



# Sample substitution can be an acceptable data-collection strategy: the case of the Belgian Health Interview Survey

Stefaan Demarest<sup>1</sup> · Geert Molenberghs<sup>2</sup> · Johan Van der Heyden<sup>1</sup> ·  
Lydia Gisle<sup>1</sup> · Herman Van Oyen<sup>1</sup> · Sandrine de Waleffe<sup>3</sup> · Guido Van Hal<sup>4</sup>

Received: 11 September 2016/Revised: 24 April 2017/Accepted: 26 April 2017/Published online: 3 May 2017  
© Swiss School of Public Health (SSPH+) 2017

## Abstract

**Objectives** Substitution of non-participating households is used in the Belgian Health Interview Survey (BHIS) as a method to obtain the predefined net sample size. Yet, possible effects of applying substitution on response rates and health estimates remain uncertain. In this article, the process of substitution with its impact on response rates and health estimates is assessed.

**Methods** The response rates (RR)—both at household and individual level—according to the sampling criteria were calculated for each stage of the substitution process, together with the individual accrual rate (AR). Unweighted and weighted health estimates were calculated before and after applying substitution.

**Results** Of the 10,468 members of 4878 initial households, 5904 members (RRind: 56.4%) of 2707 households (RRhh: 55.5%) participated. For the three successive (matched) substitutes, the RR dropped to 45%. The composition of the net sample resembles the one of the initial samples. Applying substitution did not produce any important distorting effects on the estimates.

**Conclusions** Applying substitution leads to an increase in non-participation, but does not impact the estimations.

**Keywords** Health survey · Non-response · Sampling · Matched substitution

## Introduction

Sample substitution, in which non-responding units are replaced ('substituted') by other units in the fieldwork process, is used to assure that the net sample meets the size of the target sample (Baldissera et al. 2014; Li et al. 2014; Nishimura 2015) and, at the same time, to attenuate non-response bias in survey estimates (Chapman 1983, 2003; Chapman and Roman 1985; Chiu et al. 2005; Kish and Hess 2004; Rubin and Zanutto 2002; Vehovar 1994, 1995, 1999, 2003). Sample substitution belongs to a class of methods where no weights are introduced to correct for non-response. Such methods include sample substitution, stratification to anticipate stratum response rates, and certain analysis methods like ignorable likelihood (Molenberghs and Kenward 2007). While weights have the potential, if properly chosen, to reduce non-response bias, they do so at the expense of increased variability, which can be a seriously limiting factor. Contrary to techniques—such as weighting—substitution is regarded with a strong degree of scepticism in survey methodology literature and is rarely advocated in survey textbooks (David et al. 2014; Pickery and Carton 2008; Vehovar 1994).

The most fundamental criticism against sample substitution is that it violates the principle of probabilistic sampling, given that substitutes have different selection probabilities than the initial selected units, and should be regarded as a form of quota sampling (Nishimura 2015).

However, equating quota sampling with sample substitution strategies ignores the fact that there are many substitution techniques, ranging from basic permissive

✉ Stefaan Demarest  
stefaan.demarest@wiv-isp.be

<sup>1</sup> Scientific Institute of Public Health, Brussels, Belgium

<sup>2</sup> Hasselt University, KU Leuven Leuven, Hasselt, Leuven, Belgium

<sup>3</sup> Statistics Belgium, Brussels, Belgium

<sup>4</sup> University of Antwerp, Antwerp, Belgium

approaches—as often applied in quota sampling—to very complex, rigorous procedures. Besides, a probability sampling mechanism can be made to operate at all stages. Based on features such as how substitute units are selected (random or not random), to what extent the data collection is controlled (loose versus tight), the presence or absence of common characteristics between initial units and substitutes (unmatched or matched substitution) and the level of substitution (substitution at the lowest unit level, i.e., the individual or substitution at intermediate levels such as schools, enterprises,...), at least several dozen types of substitution designs can be distinguished even without getting into more exotic hybrids such as using earlier cross sections or panel designs (Kish and Hess 2004; Lynn 2004; Nishimura 2015; Smith 2007).

Although not acknowledged as such, inflating the initial sample size according to the expected response rates to achieve a stated power—a common practice and advocated strategy in survey methodology—can also be regarded as a very basic substitution design: extra units are added to the initial sample that will serve as uncontrolled and unmatched replacements for non-respondents (Smith 2007). From this perspective, it seems doubtful to state that one of the most important disadvantages of applying substitution is that it results in higher levels of non-response. The latter is, indeed, inevitable, since, on the whole, non-respondents among the substitutes will be added to the initial non-respondents. It can even be presumed that non-response is higher in substitutes than in the initial units which might be seen as a sign that the substitutes were representing the initial non-respondents (Nishimura 2015; Smith 2007). A regrettable practice is that the initial non-respondents are sometimes not added to the denominator when calculating the response rate creating “the illusion that a non-response problem has been solved” (Vehovar 1999). Another shortcoming associated with sample substitution is that it may attenuate the interviewers’ efforts to contact units, although this remains controversial: some interviewers may limit their efforts to obtain participation from an initial case, since they know that this case will be substituted, other will do whatever is possible to obtain participation from an initial case, since they know that, in case of substitution, the whole process of contacting the substitute will restart again (Dorsett 2010; Vehovar 1999, 2003). Yet, these arguments are only relevant if one assumes that the interviewer plays a predominant role in the substitution process, while in more advanced applications of substitution, the decision to substitute a non-responding unit is made by the central office to the extent that interviewers are not even aware if an unit is an initial unit or a substitute (Smith 2007).

Sample substitution was applied in the first Belgian Health Interview Survey (BHIS) from its start in 1997 and

in all the surveys since then (2001, 2004, 2008, 2013) (Demarest et al. 2013a; Van der Heyden et al. 2014; Van Oyen et al. 1997). The basic reason for applying substitution in the BHIS is to ensure that at the end of the fieldwork, the predefined number of interviews and the initial composition of the sample in terms of sex, age group, and household size is met (Demarest et al. 2007; Demarest et al. 2013a; Van Oyen et al. 1997).

The first objective of this manuscript is to verify if substitution, as applied in the BHIS 2018, contributes in reaching the predefined net-sample size target, both in size and composition. The second objective is to assess, to the extent possible, the impact of substitution on non-response bias for a number of health estimates.

## Methods

The BHIS is a national cross-sectional survey organized thus far in 1997, 2001, 2004, 2008, and 2013, and directed towards the general population. Details on the methodology can be found elsewhere (Demarest et al. 2013a). The net-sample size for the BHIS 2013 was predefined at 10,600 face-to-face interviews: 3500 in the Flemish Region, 4100 in the Walloon Region, and 3000 in the Brussels Capital Region, i.e., the three geopolitical entities of the country.

## Sampling procedure

The addresses of households to be contacted in the BHIS 2013 were selected in the national register by applying a multistage, clustered procedure. In a first step, 12 geographical strata (or provinces) were defined. In the Flemish and Walloon Regions, the number of interviews per province was proportional to the number of inhabitants in each province. Within each stratum, municipalities were selected according to a probability proportional by size method. In each selected municipality, one or more groups of 50 individuals had to be sampled. This was first done via a systematic sampling of households. The step size for the systematic sampling was calculated by dividing the total number of households in the selected municipality by the required number of households taking into account the average household size in the municipality. As the sampling was spread over four quarters, it means that in each group (on average), 12.5 individuals had to be interviewed per quarter.

Systematic sampling was based on households ordered by statistical sector (a subdivision of a municipality), household size, and age category of the reference person. For each selected household, three consecutive households in the ordered list were selected too, creating a cluster of

four households living in the same statistical sector, with the same household size and with the same age category of the reference person. To account for the fact that all households within one quadruple of households would refuse, the step size was set to half its original value, which doubled the number of selected clusters.

Once the clusters of households were selected for each group, a procedure of vertical and horizontal scrambling was applied. Vertical scrambling of the clusters randomized the original order of the clusters. Horizontal scrambling randomized the order in which households in a particular cluster had to be activated, that is as initial household or as first, second, or third substitutes.

### Substitution procedure

Before the start of the fieldwork phase, all interviewers had to attend a training session during which the global approach used in the BHIS was explained. In particular, it was stressed that in case a household invited for participation turned out to be a non-participant, a substitute household would be activated with similar characteristics as the initial activated household. It was also made clear for the interviewers that this substitution process would continue with second or third substitutes and, if necessary, with a substitute cluster.

At the start of every trimester, a prenotification letter was sent to all initial selected households. At the same time, the interviewers could access the preloaded information on these households (first name, sex, and date of birth of all household members according to the National Register) and the contact form on their laptop (CAPI application). Every contact attempt had to be registered using the contact form in terms of date of contact attempt, time of the day, attempt mode (by phone, at doorstep), and result of the contact (no contact, incorrect address, refusal, and agreement to participate). In case no contact could be established, the interviewer had to continue trying to contact the household at least four more times, using alternative days, time of the day, and mode of contact (at least one 'at doorstep' contact attempt). After five unsuccessful attempts, the interviewer could indicate that a household was 'not contactable' and stop the contact trials. In case a household was 'not contactable', refused to participate or the info on the address was incorrect, a (first) matched substitute household was activated (that is: a prenotification letter was sent to the household). At the same time, access to the information of the initial household was blocked in the laptop management system and interviewers could access the preloaded information on the (first) substitute household instead. Exactly the same contact procedure had to be applied for the substitute household. In case

contact attempts with the first substitute household did not result in participation, the (second) substitute household was activated. This process was repeated until a household participated in the survey or until the two clusters were exhausted.

Although the BHIS is basically a survey directed towards households, its target is expressed in terms of individual participation. For each participating household, at most four members were invited for the interview: the reference person, his/her partner (if any), and two other ad random sampled members (in case of a partner) or three other members (in case of no partner). Intra-household substitution was not allowed; a selected household member that refused to participate and did not allow an interview-by-proxy was not substituted.

### Databases

Two databases were used for the analysis: one with information on the contact procedure with the activated households (that is: households to which a prenotification letter was sent) and one with data collected through the survey both at the level of the households and individual household members. Combining both databases allowed to disentangle the whole substitution process and to assess the (non-)participation for each substitution stage.

### Statistical analysis

For every stage (initial and substitutes), the response rates at both household level ( $RR_{hh}$ ) and at individual level ( $RR_{ind}$ ) were calculated, defined as the proportion of participating units among the activated units, regardless of the reason for non-participation. The substitution process is implemented to hasten the realization of a predefined number of individual interviews. Therefore, the Accrual Rate (AR), defined as the number of interviews realized after each (initial, substitute) stage of the data collection divided by the target (10,600 interviews), was also calculated. To assess the impact of substitution on the estimates, poststratified weighted estimates for the initial participants, the substitutes, and for all participants are compared for a selection of health indicators (indicators listed in Fig. 1).

Weights were based on age, gender, household size, and province, and were calculated separately for each health indicator (taking into account item non-response) and this for the initial participants, the substitutes, and for all participants. Differences between the estimates based on, respectively, initial participants, substitutes, and all participants were calculated together with their 95% confidence intervals.

### Results

Table 1 presents an overview of the results of the applied substitution process in terms of the composition of the initial activated sample of households and of the successive substitute samples (sex and age group of the reference person, and size of the household). Added to the table is the nationality of the reference person, which is not used as a criterion when selecting the sample and is to be considered a latent characteristic. The size and composition of every substitute sample are associated with the  $RR_{hh}$  of the previous phase. Table 2 presents the same overview as Table 1, yet at the individual level, that is at the level of the household members invited to participate in the BHIS. The results are presented after random sampling of four household members in households counting more than four members. Individual non-participation comprises both non-participation due to the fact that the household as such did not participate as well as to intra-household refusal.

Of the 10,468 members of 4878 initial activated households invited to participate in the survey r, 5904 members ( $RR_{ind\_in}$ : 56.4%) of 2707 households ( $RR_{hh\_in}$ : 55.5%) participated. It should be noted that the total size of the initial individual sample is somewhat lower than the number of interviews to be achieved (10,600). This is due

to the fact that in a few municipalities, no interviewer was available during the start-up of the data-collection.

As shown in Table 1, participation in the survey differed according to the characteristics of the households; the initial selected households with a young reference person (15–24 years), for example, were less prone to participate ( $RR_{hh\_in\_15-24}$ : 42.3%) than households with a reference person aged 65–74 years ( $RR_{hh\_in\_65-74}$ : 60.0%). One-person households had the lowest response rate ( $RR_{hh\_in\_size1}$ : 49.7%). More disturbing is the finding that households with a non-Belgian reference person have a much lower response rate than households with a Belgian reference person (respectively,  $RR_{hh\_in\_nbel}$  of 48.5% and  $RR_{hh\_in\_bel}$  of 56.7%), given that the nationality of the reference person is not addressed in the substitution process.

The response rates at individual level appear to be more homogeneous. What holds is the lower response rate for the members of one-person households in comparison with members of three- or four-person households (respectively  $RR_{ind\_in\_size1}$  of 57.8% and  $RR_{ind\_in\_size4}$  of 59.0%). Lower response rates for households with a non-Belgian reference person are also reflected in lower response rates at the individual level ( $RR_{ind\_in\_nbel}$  for non-Belgians is 51.0%, versus  $RR_{ind\_in\_bel}$  of 57.3% for Belgian citizens). It should be mentioned that in a limited number of cases (402

**Table 1** Relative number of activated households (Act) and response rates (RR) for the initial and the substitute samples of households according to the characteristics of the household in terms of sex, age

group, and nationality of the reference person and size of the household (HH); Belgian Health Interview Survey 2013

	Initial sample		1st substitutes		2nd–3rd substitutes		4th–7th substitutes		Total net sample		
	Act <sub>hh_in</sub>	RR <sub>hh_in</sub>	Act <sub>hh_s1</sub>	RR <sub>hh_s1</sub>	Act <sub>hh_s2-3</sub>	RR <sub>hh_s2-3</sub>	Act <sub>hh_s2-3</sub>	RR <sub>hh_s4-7</sub>	Act <sub>hh_tot</sub>	Part <sub>hh_tot</sub>	RR <sub>hh_tot</sub>
Sex ref. person											
Male	3092	56.0	1370	50.5	1058	46.6	490	50.4	6010	3162	52.6
Female	1786	54.7	862	49.4	748	41.8	312	54.5	3708	1886	50.9
Age ref. person											
15–24 years	104	42.3	64	18.8	73	34.2	20	50.0	261	91	34.9
25–34 years	690	54.8	341	46.3	293	32.1	134	47.0	1458	693	47.5
35–44 years	865	55.1	397	45.6	339	49.6	158	54.4	1759	912	51.8
45–54 years	936	57.5	407	50.9	307	49.2	141	44.0	1791	958	53.5
55–64 years	888	55.4	398	54.3	280	50.7	158	56.3	1724	939	54.5
65–74 years	635	60.0	277	62.5	180	51.1	91	58.2	1183	699	59.1
75+ years	760	52.2	348	49.1	334	40.1	100	54.0	1542	756	49.0
Household size											
1 person HH	1821	49.7	898	45.1	854	36.2	309	46.6	3882	1763	45.4
2 persons HH	1450	57.3	673	56.2	464	53.2	183	64.5	2770	1574	56.8
3 persons HH	678	59.4	273	48.0	203	48.3	115	47.8	1269	687	54.1
4+ persons HH	929	61.1	388	52.6	285	53.3	195	51.3	1797	1024	57.0
Nat. ref. person											
Belgian	4132	56.7	1839	52.6	1431	47.1	642	53.3	8044	4326	53.8
Non-Belgian	746	48.5	393	38.2	375	34.9	160	46.3	1674	717	42.8
Total	4878	55.5	2232	50.1	1806	44.6	802	52.0	9718	5048	51.9

**Table 2** Relative number of activated individuals (Act) and response rates (RR) for the initial group, and nationality, and size of their household (HH); Belgian Health Interview Survey and the substitute samples of individuals according to their characteristics in terms of sex, age 2013

	Initial sample		1st substitutes		2nd-3rd substitutes		4th-7th substitutes		Total net sample		
	Act <sub>ind_in</sub>	RR <sub>ind_in</sub>	Act <sub>ind_s1</sub>	RR <sub>ind_s1</sub>	Act <sub>ind_s2-3</sub>	RR <sub>ind_s2-3</sub>	Act <sub>ind_s4-7</sub>	RR <sub>ind_s4-7</sub>	Act <sub>ind_tot</sub> (% activated)	Part <sub>ind_tot</sub> (% participants)	RR <sub>ind_tot</sub>
Sex											
Male	5093	56.4	2200	50.1	1651	48.3	895	51.6	9839 (48.3)	5231 (48.7)	53.2
Female	5375	56.4	2414	50.7	1879	46.9	905	50.7	10,573 (51.7)	5598 (51.3)	52.9
Age											
0-14 years	1584	58.8	695	50.2	513	50.7	323	54.2	3115 (15.8)	1716 (15.1)	55.1
15-24 years	1167	55.1	499	42.3	414	47.6	211	47.4	2291 (10.6)	1151 (11.1)	50.2
25-34 years	1381	54.7	650	48.2	534	39.9	266	46.6	2831 (13.0)	1406 (13.2)	49.7
35-44 years	1481	56.1	637	47.1	507	49.1	266	53.4	2891 (14.1)	1522 (14.1)	52.6
45-54 years	1555	56.9	670	50.0	466	50.2	238	43.7	2929 (14.4)	1558 (14.9)	53.2
55-64 years	1360	56.3	592	55.6	435	51.3	236	56.4	2623 (13.4)	1450 (13.0)	55.3
65-74 years	960	59.1	422	60.9	268	51.5	127	55.1	1777 (9.5)	1032 (9.2)	58.1
75+ years	980	53.6	449	51.7	393	41.7	133	54.9	1955 (9.2)	994 (9.4)	50.8
Household size											
1 person HH	1821	49.7	898	45.1	854	36.2	309	46.6	3882 (16.3)	1763 (17.4)	45.4
2 persons HH	2900	56.3	1346	55.4	927	52.4	366	62.3	5539 (28.6)	3094 (27.7)	55.9
3 persons HH	2031	57.8	819	46.2	609	47.6	345	44.6	3804 (18.4)	1995 (19.4)	52.4
4+ persons HH	3716	59.0	1551	51.4	1140	52.0	780	50.6	7187 (36.7)	3977 (35.5)	55.3
Nationality											
Belgian	8900	57.3	3873	52.2	2846	49.6	1447	52.0	17,066 (85.8)	9291 (85.0)	54.4
Non-Belgian	1568	51.0	741	40.9	684	39.0	353	47.6	3346 (14.2)	1538 (15.0)	46.0
Total	10,468	56.4	4614	50.4	3530	47.5	1800	51.2	20,412	10,829	53.1
Accrual rate		55.7		77.6		93.5		102.2			

individuals, or 3.7%), linkage at individual level was not possible, given that these cases were not listed in the national register (new-born babies, new partners, unregistered household members,...).

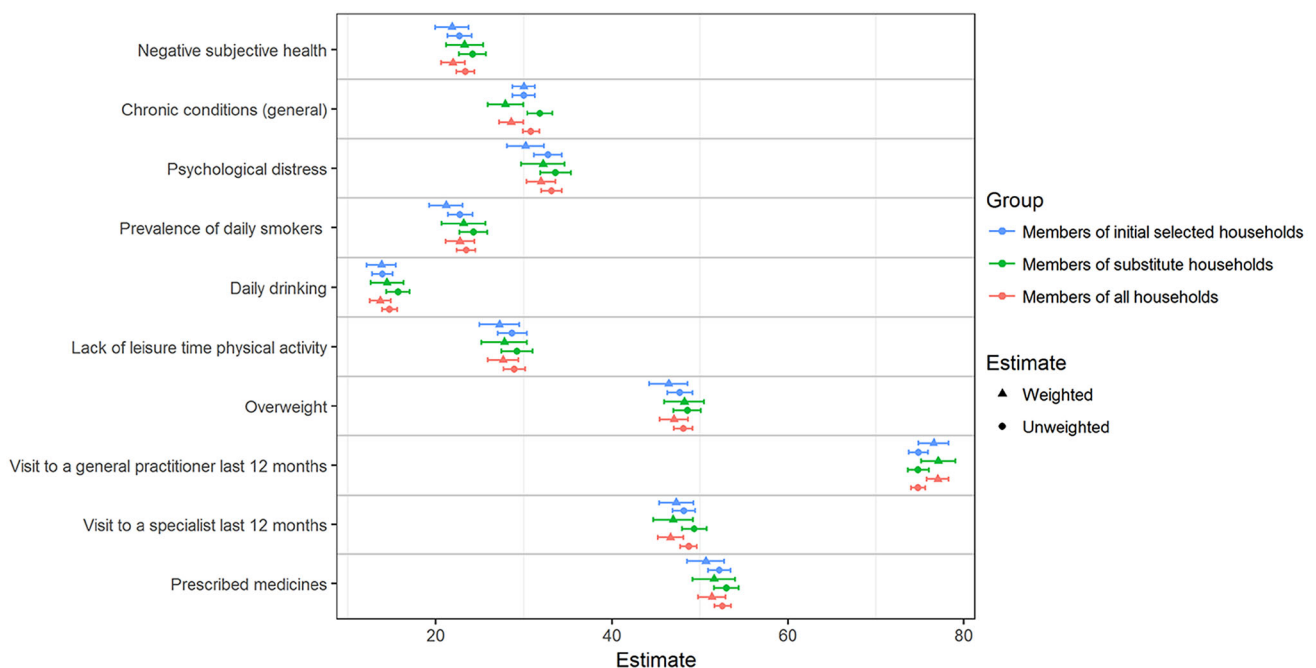
Of the initial 4878 activated households, 2171 did not participate, resulting in a shortage of 4564 individual participants. To cope with this, 2232 (matched) substitute households were activated (first substitutes), accounting for 4614 possible individual participants. One can note the overshoot in both the activated substitute households and, consequently, the number of expected individual participants. This is due to the fact that the substitution process is a permanent process throughout the data-collection period which could be avoided in case the substitution process would only start once the participation status of all initial activated households is known, that is at the end of each quarter.

Of all 2232 first substitute households, 1118 did participate ( $RR_{hh\_s1}$ : 50.1%), enabling to have an interview with 2326 household members ( $RR_{ind\_s1}$ : 50.4%). The decreasing response rates, both at household and individual level, can be found for every household or individual characteristic, with exception of the stable response rates for households with a reference person aged 55–74 years. The decline in overall response rates holds for the second and third substitute households ( $RR_{hh\_s2-3}$ : 44.6.0%). Contrary to the first three substitute households, the 4th till 7th substitute households (belonging to the second substitute cluster of 4) have common within-cluster

characteristics (statistical sector, sex, and age group of the reference person, size of the household), which are different from the initial selected household and its first three substitute households (from the initial cluster). Given the relatively low number of units in which they were activated, the results for the 4th till 7th substitute household were grouped. Even after three successive substitutions, 802 households were involved in the continuing substitution process, resulting in an additional participation of 418 households ( $RR_{hh\_s4-7}$ : 52.0%) counting 921 participating individuals ( $RR_{ind\_s4-7}$ : 51.2%). Compared to the previous substitution stages, response rates of the 4th till 7th substitution households were relatively higher both at household and at individual level.

The initially activated households combined with the substitution process resulted in a net sample of 10,829 participating members ( $RR_{ind\_tot}$ : 53.1%) of 5048 households ( $RR_{hh\_tot}$ : 51.9%). The net sample is slightly higher than the targeted sample (surplus of 289 individuals).

When comparing the estimates for a selected number of unweighted and weighted health indicators, the results show very limited and not statistically significant differences between estimates based solely on the initial sample and estimates based only on the substitutes (Fig. 1). As a consequence, no significant difference could be found between the estimates before and after substitution: the substitution process does not produce any important distorting effects on the estimates.



**Fig. 1** Unweighted and weighted estimates of a selection of health indicators (in terms of the percentage (and 95% CI) of the population that present the health-related condition) in the initial selected sample,

the substitutes, and the total net sample, Belgium, Health Interview Survey 2013

## Discussion

In the BHIS, a substitution approach is applied, in which the decision to substitute a non-participating unit follows a clear algorithm which cannot be controlled by the interviewers. It has not more than its name in common with very basic, uncontrolled and uncontrollable substitution schemes of mere quota sampling.

In this manuscript, it is shown that sample substitution in the BHIS 2013 can be used to preserve the size and composition of the initial sample with little consequences on the survey estimates. Of course, the ability to apply sample substitution in such a way heavily depends on the availability of several matching variables, such as the statistical sector, household size, nationality, sex, and age category of the reference person.... The absence of such variables—as in many sampling frames—implicates that sample substitution, as described in this manuscript, cannot be applied. The results show that in almost 60% of all cases, the initial selected units did participate, making the rather complex substitution process, as applied in the BHIS 2013, superfluous. Doubling the initial sample appears to be a more adequate approach than multiplying the sample size with factor 8, since the substitution process only affects a minority of units. Yet, different response rates at both household and individual levels, although limited, can be accounted for using substitution. Artificially increasing the sample size of initial selected households would probably yield the same differences in response rates. Applying substitution can be linked to a decline in participation rates. It seems that ‘hard to reach’ households are substituted with similar households. Their relatively increasing share in the groups of successive substitutes does impact the participation rate. The main advantage of substitution is that it increases the chance that the composition of the net sample reflects the composition of the initial sample regarding the variables used in the construction of the sample.

Based on a selection of health indicators, it was shown that the estimates derived solely on the initial participants or solely on the substitutes coincide with those calculated by taking the total net sample into account. This finding is in line with that of a similar study showed that using sample substitution enables to maintain the characteristics of the original sample without affecting the results, independently of the magnitude of substitution (Baldissera et al. 2014).

As shown, the assumption of an added value of using substitute clusters is disputed. In overall terms, 96.7% of the target sample size was realized after, at most, the third substitute household was activated. The availability of substitute clusters serves as an ultimate mean to realize the prefixed number of interviews, yet activating substitute

clusters negatively impacts the resemblance of the composition of both the initial and the net sample.

It is important to note that the substitution method is predominantly used to reduce bias as much as possible. As such, the focus is more on bias and less on precision. At the same time, the method enables us to reach the predetermined sample sizes per stratum (e.g., the number of interviews to be realized per region) much more easily than when a simple oversampling is applied. As such, the method helps to reach the prespecified precision, without undershooting or overshooting. In addition, there is evidence that the method works well, because the initial and substitute sub-samples do not significantly differ. Of course, it is not the only method that exists to this effect. In addition, one can never be sure, and, therefore, not claim that non-response bias is totally removed. Using data alone, it is not possible to prove that bias has been removed or reduced, for the simple reason that one does not have the missing data. That said, useful evidence is provided by comparing the first-wave respondents with the substitutes. While this is no definitive prove, it provides useful information.

Sample substitution should not be treated as a suspicious survey method but rather as a method that is suitable in specific circumstances, e.g., in case the net sample is pre-defined by the commissioners, the data-collection phase is rather long, the sample frame is ‘rich’ to enable high-quality matching, and response rates are uncertain. A comparison of the demographic characteristics of the initial participants and the substitutes shows very limited differences, while also the health estimates based on the initial participants are not significantly different of those based on the substitutes. It can be argued that a simple increase of the sample size to account for an expected non-participation share would yield in the same results, without the necessity of a complex, time-consuming substitution process. As stated in the introduction of this paper, using inflated sample sizes can be viewed as applying uncontrolled and unmatched substitution that would probably favor the ‘easy to reach’ units. The declining participation rates throughout the substitution process clearly indicate that substitution, as applied in BHIS, at least incorporates some assurance that also ‘hard to reach’ units are actually reached in the survey.

## Limitations

The matching variables of initial households and substitutes are limited to what is available in the sampling frame. In case of the BHIS only the statistical sector, the age of the reference person and the household size served as matching variables. Information on, e.g., socio-economic status and health characteristics which would, in the context of a

health survey, is more fruitful in matching the initial households and substitute households which were not available. The previous studies have shown for instance that lower educated people and less healthy people are more reluctant to participate in the BHIS (Demarest et al. 2013b; Lorant et al. 2007). This can lead to the practice that such people are substituted by more highly educated and healthier people. Nevertheless, the finding that health estimates before and after substitution is very alike provides evidence that the process of expelling less healthy people is very limited.

The BHIS is a survey directed to (selected members of) households. The participation status (participation, refusal, and non-contactable) is attributed to all members, while the decision to participate or not might be taken by just one household member, probably the reference person. An analysis of the consequences of applying substitution at the individual level (level of household members) is as such artificial. Yet, the analysis at the household member participation was found necessary, given that the substitution process in BHIS was applied to assure a prefixed net sample of individuals.

### Strengths

As the substitution process was completely managed by a central secretariat using a specifically developed ICT tool, the whole substitution process could be re-constructed. Every participating household in the BHIS 2013 could be attributed to a specific stage in the substitution process. Such level of detail enabled to check if the substitution process progressed as scheduled (that is, if a non-participating household is effectively substituted by a similar household).

### Conclusion

Although the results of the analysis presented in this manuscript advocate the use of sample substitution as a means to achieve prefixed net sample sizes, further research is necessary to address potential differences in the use of substitution for refusing versus non-contactable households. Indeed, for refusing households—that is, households effectively contacted by an interviewer—effective substitution was legitimate. Yet, for non-contactable households, the requested number of contacts can alter the necessity to apply substitution. Using information on the number of (unfruitful) contact attempts to the analysis will enable a further fine-tuning of the substitution process in the BHIS.

**Acknowledgements** The BHIS is a project conducted in collaboration with Statistics Belgium, responsible for drawing the sample and the fieldwork management.

### Compliance with ethical standards

**Conflict of interest** The authors declare that there is no competing interest.

### References

- Baldissera S, Ferrante G, Quarchioni E, Minardi V, Possenti V, Carrozzini G, Masocco M, Salmaso S, The PASSI Coordinating group (2014) Field substitution of nonresponders can maintain sample size and structure without altering survey estimates. The experience of the Italian behavioral risk factors surveillance system (PASSI). *Ann Epidemiol* 24:241–245
- Chapman DW (1983) The impact of substitution on survey estimates. In: Madow WG, Olkin I, Rubin DB (eds) *Incomplete data in sample surveys*, 2nd edn. Academic, New York, pp 45–61
- Chapman DW (2003) To substitute or not to substitute that is the question. *Surv Stat* 48:32–34
- Chapman DW, Roman AM (1985) An investigation of substitution for an RDD survey. In: *JSM Proceedings, survey research methods section*. American Statistical Association, Alexandria, VA, pp 269–274. [http://www.amstat.org/sections/srms/Proceedings/papers/1985\\_048.pdf](http://www.amstat.org/sections/srms/Proceedings/papers/1985_048.pdf). Accessed 29 Apr 2017
- Chiu WF, Yucel RM, Zanutto E, Zaslavsky AM (2005) Using matched substitutes to improve imputations for geographically linked databases. *Surv Methodol* 31:65–72
- David MC, Ware RS, Alati R, Dower J, Donald M (2014) Assessing bias in a prospective study of diabetes that implemented substitution sampling as a recruitment strategy. *J Clin Epidemiol* 67:715–721
- Demarest S, Gisle L, Van der Heyden J (2007) Playing hard to get: field substitutions in health surveys. *Int J Public Health* 52:188–189
- Demarest S, Van der Heyden J, Charafeddine R, Drieskens S, Gisle L, Tafforeau J (2013a) Methodological basics and evolution of the Belgian Health Interview Survey 1997–2008. *Arch Public Health* 71:24
- Demarest S, Van der Heyden J, Charafeddine R, Tafforeau J, Van Oyen H, Van Hal G (2013b) Socio-economic differences in participation of households in a Belgian national health survey. *Eur J Public Health* 23:981–985
- Dorsett R (2010) Adjusting for nonignorable sample attrition using survey substitutes identified by propensity score matching: an empirical investigation using labour market data. *J Off Stat* 26:105
- Kish L, Hess I (2004) A “replacement” procedure for reducing the bias of nonresponse. *Am Stat* 58:295–297
- Li L, Krenzke T, Mohadjer L (2014) Considerations for selection and release of reserve samples for in-person surveys. *Surv Methodol* 40:105–123
- Lorant V, Demarest S, Miermans PJ, Van Oyen H (2007) Survey error in measuring socio-economic risk factors of health status: a comparison of a survey and a census. *Int J Epidemiol* 36:1292–1299
- Lynn P (2004) The use of substitution in surveys. *Surv Stat* 49:14–16
- Molenberghs G, Kenward M (2007) *Missing data in clinical studies*. Wiley, Hoboken, NJ, p 61
- Nishimura R (2015) *Substitution of nonresponding units in probability sampling*. Diss. University of Maryland
- Pickery J, Carton A (2008) Oversampling in relation to differential regional response rates. *Surv Res Methods* 2:83–92
- Rubin DB, Zanutto E (2002) Using matched substitutes to adjust for nonignorable nonresponse through multiple imputations. In:

- Groves RM, Dillman DA, Little RJA, Eltinge J (eds) Survey nonresponse. Wiley, New York, pp 389–402
- Smith (2007) Notes on the use of substitutions in surveys. Unpublished NORC report, August
- Van der Heyden J, Demarest S, Van Herck K, De Bacquer D, Tafforeau J, Van Oyen H (2014) Association between variables used in the field substitution and post-stratification adjustment in the Belgian health interview survey and non-response. *Int J Public Health* 59:197–206
- Van Oyen H, Tafforeau J, Hermans H, Quataert P, Schiettecatte E, Lebrun L, Bellamammer L (1997) The Belgian health interview survey. *Arch Public Health* 55:1–13
- Vehovar V (1994) Field substitutions-a neglected option. In: Proceedings of the survey research methods section, American Statistical Association, pp 589–594
- Vehovar V (1995) Field substitutions in slovene public opinion survey. In: Ferligoj A, Kramberger A (eds) Contributions to methodology and statistics. Proceedings of the international conference on statistics and methodology, Fakulteta za družbene vede, vol 10. Bled, Slovenia, 13–15 September 1993, pp 39–66
- Vehovar V (1999) Field substitution and unit nonresponse. *J Off Stat* 15:335–350
- Vehovar V (2003) Field substitutions redefined. *Surv Stat* 48:32–34