Int J Public Health 52 (2007) 188–189
1661-8556/07/030188-2
DOI 10.1007/s00038-007-6089-7
© Birkhäuser Verlag, Basel, 2007

Playing hard to get: field substitutions in health surveys

Stefaan Demarest, Lydia Gisle, Johan Van der Heyden

Scientific Institute of Public Health, Belgium, Brussels

Submitted: 6 July 2006; Revised: 5 September, 18 November 2006, 7 March 2007; Accepted: 7 March 2007

Background and objectives

Field substitution occurs when a non-responding sample unit is replaced with a substitute (reserve) unit during the fieldwork phase of the survey process (Chapman 2003; Vehovar 1999). Field substitution is a contested tool in survey methodology. The main criticism is that substitutes may differ systematically from the non-respondents and consequently may bias survey estimates if used literally as substitutes (Rubin & Zanutto 1985). Furthermore, substitution can result in a reduction of the interviewers' efforts to contact initially selected units (as they know that a substitute can be used instead) and it may prolong the fieldwork since substitutes are contacted only after some effort has been made to obtain the collaboration of the originally selected unit (Chapman 1985). A plea for substitution is that it can help to ensure an appropriate representation of each segment originally distinguished in the sample (Vehovar 1994). By substituting non-participants with matching units, the probability to have a realized sample at the end of the fieldwork with the same composition as the initial sample is larger (Chapman 2003). As such, substitution may contribute to the effort also to include the "hard to reach population" in the realized sample.

To assess the consequences of applying fieldwork substitution, data from the Belgian Health Interview Survey 2004 is used. Belgium is indeed one of the rare countries that uses field substitution in conducting the survey (Demarest et al. 2002).

Data and methods

Although the units defined in the Belgian health survey are individuals and the target of the survey is expressed as a predefined number of individuals to be interviewed within a one-year period, households in selected municipalities are used as the basis of the fieldwork. The National Register, comprising administrative data on all citizens, is used as the sampling frame for the survey.

A combination of matched and random substitutes for refusing households is used in the survey. Substitution households are matched to those initially selected on the basis of several characteristics; the age of the reference person, the household size (number of members) and the area of residence (statistical sector). For every initial selected household, three matched substitutes are selected, creating a cluster of four households with common characteristics. For every cluster, a substitute cluster is created, that will be activated in case the initial cluster of households is exhausted (i.e. if all four households of the initial cluster turn out to be non-participants). Substitution clusters are selected at random, and have no common characteristics with the initial clusters. Limiting the cluster size to four (matched) households using at random substitution clusters is based on the presumption that the number of households with matching characteristics within the same area is rather limited. Interviewers were not aware if households they are dealing with were substituting households or not. They were obliged to document every contact attempt (day, hour, mode of contact) and to respect strict time frames (maximal 6 weeks between 'activation' of the household and its final status).

The 2004 survey aimed at the participation of 12 050 individuals. In order to achieve this target, a total of 6 705 eligible households[1] was initially invited to participate.

– After wave 1 all initial selected households receive their final status (participation or refusing), and the participation rate without substitution can be evaluated..
– After wave 2 (and also 3 and 4) the effect of substitution of refusing households can be estimated. After wave 2 all households used as substitutes for the initial refusing households receive their final status and a participation

[1] This total comprised an over sampling of 10%. This over sampling was intended to serve as an accelerator to assure the realization of the scheduled number of participating individuals at the end of the fieldwork phase.

Int J Public Health 52 (2007) 188–189

© Birkhäuser Verlag, Basel, 2007

Table 1 Participation rate (in %) by wave and cumulative participation rate fieldwork, Health Interview Survey, Belgium, 2004

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5*
Number of households	6,056	2,394	1,065	433	679
Participation rate by wave	64.7	60.4	54.6	54.3	50.1
Cumulative participation rate	64.7	63.5	62.5	62.2	61.4
Odds of refusing participation by wave	1	1.2	1.5	1.5	1.8
95 % Confidence Intervals		(1.1–1.3)	(1.3–1.7)	(1.2–1.8)	(1.5–2.1)

* Wave 5 refers to all contacted households of the successive waves in the replacement clusters

rate can be calculated for this group. A similar approach is followed for wave 3 and wave 4. In case wave 4 does not result in a participating household, a random selected (substitute) cluster will be activated.

– For this cluster the same procedure of substitution is applied. Since the number of cases for which a substitute cluster has to be used is limited, they are considered together and labeled as wave 5

Results

The successive waves suffer from a declining participation rate (Tab. 1); while 64.7 % of the initially selected households participated (wave 1), this percentage drops to 60.4 % for first substitutes (wave 2), 54.6 % for the second substitutes (wave 3) and 54.3 % for the third substitutes (wave 4). This phenomenon was not unexpected, since substitutes have common characteristics with previous refusing households. The overall participation rate for the substituting clusters is 50.1 % (wave 5). This is surprising since these clusters have nothing in common with the clusters they replace. An important explanation for this phenomenon is that substituting clusters are only activated once the original clusters are exhausted, mostly towards the end of fieldwork, when the pressure to realize prescheduled numbers of interviews on time is rising. In this phase, interviewers opt more than in previous waves to ask households to participate by telephone than by time consuming contacts at doorstep, leaving fewer opportunities to convince households to participate. The declining participation rates do negatively affect the overall participation rate (from 64.7 % after wave 1 to 61.4 % after wave 5); although this effect remains limited given the fact that the number of households inclined in the successive waves is shrinking.

The application of a logistic regression model shows that the odds of refusing to participate for households belonging to the fourth wave is 1.5 times higher than for initial selected households (after correction for age, nationality of the reference person and size of household).

Although confronted with declining participation rates, substitution of refusing households assured that the composition of the initial selected sample (in terms of age-category and household size) was partly reconstituted throughout the fieldwork procedure. With the process of substituting, selective dropout is at least partially countered since substitutes have common essential characteristics with the household they replace. Where e. g. after the first wave one-person-households were underrepresented in comparison to the initial sample (39.7 % instead of the 45 % as scheduled), this under representation diminishes throughout the fieldwork phase.

Conclusion

Field substitution must be considered as valuable alternative when conducting surveys. It allows an optimal control during the fieldwork, optimizes the preservation of the designed structure of the sample and is especially fruitful when targeting groups which are hard to reach. A major disadvantage is that it induces higher levels of refusal which result in a supplementary bias in the estimates.

References

Chapman D (2003). To substitute or not to substitute – That is the question. The Survey Statistician: 32–24.

Chapman D (1982). Substitution for missing units. Proceedings of the Survey Methods Section. Alexandria (VA): American Statistical Association, p. 76–84.

Chapman D, Roman A (985). An investigation of substitutions for an RDD survey. Proceedings of the Survey Methods Section. Alexandria (VA): American Statistical Association, 269–74.

Demarest S, et al. (2002). Gezondheidsenquête door middel van interview, België, 2001. Brussel: IPH/EPI reports n° 2002–22.

Rubin D, Zanutto E (2006). Using matched substitutes to adjust for nonignorable nonresponse through multiple imputations. In: R Groves, R Little, J Eltinge, eds. Survey nonresponse. New York: John Wiley.

Vehovar V (1994). Field substitution – a neglected option? Proceedings of the Survey Methods Section. Alexandria (VA): American Statistical Association, 589–94.

Vehovar V (1999). Field substitution and unit non response. Journals of Official Statistics, Statistics Sweden, 335–50

Address for correspondence

Stefaan Demarest
Scientific Institute of Public Health
Epidemiology Unit
J. Wytsmanstraat 14
1050 Brussel
Belgium
e-mail: stefaan.demarest@iph.fgov.be