



**HAL**  
open science

## Selection by socioeconomic factors into the Danish National Birth Cohort

Tine Neermann Jacobsen, Ellen Aagaard Nohr, Morten Frydenberg

► **To cite this version:**

Tine Neermann Jacobsen, Ellen Aagaard Nohr, Morten Frydenberg. Selection by socioeconomic factors into the Danish National Birth Cohort. *European Journal of Epidemiology*, 2010, 25 (5), pp.349-355. 10.1007/s10654-010-9448-2 . hal-00580341

**HAL Id: hal-00580341**

**<https://hal.science/hal-00580341>**

Submitted on 28 Mar 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Selection by socioeconomic factors into the Danish National Birth Cohort

Tine Neermann Jacobsen · Ellen Aagaard Nohr ·  
Morten Frydenberg

Received: 10 September 2009 / Accepted: 11 March 2010 / Published online: 27 March 2010  
© Springer Science+Business Media B.V. 2010

**Abstract** *Background* Low participation at recruitment to the Danish National Birth Cohort (DNBC) has raised concern about non-participation bias. *Objective* To study the socioeconomic pattern of participation to the DNBC. *Methods* Independently of the DNBC, we identified the DNBC source population in two geographical areas of Denmark by means of local birth registers with full coverage. Socioeconomic information came from national registers, and the source population consisted of 48,560 births including 15,290 participating women. For every socioeconomic characteristic, we estimated the prevalence ratio [prevalence (participants)/prevalence (source population)] which corresponds to the relative representation of the group (presented in percentages with 95% confidence intervals). *Results* The overall participation rate was 31%. Women outside the work force or with no further education than compulsory school were underrepresented in the DNBC by 62% (59%; 64%) and 43% (41%; 45%), respectively. Also, women were underrepresented by 18% (13%; 23%) if they were unemployed, by 22% (20%; 24%) if they were in the lowest income group, 38% (35%; 40%) if they received a high proportion of social benefits, and 28% (24%; 31%) if they were singles. Particularly women with low resources according to two socioeconomic factors were strongly underrepresented, typically by 50–67%.

*Conclusion* Groups with low socioeconomic resources in terms of education, occupation, income and civil status are underrepresented in the DNBC compared to the background population. These discrepancies must be taken into account when results from the DNBC and other cohorts of pregnant women are interpreted—especially when descriptive results are presented.

**Keywords** Cohorts · Non participation · Non-response bias · Pregnant women · Selection bias · Socioeconomic status

## Abbreviations

CPR-number	Danish unique personal identifier which is based on “Det Centrale Personregister” (“The Danish Civil Registration System”)
DNBC	Danish National Birth Cohort
GP	General Practitioner
PR	Prevalence ratio

## Background

Declining participation rates in epidemiological studies have been observed in many countries [1, 2]. This poses a threat to the research results as non-participation may introduce bias of the prevalence of exposures, outcomes and confounders as well as effect measures [2–5].

Low participation at recruitment was also found in the Danish National Birth Cohort (DNBC), a nation-wide cohort study with data from 100,000 women and their offspring. The DNBC was mainly established to enable investigations of short and long term consequences of intrauterine exposures [6]. Many of the studied associations are likely to be

T. N. Jacobsen (✉) · M. Frydenberg  
Department of Biostatistics, Institute of Public Health,  
Aarhus University, Vennelyst Boulevard 6, Building 1260,  
8000 Aarhus C, Denmark  
e-mail: tinenj@gmail.com

E. A. Nohr  
Department of Epidemiology, Institute of Public Health,  
Aarhus University, Vennelyst Boulevard 6,  
8000 Aarhus C, Denmark

confounded by socioeconomic factors raising concerns about selection bias related to lower participation rates among women with few socioeconomic resources.

Various studies have shown lower participation among less privileged groups of society. People with a low income or low educational level have typically been underrepresented in cohort studies in Western societies [5, 7–21] and participation rates have varied according to age and sex [5, 9–18]. The socioeconomic pattern of participation has to our knowledge not been studied in a population of pregnant women like the DNBC. Given the Danish national registers, we had a unique opportunity to do so, and the aim of the present study was to examine the associations between socioeconomic factors and participation in the DNBC. Furthermore, we studied if associations between the individual socioeconomic factors and participation were modified by other socioeconomic factors.

## Methods

### Recruitment to the Danish National Birth Cohort

The DNBC source population included all clinically recognized pregnancies among Danish speaking women who wanted to carry the pregnancy to term and who were residing in Denmark during the time of recruitment. The first Danish counties began enrolment in 1996 and from 1999 to 2002 the DNBC included all Danish counties. The Danish general practitioners (GP) were asked to take part in the recruitment of pregnant women as they would see almost all eligible women at least once during pregnancy and were usually their first contact with the antenatal health care system. Participants were therefore invited to the study at the first antenatal visit to their GP [6]. However, a large part of the GPs did not participate in the recruitment and consequently, about 40% of the eligible women never received an invitation. It has been suggested that 60% of the invited women participated in the DNBC but the actual participation among invitees remains unknown as the GPs did not accept reporting of their own participation in the recruitment [6]. Since we had no data to separate the two sources of selection, the study was carried out for all eligible women.

### Identification of the source population

We studied the non-participation in the DNBC in two well defined geographic areas of Denmark by means of two birth registers established independently of the DNBC. The birth register of the North Jutland County contains information about all births in this area and in this data source, we identified all singleton births that were eligible for

inclusion to the DNBC during the recruitment period in this county (1997–2002) ( $n = 30,628$ ). Likewise, the Aarhus Birth Cohort—an ongoing data collection on all infants scheduled to delivery at Aarhus University Hospital—provided information about all eligible births in the Aarhus Municipality (1997–2002) ( $n = 19,123$ ) as this is the only hospital which handles deliveries in the Aarhus Municipality including the very few home deliveries in the area. From these two registers, we retrieved information about gestational age at birth and CPR-numbers of the mother and the child which is a unique personal identifier given to all Danish citizens and permanent residents for life [22].

### Socioeconomic data

Socioeconomic information was derived from national registers at Statistics Denmark by linking to the women's CPR-numbers. The registers contain historic information which is collected continuously for the entire Danish population from the Civil Registration System, taxation authorities and all Danish worksites for whom it is compulsory to report certain information about their employees [23]. Socioeconomic variables were chosen a priori and were identified for every birth for the calendar year of conception except for the women's income and social benefits which were identified for the calendar year previous to the year of conception.

### Exclusions and identification of the participants

In 1191 cases, no socioeconomic information was found in Statistics Denmark and these births (2.4%) were excluded from the analyses. Following this exclusion, the source population of the present study included 48,560 singleton births equivalent to 16% of the entire DNBC source population. The participants were identified by the CPR-number of the infant which was the only information obtained directly from the DNBC.

### Definitions of socioeconomic groups

Categorization of the socioeconomic variables is displayed in Table 1. Educational groups were defined according to highest level of education. Students were included at the level of education they were studying at as opposed to the highest level they had completed. The Danish compulsory school system includes grade 10. 'Beyond compulsory school' included upper secondary education and university-preparatory schools corresponding to grade 11–12 as well as vocational education whereas 'academic level' referred to tertiary education exceeding grade 12. Health professionals were either students of or had a health-related profession—approximately 60% in this group were nurses.

**Table 1** Participation in socioeconomic groups and estimate of participation bias

Socioeconomic groups	Source population		DNBC		Participation (%)	Prevalence ratio		
	<i>n</i>	Prevalence (%)	<i>n</i>	Prevalence (%)		Estimate	95% CI	
All	48,560		15,290		31	1	–	–
Level of education								
No or unknown	898	1.8	50	0.3	6	0.18	0.14	0.23
Compulsory school	9,254	19.1	1,656	10.8	18	0.57	0.55	0.59
Beyond compulsory school	19,932	41.0	6,162	40.3	31	0.98	0.97	1.00
Academic level	18,476	38.0	7,422	48.5	40	1.28	1.26	1.29
Health professional								
Yes	4,909	10.1	2,042	13.4	42	1.32	1.28	1.36
No	43,651	89.9	13,248	86.6	30	0.96	0.96	0.97
Occupation								
Students	6,107	12.6	2,181	14.3	36	1.13	1.10	1.17
Employees	31,273	64.4	11,179	73.1	36	1.14	1.13	1.14
Self-employed	1,100	2.3	305	2.0	28	0.88	0.80	0.97
Unemployed	2,968	6.1	766	5.0	26	0.82	0.77	0.87
Not in work force	7,109	14.6	859	5.6	12	0.38	0.36	0.41
Income, women								
Lower 25%	14,603	30.1	3,571	23.4	24	0.78	0.76	0.80
Middle 50%	22,997	47.4	7,456	48.8	32	1.03	1.02	1.04
Upper 25%	10,960	22.6	4,263	27.9	39	1.24	1.21	1.26
Income, partners								
Single women	5,088	10.5	1,158	7.6	23	0.72	0.69	0.76
Lower 25%	10,892	22.4	2,807	18.4	26	0.82	0.80	0.84
Middle 50%	21,700	44.7	7,364	48.2	34	1.08	1.06	1.09
Upper 25%	10,880	22.4	3,961	25.9	36	1.16	1.13	1.18
Social benefits, women								
More than 2/3	7,710	15.9	1,514	9.9	20	0.62	0.60	0.65
1/3 to 2/3	6,029	12.4	1,904	12.5	32	1.00	0.97	1.04
Less than 1/3	12,153	25.0	4,112	26.9	34	1.07	1.05	1.10
No social benefits	22,668	46.7	7,760	50.8	34	1.09	1.07	1.10
Social benefits, partners								
Single women	5,088	10.5	1,158	7.6	23	0.72	0.69	0.76
More than 2/3	2,949	6.1	516	3.4	17	0.56	0.51	0.60
1/3 to 2/3	2,553	5.3	762	5.0	30	0.95	0.89	1.00
Less than 1/3	9,126	18.8	2,845	18.6	31	0.99	0.96	1.02
No social benefits	28,844	59.4	10,009	65.5	35	1.10	1.09	1.11
Civil status								
Single	5,088	10.5	1,158	7.6	23	0.72	0.69	0.76
Not single	43,472	89.5	14,132	92.4	33	1.03	1.03	1.04
Age of women (at time of childbirth)								
<20	788	1.6	102	0.7	13	0.41	0.34	0.49
20–25	6,363	13.1	1,482	9.7	23	0.74	0.71	0.77
25–30	18,457	38.0	6,310	41.3	34	1.09	1.07	1.10
30–35	16,086	33.1	5,349	35.0	33	1.06	1.04	1.07
35–40	6,017	12.4	1,848	12.1	31	0.98	0.94	1.01
40+	849	1.7	199	1.3	23	0.74	0.66	0.84

95% CI 95% confidence interval

The occupation group ‘not in the work force’ mainly consisted of women on welfare who did not qualify as unemployed but the group also included women on sick leave, women on leave for other reasons and women receiving early retirement pension. A partner was a person married to, living in a registered partnership with, or cohabitating with the woman in the year of the conception. Women who did not have a partner were considered to be single. We obtained information about the women’s as well as their partners’ annual gross income which included social benefits (welfare, pensions and student grants). The distribution of gross income of all Danish women between ages 18 and 46 and the distribution of the gross income of their respective partners were used to define a low income group (<25th percentile) and a high income group (>75th percentile) which were applied in this study. The percentiles were set separately for every calendar year thereby controlling for inflation. Furthermore, the variable ‘social benefits’ identified how much of the income arose from social benefits.

### Ethics

The study was based on already existing data and did not involve any contact with the participants. All analyses were done on anonymized data and the study was approved by the Danish Data Protection Board.

### Statistical methods

In order to describe the socioeconomic pattern of participation in the DNBC we compared the distribution of socioeconomic factors in the source population and in the participating population by computing the prevalence ratio (PR) of each socioeconomic group. As the prevalence ratio is the prevalence of a given subgroup among the participants divided by the prevalence of the same subgroup in the source population it is identical to the ratio between the participation rate in the subgroup and the overall participation rate. Confidence limits of the PR were found using a simple approximate formula (1). It should be noted that (1) is asymptotically valid if the participants are a random sample of the source population. However, the formula has, in a simulation study, shown to give valid confidence intervals for odds ratios in the estimation of selection bias [2]. A two-sided significance level of 0.05 was used in all statistical tests.

$$\begin{aligned} & SE(\log(\text{PR})) \\ &= \sqrt{SE(\log(\text{Prevalence}_{\text{Participants}}))^2 - SE(\log(\text{Prevalence}_{\text{Source}}))^2}. \end{aligned} \quad (1)$$

We also studied whether these associations were modified by other socioeconomic factors by calculating

participation rates for every subgroups defined by various combinations of two factors. As part of this, we evaluated to which extent the overall socioeconomic pattern of participation could be found within individual socioeconomic groups. Groups with less than 20 participants were not taken into account.

### Results

The DNBC source population of 48,560 pregnancies included 15,290 participants corresponding to an overall participation rate at enrolment of 31% (Table 1).

The participation rate varied considerably according to the women’s education, occupation, income and civil status. A low level of education was associated with low participation and only 11% of the participants had no further education than compulsory school although this group made up 19% of the source population. Consequently, these women were underrepresented by 43% in the DNBC (Table 1: PR 0.57 (0.55; 0.59)). The participants were more often students or employees and they were less often self-employed or unemployed. In particular women outside the work force were strongly underrepresented. We also found a strong association between low income and low participation both when looking at the income of the women and of their partners. Single mothers were underrepresented to the same degree as the low income group whereas women who themselves or their partners were mainly living off social benefits were even less likely to participate. Generally, the level of education and occupation seemed to be stronger predictors of participation than other socioeconomic factors when looking at the magnitude of the PRs. The highest participation was found among health professionals and lowest participation was found among women with no or unknown education who were underrepresented by 82% (Table 1). Women younger than 20 years were underrepresented by almost 60% whereas the underrepresentations of 20–25 year old women and of women older than 40 were modest (Table 1).

### Variations in participation within socioeconomic groups

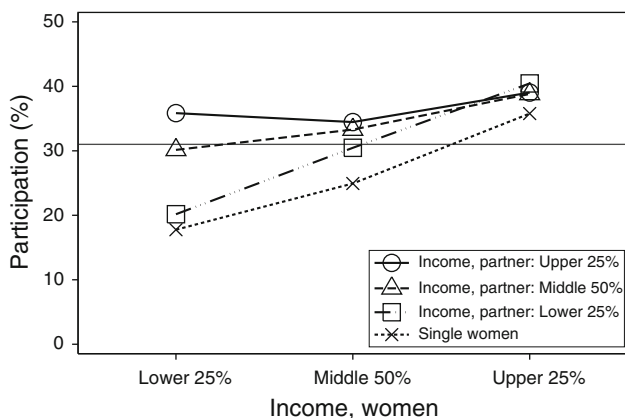
Our data suggested that the associations between individual socioeconomic factors and participation differed according to other socioeconomic factors. For instance, the association between the women’s income and participation was modified by the income of their partners as participation varied a lot according to the women’s income if their partners’ income was low or if the women were single whereas almost no variation in participation was found if

their partners' income was high (Fig. 1). Similar tendencies were found when looking at other socioeconomic factors: greater variations in participation according to other socioeconomic factors were found in groups with an overall low participation compared to groups with an overall average or high participation. So, mainly women who were part of two groups associated with low participation, for instance women with low education and low income (participation 12%), were strongly underrepresented in the DNBC (Fig. 2).

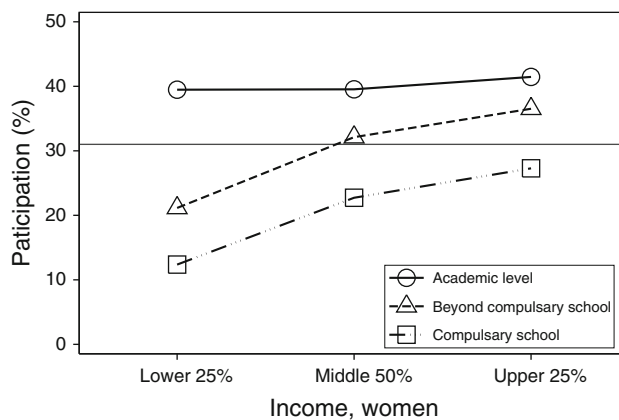
However, we observed some deviation from this pattern. The participation among students was higher than average but great variation in participation was found within the group: only university students were overrepresented whereas students in compulsory school were strongly underrepresented (participation 11%), (Fig. 3). Another exception was found among women outside the workforce who, regardless of their economic status and level of education, were considerably underrepresented (Fig. 3).

**Discussion**

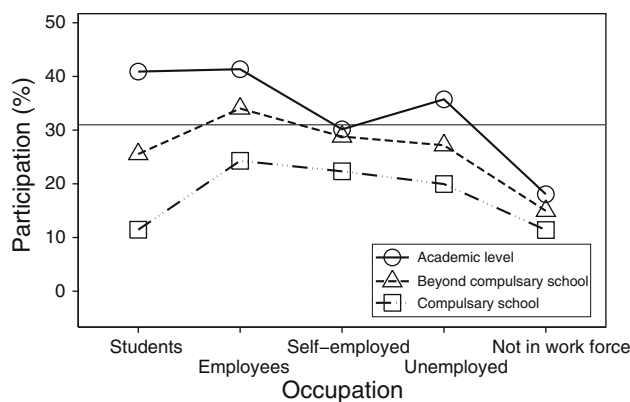
The participation rate among pregnant women invited to the DNBC was low and compatible with other contemporary population based cohort studies [5, 9, 14, 16, 17]. The analyses clearly show a social gradient in the willingness to enroll in the DNBC and corroborate that people with low socioeconomic resources tend to be underrepresented in cohort studies. Furthermore, the study provides detailed information about the socioeconomic pattern of participation in a population of pregnant women showing that women with limited education or low income, women who were mainly living off social benefits or were outside the work force, singles and very young women were strongly underrepresented in the DNBC. Health professionals and



**Fig. 1** Participation (%) according to income stratified on partners' income



**Fig. 2** Participation (%) according to income stratified on level of education



**Fig. 3** Participation (%) according to occupation stratified on level of education

students taking an academic degree were more likely to participate than any other groups and education seemed to be a stronger predictor of participation than income.

These tendencies were to some extent modified by other socioeconomic factors first and foremost showing large underrepresentations of groups with low socioeconomic resources in two areas, for example women with low education and low income, students at compulsory school, singles in the low income group or families where both parents had a low income. These groups were typically underrepresented by more than 50% (data not shown) whereas women who had higher resources in at least one socioeconomic area, for example had a partner with a high income, were only moderately underrepresented.

It is usually difficult to get information about non-participants in cohort studies [4] but the pregnancy event offered a unique opportunity to identify the source population because almost all pregnant women in Denmark get in contact with the public hospital system. This was done in two geographical areas of Denmark representing a 16%



sample of the DNBC source population. Socioeconomic information was available in national registers enabling us to describe participants and non-participants by the same socioeconomic indicators of high validity without recall bias and with low risk of misclassification compared to self-reported data [24]. The variety of data found in the registers also enabled us to give a detailed description of the selection to the cohort by studying different socioeconomic factors rather than one composite variable. Furthermore, the results were estimated with high precision given the large sample size and the quality of registers and only 2.4% (1191) were excluded due to missing data. In Scandinavian countries national registers have not before been used to study non-participation in a population of pregnant women.

One weakness of the study was the lack of evidence of similarity between the participation pattern in the present sample and that of the entire DNBC but the sample has been considered representative of the cohort although the capital city is not represented [2]. Despite of this, the study clearly shows that the DNBC does not consist of a socioeconomically representative sample of its source population. This compromises our ability to generalize descriptive results like the prevalence of exposures (e.g. folic acid intake during pregnancy) and health outcomes to the general population if these data are related to socioeconomic factors. Furthermore, concern remains about our ability to generalize results from the DNBC to low socioeconomic groups and we may have too limited information about women with really low socioeconomic resources to study this group separately. However, the aim of the DNBC was to study diseases in a life course perspective related to specific exposures. Selection on socioeconomic factors can but need not bias effect measures and cohort studies are considered to be less vulnerable to selection bias than other designs because decision to join the study can only be based on information available at baseline [25]. Nevertheless, it will always be difficult to predict whether and to what extent the socioeconomic selection will bias a given risk association [3, 26].

Although a participation rate of only 31% leaves room for substantial selection bias, it is important to note that non-participation was only partly related to the eligible women's own decision. About half of the non-participation was related to the GPs. It is possible that the GPs who decided not to take part in the recruitment had patients with a different socioeconomic distribution than the general population. It is also possible that some GPs did not provide the invitation to pregnant women whom they considered unfit or unable to participate in the study. These uncertainties weaken the generalizability to other cohort studies. This being said, a somewhat stronger socioeconomic selection should be expected in similar cohorts with

participation rates of 30% if the shown selection was primarily determined by the invitees themselves.

This study demonstrates that the socioeconomic resources of pregnant women are correlated with decision to participate in cohort studies and that the least privileged groups are likely to be underrepresented by more than 50%. It is also of interest that the health professionals, who probably had the best background for understanding the purpose of the DNBC, were the ones who were most willing to participate. These matters need to be taken into account when results from the DNBC and other pregnancy cohorts are interpreted—especially when descriptive results are presented. Also, pregnancy cohorts are now being established in many Western countries, and we expect these findings to be highly relevant when planning recruitment strategies for these studies.

**Acknowledgments** We thank Professor Jørn Olsen for valuable and constructive suggestions when writing this manuscript.

## References

- Hartge P. Raising response rates: getting to yes. *Epidemiology*. 1999;105–7.
- Nohr EA, Frydenberg M, Henriksen TB, Olsen J. Does low participation in cohort studies induce bias? *Epidemiology*. 2006;17(4):413.
- Austin MA, Criqui MH, Barrett-Connor E, Holdbrook MJ. The effect of response bias on the odds ratio. *Am J Epidemiol*. 1981;114(1):137–43.
- Stang A. Nonresponse research—an underdeveloped field in epidemiology. *Eur J Epidemiol*. 2003;18(10):929–32.
- Lundberg I, Damström Thakker K, Hällström T, Forsell Y. 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(6):475–83.
- Olsen J, Melbye M, Olsen SF, Sorensen TI, Aaby P, Andersen AM, et al. The Danish National Birth Cohort—its background, structure and aim. *Scand J Public Health*. 2001;29(4):300–7.
- Shahar E, Folsom AR, Jackson R. The effect of nonresponse on prevalence estimates for a referent population: insights from a population-based cohort study. *Ann Epidemiol*. 1996;6(6):498–506.
- Boeing H, Korfmann A, Bergmann MM. Recruitment procedures of EPIC-Germany. *Ann Nutr Metab*. 1999;43:205–15.
- Brown WJ, Bryson L, Byles JE, Dobson AJ. Women's Health Australia: recruitment for a national longitudinal cohort study. *Women Health*. 1998;28(1):23–40.
- Goldberg M, Chastang JF, Leclerc A, Zins M, Bonenfant S, Bugel I, 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. Oxford: Oxford University Press; 2001.
- Korkeila K, Suominen S, Ahvenainen J, Ojanlatva A, Rautava P, Helenius H, et al. Non-response and related factors in a nationwide health survey. *Eur J Epidemiol*. 2001;17(11):991–9.
- Olsen A, Tjønneland A, Engholm G, Overvad K. Socio-economic determinants for participation in the Danish EPIC Diet, Cancer and Health cohort. IARC scientific publications. 2002;15655.

13. Sogaard AJ, Selmer R, Bjertness E, Thelle D. The Oslo Health Study: the impact of self-selection in a large, population-based survey. *Int J Equity Health*. 2004;3(1):3.
14. Stang A, Moebus S, Dragano N, Beck EM, Möhlenkamp S, Schmermund A, et al. Baseline recruitment and analyses of nonresponse of the Heinz Nixdorf recall study: identifiability of phone numbers as the major determinant of response. *Eur J Epidemiol*. 2005;20(6):489–96.
15. Boshuizen HC, Viet AL, Picavet HSJ, Botterweck A, van Loon AJM. Non-response in a survey of cardiovascular risk factors in the Dutch population: determinants and resulting biases. *Public Health*. 2006;120(4):297–308.
16. Fejer R, Hartvigsen J, Kyvik K, Jordan A, Christensen H, Høilund-Carlsen P. The Funen neck and chest pain study: analysing non-response bias by using national vital statistic data. *Eur J Epidemiol*. 2006;21(3):171–80.
17. Martikainen P, Laaksonen M, Piha K, Lallukka T. Does survey non-response bias the association between occupational social class and health? *Scand J Public Health*. 2007;35(2):212–5.
18. Tjønneland A, Olsen A, Boll K, Stripp C. Study design, exposure variables, and socioeconomic determinants of participation in diet, cancer and health: a population-based prospective cohort study of 57, 053 men and women in Denmark. *Scand J Public Health*. 2007;35(4):432–41.
19. van Heuvelen MJ, Hochstenbach JB, Brouwer WH, de Greef MH, Zijlstra GA, van Jaarsveld E, et al. Differences between participants and non-participants in an RCT on physical activity and psychological interventions for older persons. *Aging Clin Exp Res*. 2005;17(3):236.
20. Boesen E, Boesen S, Christensen S, Johansen C. Comparison of participants and non-participants in a randomized psychosocial intervention study among patients with malignant melanoma. *Psychosomatics*. 2007;48(6):510.
21. Huang N, Shih S, Chang H, Chou Y. Record linkage research and informed consent: who consents? *BMC Health Serv Res*. 2007; 7(1):18.
22. Det Centrale Personregister [Internet]. Det Centrale Personregister. 2009 September 6; [cited 2009 Sep 6] Available from: <http://www.cpr.dk/cpr/>.
23. Danmarks Statistik/Statistics Denmark. Danmarks Statistik [Internet]. Danmarks Statistik. 2009 September 6; [cited 2009 Sep 6] Available from: <http://www.dst.dk/>.
24. Thygesen L. The register-based system of demographic and social statistics in Denmark. *Stat J UN Econ Comm Eur*. 1995;12(1):49.
25. Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*. Philadelphia: Lippincott Williams & Wilkins; 2008.
26. Hernán MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15(5):615.