Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Randomized Comparison of Health Examination Nonparticipation and Questionnaire Nonresponse in the Healthy Finland Survey

Tommi Härkänen

tommi.harkanen@thl.fi

Finnish Institute for Health and Welfare

Tuija Jääskeläinen

Finnish Institute for Health and Welfare

Suvi Parikka

Finnish Institute for Health and Welfare

Tarja Palosaari

Finnish Institute for Health and Welfare

Annamari Lundqvist

Finnish Institute for Health and Welfare

Research Article

Keywords: Participation, Response, Questionnaire survey, Health examination survey, Random forest, Cluster analysis, Inverse probability weights

Posted Date: March 22nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8959646/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Participation rates in health examination surveys (HES), and questionnaire or interview surveys (QS) have declined over time. Although low participation does not bias results *per se*, non-participation is often selective, reducing representativeness even after adjustments on sociodemographic factors. Range of incentives have been tested to improve participation, and health examinations themselves may act as non-monetary, conditional incentives by offering personalized feedback, but direct comparisons between HES and QS remain limited. We aimed 1) to compare participation between individuals invited to the HES versus the questionnaire only in a randomized setting and using register-based characteristics, and 2) to examine how this information affects estimates.

Methods

The Healthy Finland Survey (2022–2023) was based on a stratified, nationally representative QS sample (n=61,600). For the HES, a random subsample was selected from the QS sample, providing a randomized setting to estimate the incentive effect of the health examination. Comprehensive register data were individually linked with the full sample for the missing data analyses.

Inverse probability weights (IPW) were based on pre-2022 register data and supplemented with diagnoses recorded during and after the survey, as symptoms preceding diagnosis may influence participation. Logistic regression models and non-parametric, machine learning methods (e.g. random forests) were used to identify predictors and estimate IPWs. Clustering analysis was applied to assess differences of the estimated participation probabilities based on the parametric main-effects and random forest models.

Results

Participation rates were substantially higher in the HES sample compared to the QS-only sample among individuals aged 20–75 years, with a difference of 15–20 percentage points under age 50. Based on the Bayesian information criteria, Age*HES, Age*Sex, and Sex*Married interactions were included in the QS weighting model. Random forest variable selection identified a relatively small set of influential predictors. Age-standardized prevalence estimates of selected questionnaire variables did not differ between HES and QS participants.

Conclusion

Health examinations markedly increased participation among working-age individuals compared to questionnaire alone. However, prevalence estimates based on questionnaire variables were similar in the HES and QS only groups, suggesting that lower QS participation did not compromise validity.

Trial registration

Clinical trial number: not applicable.

Background

Declining participation rates have been consistently observed in both health examination surveys (HES), [1, 2] and questionnaire or interview surveys (QS).[2, 3] Even with low participation rates, results remain unbiased if missingness is completely at random. In reality, however, survey non-participation has often been found selective,[4, 5] compromising the representativeness of survey-based estimates even after adjusting the estimates by sociodemographic information.[6] It has been shown that additional contacts and reminders enhance higher participation rates.[7] Furthermore, various incentives[8] – conditional or unconditional, monetary[9, 10] or non-monetary, have been proposed to improve participation, and they have been compared in randomized studies; the meta-analysis of Abdelazeem et al.[9] reported that money incentives increased participation by 25%, especially, if paid in advance in a cross-sectional study.[11] Health examination themselves may improve participation acting as non-monetary incentives by providing personalized feedback to the participants. While numerous studies have examined temporal trends, incentives or survey modes, comparisons have typically been restricted to either HES or QS, leaving direct comparisons between these modes largely unexplored.

Missing data mechanisms likely differ by survey mode.[12] QSs typically suffer from lower participation rates compared to HESs. HESs may attract participants by offering personalized health information (e.g. different laboratory and physical measurements). However, HESs impose greater time and mobility demands, potentially reducing participation rates among individuals with long distances from home to the examination clinic or limited availability or physical constraints, in addition to higher costs.[13]

Register-based information on sociodemographic factors such as age, sex, marital status and education along with health service use are established predictors of participation.[14, 15] Identifying predictors of participation in the HES, but not in QS – or vice versa – can enhance overall survey quality. Nonparticipants have been shown to exhibit higher total and cause-specific mortality, suggesting that health-related factors strongly predict participation. Thus, a deeper understanding of the mechanisms and effects of nonparticipation remains crucial.[16] Consequently, incorporating hospitalization and medication data may improve models of nonparticipation. However, timely access to such data is often limited due to delays in register updates. Extending follow-up periods to include events during and after the survey may strengthen these models and provide insight to assess the importance of contemporary health register data beyond the historical data available before the sampling date.

Nonparticipation analyses often rely on parametric models, such as logistic regression, which assume linearity or a limited set of interactions – assumptions that become restrictive as the number of predictors grows. Nonparametric approaches, such as random forest methods[17] offer higher predictive accuracy especially if a larger number of predictors are available, and automatically capture nonlinearities and interactions, making them suitable for estimating participation probabilities in inverse probability weighting (IPW) methods.[18] For high-dimensional problems, several variations have been

proposed in random forest-based variable selection with different pros and cons, for example, in terms of computing time, error rate and parsimony.[19]

While random forests outperform parametric models in flexibility, they lack interpretability regarding predictor-outcome relationships. Comparing predictions from simple regression models (e.g., main effects only) with those from random forests can highlight their differences due to implicit interactions and nonlinearities, particularly by examining subsets where prediction differences are largest.

To unify the terminology, we use the term participation throughout the article, although the term response is used in questionnaire surveys.

Using the randomized allocation of individuals to health-examination (HES) versus questionnaire-only (QS) invitations in the Healthy Finland survey, we examine how participation and its determinants differ between these groups to better understand missing-data mechanisms. Specifically, we compare participation rates, evaluate the added value of follow-up data, construct inverse-probability weights based on parametric and nonparametric methods, identify subsets where the parametric models fit poorly, and assess whether questionnaire-based variables differ between HES and QS participants.

Methods

The Healthy Finland Survey was conducted in the continental Finland in 2022–23. All participants were first given a questionnaire to fill in. A random subset of the sample was invited to the health examination including measurements, blood sampling and additional questionnaires. Survey data were individually linked to national registry data. The methods for collecting the data are described in greater detail elsewhere.[20]

Sampling design

Nationally representative QS sample was drawn from 22 wellbeing services counties (WSCs) from the population aged 20 years or over. From each WSC, 2,000 individuals aged 20–74 years and 800 individuals aged 75 years or older were selected using stratified random sampling. In total, 61,600 individuals were randomly selected into the QS sample.

The HES sampling design was based on previous HES studies, the Health 2000[21] (H2000) and the FinHealth 2017 (FH2017) study[22]. The HES sample was selected as a random subsample from the QS sample from the 50 health center districts (HCDs), identical to those used in the FH2017.[22] These districts were, due to budget constraints, a subset of the 80 HCDs selected in the H2000 Survey,[21] in which the 15 largest cities were selected with probability one, thus forming 15 strata, and other 65 HCDs were selected using the probability proportional to size (PPS) sampling from five rural strata, defined by the five university hospital regions (excluding the 15 largest cities), based on the population sizes in 2000. In the FH2017 survey, the number of districts was reduced to seven per rural stratum.

For the HES, sample sizes were proportional to the population sizes of the 20 strata. In the rural strata the sample sizes were adjusted so that the individual sampling probabilities were equal after accounting for the original PPS probabilities of the HCDs in year 2000. Individuals were then sampled from the QS sample using random sampling from the selected HCDs. Due to small population sizes in some HCDs, an additional sample (n = 1,586) was selected from these HCDs. In total, 9,973 QS and additional sample members were randomly selected into the HES sample.

Variables

Analysis Subsets

The outcomes in these analyses were participation indicators with value one, if an individual participated in the survey, and zero, if not. The primary participation indicators were defined as the overall participation in the QS, and for the HES the overall participation in the examination or in the questionnaire. Three subsets of the full sample were defined:

- D1: Full QS sample (n = 61,600)
- D2: HES sample (n = 9,973)
- D3: QS sample with the additional sample (n = 61,600 + 1,586), restricted to the 50 HES areas (n = 31,275)

D1 was used to calculate weights for QS data and D2 for HES data. D3 was used in this work for comparing participation between the HES sample and the QS sample excluding the HES sample (henceforth “QS-only”) within the same geographic areas.

Register data were linked with the full sample, including the nonparticipants, using the individual identity codes. Demographic variables (age, sex, address, marital status and language) were obtained from the Digital and Population Data Services Agency (DVV)[23] along with the sample selection. Education,[24] profession[25] and socioeconomic position[26] were obtained from Statistics Finland. Hospitalization data (event dates and ICD-10 codes of the diagnoses) were retrieved from the Care Register for Health Care.[27] At the time of the survey, the register information on hospitalization was available from 1996 to 2021, data on education through 2021, and for profession and socioeconomic status through 2020.

Subsequent updates provided extended register data, including also medication information[28] from the Social Insurance Institution, available through 2023. These data were used in the analyses of D3.

To compare different statistical methods to handle missing data, we selected a variety of survey variables from the questionnaire. From the questionnaire we used the following binary variables:

current daily smoker; heavy alcohol risk consumption was evaluated with Alcohol Use Disorder Identification Test-Concise (AUDIT-C) score[29, 30] (scores of ≥ 6 for men, ≥ 5 for women and ≥ 4 for those aged 65 or more were used to indicate alcohol risk consumption); feeling lonely (sometimes or more frequently); obesity (based on self-reported BMI ≥ 30 kg/m2); quality of life (EuroHIS-8,[31] the

mean of eight items was calculated, with higher scores indicating better quality of life; physical activity below recommendations (based on intensity, duration, and muscle engagement during weekly exercises) and difficulties in walking 500 m.

Statistical methods

All analyses were conducted using the R statistical software.[32]

Inverse Probability Weighting for D1

The inverse probability (IP) weights[18] were calculated using a logistic regression model for D1. Predictors included age, gender, stratum, married, language, education, urbanization and the HES indicator. The selection of predictors was based on the previous Finnish questionnaire surveys, the Regional Health and Well-being Study (ATH)[14] and FinSote QS, thus other predictors were not included. First and second order interactions were assessed individually using the Bayesian information criterion (BIC).[33]

Extended Register Data for D3

Original IP weights were based on pre-2022 register data. To assess the impact of additional health information, we incorporated diagnoses recorded before, during and after the survey, as symptoms preceding diagnosis may influence participation. Two categorical variables were created: time of last event before sampling (“pre”) and first event after sampling (“post”), distinguishing prevalent from incident cases. In addition to health categories cardiovascular, mental and infectious diseases, birth events and injuries, also musculoskeletal, rheumatologic, cancer, and diabetes were applied.

Variable Selection for D3

For D3, predictor importance was first assessed using a stepwise main-effects model with an age–sex interaction, selected via BIC. Due to the large number of predictors, a grid search for interactions was infeasible; instead, recursive feature elimination (RFE) with random forests[17] and 10-fold cross-validation was applied. We also used the VSURF[34] and Boruta[35] methods; VSURF identifies a parsimonious model with near-optimal predictive power, addressing the minimum-optimal problem, while Boruta identifies all relevant predictors using Z-scores of variable importance, addressing the all-relevant problem.

Participation Probability Estimation for D3

Participation probabilities were estimated using main-effects logistic regression (including the interaction of age and sex) and random forests. For random forests, probabilities were computed as the observed participation rate within each terminal node across 500 trees, approximating expected values and avoiding zero-probability issues.

Clustering Analysis

Differences between logistic and random forest predictions indicated potential interactions and nonlinearities were explored by partitioning D3 into three groups: (i) logistic main effects model predicted ≥ 0.15 lower probability than random forest, (ii) logistic main effects model predicted ≥ 0.15 higher probability, and (iii) minimal difference (“adequate fit”).

Groups with large participation probability discrepancies cannot be detected using marginal distributions of the predictors alone. Therefore, clustering was applied separately within groups (i) and (ii) using the general similarity coefficient[36] for categorical predictors to calculate pairwise dissimilarities (distances) between observations in the data set. The partitioning around medoids (PAM) algorithm,[37] implemented in the cluster[38] package was used to identify representative individuals (“medoids”) for each cluster. Optimal cluster number was selected using the silhouette width, which measures how similar an observation is to its own cluster compared its closest neighbouring cluster, across 2–11 clusters, supplemented by confidence intervals of their differences estimated via subsampling bootstrap[39, 40] (subsample size 1,000 without replacement). Medoids and the corresponding cluster sizes were reported to illustrate individuals with the greatest prediction discrepancy.

Health examination survey (D2)

The statistical methods and the results for the HES sample (D2) are presented in the supplementary material (Additional file 1).

Use of Artificial intelligence (AI) tools

Microsoft 365 Copilot app was applied in improving the readability and English grammar.

Results

Descriptive statistics

Participation rates in the HES sample was higher than in the QS-only sample across ages 20–75, with maximum 73% around age 70 (Fig. 1). The difference was 15–20%-units under age 50, about 10%-units between 60 and 70 years and after age 80 participation rates were converged and decreased rapidly from over 60% to 25%. In the HES sample, participation rates in the health examination and in the questionnaire parts were almost equal whereas participation in at least one of these were about 10%pt higher, and this difference remained almost constant for all ages; participation in at least one of these exceeded 80% around age 70 years. Similar differences were also observed in participating in the questionnaire between men and women; among the youngest, participation rates were about 15%pt higher among women than men, and the difference gradually decreased, vanishing after age 70 years.

Participation rates in different groups (D1)

The overall age and sex adjusted participation rate for the questionnaire were 15 percentage points (%pt) higher in the HES sample (58.6%) than in the QS-only sample (43.7%). Furthermore, the crude difference between the QS-only questionnaire participation rate (44.1%) and the HES health examination or questionnaire participation rate (67.6%) was as large as 23.5%pt (Table 1).

Table 1
Crude and adjusted participation rates (%) for the questionnaire in the QS sample.

		QS		
		Questionnaire participation (%)		
		n	Crude	Adj
(Intercept)		61,600	45.7	
Age	20–24	3,256	26.1	26.4
	25–34	7,580	28.2	28.4
	35–44	7,945	31.9	32.2
	45–54	7,692	37.8	38.0
	55–64	8,857	51.4	51.6
	65–74	8,670	63.3	63.3
	75–84	12,992	61.1	60.8
	85–94	4,329	38.7	37.7
	95+	279	26.9	25.2
Sex	Men	29,625	41.6	42.1
	Women	31,975	49.5	49.1
WSC	(See Additional file 1: Supplementary Table 1)			
Marriage	Not married	33,566	39.2	41.3
	Married	28,034	53.5	50.8
Language	fi	55,454	46.8	46.4
	sv	3,504	45.5	45.1
	other	2,642	23.8	30.9
Education	low	15,919	38.7	33.4
	middle	26,184	41.4	43.9
	high	19,497	57.2	58.5

WSC: Welfare services county

Crude, and age and sex adjusted participation rates (%) for the questionnaire in the QS sample. The number of sample members (n) in the questionnaire survey are presented.

				QS		
				Questionnaire participation (%)		
Urban-rural	Urban			41,160	46.0	47.0
(municipality)	Dense rural			9,810	43.8	42.6
	Sparse rural			10,630	46.3	43.8
Sample	QS only			53,213	44.1	43.7
	HES			8,387	55.8	58.6
Age	20–24	**Sex**	Men	1,716	18.8	23.3
	20–24		Women	1,540	34.3	29.2
	25–34		Men	3,919	21.2	25.2
	25–34		Women	3,661	35.7	31.4
	35–44		Men	4,148	26.8	28.7
	35–44		Women	3,797	37.5	35.3
	45–54		Men	3,920	32.3	34.3
	45–54		Women	3,772	43.4	41.4
	55–64		Men	4,466	46.6	47.6
	55–64		Women	4,391	56.3	55.2
	65–74		Men	4,099	60.1	59.6
	65–74		Women	4,571	66.2	66.7
	75–84		Men	5,818	61.9	57.1
	75–84		Women	7,174	60.4	64.3
	85–94		Men	1,486	42.9	34.0
	85–94		Women	2,843	36.5	41.1
	95+		Men	53	30.2	22.2
	95+		Women	226	26.1	28.0
Sex	Men	**Marriage**	Not married	15,288	31.8	37.4

WSC: Welfare services county

Crude, and age and sex adjusted participation rates (%) for the questionnaire in the QS sample. The number of sample members (n) in the questionnaire survey are presented.

				QS		
				Questionnaire participation (%)		
	Men		Married	14,337	52.1	46.8
	Women		Not married	18,278	45.4	44.9
	Women		Married	13,697	55.0	54.5
Age	20–24	**Sample**	QS only	2,636	23.0	23.7
	20–24		HES	620	39.2	37.6
	25–34		QS only	6,189	24.4	25.8
	25–34		HES	1,391	45.1	40.2
	35–44		QS only	6,626	28.1	29.6
	35–44		HES	1,319	50.9	44.9
	45–54		QS only	6,521	34.6	35.5
	45–54		HES	1,171	55.2	51.6
	55–64		QS only	7,530	49.3	49.2
	55–64		HES	1,327	63.5	65.2
	65–74		QS only	7,415	62.2	61.2
	65–74		HES	1,255	70.4	75.4
	75–84		QS only	12,030	60.6	59.8
	75–84		HES	962	66.6	74.2
	85–94		QS only	4,007	38.7	36.5
	85–94		HES	322	38.2	52.7
	95+		QS only	259	26.3	24.2
	95+		HES	20	35.0	38.2

WSC: Welfare services county

Crude, and age and sex adjusted participation rates (%) for the questionnaire in the QS sample. The number of sample members (n) in the questionnaire survey are presented.

The crude participation rate was 8%pt higher among women than among men, with the largest difference (approximately 15%pt) observed among the youngest age groups. Not being married decreased the age and sex adjusted participation rates almost 10%pt. The adjustment increased

participation among non-Finnish and non-Swedish speakers from 23.8% to 30.9%, reflecting their lower mean age. Similarly, the adjustment increased the difference between the low and high education groups from 19% to 25%. While crude participation rates were similar in urban and sparsely populated rural municipalities (both 46%), adjustment indicated a 3.2%pt lower participation rate in the sparsely populated rural municipalities. Considerable regional variation was observed, with the highest adjusted participation rate in Helsinki (50.7%), and the lowest in South Ostrobothnia (40.7%) and in Ostrobothnia (41.0%) (Additional file 1: Supplementary Table 1).

Table 1 **approximately here.**

Model selection (D1)

In addition to the main effects, the first-order interactions Age*Sex, Sex*Married and Age*HES were included in the IP weighting model for the QS (Additional file 1: Supplementary Table 2).

Logistic regression model estimates (D1)

Participation rates in the QS were higher among the youngest sample members, who were invited in the HES (OR = 2.10, Additional file 1: Supplementary Table 3); according to the interaction of age and HES, this difference remained large until age 64. Participation was associated with age as in Fig. 1; the sex difference was largest among the youngest individual (OR = 2.06), and it diminished after age 75 years, according to the interaction estimates of age and sex. Participation was higher among those who were married among men (OR = 1.59), and slightly smaller among women (OR = 0.80). Participation was higher also among those who were Finnish or Swedish speaking, had highest education, lived in urban or sparsely populated rural areas, or were selected in the HES sample. Participation differences between the HES and QS only samples diminished after age 75 years. The highest participation rates were in the Pirkanmaa WSC (OR = 1.18), and the lowest in the South Ostrobothnia WSC (OR = 0.78).

Model selection using the QS sample in the HES areas (D3)

Boruta and random forest with 10-fold cross-validation included almost all variables in the model (27 and 28 out of 30 variables, respectively), whereas the VSURF method using the predictions in model selection included only five variables (Table 2). VSURF based on the interpretation included seven variables, and RFE 16 variables. On the other hand, the RFE method included many of the additional health event variables based on use of medications or extended follow-up of health events variables.

Table 2
Register variables selected using the random forest methods.

Variable	Boruta[1]	VSURF using interpretation[2]	VSURF using prediction[2]	RFE[3]	RF CV[4]	Sum
Sex	Yes				Yes	**2**
Age	Yes	Yes	Yes	Yes	Yes	**5**
Sample	Yes	Yes	Yes	Yes	Yes	**5**
Profession	Yes	Yes			Yes	**3**
Education	Yes	Yes	Yes	Yes	Yes	**5**
Urban-rural (municipality)						**0**
Urban-rural (squares)	Yes				Yes	**2**
Socio-economics	Yes	Yes	Yes	Yes	Yes	**5**
Language	Yes				Yes	**2**
Marital status	Yes	Yes		Yes	Yes	**4**
Musculoskeletal pre	Yes			Yes	Yes	**3**
Musculoskeletal post	Yes			Yes	Yes	**3**
Arthropathies pre	Yes			Yes	Yes	**3**
Arthropathies post	Yes				Yes	**2**
Rheuma-related pre	Yes			Yes	Yes	**3**
Rheuma-related post	Yes				Yes	**2**
Diabetes pre	Yes				Yes	**2**
Diabetes post	Yes				Yes	**2**
Cancer pre	Yes				Yes	**2**
Cancer post	Yes				Yes	**2**

Footnotes:

[1] Boruta

[2] VSURF using interpretation or prediction

[3] Recursive feature elimination (RFE)

[4] Random forest, using 10-fold cross-validation

Variable	Boruta[1]	VSURF using interpretation[2]	VSURF using prediction[2]	RFE[3]	RF CV[4]	Sum
CVD pre	Yes			Yes	Yes	**3**
CVD post	Yes			Yes	Yes	**3**
Mental pre	Yes	Yes	Yes	Yes	Yes	**5**
Mental post	Yes			Yes	Yes	**3**
Infections pre	Yes			Yes	Yes	**3**
Infections post	Yes				Yes	**2**
Birth-related pre	Yes			Yes	Yes	**3**
Birth-related post						**0**
Injuries pre	Yes			Yes	Yes	**3**
Injuries post					Yes	**1**
Sum	**27**	**7**	**5**	**16**	**28**	
Footnotes:						
[1] Boruta						
[2] VSURF using interpretation or prediction						
[3] Recursive feature elimination (RFE)						
[4] Random forest, using 10-fold cross-validation						

Table 2 **approximately here.**

Optimal number of clusters was four for those, where the participation rates were better than indicated by the main effects logistic regression model (“main effects smaller”), but in the opposite case (“main effects larger”) we selected two clusters (Additional file 1: Supplementary Fig. 1).

The largest differences between the main effects logistic regression model and a random forest occurred for medoids (representative individuals of the clusters), who were in the QS only sample and lived in urban areas. The working-age medoid did not have disease events. The logistic regression model underestimated the participation probabilities among 1344 individuals; of the four medoids three were women (a 31-year old, unmarried, technician or associate professional with high education and without diseases; the other two were 79 and 86 year old, retired, married or widowed, and with several diseases) and one was a retired man, aged 63 years (Additional file 1: Supplementary Table 4).

The logistic regression model overestimated the participation probabilities of retired individuals with several diseases (a 76-year-old married man with high education, and a 77-year-old divorced woman with middle education).

Based on the random forest model, the adjusted participation rates in the HES sample were almost 60% among individuals with no hospitalization prior to the sampling date, whereas it was approximately 45% among those with a recent hospitalization (Fig. 2). In the QS only sample, the lowest participation rates were observed among those hospitalized 1–6 years prior to the sampling date.

Models for the inverse probability weights (IPW) for the QS sample (D1)

Participation varied more among men than among women, with the highest participation observed at around age 70 for both sexes (Additional file 1: Supplementary Fig. 2). For other risk factors, participation was highest among Finnish speakers; married individuals and widows; profession; those with high education; students; and residents of rural areas.

The OR estimates based on the QS OR estimates (Additional file 1: Supplementary Table 5) indicated higher participation rates among individuals, who were women, Finnish or Swedish speaking, married or widowed, had high education, were students or had had no hospitalization events in the past. Low participation rates were among working individuals, who were craft and related trades workers, had unknown working status, or lived in sparsely populated rural areas in municipalities, which were categorized as urban or densely populated rural municipalities.

Estimates based on different weights (D1)

All weights for the QS sample produced similar estimates (Additional File 2). For the QS sample, the unweighted estimates differed in several variables, such as in difficulties in walking 500 meters, where the oversampling of the 75 –year-olds or older was 11.2% compared to the weighted estimates, which ranged between 6.9% and 7.5%. The prevalence estimates of daily smokers ranged between 9.7% and 9.9%; the unweighted QS estimate was lower as the prevalence of smokers was low (3.1%) among the 75 year old or older. Similar findings were found for several other variables. Design effects were large in the QS sample, as the sampling probabilities varied between both WSCs and age groups.

Questionnaire response differences between HES and QS-only (D1)

Age-standardized prevalences of the survey variables among the participants did not differ between the HES and QS only samples, although in the crude prevalences there were some differences due to oversampling among the 75 years old or older in the QS sample (Table 3).

Table 3
prevalences (%) of questionnaire variables in the health examination and questionnaire only samples.

		QS only			HES			Difference		
	[1]	Mean	95% CI		Mean	95% CI		Mean	95% CI	
Alcohol risk consumption (AUDIT-C) [2]	Std	29.1	28.0	30.2	30.7	29.3	32.2	1.61	-0.23	3.45
	All	27.9	26.9	28.9	31.2	29.7	32.7	3.32	1.55	5.08
Daily smoker	Std	8.98	8.29	9.67	8.74	7.90	9.59	-0.23	-1.33	0.86
	All	7.28	6.76	7.80	8.53	7.71	9.35	1.25	0.28	2.21
Difficulties in walking 500 m	Std	5.89	5.49	6.29	5.94	5.33	6.55	0.04	-0.69	0.77
	All	11.3	10.7	11.9	6.86	6.13	7.59	-4.42	-5.38	-3.46
Feeling lonely	Std	11.6	10.9	12.4	10.6	9.72	11.5	-0.97	-2.15	0.21
	All	11.2	10.6	11.8	10.3	9.41	11.2	-0.89	-1.97	0.18
Obesity (questionnaire) [2]	Std	22.9	21.9	23.9	22.9	21.6	24.1	-0.02	-1.58	1.55
	All	21.4	20.6	22.2	23.0	21.8	24.2	1.65	0.19	3.12
Physical activity below recommendation[3]	Std	58.1	56.9	59.3	56.2	54.6	57.7	-1.93	-3.88	0.01
	All	61.7	60.7	62.8	56.6	55.1	58.1	-5.11	-6.96	-3.26
Good quality of life (EuroHIS-8) [4]	Std	53.4	52.2	54.5	53.8	52.4	55.3	0.46	-1.38	2.30
	All	52.5	51.6	53.5	53.9	52.5	55.3	1.35	-0.38	3.09
Footnotes:										

[1] Crude (“all”) and directly standardized (“std”) prevalences (%) and 95% confidence intervals of questionnaire variables in the health examination and questionnaire only samples.

[2] Scores of ≥ 6 for men, ≥ 5 for women and ≥ 4 for those aged 65 or more were used to indicate alcohol risk consumption.

[2] Self-reported BMI ≥ 30 kg/m2.

[3] Based on intensity, duration, and muscle engagement during weekly exercises.

[4] The mean of eight items was calculated, with higher scores indicating better quality of life.

Discussion

We observed that participation rate for the questionnaire was approximately 20 percentage points (%pt) higher in the health examination sample compared to the questionnaire-only sample among individuals under 50 years of age; this difference diminished after age 75, suggesting that inclusion of a health examination substantially improve survey participation. Overall, health examination invitations increased participation by about 10%pt, indicating that the health examination served as a strong incentive. Other predictors of participation included several socio-demographic variables and absence of hospitalization for mental disorders. One of the highest and the lowest participation rates occurred in neighbouring areas (Pirkanmaa and South Ostrobothnia).

Analysis of contemporary and subsequent health records revealed that musculoskeletal conditions, cardiovascular disease, infections, birth-related events and injuries were also important predictors of nonresponse, as identified through recursive feature elimination. On the other hand, when the predictors were assessed simultaneously by the VSURF methods, the variables selected were included in the original HES IP weighting model, indicating that the additional health event variables based on use of medications or extended follow-up of health events after the sampling date were not important.

Traditional regression models generally performed well in predicting participation probabilities, particularly when the number of predictors was small.

Clustering results indicated subsets of the sample where a main effects model was insufficient, as the random forest model captured additional interactions and nonlinearities, yielding lower or higher participation probabilities in these subsets.

Age-adjusted prevalences based on selected questionnaire variables did not differ between the HES and QS only samples. This suggests, despite the differing participation rates, survey mode did not bias prevalence estimates.

Our findings are consistent with earlier Finnish HES studies (the Health 2000[21], Health 2011[41] surveys, and the FinHealth 2017 study[22]), which reported higher participation rates for any part of the survey than for examinations or questionnaires alone. In these surveys, the same individuals were invited both in the health examination and in the questionnaire parts. In the contemporary Finnish QSs without a health examination (the Regional Health and Well-being Survey 2010–2016[14] and the National FinSote survey 2017–2020) reported considerably lower participation rates.

The meta-analysis of Cornesse et al.[42] supports the use of mixed mode designs to reduce bias, reinforcing the value of combining health examination and questionnaire samples. Surveys based on a nonrandom samples have been generally found less valid.[43]

Several predictors identified in our study align with previous literature. For example, Markkula et al. found that the use of register data on mental health diagnoses increased the prevalence estimates of major depressive disorder and dysthymia.[44] The Health 2011 Survey, which re-invited eligible Health 2000 sample members, observed a marked decline in participation; baseline data[15] revealed that greater social activity and better perceived health predicted continued participation.

Our comparison of statistical methods showed that parametric regression models performed well relative to the random forest method when few predictors were included, suggesting limited complex interactions or nonlinearities. However, in the HES sample with extensive register-based predictors the random forest model outperformed the parametric model, indicating the presence of higher-order interactions.

Cluster analysis highlighted representative individuals (medoids), primarily from the QS only sample and urban areas, where model discrepancies were greatest.

Importantly, the prevalence estimates based on the questionnaire variables did not differ between the HES and QS only samples, suggesting that the nonparticipation bias was not driven by the health examination incentives. This is reassuring for questionnaire surveys, as the lower participation rates did not seem to compromise validity. As the budgets decline and costs rise, questionnaire surveys are attractive as they provide an inexpensive way to collect data compared to interviews or health examinations. Nonetheless, participation was particularly low among younger males, underscoring the need for targeted incentives.

Health examination surveys remain essential for collecting reliable physiological measurements (e.g., blood samples, anthropometrics) that cannot be obtained through questionnaires. The personal feedback based on the health examination is an important incentive to increase participation rates.

Strengths and limitations

Our study has several strengths. Randomized design minimized confounding, enabling valid comparisons between health examination and questionnaire-only groups. Wide range of register data, including the Population Register allowing a representative sample of the population and comprehensive health-related data, were linked with the full sample; the Finnish registers have also been found good both in coverage and quality.[45] Participation rates in Finnish surveys remain relatively high, and in the HES sample the participation rates were good also among the younger age groups.

Limitations include that our results do not reveal detailed information on how the individuals, who participated because they were invited in the health examination, differed from those who participated in any case. It is also possible, that there are individuals, who participated in the questionnaire survey because they were not invited in the health examination. Individuals in the first group might have more diseases or they could be more interested in their health. Individuals in the latter group could be more

diverse, as they are likely to have a positive attitude to the population survey, but they might not have time or need for the health examination.

Implications of findings

Questionnaire surveys are less costly than health examinations, but our findings suggest that the higher cost of examinations may be offset by improved participation and data reliability, especially among young males. Alternative incentives to boost participation remain challenging to identify.

Conclusion

Participation rates were considerably higher among the working-age individuals invited to the health examination compared to the questionnaire only sample. However, prevalence estimates based on questionnaire data did not differ between groups, indicating that lower participation rates in the questionnaire survey did not compromise validity; if the participation rates decrease more, there can be a breakpoint after which the participants of a QS do not allow representative results of the population even after applying statistical methods to correct for the effects of nonparticipation, in which case higher participation rates in HESs can provide more reliable results. Further research is needed to clarify the determinants of participation or nonparticipation.

Abbreviations

BIC
Bayesian Information Criterion
DVV
Digital and Population Data Services Agency
D1
Full QS sample
D2
HES sample
D3
QS sample with the additional sample, restricted to the 50 HES areas
FH2017
The FinHealth 2017 Study
HCD
Health Center District
HES
Health examination survey
H2000
The Health 2000 Survey
PAM

Partitioning around medoids
PPS
Probability proportional to size
QS
Questionnaire survey
RFE
Recursive feature elimination
WSC
Welfare services county

Declarations

Ethics approval and consent to participate

For the questionnaire part of the survey, the ethical review was conducted by the Institutional Review Board (IRB) of THL.[20] Filling in the questionnaire was considered to indicate consent. For the health examination part of the survey, the survey was reviewed by the Helsinki and Uusimaa Hospital District Regional Committee on Medical Research Ethics (decision number HUS/900/2022). Written informed consent was obtained from each individual participating in the health examination.

Consent for publication

Not applicable.

Funding

The authors received no specific funding for this work.

Author Contribution

T.H., T.J., S.P., T.P. and A.L. (project administration) conceptualized the research question. All authors contributed to the study design. T.H. selected the statistical methodology, implemented it in statistical software, conducted the formal (statistical) analyses and visualization, and wrote (original draft) the manuscript. All authors participated in interpretation of the results and writing (review & editing) the manuscript. All authors have read and approved the final version.

Data Availability

The Healthy Finland data are not publicly available as they include confidential information that could compromise the privacy of the participants. The data can be used for research and monitoring of health,

wellbeing, functioning, and use of services of the population at THL, and with collaborators based on collaboration agreement. The data available from the THL Biobank cover those who have participated in the health examination and have given consent to biobanking, and can be applied in accordance with the Biobank Act and THL Biobank research areas [[www.thl.fi/biobank](https:/www.thl.fi/biobank)]. The Finnish Social and Health Data Permit Authority, Findata, may grant permits in accordance with the act on the secondary use of social and health data in Finland [[www.findata.fi/en](https:/www.findata.fi/en)]. Further enquiries can be directed to the email address [tervesuomi@thl.fi].

References

1. Tolonen H, Ahonen S, Jentoft S, Kuulasmaa K, Heldal J, for the European Health Examination Pilot Project. Differences in participation rates and lessons learned about recruitment of participants – The European Health Examination Survey Pilot Project. Scand J Public Health. 2015;43:212–9. https://doi.org/10.1177/1403494814565692.
2. NHANES Response Rates and Population Totals.
3. Jabkowski P, Cichocki P. Survey response rates in European comparative surveys: a 20-year decline irrespective of sampling frames or survey modes. Qual Quant. 2025;59:635–55. https://doi.org/10.1007/s11135-024-01993-9.
4. Ekholm O, Gundgaard J, Rasmussen NKR, Hansen EH. The effect of health, socio-economic position, and mode of data collection on non-response in health interview surveys. Scand J Public Health. 2010;38:699–706. https://doi.org/10.1177/1403494810382474.
5. Reinikainen J, Tolonen H, Borodulin K, Härkänen T, Jousilahti P, Karvanen J, et al. Participation rates by educational levels have diverged during 25 years in Finnish health examination surveys. Eur J Pub Health. 2018;28:237–43. https://doi.org/10.1093/eurpub/ckx151.
6. Kelfve S, Thorslund M, Lennartsson C. Sampling and non-response bias on health-outcomes in surveys of the oldest old. Eur J Ageing. 2013;10:237–45. https://doi.org/10.1007/s10433-013-0275-7.
7. Jezek AH, Ekholm O, Thygesen LC, Christensen AI. The impact of reminders on representativeness and survey estimates among web-mode invited in the Danish National Health Survey. Eur J Pub Health. 2025;35:256–62. https://doi.org/10.1093/eurpub/ckae176.
8. Lipps O, Herzing JME, Pekari N, Ernst-Staehli M, Pollien A, Riedo G, et al. Incentives in surveys. FORS Guide N°8. 2019. https://doi.org/10.24449/FG-2019-00008. :Version 1.1.
9. Abdelazeem B, Hamdallah A, Rizk MA, Abbas KS, El-Shahat NA, Manasrah N, et al. Does usage of monetary incentive impact the involvement in surveys? A systematic review and meta-analysis of 46 randomized controlled trials. PLoS ONE. 2023;18:e0279128. https://doi.org/10.1371/journal.pone.0279128.
10. Smith MG, Witte M, Rocha S, Basner M. Effectiveness of incentives and follow-up on increasing survey response rates and participation in field studies. BMC Med Res Methodol. 2019;19:230. https://doi.org/10.1186/s12874-019-0868-8.

11. Décieux JP, Zinn S, Andreas E. Effects of Changing the Incentive Strategy on Panel Performance: Experimental Evidence From a Probability-Based Online Panel of Refugees. Surv Res Methods. 2025;19:223–39. https://doi.org/10.18148/SRM/2025.V19I2.8437.
12. Cheung KL, ten Klooster PM, Smit C, de Vries H, Pieterse ME. The impact of non-response bias due to sampling in public health studies: A comparison of voluntary versus mandatory recruitment in a Dutch national survey on adolescent health. BMC Public Health. 2017;17:276. https://doi.org/10.1186/s12889-017-4189-8.
13. Reinikainen J, Saarsalmi P, Härkänen T, Jousilahti P, Karvanen J, Männistö S, et al. Non-participation modestly increased with distance to the examination clinic among adults in Finnish health examination surveys. Scand J Public Health. 2018;46:752–4. https://doi.org/10.1177/1403494817739502.
14. Härkänen T, Kaikkonen R, Virtala E, Koskinen S. Inverse probability weighting and doubly robust methods in correcting the effects of non-response in the reimbursed medication and self-reported turnout estimates in the ATH survey. BMC Public Health. 2014;14:1150. https://doi.org/10.1186/1471-2458-14-1150.
15. Härkänen T, Karvanen J, Tolonen H, Lehtonen R, Djerf K, Juntunen T, et al. Systematic handling of missing data in complex study designs–experiences from the Health 2000 and 2011 Surveys. J Applied Statistics. 2016;43:2772–90. https://doi.org/10.1080/02664763.2016.1144725.
16. Jousilahti P, Salomaa V, Kuulasmaa K, Niemela M, Vartiainen E. Total and cause specific mortality among participants and non-participants of population based health surveys: a comprehensive follow up of 54 372 Finnish men and women. JEpidemiolCommunity Health. 2005;59:310–5. https://doi.org/10.1136/jech.2004.024349.
17. Breiman L, Random, Forests. Mach Learn. 2001;45:5–32. https://doi.org/10.1023/A:1010933404324.
18. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. J Am Stat Assoc. 1994;89:846–66.
19. Speiser JL, Miller ME, Tooze J, Ip E. A comparison of random forest variable selection methods for classification prediction modeling. Expert Syst Appl. 2019;134:93–101. https://doi.org/10.1016/j.eswa.2019.05.028.
20. Sääksjärvi K, Elonheimo HM, Ikonen J, Lehtoranta L, Parikka S, Härkänen T, et al. Cohort Profile: Healthy Finland Survey. Int J Epidemiol. 2024;54:dyae166. https://doi.org/10.1093/ije/dyae166.
21. Heistaro S. Methodology report: Health 2000 Survey. Helsinki: National Public Health Institute; 2008.
22. Borodulin K, Sääksjärvi K. FinHealth 2017 Study: Methods. THL; 2019.
23. Digital and Population Data Services Agency. Customer service for organisations. Digi- ja väestötietovirasto. https://dvv.fi/en/organisations. Accessed 1 Dec 2025.
24. Statistics Finland. Educational structure of population | Statistics Finland. 2025. https://stat.fi/en/statistics/vkour. Accessed 1 Dec 2025.

25. Statistics Finland. Classification of Occupations 2010 | Statistics Finland. 2009. https://stat.fi/en/luokitukset/ammatti/ammatti_1_20100101. Accessed 1 Dec 2025.
26. Finland S. Classification of Socio-economic Groups 1989 | Statistics Finland. 1988. https://stat.fi/en/luokitukset/sosioekon_asema/sosioekon_asema_1_19890101. Accessed 1 Dec 2025.
27. Finnish Institute for Health and Welfare. Care Register for Health Care. THL. 2023. https://thl.fi/en/statistics-and-data/data-and-services/register-descriptions/care-register-for-health-care. Accessed 1 Dec 2025.
28. Kela. Statistics on reimbursement entitlements in respect of medicines | Kela's Info Tray. https://tietotarjotin.fi/en/statistic/2855926/statistics-on-reimbursement-entitlements-in-respect-of-medicines. Accessed 1 Dec 2025.
29. Bush K. The AUDIT Alcohol Consumption Questions (AUDIT-C)An Effective Brief Screening Test for Problem Drinking. Arch Intern Med. 1998;158:1789. https://doi.org/10.1001/archinte.158.16.1789.
30. Bradley KA, DeBenedetti AF, Volk RJ, Williams EC, Frank D, Kivlahan DR. AUDIT-C as a Brief Screen for Alcohol Misuse in Primary Care. Alcoholism Clin Exp Res. 2007;31:1208–17. https://doi.org/10.1111/j.1530-0277.2007.00403.x.
31. Schmidt S, Mühlan H, Power M. The EUROHIS-QOL 8-item index: psychometric results of a cross-cultural field study. Eur J Pub Health. 2006;16:420–8. https://doi.org/10.1093/eurpub/cki155.
32. R Core Team. R: A Language and Environment for Statistical Computing. 2024.
33. Schwarz GE. Estimating the dimension of a model. Ann Stat. 1978;6:461–4.
34. Genuer R, Poggi J-M, Tuleau-Malot C. VSURF: An R Package for Variable Selection Using Random Forests. R J. 2015;7:19–33.
35. Kursa MB, Rudnicki WR. Feature Selection with the Boruta Package. J Stat Softw. 2010;36:1–13. https://doi.org/10.18637/jss.v036.i11.
36. Gower JC. A General Coefficient of Similarity and Some of Its Properties. Biometrics. 1971;27:857–71. https://doi.org/10.2307/2528823.
37. Kaufman L, Rousseeuw PJ. Finding Groups in Data: An Introduction to Cluster Analysis. 1st edition. Wiley; 1990. https://doi.org/10.1002/9780470316801
38. Maechler M, Rousseeuw P, Struyf A, Hubert M, cluster. Finding Groups in Data: Cluster Analysis Extended Rousseeuw 1999. https://doi.org/10.32614/cran.package.cluster
39. Politis DN, Romano JP. Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions. Annals Stat. 1994;22:2031–50. https://doi.org/10.1214/aos/1176325770.
40. Dalitz C, Lögler F. moonboot: An R Package Implementing m-out-of-n Bootstrap Methods. 2025. https://doi.org/10.48550/arXiv.2412.05032
41. Lundqvist A, Mäki-Opas T. Health 2011 Survey - Methods. Helsinki: The National Institute for Health and Welfare (THL); 2016.

42. Cornesse C, Bosnjak M. Is there an association between survey characteristics and representativeness? A meta-analysis. Surv Res Methods. 2018;12:1–13. https://doi.org/10.18148/SRM/2018.V12I1.7205.
43. Cornesse C, Blom AG, Dutwin D, Krosnick JA, De Leeuw ED, Legleye S, et al. A Review of Conceptual Approaches and Empirical Evidence on Probability and Nonprobability Sample Survey Research. J Surv Stat Methodol. 2020;8:4–36. https://doi.org/10.1093/jssam/smz041.
44. Markkula N, Suvisaari J, Saarni SI, Pirkola S, Peña S, Saarni S, et al. Prevalence and correlates of major depressive disorder and dysthymia in an eleven-year follow-up – Results from the Finnish Health 2011 Survey. J Affect Disord. 2015;173:73–80. https://doi.org/10.1016/j.jad.2014.10.015.
45. Tolppanen A-M, Taipale H, Koponen M, Lavikainen P, Tanskanen A, Tiihonen J et al. Use of existing data sources in clinical epidemiology: Finnish health care registers in Alzheimer's disease research–the Medication use among persons with Alzheimer's disease (MEDALZ-2005) study. Clin Epidemiol. 2013;5 Journal Article:277.

Figures

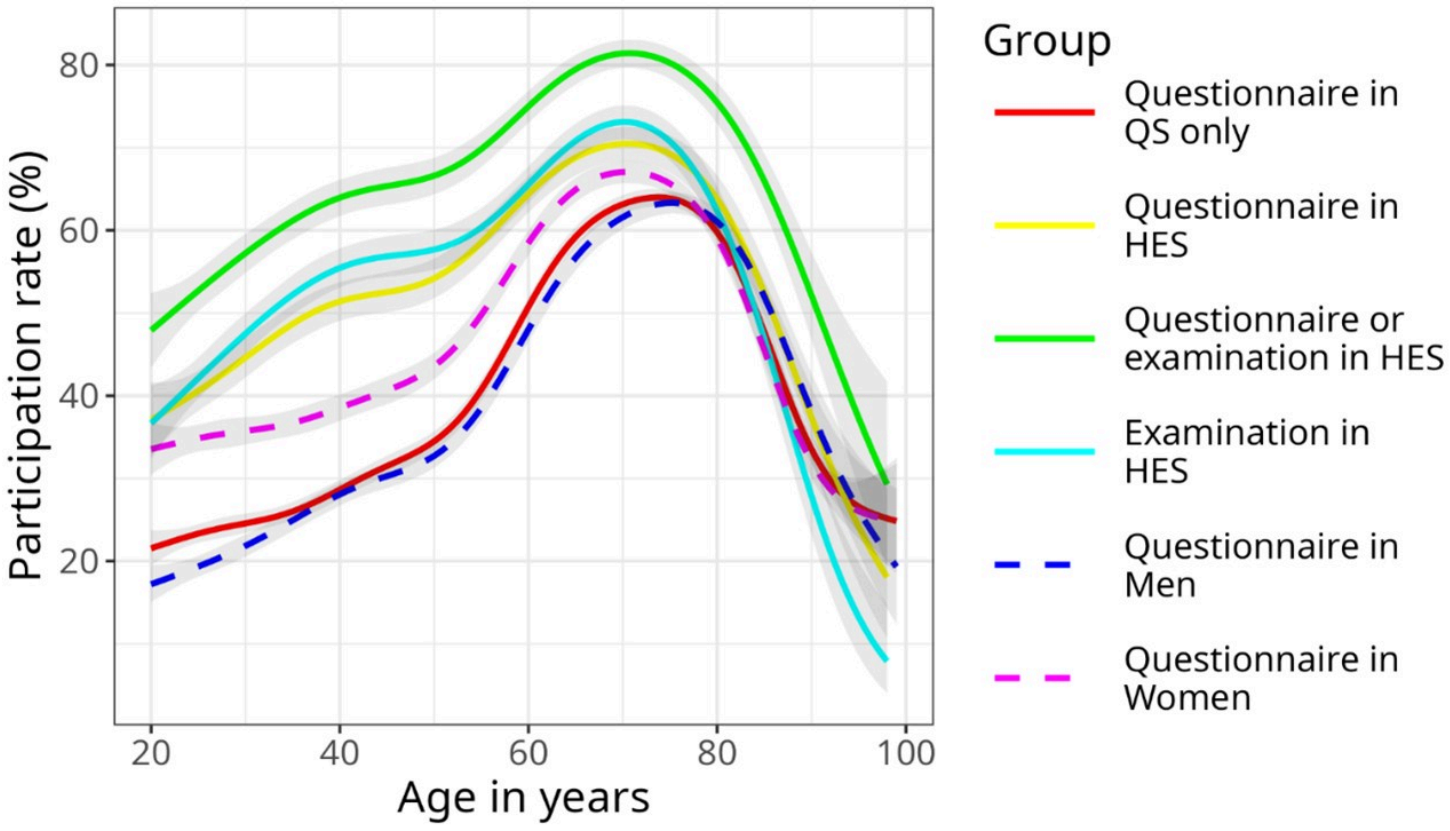


Figure 1

Crude participation rates with respect to age (in years) for the questionnaire and the health examination for the HES (health examination, and health examination or questionnaire) and QS only groups. For men and women participation in the questionnaire.

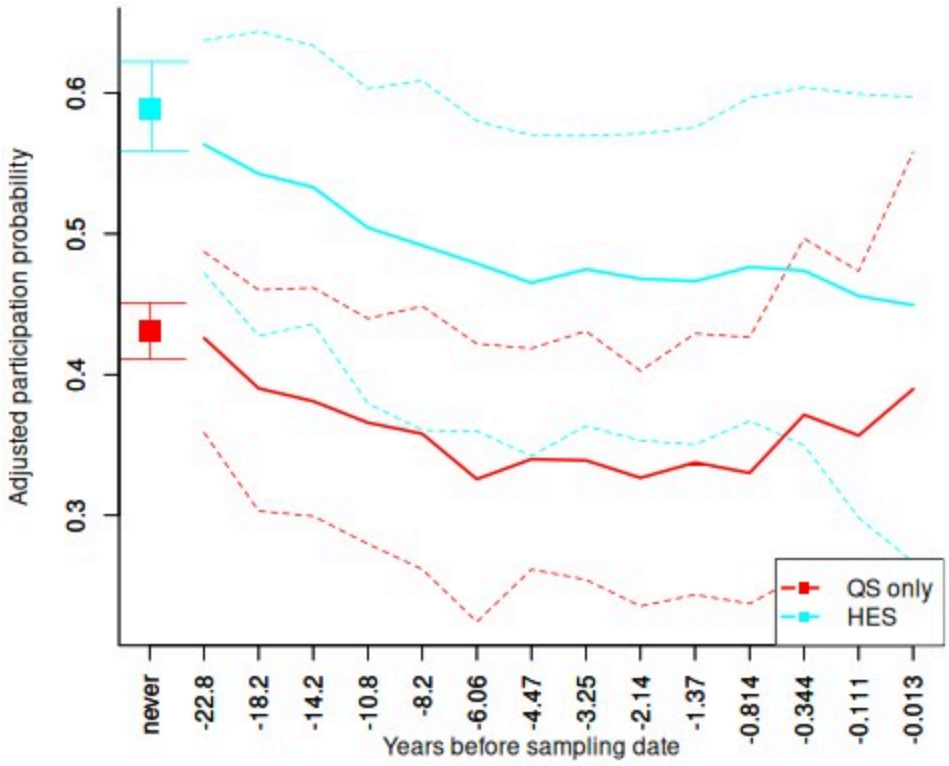


Figure 2

Adjusted QS only and HES participation rates with 95% confidence intervals for the latest hospitalization due to mental health events in years before the sampling date. Participation rates for individuals with no preciding events are denoted by filled squares and the corresponding confidence intervals with whiskers.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- suppl120260205.docx
- suppl2.html