BMJ Open

Validity of self-reported myocardial infarction and stroke in regions with Sami and Norwegian populations: the SAMINOR 1 Survey and the CVDNOR project

Bent-Martin Eliassen,[1] Marita Melhus,[1] Grethe S Tell,[2,3] Kristin Benjaminsen Borch,[4] Tonje Braaten,[4] Ann Ragnhild Broderstad,[1,5] Sidsel Graff-Iversen[4,6]

For numbered affiliations see end of article.

Correspondence to
Dr Bent-Martin Eliassen;
bent-martin.eliassen@uit.no

ABSTRACT

Objective: Updated knowledge on the validity of self-reported myocardial infarction (SMI) and self-reported stroke (SRS) is needed in Norway. Our objective was to compare questionnaire data and hospital discharge data from regions with Sami and Norwegian populations to assess the validity of these outcomes by ethnicity, sex, age and education.

Design: Validation study using cross-sectional questionnaire data and hospital discharge data from all Norwegian somatic hospitals.

Participants and setting: 16 865 men and women aged 30 and 36–79 years participated in the Population-based Study on Health and Living Conditions in Sami and Norwegian Populations (SAMINOR) 1 Survey in 2003–2004. Information on SMI and SRS was available from self-administered questionnaires for 15 005 and 15 088 of these participants, respectively. We compared this information with hospital discharge data from 1994 until SAMINOR 1 Survey attendance.

Primary and secondary outcomes: Sensitivity, specificity, positive predictive value (PPV), negative predictive value and κ.

Results: The sensitivity and PPV of SMI were 90.1% and 78.9%, respectively; the PPV increased to 93.1% when all ischaemic heart disease (IHD) diagnoses were included. The SMI prevalence estimate was 2.3% and hospital-based 2.0%. The sensitivity and PPV of SRS were 81.1% and 64.3%, respectively. The SRS prevalence estimate was 1.5% and hospitalisation-based 1.2%. Moderate to no variation was observed in validity according to ethnicity, sex, age and education.

Conclusions: The sensitivity and PPV of SMI were high and moderate, respectively; for SRS, both of these measures were moderate. Our results show that SMI from the SAMINOR 1 Survey may be used in aetiological/analytical studies in this population due to a high IHD-specific PPV. The SAMINOR 1 questionnaire may also be used to estimate the prevalence of acute myocardial infarction and acute stroke.

Strengths and limitations of this study

- The overall results are considered representative for the general population aged 36–79 years in the SAMINOR 1 region.
- Limited selection bias with regard to hospital discharge records.
- Imperfect reference standard.

INTRODUCTION

In the absence of registries, epidemiological studies often use questionnaires to obtain information on cardiovascular disease (CVD). Validation studies published within the past decade and onwards used different instruments and reference standards, and indicated moderate-to-high sensitivity for self-reported myocardial infarction (SMI) (78–98%)[1–4] and self-reported stroke (SRS) (69–81%),[1–5] whereas reported positive predictive values (PPVs) were lower (SMI: 43–74%;[1–4 6 7] SRS: 22–79%[1–8]).

There have been a few validity and reliability studies of SMI and SRS in Norway, but recent studies are lacking.[8–10] Moreover, no studies until now have assessed the validity of these questions in the indigenous Sami population of Norway. We used questionnaire data from the SAMINOR 1 Survey, the first survey of the Population-based Study on Health and Living Conditions in Sami and Norwegian Populations,[11] and hospital discharge data from the Cardiovascular Disease in Norway (CVDNOR) project[12] to assess the validity of SMI and SRS by ethnicity, sex, age and education.

METHODS
Study population

In 2003–2004, in collaboration with the Norwegian Institute of Public Health, the Centre for Sami Health Research conducted the SAMINOR 1 cross-sectional survey.[11] Men and women aged 30 (born 1973–1974) and 36–79 years (born 1925–1968) residing in selected regions[11] were identified through the National Registry and invited to participate in the survey, regardless of their ethnic background (N=27 987). Of those invited, 16 865 (60.3%) agreed to participate. Apart from those recruited from the municipality of Alta (n=4741), the population was exclusively rural (ibid.).

The design of the SAMINOR 1 Informed consent was obtained from all participants. Survey changed during the data collection. In the original design, which was used in the predominantly Sami municipalities of Tana, Nesseby, Karasjok and Kautokeino in Finnmark County, a two-paged initial questionnaire (Q1) was first sent. Those who replied and stated that they wanted to take part in the screening received an invitation to a clinical examination. The invitation included a three-paged screening questionnaire—the Q2—which they were asked to fill in and bring to the examination. The clinical examinations took place in two buses travelling between the municipalities. This two-stage design led to a very low attendance at the clinical examination. Hence, after the first four municipalities, the two-stage sampling design was dropped. In the rest of the survey, the Q1 and Q2 were combined into a five-paged questionnaire which was sent together with the invitation to the clinical examination. In Finnmark and Troms, the buses returned after 2 or 3 months, and non-responders got a second invitation to the screening. In Nordland and Trøndelag, there was only one screening round and no reminder. Owing to the original design, not all participants provided the Q1 and the Q2. The data collection is described in detail by Lund et al (ibid.). The questionnaires in Norwegian, Sami and English languages may be accessed at http://www.saminor.no.

Of the 16 865 participants, 815 did not complete the Q2—which includes the questions on CVD to be validated in this paper—and another 87 did not consent to having their information linked to national registries. Thus, a total of 15 963 participants were eligible for inclusion in this validation study.

Separate analyses were conducted for SMI and SRS; participants were included in both analyses if both events were reported. Excluded were those with missing information on SMI (n=676) and/or SRS (n=720). Participants who did not report their age at the time of myocardial infarction (n=54) and/or stroke (n=46) were also excluded. The CVDNOR project does not have hospital discharge data before 1994, and thus participants with SMI (n=228) and/or SRS (n=109) before this year were also excluded. The total numbers included in the final SMI and SRS analyses were 15 005 and 15 088, respectively.

Questionnaire data

Self-administered questionnaires were used to collect information on disease, ethnicity and education. SMI and SRS were measured by the question: *Do you have, or have you had*: 'a myocardial infarction' and/or 'a stroke/ brain haemorrhage?' Response options were 'yes' or 'no', and participants who answered yes were asked to report their age at the time of the first event.

Ethnicity was determined by the questions: *what language(s) do/did you, your parents, and your grandparents use at home*? And *what is your, your father's and your mother's ethnic background*? The response options for both questions were 'Norwegian', 'Sami', 'Kven' or 'other'. The respondents also reported whether they considered themselves to be Norwegian, Sami, Kven or other (self-perceived ethnicity). On all the aforementioned questions, multiple answers were allowed, and participants were categorised as Sami if they answered 'Sami' to one or more of the questions above. Respondents who were born abroad and answered 'other' to questions on ethnicity (n=263) were considered missing values. The remaining participants were categorised as non-Sami.

Education was assessed by the question: *how many years of schooling/education have you completed? (including all the years you attended school or have been studying.)* Information on age and sex was collected from the National Registry.

Hospital discharge data

Hospital discharge data were taken from the CVDNOR project (http://www.cvdnor.no), which is a collaborative project between the University of Bergen and the Norwegian Knowledge Centre for the Health Services. The CVDNOR project retrieved information on all hospital stays with a primary or secondary diagnosis of CVD or diabetes mellitus (International Classification of Diseases, Ninth Revision (ICD-9) codes 250, 390–459, 745–747; and ICD-10 codes E10–E14, I00–I99, Q20–Q28), as well as all related procedures, from the Patient Administrative Systems (PAS) of all Norwegian somatic hospitals from 1994 (the year from which all hospitals adopted an electronic PAS) through 2009.[13] The CVDNOR data contain information on patients' date and time of hospital admission and discharge, and up to 21 discharge diagnoses.[12]

Reference standards

The validity of SMI and SRS was assessed against hospital discharge data by linking SAMINOR 1 and CVDNOR data using the unique 11-digit national identification number assigned to each Norwegian resident. We used main and secondary hospital discharge diagnosis codes as reference standards including ICD-9: 410, or ICD-10: I21 or I22 for acute myocardial infarction (AMI); and ICD-9: 430, 431, 434 or 436, or ICD-10: I60, I61, I63 or I64 for cerebral infarction, intracerebral haemorrhage and subarachnoid haemorrhage (acute stroke). SMI and SRS were compared with hospital discharge data from 1994 until SAMINOR 1 Survey attendance.

Statistical analyses

SMI-specific and SRS-specific analyses were performed using STATA V.14.1 (StataCorp., College Station, Texas, USA). Sensitivity was the proportion of people hospitalised for CVD, who self-reported CVD a/(a+c). Specificity was calculated as the proportion of non-hospitalised people who did not report a CVD d/(d+b). PPV was the proportion of people who reported CVD who were hospitalised for CVD a/(a+b). Negative predictive value (NPV) was the proportion of people who did not report CVD who were not hospitalised for CVD d/(c+d) (figure 1).

The prevalence of self-reported CVD was calculated using this formula: (a+b)/(a+b+c+d). The prevalence of CVD based on hospital data was calculated by the following formula: (a+c)/(a+b+c+d).

Since the sensitivity, specificity, NPV and PPV are proportions, statistical methods based on the binomial distribution are used. The 95% CIs are exact binomial (Clopper-Pearson intervals) and were computed using the 'diagt' command in STATA. Owing to large groups, the Pearson's χ^2 test is used for comparisons.[14]

Cohen's κ was calculated to determine the agreement between self-reported information in the SAMINOR 1 questionnaire and hospital discharge data from CVDNOR. The 'kapci' command in STATA was used. Differences in κ statistics were tested using the following test statistic:[15]

$$Z = \frac{|\kappa_A - \kappa_B|}{\sqrt{\mathrm{Var}(\kappa_A) + \mathrm{Var}(\kappa_B)}} \sim N(0,1).$$

An alternative categorisation of ethnicity was assessed by comparing those who reported speaking Sami language at home with those who did not speak Sami. Potential selection bias was assessed by checking how many of those with SMI and SRS and no information on age at the time of the event were found in the hospital discharge data. We also checked how many of those with missing self-reported disease status that were found with a relevant hospital discharge code.

RESULTS

Non-responders tended to be male, younger, non-married and from Nordland County (table 1).

	Reference standard	
Self-report of CVD	Yes	No
Yes	a	b
No	c	d

Figure 1 Calculation of sensitivity, specificity, PPV, NPV and prevalence. CVD, cardiovascular disease; NPV, negative predictive value; PPV, positive predictive value.

Poorer response to Q2 was observed in Sami compared with non-Sami; this was due to initial design issues in predominantly Sami municipalities in Finnmark.[11]

In the total sample, the sensitivity and PPV for SMI were 90.1% and 78.9%, respectively; specificity and NPV were >99.4%. The prevalence of SMI was 2.3%, while the hospital discharge data showed a prevalence of 2.0%. The corresponding κ value was 0.84 (table 2). In sensitivity analyses, only 3 of the 676 participants excluded due to missing self-reported disease status were found to have had an AMI in the hospital discharge data, as were 16 of the 54 participants who failed to report their age at the time of the event (data not shown).

Owing to the small numbers of SMI and hospital discharge events, precision is lacking and only very large differences in the sensitivity, PPV with regard to ethnicity, sex, age, and education would have been statistically significant in our study. Higher sensitivity (96.6% vs 87.4%) was observed in those aged $\leq$59 years compared with those aged $\geq$60 years (p<0.05). Small but statistically significant differences (p<0.05) were found for sex, age and education in terms of specificity and NPV which were consistently $\geq$99.0%. No significant ethnic variation in the validity of SMI was observed (table 2).

In the total sample, the sensitivity and PPV for SRS were 81.1% and 64.3%, respectively; specificity and NPV were >99.4. The prevalence of SRS was 1.5%, while the hospital discharge data gave a prevalence of 1.2%. The κ was 0.71 (table 3). Sensitivity analyses showed that only 4 of the 720 participants excluded due to missing SRS status had an acute stroke event as of the hospital discharge data (data not shown). Furthermore, only 10 of the 46 participants who failed to report their age at disease onset had a relevant discharge code (data not shown).

As with SMI, statistical power and precision are low when comparing the sensitivity, PPV and κ between categories of the selected demographic variables. A substantially higher PPV (81.8% vs 59.1%) was found among participants with $\geq$12 years compared with 0–11 years of education (p<0.05). Small but statistically significant (p<0.05) differences were observed for age and education in terms of specificity and NPV, as well as for sex in terms of NPV; all were >99.1. Statistically significant (p<0.05) stronger κ coefficients were observed in participants aged 30–59 years compared with those aged 60–79 years (0.76 vs 0.68), and for education $\geq$12 years compared with 0–11 years (0.83 vs 0.67) (table 3).

Of the 346 SMIs, 73 could not be verified in the hospital discharge data. However, 43 of these (12% of all SMIs) had a discharge code of angina pectoris (ICD-9: 411, 413; ICD-10: I20), and 6 (2% of all SMIs) had a discharge code of ischaemic heart disease (IHD) (ICD-9: 412, 414; ICD 19: I24–I25) other than AMI and angina pectoris (table 4). This gave an IHD-specific PPV of 93.1%.

Among the 221 participants with SRS, we were unable to find a relevant stroke discharge code for 79; among these, we found that 7 (3% of all SRSs) and 12 (5% of

Table 1 Characteristics of the invited cohort in the SAMINOR 1 Survey (2003–2004) aged 30 and 36–79 years, total participants,* respondents to the Q2† and included subsamples

Number Attendance (%)	Invited (%) 27 987 100	Total participants (%)* 16 865 60.3	Completed the Q2† 16 050 57.4	SMI (%)‡ 15 005 53.6	SRS (%)§ 15 088 53.9
Sex					
Women	13 446 (48)	8755 (52)	8389 (52)	7872 (52)	7867 (52)
Men	14 541 (52)	8110 (48)	7661 (48)	7133 (48)	7221 (48)
Age (years)					
30–59	19 323 (69)	11 430 (68)	10 878 (68)	10 381 (69)	10 361 (69)
60–79	8664 (31)	5435 (32)	5172 (32)	4624 (31)	4727 (31)
County					
Trøndelag	1526 (5)	994 (6)	994 (6)	957 (6)	955 (6)
Nordland	2666 (10)	1223 (7)	1223 (8)	1146 (8)	1151 (8)
Troms	6725 (24)	3998 (24)	3996 (25)	3544 (24)	3562 (24)
Finnmark	17 070 (61)	10 650 (63)	9837 (61)	9358 (62)	9420 (62)
Marital status					
Single	7057 (25)	3415 (20)	3159 (20)	2981 (20)	2980 (20)
Married	15 394 (55)	10 354 (62)	9953 (62)	9327 (62)	9388 (62)
Widow(er)	1826 (7)	1066 (6)	1015 (6)	905 (6)	924 (6)
Divorced	3071 (11)	1712 (10)	1624 (10)	1506 (10)	1510 (10)
Separated	638 (2)	318 (2)	299 (2)	286 (2)	286 (2)
Ethnicity					
Non-Sami		10 390 (62)	10 222 (64)	9546 (64)	9618 (64)
Sami		5915 (35)	5297 (33)	4963 (33)	4982 (33)
Missing/born abroad		560 (3)	531 (3)	496 (3)	488 (3)
Education					
0–11 years			8 635 (54)	7959 (53)	8034 (53)
≥12 years			6424 (40)	6182 (41)	6181 (41)
Missing			991 (6)	864 (6)	873 (6)

*Participants who consented to medical research, and completed at least one questionnaire or attended the clinical examination.
†Participants who completed the screening questionnaire (Q2).
‡Participants who consented to having their information linked to national registries, and responded to the question: *Do you have, or have you had*: 'a myocardial infarction?' (SMI).
§Participants who consented to having their information linked to national registries, and responded to the question: *Do you have, or have you had*: 'a stroke/brain haemorrhage?' (SRS).
SMI, self-reported myocardial infarction; SRS, self-reported stroke.

all SRSs) had discharge codes referring to a transient ischaemic attack (TIA) (ICD-9: 435.9; ICD-10: G45.9) or other cerebrovascular diseases (ICD-9: 432–433, 435, 437–438; ICD-10: I65–I69), respectively (table 4). None had a code referring to intracranial injury (850–854, S06) or amaurosis fugax (362.34, G45.3) (data not shown).

The alternative definition of ethnicity did not affect results.

DISCUSSION

The sensitivity (90.1%) and PPV (78.9%) of SMI were high and moderate, respectively; the PPV increased to 93.1% when all IHD diagnoses were considered. For SRS, both of these measures were moderate (81.1% and 64.3%). The self-reported prevalence estimates were almost identical to those based on hospital discharge data for both SMI and SRS. Moderate to no variation was observed in validity according to ethnicity, sex, age and education.

The validity of a questionnaire may vary across populations, time periods and social circumstances.[16] Furthermore, when comparing results from validation studies, it is important to consider the reference standard used, and which information was obtained from questionnaires. Indeed, these variables vary widely between studies, which may contribute to the disparity that has been observed in this field of research.[1] Nevertheless, we have identified some common findings between our study and recent studies based on population-based survey data that included both sexes and middle-aged and older participants.[1–8]

A typical AMI is easily recalled, as it usually involves severe pain and hospitalisation, which enhance the validity of SMI.[3] Only a few studies have provided κ values for SMI, but the agreement they[1 3 4] reported was lower than that found in our study (κ=0.84). However, κ measurements may be affected by marginal totals and disease prevalence.[17 18] While the corresponding sensitivity estimate for SMI in our study was high (90.1%), which is in accordance with recent literature,[1–4] our PPV

Table 2 Validity of self-reported myocardial infarction in participants aged 30 and 36–79 years by ethnicity, sex, age and education. The SAMINOR 1 Survey (2003–2004) and the CVDNOR project, n=15 005.

Self-reported myocardial infarction		PAS† Yes	No	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	κ (95% CI)
Total								
	Yes	273	73	90.1%	99.5%	78.9%	99.8%	0.84
	No	30	14 629	(86.2 to 93.2)	(99.4 to 99.6)	(74.2 to 83.1)	(99.7 to 99.9)	(0.81 to 0.87)
Ethnicity‡								
Sami	Yes	91	28	90.1%	99.4%	76.5%	99.8%	0.82
	No	10	4834	(82.5 to 95.1)	(99.2 to 99.6)	(67.8 to 83.8)	(99.6 to 99.9)	(0.77 to 0.88)
Non-Sami§	Yes	174	44	91.1%	99.5%	79.8%	99.8%	0.85
	No	17	9311	(86.1 to 94.7)	(99.4 to 99.7)	(73.9 to 84.9)	(99.7 to 99.9)	(0.81 to 0.89)
Sex								
Women	Yes	73	19	89.0%	99.8%*	79.3%	99.9%*	0.84
	No	9	7771	(80.2 to 94.9)	(99.6 to 99.9)	(69.6 to 87.1)	(99.8 to 99.9)	(0.78 to 0.90)
Men	Yes	200	54	90.5%	99.2%	78.7%	99.7%	0.84
	No	21	6858	(85.8 to 94.0)	(99.0 to 99.4)	(73.2 to 83.6)	(99.5 to 99.8)	(0.80 to 0.87)
Age								
30–59 years	Yes	85	28	96.6%*	99.7%*	75.2%	100%*	0.84
	No	3	10 265	(90.4 to 99.3)	(99.6 to 99.8)	(66.2 to 82.9)	(99.9 to 100)	(0.79 to 0.90)
60–79 years	Yes	188	45	87.4%	99.0%	80.7%	99.4%	0.83
	No	27	4364	(82.3 to 91.6)	(98.6 to 99.3)	(75.0 to 85.6)	(99.1 to 99.6)	(0.79 to 0.87)
Education¶								
0–11 years	Yes	195	49	87.8%	99.4%*	79.9%	99.7%*	0.83
	No	27	7688	(82.8 to 91.8)	(99.2 to 99.5)	(74.3 to 84.8)	(99.5 to 99.8)	(0.80 to 0.87)
≥12 years	Yes	49	17	94.2%	99.7%	74.2%	100	0.83
	No	3	6113	(84.1 to 98.8)	(99.6 to 99.8)	(62.0 to 84.2)	(99.9 to 100)	(0.76 to 0.90)

*p<0.05, Column-specific Pearson's χ^2 test for difference between the two categories of each variable.
†Includes primary or secondary discharge diagnoses of ICD-9: 410; ICD-10: I21, I22 from the PAS of all Norwegian somatic hospitals from 1994 through SAMINOR 1 attendance.
‡Smaller n due to missing values: n=496.
§No Sami identity marks reported.
¶Smaller n due to missing values: n=864.
CVDNOR, Cardiovascular Disease in Norway; ICD-9, International Classification of Diseases, Ninth Revision; ICD-10, International Classification of Diseases, Tenth Revision; NPV, negative predictive value; PAS, Patient Administrative Systems; PPV, positive predictive value.

Table 3 Validity of self-reported stroke in participants aged 30 and 36–79 years by ethnicity, sex, age and education. The SAMINOR 1 Survey (2003–2004) and the CVDNOR project, n=15 088.

Self-reported stroke		PAS† Yes	No	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	κ (95% CI)
Total								
	Yes	142	79	81.1%	99.5%	64.3%	99.8%	0.71
	No	33	14 834	(74.5 to 86.6)	(99.3 to 99.6)	(57.6 to 70.6)	(99.7 to 99.9)	(0.66 to 0.76)
Ethnicity‡								
Sami	Yes	50	25	79.4%	99.5%	66.7%	99.7%	0.72
	No	13	4894	(67.3 to 88.5)	(99.3 to 99.7)	(54.8 to 77.1)	(99.5 to 99.9)	(0.64 to 0.81)
Non-Sami§	Yes	90	52	84.1%	99.5%	63.4%	99.8%	0.72
	No	17	9459	(75.8 to 90.5)	(99.3 to 99.6)	(54.9 to 71.3)	(99.7 to 99.9)	(0.66 to 0.78)
Sex								
Women	Yes	58	38	84.1%	99.5%	60.4%	99.9%*	0.70
	No	11	7760	(73.3 to 91.8)	(99.3 to 99.7)	(49.9 to 70.3)	(99.7 to 99.9)	(0.62 to 0.78)
Men	Yes	84	41	79.2%	99.4%	67.2%	99.7%	0.72
	No	22	7074	(70.3 to 86.5)	(99.2 to 99.6)	(58.2 to 75.3)	(99.5 to 99.8)	(0.66 to 0.79)
Age								
30–59 years	Yes	54	25	87.1%	99.8%*	68.4%	99.9%*	0.76**
	No	8	10 274	(76.1 to 94.3)	(99.6 to 99.8)	(56.9 to 78.4)	(99.8 to 100)	(0.69 to 0.84)
60–79 years	Yes	88	54	77.9%	98.8%	62.0%	99.5%	0.68
	No	25	4560	(69.1 to 85.1)	(98.5 to 99.1)	(53.5 to 70.0)	(99.2 to 99.6)	(0.62 to 0.75)
Education¶								
0–11 years	Yes	91	63	77.8%	99.2%*	59.1%*	99.7%*	0.67**
	No	26	7854	(69.2 to 84.9)	(99.0 to 99.4)	(50.9 to 66.9)	(99.5 to 99.8)	(0.60 to 0.73)
≥12 years	Yes	27	6	84.4%	99.9%	81.8%	99.9%	0.83
	No	5	6143	(67.2 to 94.7)	(99.8 to 100)	(64.5 to 93.0)	(99.8 to 100)	(0.73 to 0.93)

*p<0.05, column-specific Pearson's χ^2 test for difference between the two categories of each variable. **p<0.05, test for difference in κ statistics between the two categories of each variable:

$$Z = \frac{|\kappa_A - \kappa_B|}{\sqrt{\text{Var}(\kappa_A) + \text{Var}(\kappa_B)}} \sim N(0, 1).$$

†Includes primary or secondary discharge diagnoses of ICD-9: 430, 431, 434, 436; ICD-10: I60, I61, I63, I64 from the PAS of all Norwegian somatic hospitals from 1994 through SAMINOR 1 attendance.
‡Smaller n due to missing values: n=488.
§No Sami identity marks reported.
¶Smaller n due to missing values: n=873.
CVDNOR, Cardiovascular Disease in Norway; ICD-9, International Classification of Diseases, Ninth Revision; ICD-10, International Classification of Diseases, Tenth Revision; NPV, negative predictive value; PAS, Patient Administrative Systems; PPV, positive predictive value.

Table 4 Hospital discharge codes in those with false-positive self-reported myocardial infarction and stroke. The SAMINOR 1 Survey (2003–2004) and the CVDNOR project.

Self-reported and registered events (ICD-9; ICD-10)	n
Myocardial infarction	73
Angina pectoris (411, 413; I20)	43
Other ischaemic heart diseases (412, 414; I24–I25)	6
An ischaemic heart diseases event	24
Stroke/brain haemorrhage	79
Transient ischaemic attack (435.9; G45.9)	7
Other cerebrovascular diseases (432–433, 435, 437–438; I65–I69)	12
Not a cerebrovascular disease event	60

CVDNOR, Cardiovascular Disease in Norway; ICD-9, International Classification of Diseases, Ninth Revision; ICD-10, International Classification of Diseases, Tenth Revision.

of 78.9% was higher than what has been reported previously.[1–4 6 7] The high specificities and NPVs in our total sample (>99.4%) are also in line with earlier studies.

Twelve per cent of all participants with SMI had a discharge code indicating angina pectoris, not AMI. Norwegian hospitals adopted the use of troponin during 1999–2001.[19] The troponin method is more sensitive than earlier methods and have probably increased the rates of AMI as diagnosis on the cost of angina pectoris. Some patients may have been told that an acute heart attack diagnosed as angina during 1994–2001 could be an AMI according to new diagnostics, thus leading to a subsequent false-positive SMI. In an Australian study, Barr et al[6] found that 18% of all SMIs were in fact unstable angina as per World Health Organization (WHO)/Multinational MONItoring of trends and determinants in CArdiovascular disease (MONICA) criteria.[20] In most instances, IHDs share an underlying pathophysiological condition (atherosclerosis); hence, one may conclude that the high IHD-specific PPV of 93.1% in our study justifies the use of the SAMINOR 1 questionnaire in aetiological/analytical studies of AMI in this population. This study also showed that the SAMINOR 1 questionnaire is valid in terms of estimating the prevalence of AMI.

Similar to AMI, a stroke often has an abrupt onset and causes impairment leading to hospitalisation. The sensitivity (81.1%), PPV (64.3%) and κ (0.71) estimates for SRS in this study were in the higher range of what has been previously found.[1–8] Three and five per cent of those reporting a stroke did not have a hospitalisation for acute stroke, but for TIA or other cerebrovascular diseases, respectively. Engstad et al[8] and Machón et al[1] found that about 6% of SRSs' were in reality TIAs. We observed a large proportion of false-positive self-reports. Given this and the fact that ischaemic stroke and haemorrhagic strokes have different risk factor patterns,[21] one may conclude that the use of SRS is inappropriate in aetiological/analytical studies. However, owing to the low false-negative rate, the variable is valid in terms of estimating the prevalence of acute stroke in this population.

Overall, no significant ethnic differences were observed with regard to the validity of SMI or SRS. We found no papers addressing the validity of questionnaire-assessed CVD in indigenous peoples.

Machón et al[1] found lower sensitivity and higher PPV in men compared with women with regard to both SMI and SRS. Lower PPV for SMI in women was observed by Yamagishi et al.[2] Engstad et al[8] found higher PPV for SRS in men than in women. We found no evidence of this. As mentioned earlier, however, owing to the small numbers of SMI, SRS and hospital discharge events, only very large differences would have been statistically significant in our study. Age ≤60 years was associated with increased sensitivity of SMI and of SRS in our study. Engstad et al[8] found a borderline significantly higher PPV for SRS (83% vs 73%) in those aged >60 years. In the study by Yamagishi,[2] age <65 years compared with age ≥65 years was associated with increased validity for SMI and SRS. Poorer validity in older age may be due to age-related memory loss. For SRS, stroke-induced cognitive impairment may also be a plausible explanation.

Older age in this population was related to fewer years of education, as education opportunities in rural Norway developed gradually after World War II. Increased schooling may be related to the ability to understand health-related information[22] and may potentially influence the ability to correctly report the presence or absence of ill health. SRS rendered higher PPV among those with ≥12 years of education. Few studies have examined the influence of education on the validity of SMI and SRS. Engstad et al[8] found that higher education did not influence the PPV of SRS in the Tromsø study. Similar educational attainment in Sami and non-Sami[23] may be a possible explanation for some of the ethnic homogeneity observed in this study.

Strengths
The relatively large study sample enabled an in-depth analysis of the validity of SMI and SRS. CVDNOR covers the whole population of Norway; thus, there is limited selection bias with regard to hospital discharge records in this study. The total results are considered representative for the general population aged 36–79 years in the SAMINOR 1 region (see below).

Limitations

Overall, non-responders in SAMINOR 1 were more likely to be males, of younger age (especially 30 years),[11] from Nordland County and single (table 1). This is in line with previous population-based studies in Norway having shown poorer participation in the ill, in men and in younger age groups.[24] The results specific to sex and age in this study are thus somewhat uncertain; we believe, however, that the differences in response rates are not big enough as to affect total results.

Furthermore, owing to initial design issues, poorer response to the Q2 in certain municipalities dominated by people of Sami ethnicity was observed. However, response rates were not too different; this has thus probably not affected the results specific to ethnicity and hence the overall results.

Some AMIs may have occurred without symptoms, leading to false-negative self-reporting and an absence of a registered AMI in the hospital discharge data. Furthermore, an AMI could be diagnosed without leading to hospitalisation, resulting in a true-positive SMI not being confirmed in hospital discharge data. However, we believe that this is not a major problem in this study, as hospitalisation among patients with AMI has been common for decades in Norway. In a forthcoming Norwegian study, one found a PPV of ~95% for hospital admissions with a main or secondary diagnosis of AMI (ICD-10: I21 and I22) when compared with adjudication of medical records (Ragna Elise Støre Govatsmark, personal communication, 2016). This suggests that our reference standard is more or less correct.

A recent Norwegian study reported a PPV of 79.7% for hospital admissions with a main or secondary diagnosis of acute stroke (ICD-10: I61, I63, and I64) when compared with adjudication of medical records; 65% of the false-positives were defined as previous strokes, rehabilitations after stroke or old strokes not previously diagnosed.[25] In the past, however, patients with ischaemic stroke were not consistently admitted to hospitals for exact diagnosis and specific treatment, although this practice has become more standard in the past decade.[26] Thus, stroke may have been diagnosed in primary care alone, particularly during the first decade of the CVDNOR database (1994–2004). In a community study from Norway in the late 1990s, Ellekjær et al[27] found that 41 of the 430 patients with incident stroke were not hospitalised. They also found that the sensitivity and PPV of hospital discharge ICD-9 codes 430, 431, 434 and 436 were 81% and 68%, respectively, when compared with a register also including non-hospitalised cases. Consequently, the validity of the SRS in this study may hence be either overestimated or underestimated; non-hospitalised strokes are likely to be milder than hospitalised cases. We may thus assume that by missing potentially milder stroke cases from the reference standard in our study, we could be overestimating the sensitivity of SRS; milder cases might be reported less accurately by participants, either because they are less likely to be recalled, or because they are more easily confused with TIA. If milder cases, however, are reported correctly, the PPV of SRS may be underestimated. This suggests that hospital discharge data is an imperfect reference standard with regard to SRS in our study.

Participants with SMI (n=54) and/or SRS (n=46) who did not specify their age at disease onset were excluded from the analyses, which may have introduced selection bias. However, if people failed to recall their age at disease onset, it is likely that in reality the event dates back years or even decades. Cases before 1994 could not be validated using CVDNOR data. Nonetheless, we found that 16 of those with SMI, and 10 of those with SRS who did not specify their age at disease onset, had been hospitalised with an AMI or an acute stroke during the study period. If they had been included in the analyses as true positives, the sensitivity and PPV would be even higher.

CONCLUSION

The sensitivity and PPV of SMI were high and moderate, respectively; for SRS, both of these measures were moderate. Our results show that SMI from the SAMINOR 1 Survey may be used in aetiological/analytical studies in this population due to a high IHD-specific PPV. The SAMINOR 1 questionnaire may also be used to estimate the prevalence of AMI and acute stroke.

Author affiliations
[1]Faculty of Health Sciences, Department of Community Medicine, Centre for Sami Health Research, UiT The Arctic University of Norway, Tromsø, Norway
[2]Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway
[3]Norwegian Institute of Public Health, Bergen, Norway
[4]Faculty of Health Sciences, Department of Community Medicine, UiT The Arctic University of Norway, Tromsø, Norway
[5]Department of Medicine, University Hospital of Northern Norway, Harstad, Norway
[6]Norwegian Institute of Public Health, Oslo, Norway

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

REFERENCES

1. Machón M, Arriola L, Larrañaga N, *et al*. Validity of self-reported prevalent cases of stroke and acute myocardial infarction in the Spanish cohort of the EPIC study. *J Epidemiol Community Health* 2013;67:71–5.
2. Yamagishi K, Ikeda A, Iso H, *et al*. Self-reported stroke and myocardial infarction had adequate sensitivity in a population-based prospective study JPHC (Japan Public Health Center)-based prospective study. *J Clin Epidemiol* 2009;62:667–73.
3. Okura Y, Urban LH, Mahoney DW, *et al*. Agreement between self-report questionnaires and medical record data was substantial for diabetes, hypertension, myocardial infarction and stroke but not for heart failure. *J Clin Epidemiol* 2004;57:1096–103.
4. Fourrier-Réglat A, Cuong HM, *et al*. Concordance between prescriber- and patient-reported previous medical history and NSAID indication in the CADEUS cohort. *Pharmacoepidemiol Drug Saf* 2010;19:474–81.
5. Carter K, Barber PA, Shaw C. How does self-reported history of stroke compare to hospitalization data in a population-based survey in New Zealand? *Stroke* 2010;41:2678–80.
6. Barr EL, Tonkin AM, Welborn TA, *et al*. Validity of self-reported cardiovascular disease events in comparison to medical record adjudication and a statewide hospital morbidity database: the AusDiab study. *Intern Med J* 2009;39:49–53.
7. Misirli G, Bamia C, Dilis V, *et al*. Validation of self-reported incident cardiovascular disease events in the Greek EPIC cohort study. *Ital J Public Health* 2012;9:1–9.
8. Engstad T, Bonaa KH, Viitanen M. Validity of self-reported stroke: the Tromso Study. *Stroke* 2000;31:1602–7.
9. Njølstad I, Universitetet i Tromsø. *Incidence of and risk factors for myocardial infarction, stroke, and diabetes mellitus in a general population: the Finnmark study 1974-1989*. Tromsø: Institute of Community Medicine, University of Tromsø, 1998.
10. Tretli S, Lund-Larsen PG, Foss OP. Reliability of questionnaire information on cardiovascular disease and diabetes: cardiovascular disease study in Finnmark county. *J Epidemiol Community Health* 1982;36:269–73.
11. Lund E, Melhus M, Hansen KL, *et al*. Population based study of health and living conditions in areas with both Sami and Norwegian populations—the SAMINOR study. *Int J Circumpolar Health* 2007;66:113–28.
12. Sulo G, Igland J, Vollset S, *et al*. Cardiovascular disease and diabetes mellitus in Norway during 1994-2009. CVDNOR—a nationwide research project. *Nor Epidemiol* 2013;23:7.
13. Clench-Aas J, Helgeland J, Dimoski T, *et al*. *Methodological development and evaluation of 30-day mortality as quality indicator for Norwegian hospitals*. Oslo: The Norwegian Knowledge Centre for the Health Services, 2005:198.
14. Kirkwood BR, Sterne JAC. *Essential medical statistics*. 2nd edn. Malden, MA: Blackwell Science, 2003.
15. Ientilucci E. On Using and Computing the Kappa Statistic. 2006. http://www.cis.rit.edu/~ejipci/Reports/On_Using_and_Computing_the_Kappa_Statistic.pdf
16. DeVellis RF. *Scale development: theory and applications*. 3rd edn. Thousand Oaks, CA: Sage, 2012.
17. Chen G, Faris P, Hemmelgarn B, *et al*. Measuring agreement of administrative data with chart data using prevalence unadjusted and adjusted kappa. *BMC Med Res Methodol* 2009;9:5.
18. Warrens M. On marginal dependencies of the 2×2 Kappa. *Adv Stat* 2014;2014:1–6.
19. Hagen TP, Anthun KS, Reikvam A. [Acute myocardial infarctions in Norway 1991—2007]. *Tidsskr Nor Laegeforen* 2010;130:820–4.
20. WHO. *MONICA manual—section 1: coronary event registration data components*. Geneva: World Health Organization, 1999:1.
21. Mathiesen EB, Njølstad I, *et al*. [Risk factors for cerebral stroke]. *Tidsskr Nor Laegeforen* 2007;127:748–50.
22. Montez JK, Friedman EM. Educational attainment and adult health: under what conditions is the association causal? *Soc Sci Med* 2015;127:1–7.
23. Eliassen BM, Graff-Iversen S, Braaten T, *et al*. Prevalence of self-reported myocardial infarction in Sami and non-Sami populations: the SAMINOR study. *Int J Circumpolar Health* 2015;74:24424.
24. Krokstad S, Knudtsen MS. *Folkehelse i endring. Helseundersøkelesen i Nord-Trøndelag*. HUNT forskningssenter, NTNU, 2011:214.
25. Varmdal T, Bakken IJ, Janszky I, *et al*. Comparison of the validity of stroke diagnoses in a medical quality register and an administrative health register. *Scand J Public Health* 2016;44:143–9.
26. Powers WJ, Derdeyn CP, Biller J, *et al*. 2015 American Heart Association/American Stroke Association focused update of the 2013 guidelines for the early management of patients with acute ischemic stroke regarding endovascular treatment: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* 2015;46:3020–35.
27. Ellekjær H, Holmen J, Krüger O, *et al*. Identification of incident stroke in Norway: hospital discharge data compared with a population-based stroke register. *Stroke* 1999;30:56–60.