JECH Online First, published on September 26, 2013 as 10.1136/jech-2013-202707

Initial non-participation and loss to follow-up in a Danish youth cohort: implications for relative risk estimates

Trine N Winding,[1] Johan H Andersen,[1] Merete Labriola,[1,2] Ellen A Nohr[3,4]

▸ Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/jech-2013-202707).

[1]Department of Occupational Medicine, Danish Ramazzini Centre, Regional Hospital Herning, Herning, Denmark
[2]Department of Clinical Social Medicine, Public Health and Quality Management, Central Denmark Region and Section of Clinical Social Med, and Rehabilitation, School of Public Health, Aarhus University, Aarhus, Denmark
[3]Research Unit of Gynaecology and Obstetrics, Institute of Clinical Research, University of Southern Denmark, Odense, Denmark
[4]Department of Public Health, Section for Epidemiology, Aarhus University, Aarhus, Denmark

Correspondence to
Trine Nøhr Winding, Department of Occupational Medicine, Danish Ramazzini Centre, Regional Hospital Herning, Gl. Landevej 61, Herning 7400, Denmark; trinwind@rm.dk

Received 5 April 2013
Accepted 3 September 2013

To cite: Winding TN, Andersen JH, Labriola M, *et al*. *J Epidemiol Community Health* Published Online First: [*please include* Day Month Year] doi:10.1136/jech-2013-202707

ABSTRACT

Background Initial non-participation and loss to follow-up in the Danish youth cohort Vestliv could introduce selection bias of the measured risk estimates.

Objective To investigate the impact of initial non-participation and loss to follow-up on the validity of descriptive measures and selected estimates of relative risk.

Methods Of the 3681 young people defining the source population, 83% answered a questionnaire at baseline in 2004. At follow-up waves in 2007 and 2010, the response rates were 71% and 64%, respectively. Relative ORs (RORs) were used to examine the impact of initial non-participation and loss to follow-up on the association between socioeconomic or personal risk factors at age 14/15 and educational attainment at age 20/21. RORs were calculated as OR (baseline population)/OR (source population) or OR (follow-up population)/OR (baseline population).

Results The participants had slightly better school abilities and came more often from homes with two adults, higher income or higher educational level. These differences increased at subsequent follow-ups. The effect of initial non-participation on the ORs was modest with most RORs being close to one. Loss to follow-up led to larger variations in the RORs ranging from 0.77 to 1.62 although for most estimates, the bias was minor. None of the measured RORs were statistically different from one indicating no significant bias.

Conclusions Although certain characteristics were related to those who initially chose to participate and especially to those who participated at follow-ups, it did not have any large influence on the relative risk estimates measured in the study.

INTRODUCTION

Selective response in a cohort study can be a serious threat to the validity of the study and may potentially lead to selection bias. Selection problems can be related to initial non-participation, missing data or loss to follow-up.[1] Findings have shown that participants and non-participants differ in many ways, typically with participants belonging to higher socioeconomic strata and having better health status.[2–5] Young age itself has also been found to be associated with non-participation.[2 6 7] In cohort studies of adolescents and young adults, certain characteristics such as tobacco and alcohol use, male gender and high body mass index (BMI) were more frequent among non-participants.[6 8–10]

Although certain characteristics may be related to those who decide to participate in a cohort study, it does not necessarily introduce any important selection bias for the associations under study.[11 12] Compared with initial non-participation, loss to follow-up is in general considered a much greater threat to the validity of the internal comparisons since it may be related to both exposure and outcome.[6 8 13]

Recently, a study on socioeconomic and personal predictors of educational attainment was conducted in the Danish youth cohort Vestliv. This study investigated to what degree low school performance, low sense of meaningfulness, low self-esteem and poor health status at age 14/15 were associated with completing a secondary education after compulsory school (age 20/21).[14] The cohort was established in 2004 with a participation rate of 83% at baseline and with two follow-ups in 2007 and 2010 where participation rates were 65% and 58%, respectively. Of the participants at baseline, 71% responded at follow-up in 2007 and 64% at follow-up in 2010. These selections may have affected both the constitution of the available study sample and the validity of the internal comparisons.

The aims of the present study were to investigate if participants at baseline differed from the source population and if participants at follow-ups differed from the baseline population. Furthermore, we wanted to examine if (1) initial non-participation in 2004 and (2) loss to follow-up in 2007 and 2010 affected the associations between socioeconomic or personal risk factors and educational attainment.

METHODS

The source population of the prospective cohort study Vestliv consisted of all individuals born in 1989 and living in the county of Ringkjoebing, Denmark in early April 2004, altogether 3681 adolescents. Recruitment took place at the schools within the county where a baseline questionnaire was filled out during school hours in 2004 when the participants were 14/15 years old. Those not at school on the day of collection received the questionnaire by mail. Of the potential 3681 responders, 3054 (83%) chose to participate. A follow-up was conducted in 2007 when the young people were 17/18 years old, using both emailed and postal questionnaires. A second follow-up was carried out in 2010 when the young people were 20/21 years old, using only emailed questionnaires. Of the 3054 individuals defining the initial study population (called baseline population in the following), 2181 (71%) participated in the first follow-up in 2007 and 1945 (64%) in the second follow-up in 2010 (figure 1).

Register information about the source population was retrieved from Statistics Denmark by using the

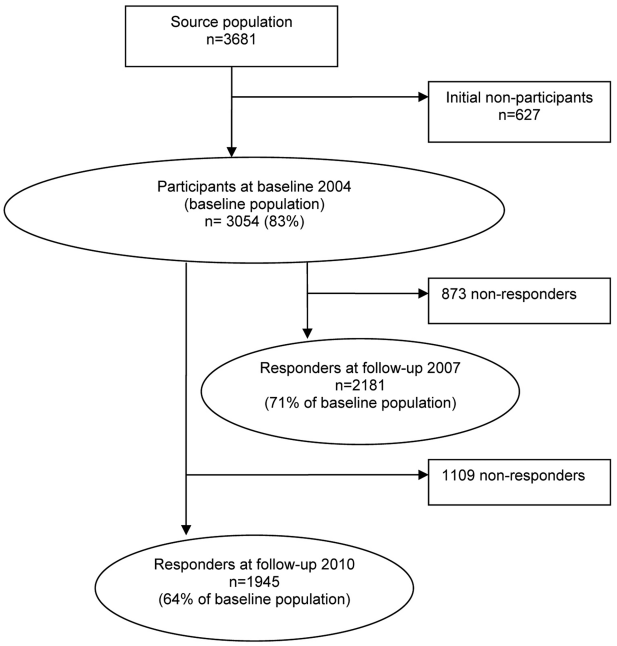

Figure 1 Distribution of participants and non-participants at baseline and responders and non-responders at follow-ups in 2007 and 2010.

personal identification number from the Central Person Register (CPR number), which is given to every inhabitant in Denmark at birth or upon entry for immigrants. To obtain information about family conditions, the young people were linked to their parents or guardians also using the CPR number.[15] The study was approved by the Danish Data Protection Agency.

The outcome was a registered information about educational attainment after compulsory school measured in year 2010 when the participants were 20/21 years old.[16] In Denmark, compulsory school ends after ninth grade, typically at the age of 16 years. The participants were categorised into one of the following four categories: (1) 'Completed' consisting of participants who had completed a secondary education; (2) 'Still studying' consisting of those who were still attaining a secondary education; (3) 'Dropped out' if they had dropped out of their last secondary education and never attained another and (4) 'Never attained' if they had never attained a secondary education.

The following exposure variables from 2004 (age 14/15) were derived from the registers. Yearly household income was recoded in tertiles corresponding to lowest (<64 540 EUR), middle (64 540–82 402 EUR) and highest tertile (>82 402 EUR).[17] Highest education in the household was categorised into three categories <10, 10–12 and >12 years.[16] Family type was dichotomised into 'living with one adult or not living with adults' or 'living with two adults'.[15] Register-based information on the oral Danish and written math examination grades when finalising ninth grade were used, and a dichotomous variable indicating 'high grades' (8–13 equivalent to B+ and above) and 'low grades' (00–7 equivalent to B and below) was generated.[16] Drug use was based on information about prescription medication for pain and nervous symptoms in the period 2005–2006 using the standardised international Anatomical Therapeutic Chemical (ATC) classification system.[18] Four ATC classifications were included (painkilling drugs (N02), drugs for anxiety and sleep disturbance (N05B and N05C) and drugs against depression (N06A)) and combined into one variable defined as 'no drug use' or 'drug use'.

The following exposure variables were based on questionnaire information from 2004 (age 14/15). Depressive symptoms were measured using the four-item validated version of 'The Center for Epidemiologic Studies Depression Scale for Children' and the responses were dichotomised into 'depressive symptoms' and 'no depressive symptoms' at the recommended cut-off at 3.[19] General health status was measured using a single item from SF-36 and the variable was dichotomised into 'high' and 'low' general health status by adding a cut point after the two highest categories.[20] BMI was categorised into low weight (BMI<17 kg/m^2 for both boys and girls), normal weight (17–23.29 kg/m^2 for boys and 17–23.94 kg/m^2 for girls) and overweight (BMI>23.29 kg/m^2 for boys and BMI>23.94 kg/m^2 for girls), using thresholds for 15-year-old children.[21] The four items about meaningfulness from the 'Sense of coherence-revised short version for children' (13 items) were used to categorise data with cut-off point at the 25% centile as normal/high and low scores.[22] Self-esteem was measured by using six items from Rosenberg's 10 items self-esteem scale and also categorised into normal/high and low self-esteem by applying a cut-off point at the 25% centile.[23]

Statistical analysis

First, the pattern of participation was examined by calculating prevalences (P) of register-based risk factors in the source population and the baseline population and comparing them by estimating prevalence ratios PR=(PBaseline population/PSource population).[4] The same was carried out for the follow-ups in 2007 and 2010 where prevalences of register-based and questionnaire-based risk factors for responders were compared with those for the baseline population.

Second, multinomial logistic regression was used to estimate ORs for associations between different risk factors and educational attainment in the source population, the baseline population and the follow-up populations in 2007 and 2010. Individuals who had completed a secondary education were used as reference group. Gender and age when completing ninth grade were included in all models. To examine the impact of the initial non-participation, we computed relative ORs (ROR) as the ratio of the OR for the baseline population and the OR in the source population (ROR=ORBaseline population/ORSource population). Similarly, loss to follow-up was examined by computing RORs as the ratio of the OR for the follow-up populations in 2007 or 2010 and the OR of the baseline population (ROR=ORFollow-up population/ORBaseline population).[11] In the loss to follow-up analyses, the outcome category 'never attained' contained too few participants to be able to calculate valid ROR estimates and therefore the category was excluded from these analyses. To calculate 95% CI of the PR and the ROR estimates, the following approximate formula was applied:$\mathrm{se}(\widehat{\theta}_{\mathrm{Sub}} - \widehat{\theta}_{\mathrm{Tot}}) = \sqrt{\mathrm{se}(\widehat{\theta}_{\mathrm{Sub}})^2 - \mathrm{se}(\widehat{\theta}_{\mathrm{Tot}})^2}$,[11] where $\widehat{\theta}_{\mathrm{Tot}}$ is the estimate of the total sample and $\widehat{\theta}_{\mathrm{Sub}}$ is the estimate in a subsample. This formula has been used in previous studies [4 11 12] and has in a Danish simulation study shown to give valid CIs for RORs when the expected bias related to the selection is modest.[11] STATA statistical package (V.12.0; Stata, College Station, Texas, USA) was used for all analyses.

Sensitivity analyses were performed with exposure variables included as continuous values or with alternative categorisations. These analyses did not reveal RORs that differed from those found in the main analysis; neither did they provide higher statistical power.

Table 1 Participation at baseline and response at follow-ups in 2007 and 2010 according to socioeconomic and personal risk factors

	Source population (N=3681)		Baseline 2004 (N=3054)					Follow-up 2007 (N=2181)					Follow-up 2010 (N=1945)				
		Prevalence		Prevalence	Participation*	Prevalence ratio†			Prevalence	Response*	Prevalence ratio‡			Prevalence	Response*	Prevalence ratio§	
	n	Per cent	n	Per cent	Per cent	Estimate	95% CI	n	Per cent	Per cent	Estimate	95% CI	n	Per cent	Per cent	Estimate	95% CI
All	3681		3054	100	83			2181	100	71			1945	100	64		
Register-based risk factors																	
Grades—oral Danish	3340		2889		86			2086		72			1865		65		
8 or above	2358	71	2102	73	89	**1.03**	1.01 to 1.06	1596	77	76	**1.05**	1.04 to 1.06	1424	76	68	**1.05**	1.04 to 1.06
7 or below	982	29	787	27	80	**0.93**	0.87 to 0.98	490	23	62	**0.86**	0.82 to 0.91	441	24	56	**0.87**	0.82 to 0.92
Grades—written math	3327		2874		86			2078		72			1857		65		
8 or above	2176	65	1962	68	90	**1.04**	1.02 to 1.07	1520	73	77	**1.07**	1.06 to 1.08	1385	75	71	**1.09**	1.08 to 1.10
7 or below	1151	35	912	32	79	**0.92**	0.87 to 0.97	558	27	61	**0.85**	0.81 to 0.89	472	25	52	**0.80**	0.76 to 0.85
Drug use	3681		3054		83			2181		71			1945		64		
No	3500	95	2905	95	83	1.00	0.99 to 1.01	2080	95	72	1.00	1.00 to 1.01	1838	95	63	0.99	0.99 to 1.00
Yes	181	4.9	149	4.9	82	0.99	0.85 to 1.16	101	4.6	68	0.95	0.85 to 1.06	107	5.5	72	**1.13**	1.02 to 1.24
Household income	3678		3053		83			2180		71			1945		64		
Highest	1138	31	1017	33	89	**1.08**	1.02 to 1.13	787	36	77	**1.08**	1.06 to 1.11	692	36	68	**1.07**	1.03 to 1.10
Middle	1177	32	1018	33	86	1.04	0.99 to 1.10	754	35	74	**1.04**	1.01 to 1.07	693	36	68	**1.07**	1.03 to 1.10
Lowest	1363	37	1018	33	75	**0.90**	0.86 to 0.95	639	29	63	**0.88**	0.84 to 0.92	560	29	55	**0.86**	0.82 to 0.91
Highest household education (years)	3582		3001		84			2153		72			1927		64		
>12	1219	34	1094	36	90	**1.07**	1.02 to 1.12	833	39	76	**1.06**	1.04 to 1.09	753	39	69	**1.07**	1.04 to 1.10
10–12	1867	52	1548	52	83	0.99	0.96 to 1.02	1113	52	72	1.00	0.98 to 1.02	973	50	63	0.98	0.95 to 1.01
<10	496	14	359	12	72	**0.86**	0.78 to 0.95	207	10	58	**0.80**	0.74 to 0.88	201	10	56	**0.87**	0.80 to 0.95
Family type	3681		3054		83			2181		71			1945		64		
Two adults	3130	85	2652	87	85	**1.02**	1.01 to 1.04	1957	90	74	**1.03**	1.03 to 1.04	1744	90	66	**1.03**	1.03 to 1.04
One adult	551	15	402	13	73	**0.88**	0.80 to 0.96	224	10	56	**0.78**	0.72 to 0.85	201	10	50	**0.78**	0.71 to 0.86
Questionnaire-based risk factors																	
Self-esteem			2974					2126		71			1901		64		
Normal/high	–	–	2237	75	–	–	–	1614	76	72	1.01	1.00 to 1.02	1410	74	63	0.99	0.97 to 1.00
Low	–	–	737	25	–	–	–	512	24	69	0.97	0.93 to 1.01	491	26	67	1.04	1.00 to 1.09
Sense of meaningfulness			3023					2164		72			1933		64		
Normal/high	–	–	2439	81	–	–	–	1768	82	72	1.01	1.00 to 1.02	1578	82	65	1.01	1.00 to 1.02
Low	–	–	584	19	–	–	–	396	18	68	0.95	0.90 to 1.00	355	18	61	0.95	0.90 to 1.01
General health status			3034					2168		71			1932		64		
High	–	–	2905	96	–	–	–	2086	96	72	1.00	1.00 to 1.01	1859	96	64	1.00	1.00 to 1.01
Low	–	–	129	4.3	–	–	–	82	3.8	64	0.89	0.78 to 1.01	73	3.8	57	0.89	0.77 to 1.03
Depressive symptoms			3000					2152		72			1924		64		
No	–	–	1959	65	–	–	–	1408	65	72	1.00	0.99 to 1.02	1271	66	65	1.01	0.99 to 1.03

Continued

Table 1 Continued

	Source population (N=3681)		Baseline 2004 (N=3054)				Follow-up 2007 (N=2181)				Follow-up 2010 (N=1945)			
		Prevalence		Prevalence	Participation*	Prevalence ratio†		Prevalence	Response*	Prevalence ratio‡		Prevalence	Response*	Prevalence ratio§
	n	Per cent	n	Per cent	Per cent	Estimate 95% CI	n	Per cent	Per cent	Estimate 95% CI	n	Per cent	Per cent	Estimate 95% CI
Yes	–	–	1041	35	–	– –	744	35	71	1.00 0.97 to 1.03	653	34	63	0.98 0.94 to 1.02
Body mass index			2879				2067		72		1831		64	
Low weight	–	–	273	9.5	–	– –	198	9.6	73	1.01 0.94 to 1.08	162	8.9	59	0.93 0.85 to 1.03
Normal weight	–	–	2301	80	–	– –	1669	81	73	1.01 1.00 to 1.02	1481	81	64	1.01 1.00 to 1.03
Overweight	–	–	305	11	–	– –	200	9.7	66	**0.91** 0.85 to 0.99	188	10	62	0.97 0.89 to 1.06

*Different participation rates/response rates caused by missing register information.
†Prevalence ratio=prevalence in the baseline population/prevalence in the source population.
‡Prevalence ratio=prevalence in the 2007 follow-up population/prevalence in the baseline population.
§Prevalence ratio=prevalence in the 2010 follow-up population/prevalence in the baseline population.
The significant prevalence ratios are written in bold.

RESULTS

Table 1 shows participation rates at baseline, response rates at follow-ups and the distribution of risk factors in the source, baseline and follow-up populations. Participants at baseline did not differ much from the source population except that adolescents from families in the highest income and educational group, those brought up by two adult and those with high grades in math and Danish were slightly over-represented. At follow-up in 2007, these over-representations increased. The lowest response rates in specific groups were 56–63% and the highest 76–77% as compared with the general participation rate of 71%. This led to prevalence ratios ranging from 1.08 to 0.78 indicating only modest over-representation and under-representation. The pattern of response at follow-up in 2010 was identical to the pattern in 2007. None of the risk factors measured in the baseline questionnaire were related to the loss to follow-up (table 1).

Impact of initial non-participation on relative estimates

Table 2 displays ORs and RORs for educational attainment measured in the baseline population and the source population. For most associations, we found no or minor differences in the ORs for the two populations leading to RORs close to one. Only for the associations between 'low income' or 'living with one adult' and 'never attained a secondary education', underestimation or overestimation of about 20% were observed, but the CIs of the RORs were wide, and all included the value 1.

Impact of loss to follow-up on relative estimates

The bias estimates related to the loss to follow-up in 2007 and 2010 are displayed in table 3. Compared with the baseline population, stronger associations were observed for the participants in the 2007 follow-up for 'low general health status' and 'still studying' (ROR 1.55) and for 'low grades in oral Danish' or 'low household education' and 'dropping out' (RORs 1.36 and 1.26), however all with CIs including the value 1. The rest of the RORs did not reveal any underestimation or overestimation exceeding 23%, and most RORs were close to 1.

Among participants in the 2010 follow-up, we observed stronger associations between 'low weight' and 'still studying' (ROR 1.49) and between 'drug use' or 'overweight' and 'dropping out' (RORs 1.59 and 1.62) as compared with the baseline population. None of the remaining RORs, revealed underestimation or overestimation exceeding 23%, and most RORs were close to 1. Again, CIs were wide showing no statistically significant bias.

A supplementary analysis comparing the follow-up populations with the source population was conducted for register variables. It revealed increasing bias for some RORs related to grades, drug use and low household education, which changed from underestimation or overestimation below 23% in the main analysis to underestimation or overestimation of 24–29%. All but one of the CIs was not statistically significant (see online supplementary table S1).

DISCUSSION

In this Danish cohort of young people, the initial participation only involved minor selection on socioeconomic and personal risk factors. The participants had slightly better school abilities and came more often from homes with two adults or with higher income and educational level. This selection pattern became more pronounced at the first follow-up, but was not further strengthened at the second follow-up.

The impact of the initial participation on relative risk estimates for educational attainment was modest with no indication

Table 2 Relative ORs (ROR) examining the impact of initial non-participation on ORs for educational attainment

	ORs for educational attainment											
	Source population (n=3668)*			Participants at baseline (n=3053)†			RORs‡					
	Still studying	Dropping out	Never attained	Still studying	Dropping out	Never attained	Still studying		Dropping out		Never attained§	
	(n=418)	(n=286)	(n=129)	(n=312)	(n=202)	(n=72)						
	OR¶	OR¶	OR¶	OR¶	OR¶	OR¶	ROR¶	95% CI	ROR¶	95% CI	ROR¶	95% CI
Register-based risk factors												
Grades—oral Danish												
8 or above	Ref	Ref	Ref	Ref	Ref	Ref						
7 or below	**2.34**	**3.56**	**3.03**	**2.53**	**3.28**	**2.62**	1.08	0.75 to 1.55	0.92	0.70 to 1.22	0.87	0.60 to 1.26
Grades—written math												
8 or above	Ref	Ref	Ref	Ref	Ref	Ref						
7 or below	**3.24**	**4.19**	**3.85**	**3.22**	**3.95**	**3.98**	0.99	0.70 to 1.40	0.94	0.68 to 1.31	1.03	0.26 to 4.09
Drug use												
No	Ref	Ref	Ref	Ref	Ref	Ref						
Yes	**1.69**	1.44	**4.55**	**1.96**	1.21	**5.15**	1.16	0.64 to 2.08	0.84	0.67 to 1.05	1.13	0.08 to 16.5
Household income												
Highest	Ref	Ref	Ref	Ref	Ref	Ref						
Middle	**1.52**	**1.46**	1.49	**1.49**	1.50	1.43	0.98	0.81 to 1.18	1.03	0.74 to 1.42	0.96	0.54 to 1.71
Lowest	**2.28**	**3.46**	**4.62**	**2.27**	**3.56**	**3.76**	1.00	0.73 to 1.37	1.03	0.50 to 2.11	0.81	0.62 to 1.07
Highest household education (years)												
>12	Ref	Ref	Ref	Ref	Ref	Ref						
10–12	**1.39**	**1.53**	1.50	**1.52**	**1.50**	0.98	1.09	0.86 to 1.39	0.98	0.78 to 1.24	–	–
<10	**2.26**	**3.59**	**3.00**	**2.23**	**3.77**	**2.69**	0.99	0.63 to 1.54	1.05	0.42 to 2.60	0.90	0.37 to 2.16
Family type												
Two adults	Ref	Ref	Ref	Ref	Ref	Ref						
One adult	**1.66**	**2.15**	**3.16**	**1.67**	**2.36**	**2.46**	1.01	0.75 to 1.35	1.10	0.65 to 1.86	0.78	0.42 to 1.44

*Includes 2835 participants who had completed a secondary education (reference group). Information about age when completing ninth grade was missing for 13 persons.
†Includes 2467 participants who had completed a secondary education (reference group). Information about age when completing ninth grade was missing for one person.
‡ROR=OR in the baseline population/OR in the source population.
§Not possible to calculate an ROR with CIs for 'parental education 10–12 years' and 'never attained'.
¶Adjusted for gender and age when completing ninth grade.
The significant ORs and RORs are written in bold.

Table 3 Relative ORs (ROR) examining the impact of loss to follow-up on ORs for educational attainment

	ORs for educational attainment													
	Participants at baseline (n=2981)*		Participants at follow-up in 2007 (n=2146)†		Participants at follow-up in 2010 (n=1915)‡		Relative ORs (ROR)§							
							Follow-up in 2007				Follow-up in 2010			
	Still studying	Dropping out	Still studying	Dropping out	Still studying	Dropping out	Still studying		Dropping out		Still studying		Dropping out	
	(n=312)	(n=202)	(n=199)	(n=87)	(n=167)	(n=85)								
	OR¶	OR¶	OR¶	OR¶	OR¶	OR¶	ROR¶	95% CI	ROR¶	95% CI	ROR¶	95% CI	ROR¶	95% CI
Register-based risk factors														
Grades—oral Danish														
8 or above	Ref	Ref	Ref	Ref	Ref	Ref								
7 or below	**2.52**	**3.28**	**2.79**	**4.47**	**2.97**	**2.88**	1.11	0.60 to 2.05	1.36	0.22 to 8.54	1.18	0.53 to 2.64	0.88	0.35 to 2.20
Grades—written math														
8 or above	Ref	Ref	Ref	Ref	Ref	Ref								
7 or below	**3.22**	**3.95**	**3.30**	**4.79**	**3.72**	**3.94**	1.02	0.53 to 1.97	1.21	0.18 to 8.29	1.15	0.42 to 3.14	1.00	0.24 to 4.12
Drug use														
No	Ref	Ref	Ref	Ref	Ref	Ref								
Yes	**1.92**	1.21	**1.87**	1.08	**2.03**	1.92	0.97	0.50 to 1.89	0.89	0.41 to 1.93	1.06	0.46 to 2.45	1.59	0.42 to 5.98
Household income														
Highest	Ref	Ref	Ref	Ref	Ref	Ref								
Middle	**1.48**	1.50	1.46	1.44	1.37	1.49	0.98	0.73 to 1.32	0.96	0.53 to 1.75	0.92	0.66 to 1.28	0.99	0.48 to 2.04
Lowest	**2.25**	**3.55**	**2.08**	**3.13**	**2.28**	**4.10**	0.92	0.63 to 1.34	0.88	0.30 to 2.57	1.01	0.54 to 1.91	1.15	0.16 to 8.23
Highest household education (years)														
>12	Ref	Ref	Ref	Ref	Ref	Ref								
10–12	**1.51**	**1.50**	**1.51**	1.57	**1.59**	1.53	1.00	0.76 to 1.32	1.04	0.55 to 1.99	1.05	0.71 to 1.55	1.02	0.54 to 1.91
<10	**2.20**	**3.77**	1.69	**4.74**	1.69	**4.54**	0.77	0.58 to 1.02	1.26	0.10 to 15.4	0.77	0.51 to 1.16	1.20	0.12 to 12.6
Family type														
Two adults	Ref	Ref	Ref	Ref	Ref	Ref								
One adult	**1.64**	**2.36**	1.55	**1.90**	**1.73**	**2.80**	0.94	0.62 to 1.43	0.80	0.39 to 1.67	1.05	0.58 to 1.93	1.18	0.33 to 4.19
Questionnaire-based risk factors														
Self-esteem														
Normal/high	Ref	Ref	Ref	Ref	Ref	Ref								
Low	**2.01**	**1.73**	**1.85**	**1.97**	**2.18**	**2.05**	0.92	0.65 to 1.31	1.14	0.54 to 2.40	1.09	0.60 to 1.99	1.18	0.52 to 2.68
Sense of meaningfulness														
Normal/high	Ref	Ref	Ref	Ref	Ref	Ref								
Low	**1.71**	**2.10**	**1.73**	**2.44**	1.44	**2.02**	1.02	0.70 to 1.47	1.16	0.46 to 2.94	0.84	0.63 to 1.13	0.96	0.47 to 1.97

General health status															
High	Ref	Ref	Ref	Ref	Ref	Ref									
Low	1.73	**2.63**	**2.68**	**2.65**	1.88	**2.50**	1.55	0.36 to 6.72	1.01	0.19 to 5.26	1.09	0.33 to 3.59	0.95	0.18 to 5.15	
Depressive symptoms															
No	Ref	Ref	Ref	Ref	Ref	Ref									
Yes	1.21	**1.40**	1.28	**1.70**	1.30	1.53	1.06	0.82 to 1.37	1.21	0.65 to 2.26	1.07	0.78 to 1.47	1.10	0.63 to 1.91	
Body mass index															
Low weight	0.88	0.82	0.83	0.83	1.31	0.91	0.94	0.75 to 1.19	1.02	0.62 to 1.67	1.49	0.83 to 2.66	1.11	0.58 to 2.12	
Normal weight	Ref	Ref	Ref	Ref	Ref	Ref									
Overweight	0.94	1.42	1.03	1.25	1.08	**2.29**	1.10	0.76 to 1.57	0.88	0.46 to 1.69	1.15	0.71 to 1.86	1.62	0.47 to 5.52	

*Includes 2467 participants who had completed a secondary education (reference group). Information about age when completing ninth grade was missing for one person.
†Includes 1860 participants who had completed a secondary education (reference group).
‡Includes 1663 participants who had completed a secondary education (reference group).
§RORs, comparing ORs for participants at baseline in 2004 with ORs of participants at follow-up in 2007 or 2010, respectively.
¶Adjusted for gender and age when completing ninth grade.
The significant ORs and RORs are written in bold.

of underestimation or overestimation exceeding 22%, and most estimates were similar in the source and baseline population. Also the examination of loss to follow-up showed in general only small differences on relative risk estimates when comparing the follow-up populations with the baseline population. For six relative risk estimates (three at each follow-up), we observed underestimation or overestimation exceeding 23%, but they were all based on small numbers and CIs were wide. At no point in time did the selection lead to statistically significant bias. These findings are in accord with previous studies based on adult cohorts that also calculated relative risk estimates and did not find any considerable selection bias, neither in relation to the initial non-participation,[11 12] nor in relation to the loss to follow-up.[13]

It is well known that socioeconomic measures are strong predictors of non-participation both initially and at follow-up among adult populations.[2–4 7 24 25] This study shows that adolescents to some degree follow the response pattern of their parents, as low household income and low parental education were associated with both initial non-participation and loss to follow-up. This is also consistent with the results from two young adult cohorts where people were mainly in their 20s when recruited.[8 26]

At the follow-ups, we found health and well-being of the participants to be almost identical to that of the baseline population. Previous studies have found high mental distress,[5 24] low general health status[5 27 28] and high BMI[5] to be related to initial non-participation and loss to follow-up whereas others did not.[25 29 30] It remains unclear why the present study could not replicate these response patterns, but one explanation may be that young people have less severe morbidity in general which may have a relatively smaller impact on their ability to participate. Also, they may be more willing to tell about their, relatively minor, health problems as compared with an adult population.

The Vestliv cohort consists of a large sample of young Danes with a high initial participation rate of 83%. Register information was almost complete for the source population, and studies have shown these data to have a high validity.[15–17] Additionally, we had high completeness of the questionnaire that was included in the study. The Vestliv cohort has multiple research aspects with the overall focus of studying inequality in a lifecourse perspective. Inspired by a recent study,[14] we investigated exposures related to educational attainment after compulsory school. These are, however, only a small subset of all the relative risk estimates that might be studied in the Vestliv cohort study, and larger bias could be observed for other associations. It is also important to emphasise that while other researchers can benefit from applying the presented method of estimating selection bias, the results cannot be generalised to other research areas.

In this study, we used ROR estimates with 95% CIs to estimate the degree of initial non-participation and loss to follow-up. If large bias had been found, it would have been possible to use inverse probability weighting, but this method requests some kind of information about the missing variables, which can be difficult to achieve.[31] Multiple imputation techniques is another way of addressing problems with missing data in order to avoid selection bias and is becoming increasingly used in epidemiological studies.[32 33]

Although certain characteristics were related to those who participated initially and especially to those who responded at follow-ups, our study showed that it did not have any large influence on the relative risk estimates in the study, as contrasts sustained within the study population. For some of the bias

estimates, CIs were wide limiting our ability in detecting important differences in the relative risk estimates among source, baseline and follow-up populations. Our results are context specific, and it would be useful to obtain data from other populations as well. In combination, this could provide useful information for ongoing and future epidemiological studies involving young people. For forthcoming analyses in this cohort, we find the results from this study reassuring for obtaining valid risk estimates.

What is already known on this subject

- Previous studies of adult cohorts have found low socioeconomic status and poor health to increase non-participation and loss to follow-up.
- This study examines if certain characteristics were related to participation initially and at follow-ups in a cohort of young people and if any important selection bias of the associations under study was introduced.

What this study adds

- Initial non-participation was minor and did not have important impact on the validity of the measured relative risk estimates.
- Loss to follow-up was associated with more pronounced selection patterns, but no statistically significant effects on the relative risk estimates were observed.

Contributors JHA initiated the study. TNW designed and performed the analyses, and wrote the main paper. EAN helped designing the study and analysing the data. All authors interpreted the results and their implications and commented on the manuscript at all stages.

Funding This study was funded by The Danish Working Environment Research Fund (grant number 5253).

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Rothman K, Greenland S, Lash T. Validity in Epidemiologic Studies. In: Rothman K, Greenland S, Lash T, eds. Modern Epidemiology. 3rd ed. Philadelphia, PA: Lippincott Williams & Williams; 2008:128–150.
2 Bergman P, Ahlberg G, Forsell Y, *et al*. Non-participation in the second wave of the PART study on mental disorder and its effects on risk estimates. *Int J Soc Psychiatry* 2010;56:119–32.
3 Goldberg M, Chastang JF, Leclerc A, *et al*. Socioeconomic, demographic, occupational, and health factors associated with participation in a long-term epidemiologic survey: a prospective study of the French GAZEL cohort and its target population. *Am J Epidemiol* 2001;154:373–84.
4 Jacobsen TN, Nohr EA, Frydenberg M. Selection by socioeconomic factors into the Danish National Birth Cohort. *Eur J Epidemiol* 2010;25:349–55.
5 Torvik FA, Rognmo K, Tambs K. Alcohol use and mental distress as predictors of non-response in a general population health survey: the HUNT study. *Soc Psychiatry Psychiatr Epidemiol* 2012;47:805–16.
6 Alonso A, Segui-Gomez M, de Irala J, *et al*. Predictors of follow-up and assessment of selection bias from dropouts using inverse probability weighting in a cohort of university graduates. *Eur J Epidemiol* 2006;21:351–8.
7 Carter KN, Imlach-Gunasekara F, McKenzie SK, *et al*. Differential loss of participants does not necessarily cause selection bias. *Aust N Z J Public Health* 2012;36:218–22.
8 Powers J, Loxton D. The impact of attrition in an 11-year prospective longitudinal study of younger women. *Ann Epidemiol* 2010;20:318–21.
9 Morrison TC, Wahlgren DR, Hovell MF, *et al*. Tracking and follow-up of 16,915 adolescents: minimizing attrition bias. *Control Clin Trials* 1997;18:383–96.
10 Cunradi CB, Moore R, Killoran M, *et al*. Survey nonresponse bias among young adults: the role of alcohol, tobacco, and drugs. *Subst Use Misuse* 2005;40:171–85.
11 Nohr EA, Frydenberg M, Henriksen TB, *et al*. Does low participation in cohort studies induce bias? *Epidemiology* 2006;17:413–18.
12 Kaerlev L, Kolstad HA, Hansen AM, *et al*. Are risk estimates biased in follow-up studies of psychosocial factors with low base-line participation? *BMC Public Health* 2011;11:539.
13 Greene N, Greenland S, Olsen J, *et al*. Estimating bias from loss to follow-up in the Danish National Birth Cohort. *Epidemiology* 2011;22:815–22.
14 Winding TN, Nohr EA, Labriola M, *et al*. Personal predictors of educational attainment after compulsory school: influence of measures of vulnerability, health, and school performance. *Scand J Public Health* 2013;41:92–101.
15 Pedersen CB. The Danish Civil Registration System. *Scand J Public Health* 2011;39:22–5.
16 Jensen VM, Rasmussen AW. Danish Education Registers. *Scand J Public Health* 2011;39:91–4.
17 Baadsgaard M, Quitzau J. Danish registers on personal income and transfer payments. *Scand J Public Health* 2011;39:103–5.
18 WHO Collaborating Centre for Drug Statistics Methodology. Guidelines for ATC classification and DDD assignment 2013, 2012. http://www.whocc.no/filearchive/publications/1_2013guidelines.pdf, (accessed 21 Feb 2013).
19 Fendrich M, Weissman MM, Warner V. Screening for depressive disorder in children and adolescents: validating the Center for Epidemiologic Studies Depression Scale for Children. *Am J Epidemiol* 1990;131:538–51.
20 Idler EL, Benyamini Y. Self-rated health and mortality: a review of twenty-seven community studies. *J Health Soc Behav* 1997;38:21–37.
21 Cole TJ, Bellizzi MC, Flegal KM, *et al*. Establishing a standard definition for child overweight and obesity worldwide: international survey. *BMJ* 2000;320:1240–3.
22 Eriksson M, Lindstrom B. Validity of Antonovsky's sense of coherence scale: a systematic review. *J Epidemiol Community Health* 2005;59:460–6.
23 Turner RJ, Lloyd DA, Roszell P. Personal resources and the social distribution of depression. *Am J Community Psychol* 1999;27:643–72.
24 Lundberg I, Damstrom Thakker K, Hallstrom T, *et al*. Determinants of non-participation, and the effects of non-participation on potential cause-effect relationships, in the PART study on mental disorders. *Soc Psychiatry Psychiatr Epidemiol* 2005;40:475–83.
25 Korkeila K, Suominen S, Ahvenainen J, *et al*. Non-response and related factors in a nation-wide health survey. *Eur J Epidemiol* 2001;17:991–9.
26 Littman AJ, Boyko EJ, Jacobson IG, *et al*. Assessing nonresponse bias at follow-up in a large prospective cohort of relatively young and mobile military service members. *BMC Med Res Methodol* 2010;10:99.
27 Van Loon AJ, Tijhuis M, Picavet HS, *et al*. Survey non-response in the Netherlands: effects on prevalence estimates and associations. *Ann Epidemiol* 2003;13:105–10.
28 Goldberg M, Chastang JF, Zins M, *et al*. Health problems were the strongest predictors of attrition during follow-up of the GAZEL cohort. *J Clin Epidemiol* 2006;59:1213–21.
29 Gustavson K, von Soest T, Karevold E, *et al*. Attrition and generalizability in longitudinal studies: findings from a 15-year population-based study and a Monte Carlo simulation study. *BMC Public Health* 2012;12:918.
30 Ekholm O, Gundgaard J, Rasmussen NK, *et al*. 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.
31 Weuve J, Tchetgen Tchetgen EJ, Glymour MM, *et al*. Accounting for bias due to selective attrition: the example of smoking and cognitive decline. *Epidemiology* 2012;23:119–28.
32 Donders AR, van der Heijden GJ, Stijnen T, *et al*. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol* 2006;59:1087–91.
33 van der Heijden GJ, Donders AR, Stijnen T, *et al*. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. *J Clin Epidemiol* 2006;59:1102–9.