



Diverse recruitment strategies result in different participation percentages in a web-based study, but in similar compliance

Manas K. Akmatov · Nicole Rübsamen · Anja Schultze ·
Yvonne Kemmling · Nadia Obi · Kathrin Günther · Wolfgang Ahrens ·
Frank Pessler · Gérard Krause · Rafael T. Mikolajczyk

Received: 30 June 2014/Revised: 1 September 2015/Accepted: 2 September 2015/Published online: 14 September 2015
© Swiss School of Public Health (SSPH+) 2015

Abstract

Objectives We compared participation and compliance with a web-based data collection on infections among population-based samples recruited in different ways.

Methods Individuals were recruited from participants in the German National Cohort study (Group A, $n = 279$) or persons who were invited to this study but did not participate (Group B, $n = 53$). A third group was invited to the web-based study only (Group C, $n = 145$).

Results Response varied among groups between 3 % (B), 11 % (C) and 61 % (A), but compliance was similar (81–85 %). Response did not differ by age and sex. Compliance was lower among the youngest and oldest participants. In addition, participants currently not employed were more likely to have better compliance. Semi-parametric group-based modelling identified three distinct compliance trajectories; “poor compliance” (8 %),

“improving compliance” (14 %) and “very good compliance” (78 %).

Conclusions Participation differed among modes of recruitment, but compliance was similar among groups and notably high. Different recruitment approaches can be used and collected data can be combined to achieve greater sample sizes for longitudinal web-based studies.

Keywords Response · Compliance · Recruitment strategies · Longitudinal study · Non-responders · Population-based study · Second-stage non-response · Web-based study · German National Cohort

Introduction

Web-based tools are increasingly used in epidemiological studies. They may facilitate participation in population-based studies because they do not require physical presence in the study center. However, systematic recruitment of participants directly over the Internet is not feasible as no defined sampling frame exists (Fan and Yan 2010). Studies that use web-based tools were often conducted in specific population groups, e.g. university students (Pealer et al. 2001), doctors and nurses (Raziano et al. 2001) or self-recruited users of the Internet (Owen et al. 2013).

In recent years, a number of studies have investigated the effect of recruitment approach on participation percentage in web-based studies with formal population-based sampling. Different methods were used to obtain a sample that is representative of the general population, e.g. random digit dialling (Midanik and Greenfield 2010), a random sample from a large probability-based database with potential web respondents (Nagelhout et al. 2010), or a

M. K. Akmatov and N. Rübsamen contributed equally to this work.

M. K. Akmatov (✉) · N. Rübsamen · A. Schultze ·
Y. Kemmling · G. Krause · R. T. Mikolajczyk
Department of Epidemiology, Helmholtz Centre for Infection
Research, Brunswick, Germany
e-mail: manas.akmatov@twincore.de

N. Obi
University Cancer Center Hamburg, University Medical Center
Hamburg-Eppendorf, Hamburg, Germany

K. Günther · W. Ahrens
Department of Epidemiological Methods and Etiologic
Research, Leibniz Institute for Prevention Research and
Epidemiology-BIPS, Bremen, Germany

M. K. Akmatov · F. Pessler
TWINCORE Centre for Experimental and Clinical Infection
Research, Hannover, Germany

stratified random sample from population registries (Turner et al. 2005). These studies found different participation percentages dependent on recruitment approach. However, participants were invited to answer the web-based questionnaire only once. Whether compliance with longitudinal web-based data collection differs depending on the mode of recruitment cannot be inferred from these results. Bexelius et al. investigated the possibility to use web-based questionnaires in population-based infectious disease surveillance (Bexelius et al. 2010), but as participants only had to fill in the questionnaire when they had symptoms of infectious disease, the setting differs from a web-based longitudinal study.

Furthermore, only limited data about differences between participants and non-participants of web-based studies are available (Russell et al. 2010). This question can be addressed in a study employing a web-based data collection embedded in an existing study with a different data collection mode (second-stage studies). So far, data on second-stage participation are only available for paper-based questionnaires (Boshuizen et al. 2006; van Loon et al. 2003; Volken 2013). These studies found associations between second-stage participation and sex, socio-economic status, current employment, and subjective health. A web-based study embedded in a currently conducted large study in Germany including physical examinations offers the relatively rare opportunity to study second-stage participation in this context.

We integrated a longitudinal web-based data collection on acute infections of 3-month duration in a study requiring visits to a study center in a sample drawn from the local population registries. In addition, we recruited persons from a sample with a similar age distribution directly for participation in the web-based study. We aimed to compare participation percentages and compliance with the longitudinal data collection among those who (a) participated in the original study including the assessments in the study center (Group A), (b) did not participate in the original study, but completed the non-responder questionnaire and in this context agreed to participate in the web-based study (Group B), and (c) were invited to the web-based study only (Group C).

Methods

This study was part of the pretest studies of the German National Cohort (GNC) (Wichmann et al. 2012) and was conducted during Pretest 2. In brief, the GNC is a large multicenter prospective population-based study beginning 2014 to examine risk factors for common diseases (e.g. cardiovascular, neoplastic, metabolic, pulmonary, neuropsychiatric and infectious diseases). The study will include about 200,000 individuals between 20 and 69 years of age

in Germany. The older age groups will be oversampled (20 % of individuals younger than 40 years and 80 % aged 40 years and more). The participants will be recruited in 18 study centers across Germany. During the Pretest 2 phase 2896 participants were examined in the study centers and various feasibility aspects of the final study were investigated. In all participating study centers, random samples were drawn from the population registries in the corresponding municipalities. Potential participants were contacted via land mail. Up to two reminders and up to ten telephone calls (provided that phone numbers could be identified) were used to contact non-responders.

Three study centers (Bremen, Hamburg, and Hannover) implemented the web-based data collection on acute symptoms of infections. The response proportions to the initial study among the GNC cohort were 19 % (study center Bremen), 22 % (study center Hamburg), and 16 % (study center Hannover). Subjects who agreed to participate in the pretest of the GNC (including a 4 h long examination at the study center) and had an e-mail address were asked during their visit in the study center to participate in the prospective web-based study (Group A). Those who could not be recruited for the examination in the study center received a non-responder questionnaire. In the non-responder questionnaire an invitation to participate in the web-based study was included (Group B). Finally, persons from a population-based sample (with the same sex and age distribution as in the GNC) in Hannover who were not initially contacted regarding the participation in the pretest of the GNC were directly invited to participate in the web-based study by land mail (Group C). Group C was invited once per land mail and did not receive any reminders. The flow chart of the recruitment process is shown in Fig. 1.

Web-based data collection

The web-based data collection lasted 3 months (February to April 2013 in Groups A and B and June to August 2013 in Group C). An open-source online survey application—<http://www.limesurvey.org/> (Schmitz 2007)—was used for this purpose. Weekly emails containing the link to the web questionnaire were sent to the participants on Mondays. They were asked whether they had developed new symptoms of respiratory or gastrointestinal tract infections or conjunctivitis in the last week. If so, they were asked to provide additional information. Participation was defined as filling in the online questionnaire at least once during the study period.

Statistical analysis

Participation percentages stratified by sex and age were calculated separately for all three groups and for the

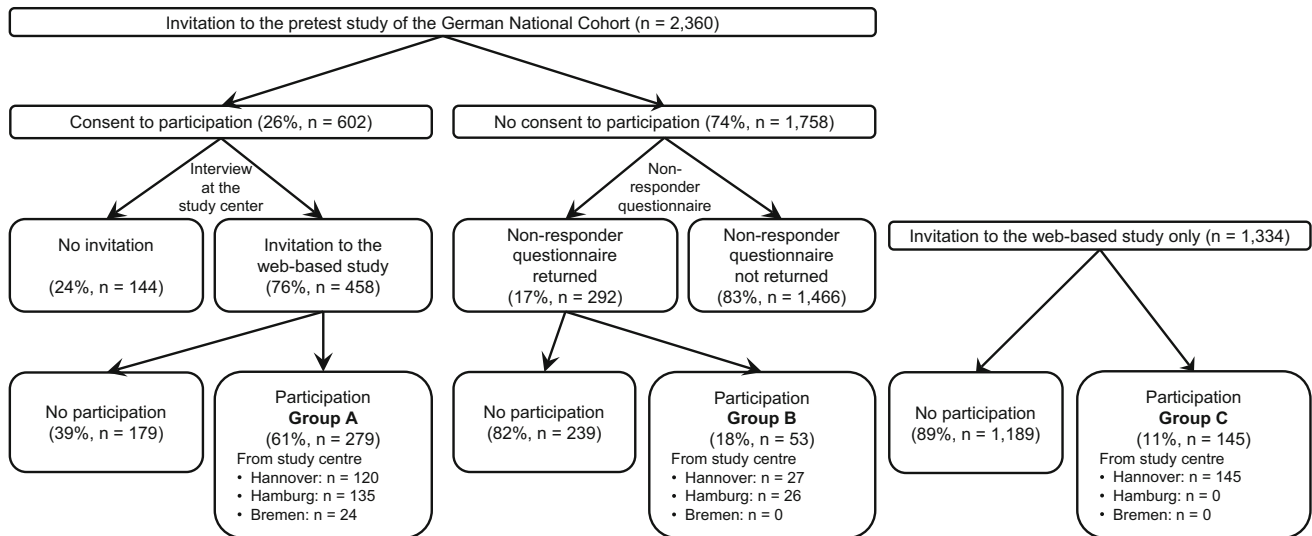


Fig. 1 Flow chart of the recruitment process in the prospective web-based study on acute infections conducted in three study centers of the German National Cohort from February to August 2013

combined Groups A and B. For the examination of variables associated with second-stage participation, data on participants and those who did not participate in the web-based study among those who attended the study center were described as percentages for categorical variables and medians for continuous variables. Differences between groups were tested with the Chi-square test (for categorical variables) and the Mann–Whitney *U* test (for continuous variables). All tests were two-sided. In addition, multivariable logistic regression was performed to identify independent associations. Compliance with the web-based data collection was defined as providing a response about the presence of symptoms in a given week, and multivariable logistic regression for correlated data based on generalized estimating equations (GEE) was used to identify variables associated with compliance (Hosmer and Lemeshow 2000). Furthermore, trajectories of compliance over the study period were studied using semi-parametric group-based modelling implemented in PROC TRAJ in SAS 9.2 (Jones et al. 2001). With PROC TRAJ unobserved groups within the sample that have distinct patterns of change in compliance over time are modelled. Trajectories are described using polynomial functions for the groups. Participants are classified as belonging to one of the groups based on individual trajectories. We started with a single trajectory and subsequently increased the number of trajectories; the selection of the final number of trajectories was based on the Bayesian Information Criterion (BIC). BIC values were -2278.22 for one trajectory, -1918.19 for two trajectories, -1891.60 for three trajectories, -1895.10 for four trajectories, and -1904.57 for five trajectories. We selected the model with three trajectories as it had the best BIC value (Jones et al. 2001).

Table 1 Descriptive overview on the distribution of the number of weeks with responses in the web-based study

Number of weeks with responses	<i>n</i>	% (<i>n</i> = 570)	% without non-participants (<i>n</i> = 477)
0	93	16.3 %	/
1	8	1.4 %	1.7 %
2	7	1.2 %	1.5 %
3	11	1.9 %	2.3 %
4	10	1.8 %	2.1 %
5	16	2.8 %	3.4 %
6	15	2.6 %	3.1 %
7	15	2.6 %	3.1 %
8	30	5.3 %	6.3 %
9	38	6.7 %	8.0 %
10	92	16.1 %	19.3 %
11	235	41.2 %	49.3 %

Results

Participation percentages

In total, 570 persons agreed to participate in the web-based study, but only 477 actually filled in at least one web questionnaire during the study period (Table 1). Given that percentages of those who responded to only few weekly questionnaires were low and that the majority provided more complete records, we defined participation as at least one completed questionnaire. The participation percentage varied in the three study groups: 61 % in Group A, 18 % in Group B if considering only those who returned the non-responder questionnaire and 3 % if all non-responders were considered, and 11 % in Group C (Table 2). Since Groups

Table 2 Participation percentages stratified by sex and age group in the prospective web-based study on acute infections conducted in three study centers of the German National Cohort from February to August 2013 (analyses based on Groups A, B, and C)

	Group A % (participated/ invited)	Group B ^a % (participated/ provided non- responder questionnaire)	Group B % (participated/ invited)	Group A + B ^b % (participated/ invited)	Group C % (participated/ invited)
Sex					
Female	62 % (146/237)	18 % (27/150)	3 % (27/870)	16 % (173/1078)	12 % (78/668)
Male	60 % (133/221)	18 % (26/142)	3 % (26/885)	15 % (159/1075)	10 % (64/666)
Age group					
20–29 years	67 % (35/52)	20 % (3/15)	2 % (3/127)	21 % (38/178)	6 % (7/110)
30–39 years	69 % (29/42)	31 % (5/16)	3 % (5/184)	15 % (34/224)	7 % (10/139)
40–49 years	64 % (72/112)	21 % (14/67)	4 % (14/386)	18 % (86/486)	11 % (32/301)
50–59 years	61 % (79/130)	17 % (16/93)	3 % (16/476)	16 % (95/576)	15 % (53/357)
60–69 years	52 % (64/122)	15 % (15/101)	3 % (15/582)	11 % (79/689)	9 % (40/427)
Total	61 % (279/458)	18 % (53/292)	3 % (53/1758)	15 % (332/2153)	11 % (145/1334)

^a Considering only those who returned the non-responder questionnaire

^b Considering all people who were initially invited to the pretest study

Table 3 Socio-demographic characteristics of the second-stage participants and non-participants in the prospective web-based study on acute infections, stratified by response to the pretest study of the German National Cohort (analyses based on Group A)

	Participants of the original study with physical assessments (i.e. Group A)				
	Participants of the web-based study (<i>n</i> = 279) %	Non-participants of the web-based study ^a (<i>n</i> = 179) %	<i>p</i> value (Chi-square test)	Odds ratio	95 % confidence interval
Sex					
Female	52	51	0.76	1	
Male	48	49		0.99	0.65, 1.50
Median age (25th percentile; 75th percentile) ^b	50 years (41; 59)	53 years (45; 62)	0.01	0.99 per 1 year increase	0.97, 1.01
Marital status					
Single	36	28	0.16	1.01	0.59, 1.75
Married	52	55		1	
Other (widowed, divorced)	12	17		0.90	0.48, 1.66
Level of school leaving qualification					
Less than 10 years school education	34	58	<0.001	0.51	0.33, 0.79
More than 10 years school education	66	42		1	
Currently employed					
Yes	79	68	<0.006	1	
No	21	32		0.67	0.41, 1.10
Smoking status					
Non-smoker	41	37	0.18	1	
Ex-smoker	26	34		0.86	0.51, 1.44
Current smoker	33	29		1.21	0.73, 2.01

^a Non-participants of the web-based study include both persons who did not have an e-mail address and were thus not eligible to participate and persons who had an e-mail address, but were not willing to participate

^b Mann–Whitney *U* test

A and B were recruited stepwise from the same initial sample their combined participation percentage to be compared with participation in group C was 15 %.

Second-stage participation

Among participants of the original study with a personal assessment in the study center, we were able to assess second-stage participation for the web-based study because we had the socio-demographic information from the original study (Table 3). Those not participating in the additional web-based study had lower education and were more frequently not employed (most of those not employed were in retirement age).

Compliance

The average proportion of web questionnaires completed in each week was similar across Groups A, B, and C, i.e. 83, 85, and 81 %, respectively.

Multivariable logistic regression for correlated data suggested that age and employment status influenced compliance, but compliance was not associated with recruitment strategy (Table 4). Compared to individuals aged 40–49 years, younger (20–39 years) and older (60–69 years) individuals showed poorer compliance.

Semi-parametric group-based modelling identified three distinct compliance trajectories over the study period: “poor compliance”, “improving compliance” and “very good compliance” (Fig. 2). Most participants (78 %) were in the trajectory “very good compliance” where compliance increased slightly at the beginning of the study and remained high over the study period. In the “improving compliance” trajectory (14 %) compliance was low at the beginning of the study, but increased sharply. Finally, in the “poor compliance” trajectory (8.5 %) compliance was moderate at the beginning and declined over the study period. The distribution of the three compliance trajectory groups in the groups A, B, and C was similar (Table 5).

Discussion

We examined participation percentages and compliance in a population-based study that used web-based data collection among participants recruited via different modes. Participation in the web-based study was highest among those who participated in the original study (Group A) and lowest among those who did not agree to participate in the original study (Group B). Other studies also report a higher participation percentage among people recruited from large panels (i.e. people who indicated their willingness to participate in research) compared to those who did not (Duffy

Table 4 Factors that influence compliance in the prospective web-based study on acute infections conducted in three study centers of the German National Cohort from February to August 2013 (analyses based on Groups A, B, and C)

	Odds ratio ^a	95 % confidence interval
Group		
Group A	1	
Group B	1.03	0.63, 1.70
Group C	1.08	0.76, 1.54
Sex		
Female	1	
Male	0.90	0.67, 1.20
Level of school leaving qualification		
Less than 10 years school education	0.97	0.71, 1.33
More than 10 years school education	1	
Marital status		
Single	1.07	0.69, 1.66
Married	1	
Other (widowed, divorced)	1.12	0.68, 1.84
Currently employed		
Yes	1	
No	1.83	1.23, 2.73
Age		
20–29 years	0.40	0.21, 0.73
30–39 years	0.55	0.31, 0.99
40–49 years	1	
50–59 years	1.03	0.67, 1.57
60–69 years	0.61	0.38, 0.98
Change in compliance over time (increase per one week)	1.12	1.09, 1.15

^a Multivariable logistic regression for correlated data with outcome providing response in a given week vs. missing response

et al. 2005; Nagelhout et al. 2010), indicating that some people are generally more willing to participate in studies. In our current study, it seems that participation in the initial study might have even enhanced willingness to participate in the additional web-based study, resulting in a higher participation in the combined Groups A + B than in Group C. We also showed that it was possible to recruit some participants for the web-based study among those who could not be recruited for the initial study including medical examinations in the study center. However, the participation percentage in the additional data collection among non-responders of the original study was only 3 %. This suggests that while examinations in the study center can be the reason for non-participation in some cases, the

Fig. 2 Compliance over the study period with the prospective web-based data collection on acute infections conducted in three study centers of the German National Cohort from February to August 2013

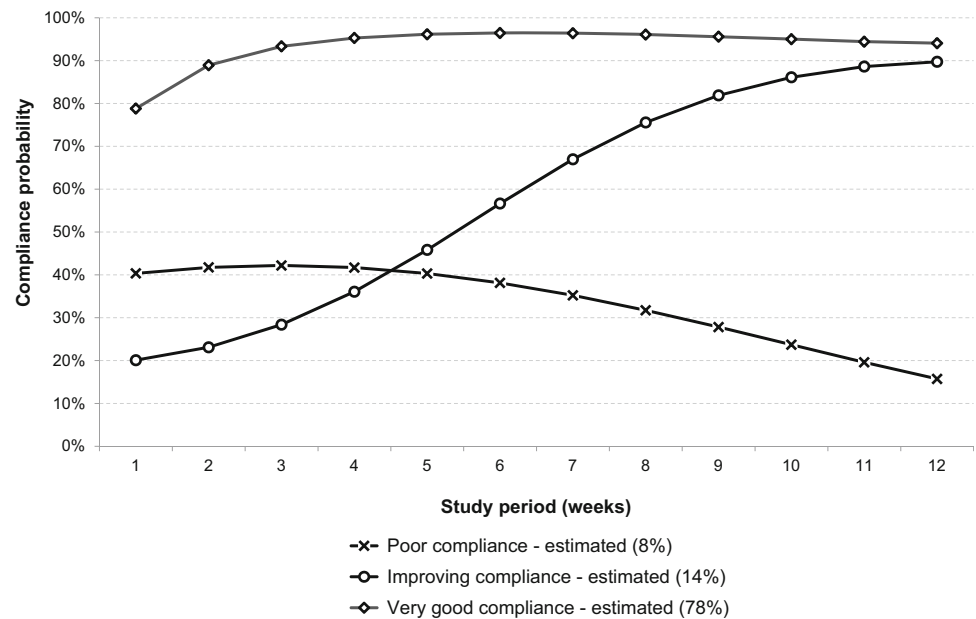


Table 5 Differences across the three compliance trajectory groups and the Groups A, B, and C

	Poor compliance <i>n</i> (%)	Improving compliance <i>n</i> (%)	Very good compliance <i>n</i> (%)	<i>p</i> value (Fisher's exact test)
Group				0.32
Group A	28 (10.0)	33 (11.8)	218 (78.1)	
Group B	4 (7.6)	8 (15.1)	41 (77.4)	
Group C	6 (4.4)	20 (14.8)	109 (80.7)	

proportion of those who can additionally be recruited for a study not involving an appointment in the study center but using web-based data collection is quite low.

Compared to the 16–22 % response in the initial study with physical examination and within this group a 61 % participation in the web-based study, the response in the primarily web-based study was only 11 %. Even if the lack of access to the internet might play some role, the interest to participate in the initial study appears to be somewhat higher, possibly due to the perception of the initial study as being more interesting or because of the feedback about health status offered to its participants. However, the proportions are not directly comparable, since potential participants of the initial study received up to two reminder letters and ten phone calls while the recruitment of Group C only included one invitation letter.

With respect to second-stage participation we found that participants of the web-based study were more educated and less likely to be unemployed (with most unemployed persons being in retirement age) than those who did not agree to participate in the additional study. These findings are consistent with the results of second-stage participation

analyses in paper-based studies (Boshuizen et al. 2006; van Loon et al. 2003; Volken 2013).

Despite the differences in participation percentages, participants in all three groups showed a similarly high compliance with web-based data collection. Even in Group B, where the participation percentage was low, compliance was as high as in the other groups. This is generally a positive finding which can encourage creation of web-based cohorts or panel studies. Taken together these results suggest that different recruitment approaches can be used and collected data can be combined to achieve greater sample sizes for longitudinal web-based studies.

Acknowledgments This project was conducted in the context of the Pretest studies (Pretest 2) of the German National Cohort (<http://www.nationale-kohorte.de>). These were funded by the Federal Ministry of Education and Research (BMBF), project number 01ER1203, and supported by the Helmholtz Association as well as by the participating universities and Institutes of the Leibniz Association. We also gratefully acknowledge the contribution to data collection by the study personnel in Hamburg, Hannover and Bremen.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical standards The study was approved by the Ethics Committees of the State Boards of Physicians of the German Federal States of Bremen, Hamburg, and Lower Saxony. All subjects gave written informed consent before entering the study.

References

Bexelius C, Merk H, Sandin S, Nyren O, Kuhlmann-Berenzon S, Linde A, Litton JE (2010) Interactive voice response and web-

- based questionnaires for population-based infectious disease reporting. *Eur J Epidemiol* 25(10):693–702
- Boshuizen H, Viet A, Picavet H, Botterweck A, van Loon A (2006) Non-response in a survey of cardiovascular risk factors in the Dutch population: determinants and resulting biases. *Public Health* 120(4):297–308
- Duffy B, Smith K, Terhanian G, Bremer J (2005) Comparing data from online and face-to-face surveys. *Int J Market Res* 47(6):615–639
- Fan W, Yan Z (2010) Factors affecting response rates of the web survey: a systematic review. *Comput Hum Behav* 26(2):132–139
- Hosmer D, Lemeshow S (2000) Logistic regression models for the analysis of correlated data. *Applied logistic regression*. Wiley, New York
- Jones B, Nagin D, Roeder K (2001) A SAS procedure based on mixture models for estimating developmental trajectories. *Socio Meth Res* 29(3):374–393
- Midanik LT, Greenfield TK (2010) Reports of alcohol-related problems and alcohol dependence for demographic subgroups using interactive voice response versus telephone surveys: the 2005 US National Alcohol Survey. *Drug Alcohol Rev* 29(4):392–398
- Nagelhout GE, Willemsen MC, Thompson ME, Fong GT, van den Putte B, de VH (2010) Is web interviewing a good alternative to telephone interviewing? Findings from the International Tobacco Control (ITC) Netherlands survey. *BMC Public Health* 10:351
- Owen J, Bantum E, Criswell K, Bazzo J, Gorlick A, Stanton A (2013) Representativeness of two sampling procedures for an internet intervention targeting cancer-related distress: a comparison of convenience and registry samples. *J Behav Med* 37(4):630–641
- Pealer L, Weiler R, Pigg R Jr, Miller D, Dorman S (2001) The feasibility of a web-based surveillance system to collect health risk behavior data from college students. *Health Educ Behav* 28(5):547–559
- Raziano D, Jayadevappa R, Valenzula D, Weiner M, Lavizzo-Mourey R (2001) E-mail versus conventional postal mail survey of geriatric chiefs. *Gerontologist* 41(6):799–804
- Russell C, Boggs D, Palmer J, Rosenberg L (2010) Use of a web-based questionnaire in the Black Women's Health Study. *Am J Epidemiol* 172(11):1286–1291
- Schmitz, C (2007) LimeSurvey. The open source survey application. <http://www.limesurvey.org>. Accessed 5 February 2015
- Turner CF, Villarroel MA, Rogers SM, Eggleston E, Ganapathi L, Roman AM, Al-Tayyib A (2005) Reducing bias in telephone survey estimates of the prevalence of drug use: a randomized trial of telephone audio-CASI. *Addiction* 100(10):1432–1444
- van Loon A, Tjhuis M, Picavet H, Surtees P, Ormel J (2003) Survey non-response in the Netherlands: effects on prevalence estimates and associations. *Ann Epidemiol* 13(2):105–110
- Volken T (2013) Second-stage non-response in the Swiss health survey: determinants and bias in outcomes. *BMC Public Health* 13:167
- Wichmann HE, Kaaks R, Hoffmann W, Jockel KH, Greiser KH, Linseisen J (2012) The German National Cohort. *Bundesgesundheitsbl* 55(6–7):781–787