal und Luxemburg stammte. Dies entspricht eher für ein Problem der Wiederkontaktierung im Haushalt und/oder des Call-Managements.

Von besonderer Bedeutung bei einer Nonresponseanalyse sind natürlich die "final disposition codes" aller eingesetzten Telefonnummern 2004 (vgl. Tabelle 2).

**Tabelle 2: "Final disposition codes" aller Telefonnummern 2004**

wrong/invalid number	1	463	3.58
no answer	2	1 519	11.73
fax/modem	3	59	0.46
hang up	4	70	0.54
busy/AM	5	689	5.32
non interested (age)	6	1 013	7.83
not interested (not justified)	7	1 244	9.61
call back	8	815	6.30
interrupted	9	10	0.08
completed	10	6 593	50.93
language selection	11	37	0.29
7 trials no success	12	433	3.34
<b>Total</b>		<b>12 945</b>	<b>100.00</b>

**Abbildung 1: "final disposition codes" aller Telefonnummern pro Interviewer 2004**



Hier fällt für zahlreiche Kategorien ein deutlicher Unterschied zwischen den Interviewern auf (vgl. Abbildung 1). So spricht der Unterschied in den Anteilen für eine falsche Nummer, keinen Kontakt, besetzte Anschlüsse, Callbacks und in den Nicht-Erreichten entweder für ein Problem der Interviewer-Allokation zu den Fällen, des Nicht- oder Falscheinsatzes eines Call-Management-Systems<sup>1</sup> oder unzureichendes Interviewer-Training bzw. Supervision. Die Verweigerungsraten allgemein pro Interviewer variieren zwischen 5.9%-16.2%, die Verweigerung durch Alter zwischen 4.4% und 10.1%. Betrachtet man den Gesamtanteil der Verweigerer (Codes 6+7), so schwankt dieser zwischen 16.5% und 22.5%.<sup>2</sup> Dies zeigt systematische Unterschiede in den Verweigerungsraten zwischen den Interviewern. Insbesondere der Code 6 (Alter) ist bei einem LFS nicht akzeptabel: Entweder gehört die Zielperson zur Population oder nicht, und ein hohes Alter führt dann zum Ausschluss aus der Befragung; ansonsten wäre eine Konvertierung erforderlich.

### 3.2.3 Andere Nonsampling-Errors

Die demographischen Variablen werden für alle Haushaltsmitglieder erfasst; alle anderen Variablen nur für die 15-74-jährigen. Befragt wird eine

kompetente Auskunftsperson im Haushalt; die meisten Angaben werden als Proxy-Angabe erhoben. In Hinsicht auf den LF-Status ergibt sich empirisch ein Anteil von Proxy-Antworten von 61.2% (2004).<sup>3</sup> In Europa liegt der Anteil der Proxy-Antworten im LFS in Luxemburg über den Werten vergleichbarer Länder (2.9% in Schweden, BRD 30%, Niederlande 46%). Angesichts der vergleichsweise niedrigen Qualität von Proxy-Angaben sowie der vermutlich stärkeren Interviewer-Effekte bei Proxy-Interviews sollten besondere Anstrengungen unternommen werden, um die Proxy-Rate zu senken. Dies dürfte durch ein verbessertes Call-Management erreichbar sein.

### 3.2.4 Sampling-Error

Insgesamt werden pro Jahr ca. 8 500 Haushalte mit insgesamt ca. 20 400 Personen befragt. Entsprechend klein sollten die Standardfehler und die resultierenden Variationskoeffizienten sein. So gibt STATECs Quality-Report für 2004 für die Zahl der Arbeitslosen einen Variationskoeffizienten von 5.04 an, für die Arbeitslosenrate 4.93.<sup>4</sup> Berechnet man diesen Wert aus den Daten unter Annahme einer einfachen Zufallsstichprobe, erhält man einen sehr ähnlichen Wert:

```
.cii 8851 410,exact
```

Variable	Obs	Mean	Std. Err.	(95% Conf. Interval)	Binomial Exact
	8 851	.0463224	.0022341	.0420372	.0509094

```
. di 0.002341/0.463224
.00505371
```

<sup>1</sup> Ein Interviewer weist 42 nicht erledigte Callbacks bei 413 Kontakten auf, selbst im Mittel über die Interviewer werden 6.2% der Callbacks nicht erledigt.

<sup>2</sup> Das 95%-Konfidenzintervall im letzten Fall reicht von 18.6-26.9%.

<sup>3</sup> 2003 lag der Anteil bei 38.3%. Eine plausible Erklärung für den Unterschied konnten wir nicht finden.

<sup>4</sup> Der Quality-Report weist ausdrücklich auf die Notwendigkeit der Berücksichtigung der Design-Effekte hin. Es bleibt aber unklar, mit welcher Methode und anhand welche Cluster-bestimmenden Variablen Design-Effekte berechnet wurden.

Berücksichtigt man hingegen die Klumpeneffekte durch die Interviewer, die Schichtung und die finite Populationskorrektur<sup>1</sup>, werden die Standardfehler etwas größer:

. svy: prop rsalu (running proportion on estimation sample)

Survey: Proportion estimation

Number of strata = 52      Number of obs = 8804  
 Number of PSUs = 300    Population size = 194970  
 Design df = 248

	Proportion	Linearized Std. Err	Binomial Wald (95% Conf. Interval)	
rsalu				
0	.9521394	.002863	.9465004	.9577784
1	.0478606	.002863	.0422216	.0534996

. di 0.002863/0.0478606  
 .05981956

Damit liegt der tatsächliche Variationskoeffizient vermutlich näher an 6% als an 5%, ist also 1.2 mal größer als bisher angenommen. Dies geht auf die Unterschätzung der Designeffekte zurück:

.estat effects, deft

	Proportion	Linearized Std. Err.	Deft
rsalu			
0	.9521394	.002863	1.25836
1	.0478606	.002863	1.25836

Die Design-Effekte liegen bei 1.26 für das Merkmal "Arbeitslosigkeit" und damit geringfügig über den bei komplexen Stichproben üblichen Werten des Design-

Effekts. Allerdings ist die Stichprobe eigentlich nur geschichtet, daher sollte der Design-Effekt eher niedriger liegen.<sup>2</sup> Häufig ist die Umrechnung der Variationskoeffizienten in Größe der Konfidenzintervalle der geschätzten Gesamtzahlen anschaulicher:

svy: total rsalu  
 (running total on estimation sample)

Survey: Total estimation

Number of strata = 52    Number of obs = 8804  
 Number of PSUs = 300    Population size = 194970  
 Design df= 248

	Total	Linearized Std. Err.	(95% Conf. Interval)	
rsalu	9331.405	1087.104	7190.272	11472.54

Entsprechend liegt für ein 95%-Konfidenzintervall die geschätzte Zahl der Arbeitslosen bei ca. 9300 ± 2100. Es wäre wünschenswert, die Größe dieses Intervalls zu verringern. Dies lässt sich durch eine Verringerung der Interviewer-Effekte und durch die Verwendung von Kalibrierungsschätzern vermutlich erreichen.

<sup>1</sup> In STATA-9.2 bedeutet dies: iweight: weight, Strata 1: stratum, SU 1: agent, FPC 1: pop

<sup>2</sup> Verwendet man die Haushalte als Klumpen, dann sind die geschätzten Design-Effekte tatsächlich niedriger. Dies ist aber inkorrekt.



## 4. Der LFS als RDD-CATI

Der folgende Teil des Gutachtens wird sich ausschließlich den eventuellen Problemen und Konsequenzen der Umstellung des LFS von einem Register-Telefon-Listen-CATI auf ein RDD-CATI widmen.

### 4.1 Probleme der Adaption des Erhebungsinstruments auf CATI-RDD

Da die Erhebung bislang bereits als CATI-Erhebung durchgeführt wird, werden keinerlei zusätzliche Probleme durch den Fragekatalog oder Erhebungsmodus erwartet.<sup>1</sup> Bei der tatsächlichen Durchführung als CATI-RDD sind zusätzliche Schulungen und eventuelle neue Hilfsmittel (Hilfsbildschirme im CATI-Programm) für die Interviewer zur Erklärung des Auswahlverfahrens von Befragten vor allem im Fall nicht eingetragener Nummern notwendig. Da RDD bei vielen Erhebungen den Standard darstellt, kann auf vorhandenes Material zurückgegriffen werden. Damit sind durch die Umstellung von Seiten des Instruments weder Probleme noch Kosten erwartbar.

### 4.2 Sampling-Frame-Konstruktion für RDD

Basis für einen Sampling-Frame eines RDD-Samples bildet entweder eine bekannte Systematik der Vergabe der Telefonnummern über aktivierte Blöcke, in denen dann Zufallszahlen anstelle der Endziffern generiert werden oder eine vollständige Liste der eingetragenen Nummern, bei denen dann die Endziffern durch Zufallszahlen ersetzt werden. In den meisten Ländern stehen keine Angaben über eine Systematik zur Verfügung, so dass in der Praxis häufig Varianten der Verwendung bereits bekannter Nummern zum Einsatz kommen. Dies würde auch für RDD-Verfahren in Luxemburg gelten.

#### 4.2.1 Verfügbarkeit der Telefonnummern

Als Ausgangsdatenbasis kommt die Liste der eingetragenen Telefonnummern in Frage. Diese sind

über eine CD-ROM des Herstellers "Data editus" verfügbar.

#### 4.2.2 Aufbereitung des Sampling-Frames

Die Datei der Ausgabe 2006 der Telefon-CD enthält 181.277 Textzeilen, von denen aber nur 162.710 Telefonnummern enthalten. Bereinigt man diese Liste um Mehrfacheinträge ergeben sich 140.920 Nummern.

#### 4.2.3 Abschätzung der Zahl der privaten Anschlüsse

Durch umfangreiche Listen von Schlüsselbegriffen sowie technische Merkmale der Einträge ("embedded dash", "multiple dots", Fax-Eintrag) wurde für die bereinigte, aber nicht deduplizierte CD ein Anteil von ca. 20% nicht-privater Anschlüsse geschätzt. Dieser liegt erheblich über dem Anteil in anderen europäischen Ländern. Nach Deduplizieren liegt der Anteil bei ca. 16%. Eine detaillierte Untersuchung zeigt, dass ein großer Teil der Differenz auf mehrfache Einträge der gleichen Nummer, einerseits mit kommerziellen Merkmalen, andererseits ohne kommerzielle Merkmale zurückzuführen ist. Der Anteil ausschließlich privater Nummern liegt bei 85% aller deduplizierten Nummern. Rein kommerzielle Nummern sind 13% der Nummern, ca. 1% der Nummern sind exakt einmal kommerziell, einmal privat eingetragen. Um dieses Ergebnis weiter zu erhärten, haben wir die "Gelben Seiten" ebenfalls exportiert, dedupliziert und mit den gereinigten "Weißen Seiten" gematched. Wertet man alle gefundenen Matches als Geschäftsanschluss, ergibt sich folgendes Bild:

Company	Freq.	Percent	Cum.
Privat	119,529	84.82	84.82
Business	21,392	15.18	100.00
Total	140,921	100.00	

Nach dieser Methode können insgesamt ca. 15% der Telefonnummern in den "Weißen Seiten" als Geschäftsanschlüsse identifiziert werden. Dies sollte die Obergrenze der Geschäftsanschlüssen darstellen.

<sup>1</sup> Selbst bei einem Modus-Wechsel sind die Effekte in einem LFS eher klein, vgl. Wilson/Blackshaw/Norris (1988).

Bei einem RDD in Luxemburg wird daher der Anteil der Kontakte, bei denen reine Geschäftsanschlüsse gewählt werden, maximal bei 15% liegen. Dieser Anteil liegt höher als in anderen europäischen Ländern und wird daher auch (minimal) höhere Kosten verursachen. Dies ließe sich aber über einen vorherigen Abgleich der generierten RLD-Nummern mit den "Gelben Seiten" verhindern.<sup>1</sup>

#### 4.2.4 Abschätzung der Blockdichten

Bei RDD-Verfahren werden Blöcke von Telefonnummern als Ausgangsbasis verwendet, wobei die Länge der Blöcke zwischen den Ländern je nach

den technischen Erfordernissen eines Landes variiert. So werden im Vereinigten Königreich Blöcke von 10000 Nummern, in der Bundesrepublik üblicherweise Blöcke mit jeweils 100 potentiellen Nummern verwendet (vgl. Gabler/Häder (1997), Nicolaas/Lynn 2002). Entscheidend für die Länge der Blöcke ist hierbei die Verteilung der Häufigkeiten der Nummern innerhalb der Blöcke, die sogenannte "Blockdichte". Anhand der bereinigten und deduplizierten Ausgangsliste wurden die Blockdichten für Luxemburg 2006 bestimmt. Die Abbildungen 2 bis 4 zeigen die Verteilung der Zahl der eingetragenen Nummern für alle 100er-Blöcke für alle Anschlüsse, für die privaten Anschlüsse und für die kommerziellen Anschlüsse.

Abbildung 2: Zahl der Anschlüsse pro Block

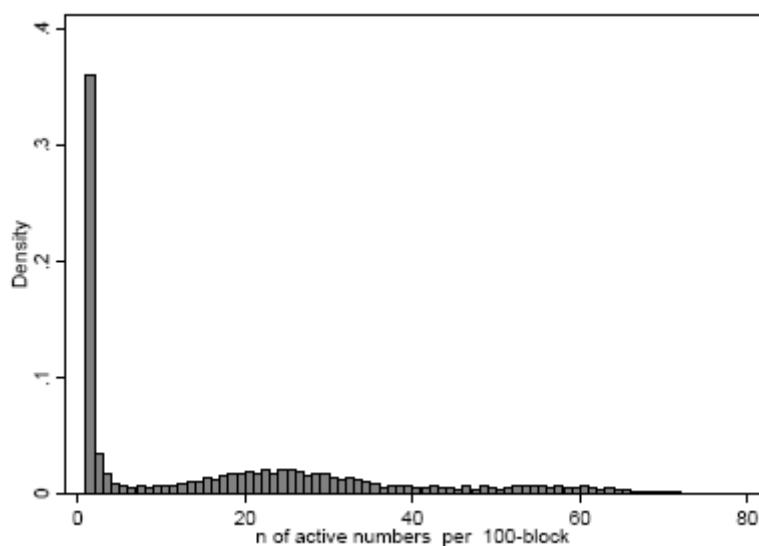
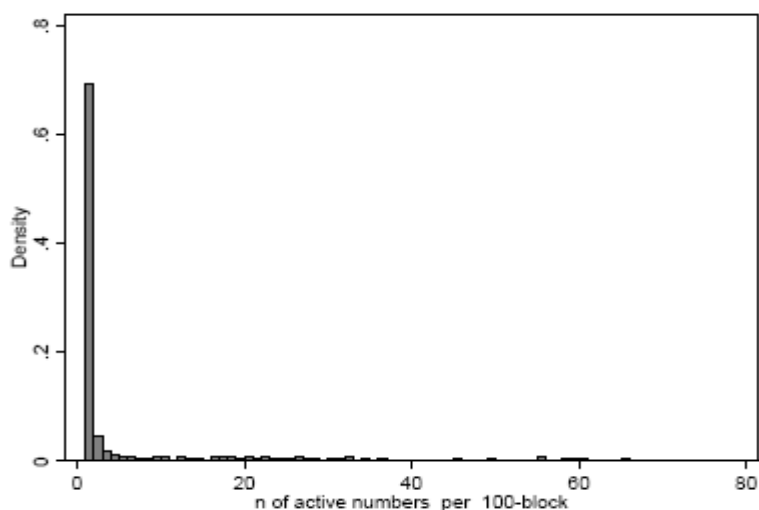
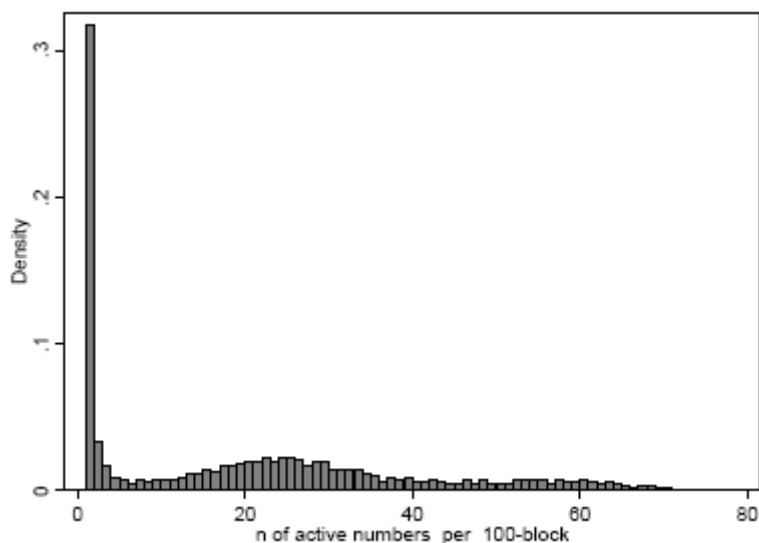


Abbildung 3: Zahl der Privatanschlüsse pro Block



<sup>1</sup> Durch diesen Abgleich wird die Bestimmung der Gewichte deutlich komplizierter. Allerdings wird der Unterschied zwischen korrekten und naiven Gewichten vermutlich vernachlässigbar klein sein.

Abbildung 4: Zahl der kommerziellen Anschlüsse pro Block



## Zusammenfassung 100er Blocks

- Mehr als 60% aller 100er-Blocks enthalten nur eine Nummer.
- Im Gegensatz dazu liegt die Blockdichte für die kommerziellen Anschlüsse sehr deutlich höher.
- Für eine Unternehmensbefragung würde sich für Luxemburg daher die Verwendung von 100er-Blöcken anbieten, wobei die Blöcke als Auswahlinheit verwendet werden könnten.
- Die Verwendung von 100er-Blöcken für die Befragung privater Haushalte würde zu

häufig zu unbesetzten Nummern führen; dies wäre ökonomisch nicht tragbar.

- Daher scheint die Verwendung von 10er-Blocks zur Auswahl der Haushalte geeignet.

Wie aus den Abbildungen 5 bis 7 ersichtlich, zeigen die Verteilungen zwischen kommerziellen und privaten Anschlüssen auch auf der Ebene der 10er-Blocks deutliche Differenzen: Die Konzentration privater Anschlüsse pro Block ist höher. Trotzdem sollte auf die Verwendung dieser Eigenschaften für eine Klumpenstichprobe verzichtet werden, da dies zu einer Vergrößerung der Design-Effekte führen würde.

Abbildung 5: Zahl der Anschlüsse pro 10er-Block

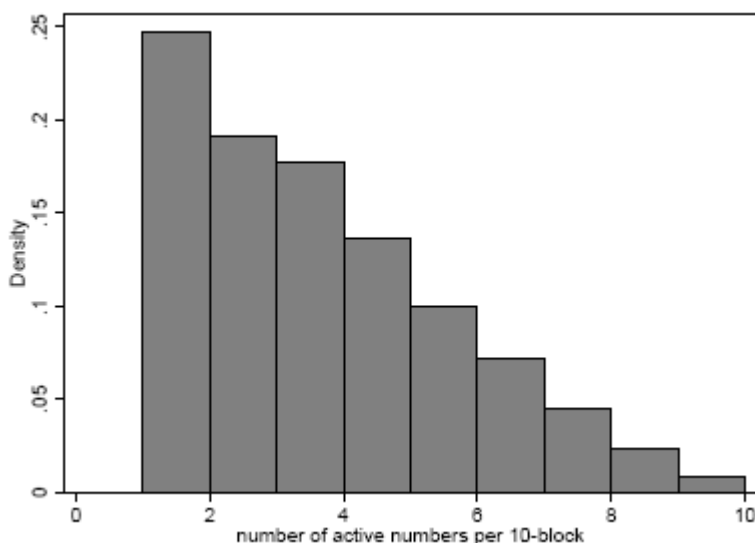


Abbildung 6: Zahl der Privatanschlüsse pro 10er-Block

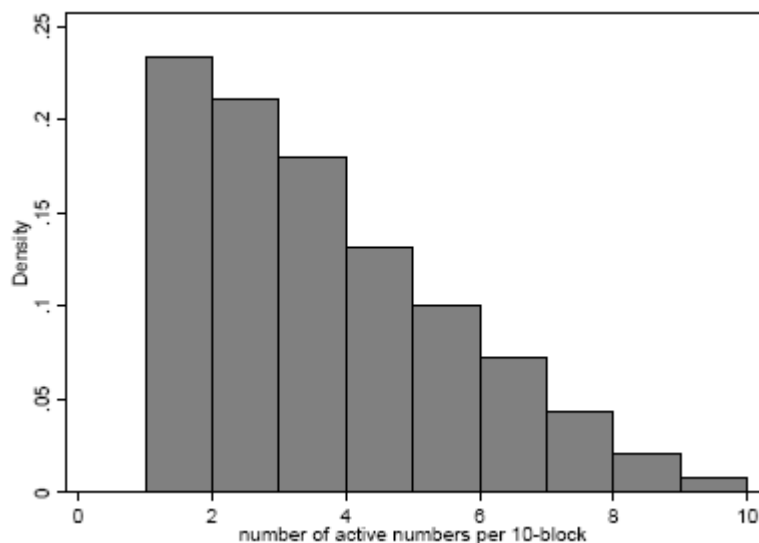
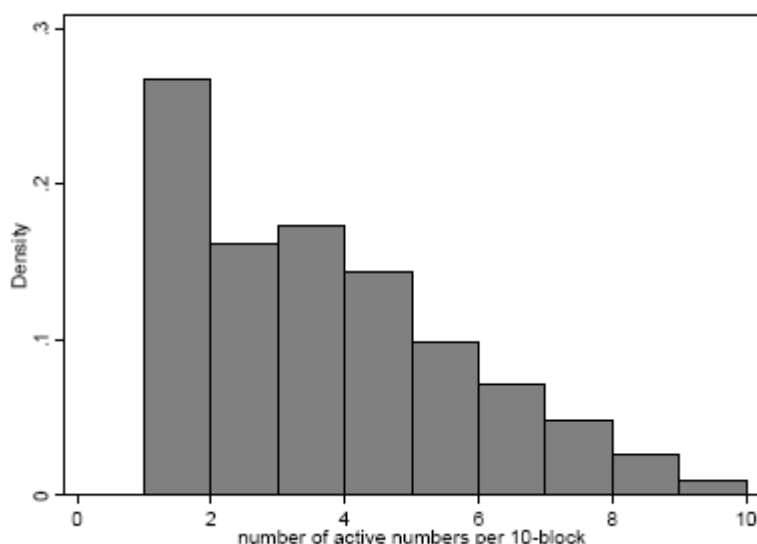


Abbildung 7: Zahl der kommerziellen Anschlüsse pro 10er-Block

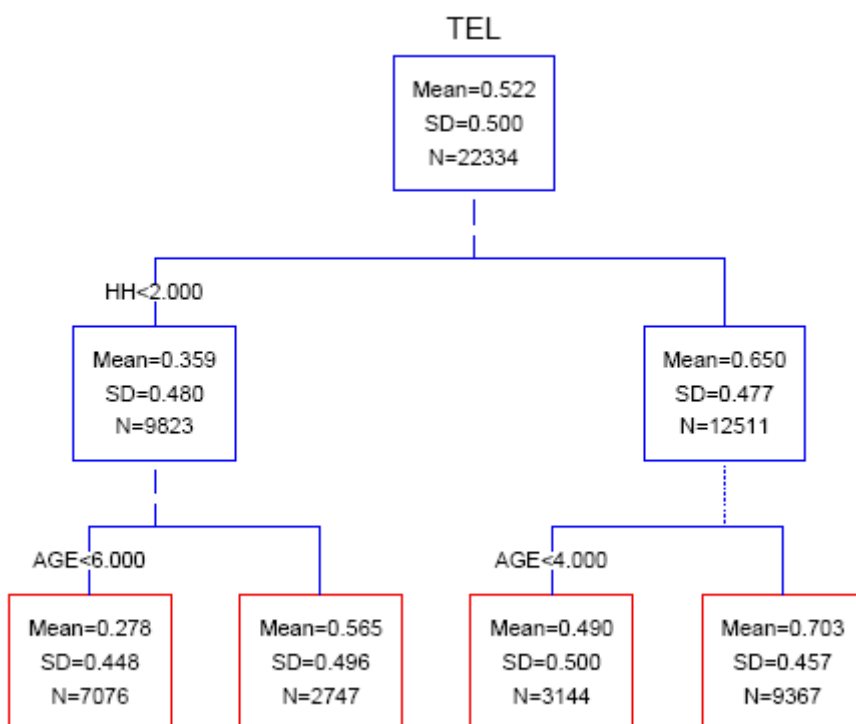


Als Design einer RDD-Stichprobe für Privathaushalte liegt daher die Verwendung einer RLD-Stichprobe nahe: Regional geschichtete Auswahl der 10er Blöcke, Bestimmung einer Zahl pro Block. Die Berechnung der Gewichte ist vergleichsweise arbeitsaufwändig, aber unproblematisch (vgl. Gabler/Häder 1997).

#### 4.2.5 Abschätzung des Effekts nicht eingetragener privater Anschlüsse

Die bisherigen Datenerhebungen bieten anhand des Vergleichs der realisierten Stichprobe mit den Registerdaten der Brutto-Stichprobe die Möglichkeit der Analyse der nicht-eingetragenen Nummern auf Haushaltsebene.

Abbildung 8: CART für eingetragene Telefonnummern 2004



Der mit Systat 11 berechnete CART (Abbildung 8) zeigt zunächst, dass Einpersonenhaushalte eine geringere Eintragungswahrscheinlichkeit gegenüber größeren Haushalten besitzen; innerhalb der Gruppe der Einpersonenhaushalte sind vor allem die jüngeren Personen nicht eingetragen.<sup>1</sup> Weitere Analysen zeigen, dass über alle Haushaltstypen hinweg sich festhalten lässt, dass die Eintragungswahrscheinlichkeit mit steigendem Alter abnimmt und ab ca. 70 Jahren wieder zunimmt. Weiterhin besitzen Haushalte mit einem ausländischen Haushaltsvorstand eine geringere Eintragungswahrscheinlichkeit gegenüber Haushalten mit einem inländischen Haushaltsvorstand. Für Personen innerhalb der Zielpopulation des LFS wird daher eine geringe Überschätzung der Arbeitslosigkeit aufgrund der nicht eingetragenen Nummern vermutet; dies wäre aber im Detail anhand der uns nicht vorliegenden

Registerdaten noch zu prüfen. Sollte hier ein Effekt nachweisbar sein, dann könnte dieser durch das RDD weitgehend beseitigt werden.

### 4.3 Abschätzung des Coverage-Errors

Da über das RDD im Gegensatz zum bisherigen Verfahren die Wahrscheinlichkeit, dass ein nicht-ingetragener Telefonanschluss doch in die Stichprobe gelangt, nahezu identisch mit der eingetragener Nummern ist,<sup>2</sup> sollte der Coverage-Error durch RDD deutlich kleiner werden. Der verbleibende Coverage-Error bestimmt sich nahezu ausschließlich über die Anzahl der luxemburgischen Privathaushalte, die nicht über einen Festnetzanschluss verfügen. Verschiedene Quellen geben unterschiedliche Telefondichten für Luxemburg<sup>3</sup> an (vgl. Tabelle 3)

<sup>1</sup> Diese Struktur lässt sich auch durch die Analyse der Gruppe der nicht eingetragenen Personen, die ihre Telefonnummer auf Anfrage zur Verfügung stellen, zeigen. Hier handelt es sich um Personen, die eher in Städten als auf dem Land zu finden sind, um Singlehaushalte und um tendenziell jüngere Personen.

<sup>2</sup> Die Ausnahme besteht aus Blocks, die keine einzige eingetragene Nummer besitzen; die Zahl solcher Blocks dürfte vernachlässigbar sein.

<sup>3</sup> PSELL ist die Studie "Panel Socioeconomique Liewen in Luxembourg".

**Tabelle 3: Telefondichte in % der Haushalte**

Jahr	%	Datenquelle
1999	99.0	Survey on telecommunication by the EU Commission
1999	94.8	PSELL
2000	93.6	PSELL

Derzeit dürfte die maximale Noncoverage-Rate bei ca. 5% liegen. Im Allgemeinen werden die Probleme aus dem Undercoverage für CATI-Surveys als gering betrachtet (vgl. Brick/Waksberg/Kulp/Starer(1995)). Zwar werden hier vermutlich an beiden Enden der Schichtungs-Skala Effekte in Hinsicht auf die interessierende Variable auftreten, die Effekte sind allerdings zeitlich konstant. Vor allem werden in dieser Hinsicht keine Unterschiede in Hinsicht auf den bisherigen Survey auftreten.

Problematisch werden langfristig vermutlich nur zwei Aspekte: Möglicherweise wird zukünftig durch die Auflösung der Bindung der Telefonnummern an eine Adresse eine (geringfügig fehlerhafte) Erhebung des Wohnortes durch die Befragung notwendig; gegebenenfalls sind dann regionale Schichtungen etwas problematischer als bisher. Das größte Problem dürften aber Haushalte darstellen, die ausschließlich über mobile Telefone erreichbar sind (vgl. Grund 2002, Kim/Lepkowski 2002). Derzeit handelt es sich dabei tendenziell um mobile, jüngere meist männliche Personen mit eher geringem Einkommen, aber diese Merkmale werden nicht lange stabil bleiben. Solange von den Telekommunikationsbetreibern keine Listen aktiver Blöcke zur Verfügung gestellt werden oder die (derzeit) beträchtlichen Kosten der -- technisch trivialen -- Absuchung des gesamten möglichen Nummernraumes akzeptiert werden, wird die Konstruktion eines geeigneten Sampling-Frames nahezu unmöglich.<sup>1</sup> Auch dieses Problem betrifft nicht das RDD allein, sondern jede CATI-Studie.<sup>2</sup> Sollte dieser Anteil wachsen und keine Listen der Nummern zur Verfügung stehen, dann wird jede Art der Stichprobenziehung für telefonische Befragungen schwer zu rechtfertigen.<sup>3</sup> Langfristig wäre dann eher ein Methoden-Mix aus telefonischer Befragung wie bisher und schriftlicher bzw. persönlicher Befragung geeigneter. Für einen absehbaren Zeitraum scheint dies aber noch nicht notwendig zu werden.

<sup>1</sup> Erwähnt werden muss in diesem Zusammenhang die Situation in der Schweiz: Das Statistische Zentralamt der Schweiz besitzt eine vollständige Datei aller Telefonnummern in der Schweiz, einschließlich nicht-eingetragener Nummern und Mobilnummern. Für jeden Anschluß liegt Name und Anschrift des Inhabers vor. Diese Datei wird nicht nur für CATI-Surveys sondern auch für Face-to-Face-Befragungen als Sampling-Frame verwendet. Möglich wird dies durch eine gesetzliche Auflage für die Betreiber von Telekommunikationssystemen. Es wäre für statistische Zwecke wünschenswert, wenn die EU oder ihre Mitgliedstaaten sich diese Politik zu eigen machen würden.

<sup>2</sup> Auch im bisherigen komplizierten Listenverfahren sind kaum Haushalte ohne Festnetznummer vorhanden. Für die im Kontaktprotokolldatensatz enthaltenen Telefonnummern 2004 gibt nur es in 4% der Fälle eine eingetragene Handynummer. Jedoch ist auch in diesen Fällen häufig zusätzlich eine Festnetznummer mit angegeben.

<sup>3</sup> Entsprechend dringend erforderlich ist die Aufnahme einer solchen Frage in eine Face-to-Face-Befragung, die aus dem Populationsregister gezogen wurde.

<sup>4</sup> Der Gewinn durch die Endlichkeitskorrektur ist vernachlässigbar gering.

#### 4.4 Abschätzung des Sampling-Errors

Der Sampling-Error wird durch das vorgeschlagene RDD-Design kaum beeinträchtigt: Gegenüber einer einfachen Zufallsstichprobe kommen Gewinne durch die Schichtung zustande, die durch die minimale Varianzerhöhung durch die Gewichte nicht verschwinden.<sup>4</sup> Das größte Problem der faktischen Standardfehler in jedem CATI-Survey sind die Klumpeneffekte durch die Interviewer (vgl. Schnell/Kreuter 2005). Diese lassen sich nur durch eine drastische Begrenzung der Fallzahl pro Interviewer verringern. Einzelheiten werden im Abschnitt 4.10.3 erläutert.

#### 4.5 Nonresponse

Nonresponse-Effekte entstehen durch die Größe der Differenz zwischen Respondenten und Nonrespondenten, wobei diese Differenz mit dem Ausmaß des Nonresponse multipliziert werden muss. Entsprechend sind sowohl sehr starke Nonresponse-Effekte bei kleinem Nonresponse-Ausmaß als auch sehr kleine Nonresponse-Effekte bei großem Ausmaß von Nonresponse möglich. Typischerweise ist eine Nonresponse-Population kaum homogen, sondern besteht aus zahlreichen Teilmengen mit unterschiedlichen Motivationen. Erschwerend kommt hinzu, dass je nach Population, Thema, Erhebungsmodus, Sponsor und zahlreichen erhebungstechnischen Details die Zusammensetzung der verschiedenen Anteile variieren kann, so dass eine Beurteilung ohne Kenntnis der jeweiligen Details kaum möglich ist (vgl. Schnell/Kreuter 2000). Trotzdem kann man einige Subgruppen unterscheiden, deren getrennte Betrachtung notwendig ist, vgl. Schnell (1997).

#### 4.5.1 Ausmaß bei unterschiedlichen Feldprozeduren

Das Ausmaß des Nonresponse läßt sich durch eine Reihe von Survey-Design-Variablen steuern. Für ein CATI-Survey ist dies zunächst die Zahl der Callbacks, die Frage der Ankündigung des Surveys, die Behandlung von Verweigern sowie die Möglichkeit der Verwendung von Incentives.

#### 4.5.2 Non-contacts und Callbacks

Durch das RDD allein entstehen keine Veränderungen in der Zahl der Non-contacts.<sup>1</sup> Die Rate der Nicht-Erreichten liegt derzeit bei insgesamt 15% aller kontaktierten Nummern 2004 bzw. 11.1% der ausgewählten Haushalte. Leider besteht die Vermutung, dass die Erreichbarkeit mit Erwerbstätigkeit, vor allem bei Einpersonenhaushalten, kovariiert. Daher sollte bei einer Umstellung der Stichprobe nicht nur das Call-Management verbessert werden, sondern auch die Zahl der Kontakte wesentlich über die bisherigen sieben Versuche hinaus ausgedehnt werden (bei einigen Surveys in der BRD sind bis zu 30 Versuche üblich). Bei einer zudem gleichmäßigeren Streuung der Kontaktversuche über die Zeit sollte der Anteil der Nichterreichten deutlich auf unter 2% gesenkt werden können.

#### 4.5.3 Anschreiben

Insbesondere für wissenschaftliche Surveys und Erhebungen der amtlichen Statistik wirken sich Ankündigungsschreiben positiv auf die Kontakt- und Kooperationswahrscheinlichkeiten aus. Für diejenigen Nummern, für die Adressen für ein Anschreiben ermittelt werden können, sollte in einem (mehrsprachigen) Ankündigungsschreiben ein Telefon-Termin und eine (kostenlose) Rückrufnummer genannt werden. Die Erhebungskosten dürften dadurch zwar um ca. 1 Euro pro Fall steigen, aber vermutlich in einer deutlichen Verringerung des Nonresponse resultieren.

#### 4.5.4 Refusals und Refusal-Konversion

Bislang erfolgt nach einer erstmaligen Verweigerung kein Konvertierungsversuch. Durch ein geeignetes spezielles Interviewtraining (vgl. Groves/McGonagle 2001), Anschreiben und Incentives kann in der Regel mehr als 1/3 der Verweigerer konvertiert werden. Weiterhin kann eine günstigere Allokation der Interviewer (größere Streuung über den Tag, Interviewerwechsel und Zeitwechsel bei Verweigerungen) eine Verringerung der Verweigerungsraten bewirken. Daher sollte bei einem RDD-Sample besonders auf den Einsatz eines Call-Management-Systems mit einer effizienten Allokation der Interviewer gedrängt werden.

#### 4.5.5 Diskussion der Möglichkeiten von Incentives

Bislang werden im LFS keine Incentives verwendet, obwohl die Adressen für alle ausgewählten Haushalte verfügbar sind. Für einen reinen RDD-Survey ist die Verwendung von Incentives unmöglich, falls keine Adress-Listen zur Verfügung stehen. Beim LFS böte sich die Möglichkeit, die generierten Nummern in den vorhandenen Datenbasen zu suchen und so Adressen für die eingetragenen Nummern zu ermitteln. Diese Adressen könnten sowohl zur Vorbereitung der Kontaktaufnahme (Anschreiben) als auch für den Einsatz von Incentives verwendet werden. Beide Maßnahmen dürften sich positiv auf eine Verringerung des Nonresponse auswirken.

Nimmt man für die Zukunft an, dass für ca. 1/4 der RDD-Stichprobe keine Adresse auffindbar sein wird<sup>2</sup> und Verweigerung und Nichtlistung der Adresse stochastisch unabhängig sind, dann wären ca 3/4 der Verweigerer (also 11.25% der Bruttostichprobe) für ein Incentive verfügbar. Grob gerundet wären dies 1250 Fälle für eine Nonresponse-Nachbearbeitung der Verweigerer. Für diese wird ein spezielles Anschreiben zusammen mit einem LFS-Flyer empfohlen. Nimmt man für Erstellung, Druck und Versand des Flyers Kosten in Höhe von einem Euro sowie Personalkosten in gleicher Höhe an, dann würde diese Nachbearbeitung 2500 Euro beanspruchen.

<sup>1</sup> Es wäre zwar möglich, dass die Interviewer bislang Informationen aus dem Register zu Steuerung ihrer Kontaktversuche verwenden, aber dafür liegen keine Hinweise vor.

<sup>2</sup> Es wäre erwägenswert, in der Pilotstudie (vgl. Abschnitt ) am Beginn des Interviews Name und Adresse zu erheben, was bei dem Hinweis auf eine Erhebung der amtlichen Statistik vermutlich wenig Probleme bereiten würde. Diese Angaben ließen sich sinnvoll für eine Verweigerer-Nachbearbeitung einsetzen. Der Anteil der Verweigerer ohne Adresse an der Bruttostichprobe läge dann eher niedriger als bei den folgenden Berechnungen; entsprechend höher wären die Kosten.

Möchte man ein effektives Incentive verwenden, dann legt die Literatur nahezu einhellig die Verwendung sicherer (also dem Anschreiben beiliegender) Incentives nahe. Ebenso einhellig werden monetäre Incentives in beträchtlicher Höhe<sup>1</sup> empfohlen (vgl. Brick et al. (2005)). Dies würde bei einem Incentive von 15-20 Euro die Kosten um maximal 25000 Euro erhöhen. Angesichts der gesamten Erhebungskosten pro Fall ist dies eine ökonomisch akzeptable Investition. Allerdings sollten die möglichen politischen Konsequenzen abgeschätzt werden. Es wäre denkbar, dass die Teilnahme an einer Befragung der amtlichen Statistik zumindest von einem Teil der Bevölkerung als Bürgerpflicht angesehen wird und somit die "Bezahlung" der Beantwortung eines Fragebogens als Verschwendung von Steuergeldern erscheint. Angesichts dieser möglichen problematischen Reaktion und des bislang noch nicht besorgniserregenden Ausmaßes an Nonresponse durch Verweigerung (ca. 15% der Bruttostichprobe im LFS 2004) sollte zwar mit der Nachbearbeitung der Verweigerer (Anschreiben, Neukontakt) begonnen werden, aber es sollten noch keine monetären Incentives verwendet werden.

#### 4.6 Implikationen des Auswahlsatzes

Der LFS in Luxemburg besitzt ein besonderes Problem, das ihn wissenschaftlich besonders interessant macht: Es ist eine wiederholte große Erhebung in einem Land mit kleiner Bevölkerungszahl. Für viele Berechnungen ist der Auswahlsatz (der Anteil der Elemente in der Stichprobe an der Anzahl der Elemente in der Grundgesamtheit) nicht mehr zu vernachlässigen: Legt man 12 000 kontaktierte Personen bei 452 000 Einwohnern zu Grunde, dann liegt der Auswahlsatz bei 2.65% der Einwohner. Entsprechend liegt die Wahrscheinlichkeit, mindestens einmal in die Stichprobe zu gelangen, bei ca. 12.6% in 5 Jahren und ca. 23.6% in 10 Jahren.<sup>2</sup> Die Erhebungen sind also inhaltlich kaum unabhängig voneinander, da die Befragten sich an eigene Befragungen erinnern. Nimmt man an, dass Befragte mit 2 bis 3 Freunden über ihre täglichen Erfahrungen sprechen, so ist zu erwarten, dass die Befragten über ihre Netzwerke innerhalb von 5 Jahren mit 30-50% Wahrscheinlichkeit von der Erhebung des LFS Kenntnis erlangen. Die Art der Durchführung des LFS könnte

daher langfristig für das Bild STATECs in der Bevölkerung von Bedeutung sein.

#### 4.7 Imputationen und deren Kosten

Das RDD wird gegenüber der bisherigen Praxis der CATI-Surveys im LFS keine Veränderungen der Praxis der Imputation oder seinen Kosten bedingen. Allgemein spielt mit Ausnahme der Variablen "Einkommen" Item-Nonresponse in sozialwissenschaftlichen Erhebungen kaum eine Rolle. Das gilt auch für den LFS: Soweit dies den Fragebögen und den Datensätzen zu entnehmen ist, liegt der Anteil fehlender Werte bzw. nicht dokumentierte Codes bei fast allen Variablen unter 1%. Sollten Item-Nonresponse-Korrekturen trotzdem notwendig erscheinen, dann sind diese über einmalig zu schreibende Imputationsroutinen nahezu problemlos durchführbar. Aufgrund der einfachen Macro-Implementierung (und den hier irrelevanten Möglichkeiten der Varianzschätzung) würde ich den Gebrauch von STATA und seinen Macros zur multiplen Imputation von Royston (2004) empfehlen.

#### 4.8 Diskussion der Möglichkeit von Record-Linkage und dessen Kosten

Für das RDD-Sample kommen mehrere Möglichkeiten des Record-Linkage in Betracht.<sup>3</sup> Zunächst sollten die ermittelten Nummern gegen das Telefonbuch gematched werden, um auf diese Weise einen Namen zu erhalten. Es wäre zu prüfen, ob die von den Befragten die Genehmigung zum Abgleich der Daten mit dem Zensus und dem Zentralregister eingeholt werden kann. Sollte die rechtliche Möglichkeit bestehen, so wäre aus statistischer Sicht ein Match mit Zensus und Zentralregister wünschenswert: Die Daten könnten für einen Coverage-Test sowie für Korrekturen der Stichprobe in Hinsicht auf Coverage, Nonresponse und vermutlich Meßfehler verwendet werden. Die Erhebungskosten für Namen und/oder Personennummer sind vernachlässigbar, die Kosten des Record-Linkage gering.<sup>4</sup> Der Präzisionsgewinn durch solche Abgleiche ist kaum zu überschätzen. Technisch stellt das Record-Linkage kein Problem dar. Da die Probleme des Record-Linkage für einen LFS rein juristisch sind, liegt eine Beurteilung der Möglichkeit außerhalb der Kompetenz des Gutachters.

---

<sup>1</sup> Üblich sind bei allgemeinen Bevölkerungsumfragen mehrere Dollar.

<sup>2</sup> Die Berechnungen basieren auf der Annahme einer Binomialverteilung der Versuche, also der Unabhängigkeit der Stichproben.

<sup>3</sup> Literatur und Software zur Durchführung von Record-Linkage ohne exakte Schlüssel wie Personen-Nummern findet sich bei Schnell/Bachteler/Bender (2004).

<sup>4</sup> In ähnlichen Projekten erzielten Schnell/Bachteler/Bender (2004) sehr hohe Einwilligungsraten (über 90% sind international bei solchen Abgleichen erreichbar, wenn man sie den Befragten plausibel macht). Die Durchführungskosten für einen solchen Abgleich liegen in der Höhe eines Mann-Monats.



#### 4.9 Veränderung der Datenerfassung oder des Data-Editing

Durch das RDD allein entstehen keine Veränderungen in den Datenaufbereitungen (Erfassung und Editieren) der CATI-Daten. Den zur Verfügung gestellten Unterlagen konnte nicht entnommen werden, dass Registerdaten zur Konsistenzkontrolle des LFS auf Individualebene genutzt werden. Falls dies doch der Fall sein sollte, würde diese einfache und direkte Methode zunächst unmöglich und ließe sich nur durch nachträgliches Matchen teilweise ausgleichen (vgl. Abschnitt ). In jedem Fall wäre ein solcher Abgleich wünschenswert.

#### 4.10 Konsequenzen des RDD für die statistische Nutzung der Daten

Die möglichen Konsequenzen liegen vor allem im Bereich der Gewichtungsverfahren, der Hochrechnung und der Varianzschätzungen.

##### 4.10.1 Einfluss auf Gewichtungsverfahren

Durch das RDD allein entstehen keine Veränderungen in den Gewichtungsverfahren der CATI-Daten. Allerdings liegt es nahe, bei einer Neuorganisation der Erhebung das Gewichtungsverfahren zu überarbeiten. So ist es uns nicht gelungen, die veröffentlichten Schätzungen exakt zu replizieren. Dies wird sowohl durch die (uns nicht vorliegende) Art der saisonalen Bereinigung bedingt als auch durch die Art der Gewichtung.<sup>1</sup> Der erste Schritt sollte also in einer exakten Dokumentation der Gewichtung, der saisonalen Bereinigung, der Kalibrierung mit CALMAR und einer eventuell zusätzlichen Glättung bestehen.

##### 4.10.2 Einfluss auf Hochrechnungsverfahren

Durch das RDD allein entstehen keine Veränderungen im Hochrechnungsverfahren. Obwohl für Luxemburg keine kleinräumige Ausweisung der Arbeitslosenzahl erforderlich ist, empfehle ich die Schätzung durch Verwendung zusätzlicher Informationen zu stabilisieren. Obwohl zahlreiche andere Möglichkeiten existieren, erscheint mir eine Orientierung an der

Praxis anderer Länder sinnvoll. So hat z.B. das Vereinigte Königreich alle Statistiken der amtlichen Statistik auf GREG-Schätzer umgestellt (vgl. ONS (2001)). Da aggregierte Registerdaten der gemeldeten Arbeitslosen zur Verfügung stehen, liegt die Verwendung dieser Daten auf der Ebene der Cantone als Hilfsvariable für eine Schätzung nahe. Auch ohne Interesse an kleinräumigen Daten wird sich die Schätzung der Aggregatstatistik durch GREG mit kleinräumigen Registerdaten verbessern.<sup>2</sup> Die Kosten für diese zusätzliche Anpassung sind -- nach der einmaligen Erstellung der Macros für ein geeignetes Analyseprogramm -- vernachlässigbar.

##### 4.10.3 Konsequenzen des RDD auf die Varianzschätzung

Die Varianzschätzung einer solchen Stichprobe muss die Schichtung, Klumpung, die Endlichkeit der Population, die Verteilung der interessierenden Variablen, die Poststratifizierung und Imputation berücksichtigen. Bislang werden nach meinem Wissen bei der Varianzschätzung fast alle diese Varianzquellen berücksichtigt. Die Ausnahmen sind die Poststratifizierung und Imputation, deren Beitrag zur Varianzschätzung aber vermutlich gering ist und wohl vernachlässigt werden kann. Dies gilt nicht für die Klumpung. Der LFS ist doppelt geklumpt: Wenige Interviewer führen sehr viele Interviews durch. Fast immer werden alle Informationen über die Personen eines Haushaltes durch einen Interviewer erhoben. Damit ist die wesentliche Variable, die die PSU festlegt, der Interviewer. Dies war bislang so und wird auch bei einer RDD-Stichprobe so bleiben. Allerdings muss der Interviewer-Effekt auf die Varianzschätzung berücksichtigt werden:

- Sowohl bei der Berechnung der Varianzen mit der Interviewer-ID als PSU-definierende Variable
- als auch bei der drastischen Verringerung der Fallzahl der Interviews pro Interviewer. Die tatsächlichen Varianzen werden durch die Neuorganisation der Feldarbeit sinken, nicht durch die Tatsache der RDD-Stichprobe. Zusätzliche Kosten entstehen nicht.

<sup>3</sup> Der Datensatz 2004 verwendet 52 Schichten, aber nur 28 Gewichte, die innerhalb der Schichten variieren. Dies dürfte durch die in der technischen Beschreibung des LFS als Hinweise enthaltene Zusammenfassung für Cantone bedingt sein; lässt sich aber im Detail weder aus den Daten noch aus der Beschreibung rekonstruieren.

<sup>4</sup> Coverage und Nonresponse-Probleme haben zu einer verstärkten Beschäftigung mit Small-Area-Estimation zur Korrektur von Stichprobendefiziten geführt. Aktuelle Übersichten bieten Longford (2005) und Särndal/Lundström (2005). Eine Simulation anhand der Daten des LFS zeigte deutliche Reduktionen; Einzelheiten finden sich in einem unveröffentlichten technischen Bericht von Hethey (2006).

### Designeffekte durch PSU-Homogenität

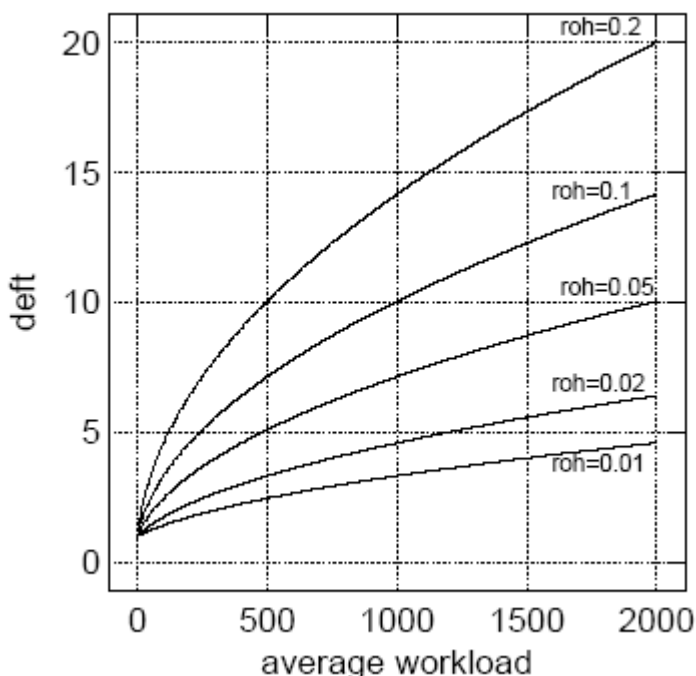
Weder die bisherige noch die zukünftige RDD-Stichprobe sind über die Haushalte hinaus geklumpt, daher werden keine Veränderungen in der Größe oder Präzision der Schätzung durch die Umstellung erwartet. Allerdings werden bislang mehr als 60% der Interviews als Proxy-Interviews geführt. Durch eine Reduktion dieses Anteils dürfte sich vor allem im Bereich minimaler Beschäftigung eine Verbesserung erzielen lassen. Dies ließe sich vermutlich durch eine Präzisierung der Regeln zur Bestimmung der auskunftserteilenden Person, der Verlagerung des Schwerpunkts der Erhebung in die Zeit außerhalb der

Hauptarbeitszeit und die Begrenzung des Workloads pro Interviewer erreichen lassen.

### Designeffekte durch Interviewer-Homogenität

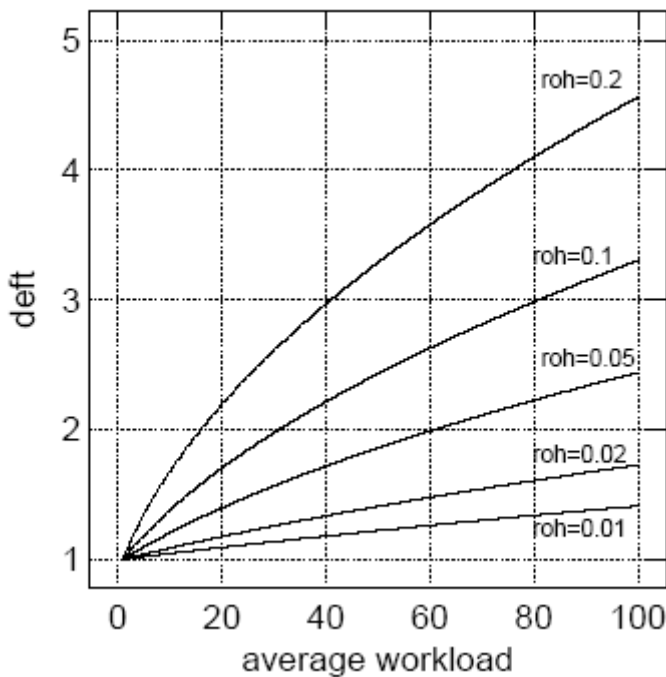
Die Umstellung auf RDD wird keinen *direkten* Einfluss auf Design-Effekte besitzen. Da aber bislang die Workload pro Interviewer sehr hoch war, sind vermutlich bislang sehr viel größere Standardfehler vorhanden, als es die bisherigen Berechnungen nahelegen, vgl. Schnell/Kreuter (2005). Die Abbildungen 9 und 10 zeigen den theoretischen Einfluss der Workload der Interviewer auf die Design-Effekte.<sup>1</sup>

Abbildung 9: Design-Effekte bei hoher Interviewer-Workload



<sup>1</sup> Der Design-Effekt ergibt sich durch  $deft = \sqrt{1 + (w-1)roh}$  wobei  $w$  die mittlere Zahl der Interviews pro Interviewer und  $roh$  die Intraklassenkorrelation ist. Der mittlere Wert für "roh" wird in der Literatur mit 0.01 berichtet; Werte über 0.1 sind nach Groves (1989:318) selten, werden aber in der neueren Literatur berichtet (Kreuter 2002:244)

Abbildung 10: Design-Effekte bei mittlerer Interviewer-Workload



Der Wechsel zu einer RDD-Stichprobe sollte daher auch eine Reorganisation der Feldarbeit nahelegen. Wünschenswert wäre dabei eine Reduktion der Workload der Interviewer auf deutlich weniger als 100 realisierte Interviews pro Interviewer. Dies allein sollte die theoretischen Design-Effekte mindestens um den Faktor 3 verkleinern. Daher ist mit einem *indirekten* aber wünschenswerten Effekt der Umstellung auf die Präzision der Schätzung zu rechnen. Die tatsächliche Größe dieses Effekts läßt sich schwer prognostizieren, da der Effekt wesentlich durch Implementierungsdetails der Feldarbeit (Art und Häufigkeit der Supervision, des Trainings, der Bezahlung, der Instrumentengestaltung etc., vgl. Schnell/Kreuter (2000)) gesteuert wird.

#### 4.11 Pilotstudie

Um jedes mögliche Risiko auszuschalten, wird die Durchführung einer Pilot-Studie empfohlen. Hierbei

sollten für die gleiche Referenzperiode ein LFS nach dem traditionellen Design sowie ein RDD-Survey durchgeführt werden. Führt man drei Monate lang kontinuierlich beide Surveys durch, dann kann durch einen solchen Vergleich eine Veränderung von 6% gegen 8% mit Alpha von 0.05 bereits mit einer Power von mehr als 0.8 (einseitig) gezeigt werden; eine Veränderung von 6% auf 4% bereits mit einer Power von 0.9. Entsprechend lassen sich Veränderungen von 45% auf 41% bzw. von 45% auf 49% mit einer Power von ca. 8.83 zeigen.<sup>1</sup> Eventuelle Differenzen sollten also mit hoher Sicherheit nachweisbar sein. Durch die parallele Erhebung bekäme man Hinweise auf eventuell notwendige Korrekturfaktoren bei Längsschnittbetrachtungen. Die Kosten der Erhebung (der Preis dürfte bei ca. 1/3 der Erhebungskosten eines LFS liegen: 1/4 der faktischen Erhebung + fixe Kosten durch Sampling und Aufbereitung) werden durch den hohen Gewinn an Sicherheit und Informationen über Bias und Korrekturfaktoren mehr als aufgewogen.

<sup>1</sup> Die Power der Tests wurden mit Systat 11 berechnet.

## 5. Langfristige Perspektiven

Mittelfristig sind RDD-CATI-Surveys sicherlich eine effektive Datenerhebungsmethode für Surveys aller Art, auch für einen LFS der amtlichen Statistik. Langfristig, d.h. mit einer Perspektive von mehr als 10-15 Jahren, dürften alle CATI-Surveys vor erheblichen Problemen stehen. Dies betrifft einerseits die Zunahme von Haushalten, die nur noch über mobile Telefone erreichbar sind. Bedeutsamer als das Noncoverage-Problem der mobilen Haushalte wird das Kooperationsproblem: Der Nonresponse bei Surveys steigt weltweit, auch bei den LFSs der amtlichen Statistik (wenn auch dort langsamer als bei anderen Surveys). Besonders stark ist der Anstieg des Nonresponse durch Verweigerung bei RDD-CATI-Surveys.<sup>1</sup> Die Ursache liegt hier vor allem in der

Zunahme des Telemarketing und der Surveys allgemein. Sowohl das Verweigerungsverhalten als auch die Exposition der Befragten gegenüber Telemarketing- und Surveyversuchen dürften systematisch mit den Variablen eines LFS zusammenhängen, wobei zumindest für einige Subgruppen an den extremen Enden der Schichtungsskalen Nonresponse aufgrund nicht beobachteter Kovariate ("not missing at random": NMAR) erwartet werden müssen. Solche Effekte lassen sich durch kein datenbasiertes Verfahren korrigieren. In 10-15 Jahren müssen CATI-Surveys (ob RDD oder nicht) entweder durch andere Erhebungsformen ersetzt oder ergänzt werden. Ohne Änderung der gesetzlichen Regelungen wird dies vermutlich ohne kostenintensive Face-to-Face-Befragungen kaum möglich sein.

---

<sup>1</sup> Da auch Anrufbeantworter oder übermittelte bzw. eben nicht übermittelte Telefonnummern von den Befragten als Screening-Hilfsmittel verwendet werden, ist die Abgrenzung zwischen Verweigerung und nicht-erreichten Telefonnummern in der Praxis unscharf.

## 6. Anhang: Konzeption und Durchführung des LFS in den Ländern Europas

Die folgende Tabelle enthält methodische Details der nationalen LFS in Europa im Jahr 2003<sup>1</sup>

**Tabelle 4: Characteristics of the Labour Force Survey in european countries in 2003**

Country	Sampling base	Sampling unit	Data collection	Compulsory (y/n)	response rate (%)
Finland	Population Register	individuals	telephone (administrative data)	n	85
Iceland	National population register	individuals	telephone (administrative data)	?	83
Switzerland	Phone number register, register of foreign persons	household, individuals	telephone	n	81
Sweden	Sweden's Register of total population	individuals	telephone, face-to-face	n	84
Norway	Central Population Register	family	telephone, face-to-face (administrative data)	y	90
Denmark	Central Population Register	individuals	mail, telephone (administrative data)	n	66
Slovenia	Central Population Register	individuals	face-to-face, telephone	n	86
Malta*	Electoral Register	households	face-to-face, telephone	y	80
United Kingdom	Royal Mail's PAF, telephone directories	households	face-to-face, telephone	n	78
Germany*	Census database (old federal states), Population register (new federal states), register of new dwellings	households	face-to-face, mail	y	96
Estonia	Population and housing census database	individuals	face-to-face	n	81
Lithuania	Population register	individuals	face-to-face	n	90
Belgium	National Register of Persons	households	face-to-face	y	78
Spain	?	households	face-to-face	y	92
Italy*	Population registers of communes	households	face-to-face	y	95
Greece	census database	households	face-to-face	?	93-94
Ireland	Census of population	households	face-to-face	n	79
Latvia	Population census	households	face-to-face	n	85
Netherlands	addresses draw up by the postal services and Population Register	households	face-to-face	n	59
Bulgaria	Population Census	households	face-to-face	n	86
Portugal	?	dwellings	face-to-face	y	91
France	Population Census	dwellings	face-to-face	y	81
Cyprus	Population Census	dwellings	face-to-face	y	99
Austria*	Census of dwellings and houses	dwellings	face-to-face	n	72
Poland	Domestic territorial Division Register	dwelling units	face-to-face	n	80
Croatia	Population Census	dwellings	face-to-face	n	87
Czech Republic	Census Register of areas and dwellings	dwellings	face-to-face	n	77
Hungary	Population and housing census	dwellings	face-to-face	n	87
Slovakia	Register of municipalities, Population census	dwellings	face-to-face	?	92
Romania	?	dwelling units	face-to-face	?	95
Turkey	Dwelling enumeration register	dwellings	?	y	89

In countries marked with a \* characteristics may have changed since 2003 while changing towards a continuous survey throughout the year.

<sup>1</sup> Quelle: Eurostat/European Commission (2005): The European Union labour force survey: main characteristics of the national surveys. *Methods and nomenclatures* (2005). Luxembourg: Office for Official Publications of the European Communities.

## 7. Literatur

- Brick, J., Waksberg, J., Kulp, D., Starer, A. (1995): Bias in list-assisted telephone surveys. In: *Public Opinion Quarterly*, 59, S. 218-235
- Brick, J. M., Montaquila, J., Hagedorn, M. C., Roth, S. B., Chapman, C. (2005): Implications for RDD-Design from an Incentive Experiment. In: *Journal of Official Statistics*, 21, S.4
- Gabler, S., Häder, S. (1997): Überlegungen zu einem Stichprobendesign für Telefonumfragen in Deutschland. In: *Zuma-Nachrichten*, 41, S. 7-19
- Groves, R. (1989): *Survey Costs and Survey Errors*. New York
- Groves, R., McGonagle, K. (2001): A Theory Guided Interviewing Training Protocol Regarding Survey Participation. In: *Journal of Official Statistics*, 17, 2, S. 249-265
- Grund, M. (2002): Mobilfunkteilnehmer: Ein Repräsentanzproblem. In: Gabler, S., Häder, S. (Herausgeber) *Telefonstichproben*, Waxmann, Münster, S. 111-120
- Hethey, T. (2006): Evaluating the quality of small area estimators in labour force surveys. Unpublished technical report, Center for Quantitative Methods and Survey Research University of Konstanz
- Kim, S. W., Lepkowski, J. M. (2002): Telephone Household Non-Coverage and Mobile Phones. In: *American Association for Public Research - Section on Survey Research Methods*, S. 1845-1850
- Kreuter, F. (2002): *Kriminalitätsfurcht: Messung und methodische Probleme*. Opladen
- Longford, N. T. (2005): *Missing Data and Small-Area Estimation*. Springer Verlag
- Nicolaas, G., Lynn, P. (2002): Random-digit dialling in the UK: viability revisited. In: *Journal of the Royal Statistical Society, Series A*, 165, S. 297-316
- Office for National Statistics UK (2001): Strategies and approaches for small area statistics. In: United Nations Economic and Social Council (Herausgeber) *Conference of European Statisticians - forty-ninth plenary session*. <http://www.unece.org/stats/documents/ces/2001/19.e.pdf>
- Royston, P. (2004): Multiple imputation of missing values. In: *The Stata Journal*, 3, S. 227-241
- Schnell, R. (1997): *Nonresponse in Bevölkerungsumfragen*. Opladen
- Schnell, R., Bachteler, T., Bender, S. (2004): A toolbox for record linkage. In: *Austrian Journal of Statistics*, 33, S. 125-133
- Schnell, R., Kreuter, F. (2000): Untersuchungen zur Ursache unterschiedlicher Ergebnisse sehr ähnlicher Viktimisierungssurveys. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 1, S. 96-117
- Schnell, R., Kreuter, F. (2005): Separating Interviewer and Sampling Point Effects. In: *Journal of Official Statistics*, 21, 3, S. 389-410
- Särndal, C.-E., Lundström, S. (2005): *Estimation in Surveys with Nonresponse*. Wiley, Hoboken
- Wilson, P., Blackshaw, N., Norris, P. (1998): An Evaluation of Telephone Interviewing on the British Labour Force Survey. In: *Journal of Official Statistics*, 4, S. 385-400