Variations in the agreement of self-reported cancer: A Danish nationwide study

Rosendahl, Heidi; Horsbøl, Trine Allerslev; Thygesen, Lau Caspar; Davidsen, Michael; Christensen, Anne Illemann; Ekholm, Ola

Published in:
International Journal of Cancer

DOI:
10.1002/ijc.34692

Publication date:
2024

Document version:
Final published version

Received: 20 March 2023 | Revised: 13 July 2023 | Accepted: 14 July 2023

DOI: 10.1002/ijc.34692

RESEARCH ARTICLE

Cancer Epidemiology

IJC INTERNATIONAL JOURNAL of CANCER

Variations in the agreement of self-reported cancer: A Danish nationwide study

Heidi Amalie Rosendahl Jensen | Trine Allerslev Horsbøl | Lau Caspar Thygesen | Michael Davidsen | Anne Illemann Christensen | Ola Ekholm

National Institute of Public Health, University of Southern Denmark, Copenhagen K, Denmark

Correspondence
Heidi Amalie Rosendahl Jensen, National Institute of Public Health, University of Southern Denmark, Studiestræde 6, 1455 Copenhagen K, Denmark.
Email: harj@sdu.dk

Abstract

Previous studies show that the agreement between self-reported and registry-documented diseases varies across diseases. Few studies have addressed these challenges across site-specific cancer diagnoses. The present study aimed to examine the sensitivity and negative predictive value (NPV) of self-reported cancer in a Danish nationwide survey among adults aged ≥16 years, using registry data as the criterion standard. Moreover, the influence of sociodemographic variables and time since diagnosis on sensitivity was explored using multiple logistic regression models. Self-reported data on cancer history of any site were derived from the Danish National Health Survey 2017 (n = 183 372). Individual-level survey data were linked to data from the Danish Cancer Registry on 10 site-specific cancer diagnoses. NPV was consistently high ≥99.5% across the included cancer diagnoses. In contrast, sensitivity varied greatly and was lowest for cancer in brain/central nervous system (CNS) among both men (25.6%) and women (23.9%) and highest for rectal cancer among men (96.9%) and for breast cancer among women (98.9%). Sensitivity was also relatively low for nonmelanoma skin cancer (41.4% among men; 44.6% among women) and urinary tract cancer (60.0% among men; 60.4% among women). When restricting diagnostic definitions for cancer in brain/CNS and urinary tract cancer to include only malignant neoplasms, sensitivity increased. For several cancer diagnoses, sensitivity decreased with increasing age and lower educational level, whereas conflicting results were observed for time from diagnosis to self-report. Future studies are encouraged to use self-reported cancer history data with caution and for example, include questions on only site-specific cancer diagnoses with high sensitivity.

KEYWORDS
cancer, epidemiology, health surveys, methods, public health

Abbreviations: AOR, adjusted odds ratio; CI, confidence interval; CNS, central nervous system; DCR, The Danish Cancer Registry; DNHS, The Danish National Health Survey; HPV, human papillomavirus; NPV, negative predictive value; PPV, positive predictive value; Y, year.

What's new?

Self-reported cancer data and data from cancer registries can vary significantly. Nonetheless, self-reported data often are used as a proxy for registry data when the latter is unavailable. Here, the authors examined the sensitivity and negative predictive value (NPV) of self-reported information on cancer history. NPV was consistently high across site-specific cancer diagnoses. Sensitivity varied and was lowest for cancers of the brain and central nervous system, nonmelanoma skin cancer, and urinary tract cancer. Thus, a relatively small proportion of respondents with registry-identified diagnoses reported cancer history. Sensitivity was affected by diagnostic definitions, however, which may impact future study methodologies.

1 | INTRODUCTION

Globally, a transition in the leading cause of death has occurred in high-income countries during the past decades, with cancer now exhibiting a higher mortality than cardiovascular disease.[1] This tendency is also observed in Denmark, and according to the Danish Health Data Authority, cancer accounted for around three out of 10 deaths in Denmark in 2020.[2] In recent years, both cancer incidence, prevalence, and survival rates have increased nationally and internationally, resulting in an overall increased societal burden of cancer.[3,4] Thus, to be able to accurately quantify the health burden of cancer and allocate health care funds properly, valid cancer data are required.

Although cancer registries are considered a valid source of cancer data,[5] acquisition of these data is not always feasible in epidemiological studies. Registry linkage with survey data is generally both time consuming and has a high administrative burden or might not even be possible. Accordingly, self-reported cancer data are often used as a reasonable proxy for registry data. However, if recall bias or misclassification in self-reports occur systematically, it may result in biased population estimates. It is therefore of great importance to investigate the validity of self-reported cancer data compared to data recorded in cancer registries.

We have previously explored the validity of self-reported cancer of any site and demonstrated a relatively low sensitivity of 65.9%,[6] which is similar to estimates from other studies.[7-10] The low overall sensitivity is, however, likely to blur the variations in the validity across site-specific cancer diagnoses. This hypothesis is supported by other studies demonstrating the sensitivity of self-reported cancer data to range between 16% and 100%, depending on cancer site. Thus, a relatively high sensitivity has been demonstrated for breast cancer (81%-96%)[7,10-16] and prostate cancer (67%-90%).[10,12-14] On the contrary, large variations have been found for colorectal cancer (17%-83%)[7,10,13,16] and lung cancer (50%-100%),[7,11-13,15,16] whereas a relatively low sensitivity is seen for cervical cancer (13%-52%)[7,10,13,15,16] and nonmelanoma skin cancer (45%-54%).[10,16] A similar pattern of variation has been documented for other validity outcomes, for example, specificity,[7,8,14-16] positive predictive value (PPV)[7,12,14,17] and negative predictive value (NPV).[7,14,17] In our previous study, we found cancer of any site to have a high specificity (99.1%), PPV (90.1%), NPV (95.9%), and total agreement (95.4%), and a moderate kappa value (0.74). Moreover, the prevalence of self-reported cancer history of any site was 8.1% compared to 11.1% based on registry data.[6]

In the literature, sociodemographic factors such as sex, age, and educational level have consistently been found to be associated with correctly self-reporting diseases, including cancer.[6,7,12,13] However, other factors have been suggested to play a substantial role, too. For example, studies have found self-reports to be considerably more sensitive for serious conditions than for less serious or transient conditions[17] and for conditions that have very clear and well-defined diagnostic criteria.[18] Also, across site-specific cancer diagnoses, invasive tumors have been found to be more accurately self-reported than in situ cancer.[7,10,16] Moreover, studies have found that longer time between diagnosis and self-report is associated with incorrect self-reports,[7,9,16] suggesting a tendency toward an increased risk of recall bias with longer recall period.

The aim of the present study was to examine whether self-reported data on cancer history could be considered a reasonable proxy for cancer registry data. Hence, the sensitivity and NPV of self-reported cancer of any site were calculated for 10 registry-based site-specific cancer diagnoses with a high prevalence. Moreover, the aim was to explore the associations between various sociodemographic variables and time from clinical diagnosis to self-reports, respectively, and the sensitivity of self-reported cancer.

2 | MATERIALS AND METHODS

2.1 | Data sources

Data for the present study were derived from the Danish National Health Survey 2017 (DNHS-2017)[19] and different administrative registries in Denmark, that is, the Danish Cancer Registry (DCR),[5] the Danish Civil Registration System,[20] and the Danish Education Registers.[21]

In Denmark, every individual with a permanent residence is assigned a unique personal identification number (ie, a CPR number), which is registered in the Danish Civil Registration System.[20] By linkage through CPR numbers, individual-level information on for example, diseases can be obtained from various registries, including the DCR.[5]

2.2 | Self-reported cancer data

The DHNS-2017 is a nationwide representative health survey among the adult population in Denmark aged 16 years or older. A total of 312 349 individuals were randomly selected using the Danish Civil Registration System and invited to participate in the survey,[20] of whom 183 372 completed the self-administered questionnaire (58.7%). The study design and data collection mode have been described in detail elsewhere.[19]

In the questionnaire, self-reported cancer was assessed by a standard checklist of various diseases and health conditions that, among others, included cancer (of any site). The question formulation was: "For each of the following diseases and health problems, we ask you to indicate if you currently have or previously have had it." There were three possible response categories: "No, I have never had this [disease or health condition]," "Yes, I have this [disease or health condition] now," and "Yes, I have had this [disease or health condition] previously." Both current and previous cancers are categorized as cancer history cases in the present article.

2.3 | Cancer registry data

The DCR was founded in 1942 and contains information on the incidence of cancer in the Danish population since 1943.[5] The coding manual in DCR is based on the international principles on coding cancer along with some Danish exceptions (on multiple primaries and the inclusion of certain benign and precancerous lesions).[5] In 2004, the DCR went through a process of modernization and now includes electronic notifications from the Danish National Patient Register,[22] general practitioners, and practicing specialists, linked with information on histology from The Danish Pathology Register,[23] using an automated cancer logic algorithm based on ICD-10 codes.

By using the CPR numbers, it is possible to determine whether an individual is already known to the DCR or is a new case. Data on cancer cases for the present study were extracted from 1 January 1978 until start date of data collection (1 February 2017), as previous time series are incomparable due to difficulties in conversion processes between ICD-7 and -10 codes.[5]

Based on data from the DCR in 2020,[5] 10 site-specific cancer diagnoses with a high prevalence among men and women, respectively, were chosen to explore the sensitivity and NPV of self-reported cancer among respondents from the DNHS-2017. The specific ICD-10 codes for each of the site-specific cancer diagnoses are provided in Table 1 and are based on definitions expressed by the Danish Health Data Authority[24] and used by the Danish Cancer Society.[3] As can be seen from the table, the included cancer diagnoses contain both C-diagnoses (C00-C97: Malignant neoplasms) and D-diagnoses (1) D00-D09: In situ neoplasms; (2) D10-D36: Benign neoplasms; and (3) D37-D48: Neoplasms of uncertain or unknown behavior).

The reason why these D-diagnoses are included in the diagnostic grouping for some site-specific cancer diagnoses is that the medical and surgical procedures following diagnosis are identical or similar to those used to treat malignancies at the same site. Thus, even though they are benign, they are to some extent considered as cancer.

TABLE 1 Diagnostic definitions for 10 site-specific cancer diagnoses with a high prevalence according to ICD-10 codes among men and women, respectively, in Denmark.

Cancer site	ICD-10 codes
Men	
Prostate	C61
Urinary tract[a]	C64-C68, D09.0-D09.1, D09.5-D09.6, D30.1-D30.9, D41.1-D41.9[b]
Melanoma skin	C43
Nonmelanoma skin	C44[c]
Colon	C18-C19
Rectum	C20
Testicle	C62
Brain and CNS	C70-C72, C75.1-C75.3, D32-D33, D35.2-D35.4, D42-D43, D44.3-D44.5
Lung[d]	C33-C34
Head and neck	C00-C14, C30-C32, C73
Women	
Breast	C50
Melanoma skin	C43
Nonmelanoma skin	C44[c]
Brain and CNS	C70-C72, C75.1-C75.3, D32-D33, D35.2-D35.4, D42-D43, D44.3-D44.5
Uterus	C54-C55
Colon	C18-C19
Cervix	C53
Urinary tract[a]	C64-68, D09.0-D09.1, D09.5-D09.6, D30.1-D30.9, D41.1-D41.9[b]
Lung[d]	C33-C34
Head and neck	C00-C14, C30-C32, C73

Abbreviation: CNS, central nervous system.
[a]Including kidneys.
[b]D-diagnoses were furthermore restricted to morphology codes 812-813.
[c]Excluding morphology code 809.
[d]Including bronchi and trachea.

Because of the completeness, high data quality, and long archival period of registrations in the DCR, a registry-documented diagnosis of cancer in this register was considered the criterion standard (ie, "the gold standard") for the presence of cancer in the present study.

2.4 | Other data sources

Sociodemographic and other data used in the present study included registry-based information on sex, age (as a continuous variable), marital status (married and not married, the latter including divorced, widowed, and unmarried),[20] highest completed educational level (basic school/no information, primary/vocational education, and

higher education),[21] and time between first cancer register-documented diagnosis and survey participation (as a continuous variable).[5]

2.5 | Statistical analyses

To examine the validity of self-reported cancer of any site compared to site-specific cancer diagnoses from the DCR, sensitivity and NPV were calculated using cancer registry data as the criterion standard. As expressed by Okura and colleagues,[25] sensitivity was calculated as the proportion of individuals with a specific registry-documented cancer diagnosis who also reported having or having had cancer in the questionnaire. NPV was calculated as the proportion of individuals who reported not having or having had cancer in the questionnaire who were also not registered in the DCR with any of the included site-specific cancer diagnoses.

Other validity studies typically also include calculations of specificity and PPV. However, as the questionnaire used in the present study only included a question on cancer history of any site, such calculations were not possible. Thus, if a respondent was *not* registered with a certain site-specific cancer diagnosis in the DCR but responded "Yes" to a cancer history of any site in the questionnaire, it would not be possible to assess whether this response was actually true (ie, the specificity), as this respondent may have a cancer history with a site-specific diagnosis not included in the present study. Similarly, if a respondent responded "Yes" to a cancer history of any site in the questionnaire but was *not* registered in the DCR with one of the included site-specific cancer diagnoses, it would not be possible to assess whether this response was actually true (ie, the PPV), as this respondent may have a cancer history with a site-specific cancer diagnosis not included in the present study.

Sensitivity analyses were conducted for cancer in brain/CNS and urinary tract cancer, as these diagnostic definitions included D-diagnoses (in addition to C-diagnoses with confirmed malignancy). It is unclear if health professionals consistently inform patients with these D-diagnoses that they have cancer or not. Thus, we explored how gradual restrictions of the diagnostic definitions affected sensitivity of self-reported cancer, ending up with only C-diagnoses. We also conducted sensitivity analyses for nonmelanoma skin cancer, in which the morphology code for basal cell carcinoma ("809") was included. As this is the most prevalent cancer diagnosis that also exhibits a very high survival rate, we explored how inclusion of this specific morphology code affected sensitivity for nonmelanoma skin cancer.

Multiple logistic regression models stratified by sex were used to identify potential variables associated with site-specific sensitivity. This means that we investigated if these variables predicted whether respondents identified with a site-specific cancer in the DCR also responded in the questionnaire that they had a cancer history.

Due to a limited number of individuals with some of the included cancer diagnoses, results from the multiple logistic regressions models are only presented for the five diagnoses with >400 respondents among men and women, respectively. As data on both highest attained educational level and marital status were included in the regression models, only individuals aged 25 years or older were included in these analyses. The significance level was set to 5%, and statistical analyses were performed using SAS Version 9.4 (SAS Institute Inc., Cary, NC).

3 | RESULTS

The characteristics of the study population, that is, the respondents from the DHNS-2017, are shown in Table 2. For instance, the study population included a slightly higher proportion of women (53.9%) than men. In the study population, the prevalence of self-reported cancer history of any site was 8.1% (data not shown).

Table 3 shows the sensitivity and NPV (in percentage) for 10 site-specific cancer diagnoses with a high prevalence in Denmark among men and women, respectively, listed in descending prevalence order. Among both men and women and across all included cancer diagnoses, NPV was consistently high (≥99.6%). In contrast, large variations were seen for sensitivity, which varied between 25.6% for cancer in brain/CNS and 96.9% for rectal cancer among men, and between 23.9% for cancer in brain/CNS and 98.9% for breast cancer among

TABLE 2 Characteristics of the respondents in the Danish National Health Survey 2017.

	Number of individuals (n)	Percentage (%)[a]
Sex		
Men	84 607	46.1
Women	98 765	53.9
Age		
16-24 y	18 913	10.3
25-44 y	44 121	24.1
45-64 y	67 221	36.7
≥65 y	53 117	29.0
Educational level[b]		
Basic school/no information	38 800	23.6
Secondary/vocational education	69 098	42.0
Higher education	56 561	34.4
Marital status[b]		
Married	99 914	60.8
Not married[c]	64 545	39.3
Ethnic background		
Danish	168 932	92.1
Western	6125	3.3
Non-Western	8315	4.5

Note: Age ≥ 16 years. Number of individuals (n) and percentage (%).
[a]Deviations in totals (100%) are due to roundings.
[b]Restricted to individuals aged ≥25 years.
[c]Divorced, widowed, or unmarried.

TABLE 3 Sensitivity and NPV for self-reported cancer compared to 10 prevalent registry-documented site-specific cancer diagnoses among Danish men and women, respectively.

	Validity outcome measure	
	Sensitivity, %	NPV, %
Men		
Prostate (n = 1906)	86.3	99.6
Urinary tract (n = 868)	60.0	99.5
Melanoma skin (n = 582)	83.7	99.9
Nonmelanoma skin (n = 579)	41.1	99.5
Colon (n = 561)	93.8	100.0
Rectum (n = 359)	96.9	99.7
Testicle (n = 331)	96.7	100.0
Brain and CNS (n = 320)	25.6	99.7
Lung (n = 196)	92.4	100.0
Head and neck (n = 185)	95.7	100.0
Women		
Breast (n = 3078)	98.9	100.0
Melanoma skin (n = 864)	87.5	99.9
Colon (n = 553)	94.6	100.0
Nonmelanoma skin (n = 462)	44.6	99.7
Brain and CNS (n = 426)	23.9	99.6
Uterus (n = 418)	89.0	100.0
Cervix (n = 366)	84.2	99.9
Urinary tract (n = 331)	60.4	99.8
Lung (n = 244)	96.7	100.0
Head and neck (n = 98)	93.9	100.0

Note: Age ≥ 16 years. Percentage (%) and number of individuals (n).

women. The sensitivity for nonmelanoma skin cancer (41.1% and 44.6% among men and women, respectively) and urinary tract cancer (60.0% and 60.4% among men and women, respectively) also exhibited a relatively low sensitivity, with the remaining cancer diagnoses exhibiting a relatively high sensitivity (>84%). Accordingly, based on sensitivity calculations, the proportion of false positive self-reports (calculated as 100%-sensitivity) varied between 3.1% for rectal cancer and 74.4% for cancer in brain/CNS among men, and between 1.1% for breast cancer and 76.1% for cancer in brain/CNS among women.

As demonstrated in Figure 1, the sensitivity for urinary tract cancer increased from 60.0% to 78.5% among men and from 60.4% to 71.9% among women when restricting the diagnostic definition of urinary tract cancer to diagnoses that were either malignant neoplasms (C64-C68) or neoplasms of uncertain or unknown behavior (D41.1-D41.9), yet with the inclusion of morphology codes 812-813 (Model II). By further restrictions to only malignant neoplasms (C64-C68), sensitivity increased to 90.4% among men and 88.7% among women (Model III).

Similarly, when restricting the diagnostic definition of cancer in brain/CNS to exclude endocrine tumors (ie, by including the following codes: C70-C72, D32-D33 + D42-D43) as expressed by NORDCAN (27), sensitivity remained stable (25.6% vs 26.7% and 23.9 vs 25.6% among men and women, respectively) (Model II). When further restricting the diagnostic definition to include only malignant neoplasms (C70-C72), sensitivity increased to 66.7% among men (n = 28) and to 65.8% among women (n = 25) (Model III), despite a relatively small number of individuals included in these analyses.

A contrasting tendency was observed for nonmelanoma skin cancer, as the sensitivity decreased from 41.1% to 31.6% among men and from 44.6% to 32.1% among women when including the morphology code for basal cell carcinoma ("809") (Model II).

The results from the multiple logistic regression models of the potential associations between sociodemographic variables and time between diagnosis and self-reported cancer, respectively, and sensitivity are shown in Table 4 (men) and 5 (women). For several of the included cancer diagnoses, increasing age was significantly associated with a decreasing sensitivity, especially among men (prostate (adjusted OR [AOR] = 0.92, 95% CI: 0.90;0.94), melanoma skin (AOR = 0.94, 0.91;0.96), nonmelanoma skin (AOR = 0.97, 0.96;0.99), and colon (AOR = 0.95, 0.91;0.99)). Among women, increasing age was negatively associated with sensitivity for breast cancer (AOR = 0.94, 0.91;0.98) and melanoma skin cancer (AOR = 0.97, 0.95;0.98). For prostate cancer (AOR = 1.06, 1.02;1.10), a one-year increase in time between diagnosis and self-report was associated with an increasing sensitivity, whereas increasing time was associated with a decreasing sensitivity for cancer in brain/CNS among women (AOR = 0.96, 0.93; 0.99). Compared to a higher educational level, lower educational level (basic school/no information and secondary/vocational education) was associated with lower sensitivity for prostate cancer (AOR = 0.58, 0.40; 0.85, and AOR = 0.54, 0.38; 0.76, respectively) and melanoma skin cancer among men (AOR = 0.78, 0.40; 1.52, and AOR = 0.58, 0.34; 0.98, respectively) and cancer in uterus (AOR = 0.20, 0.06; 0.68, and AOR = 0.22, 0.06; 0.76, respectively) and breast cancer among women (AOR = 0.22, 0.06; 0.75, and AOR = 0.27, 0.08; 0.94, respectively). Marital status was not associated with sensitivity for any of the included cancer diagnoses.

4 | DISCUSSION

In the present study using data from the DCR as the criterion standard against which self-reported data on current or previous cancer of any site were compared, we found the sensitivity to vary considerably across site-specific cancer diagnoses. In contrast, NPV was excellently high for all included diagnoses (≥99.6% among both men and women). Thus, sensitivity varied between 25.6% for cancer in brain/CNS to 96.9% for rectal cancer among men, and between 23.9% for cancer in brain/CNS to 98.9% for breast cancer among women. This means that only for some registry-identified cancer diagnoses, respondents were very likely to report a cancer history in their survey response (high sensitivity). For other diagnoses, respondents were *not* likely to report such history (low sensitivity). For most of the included site-specific cancer diagnoses, increasing age was associated with a lower

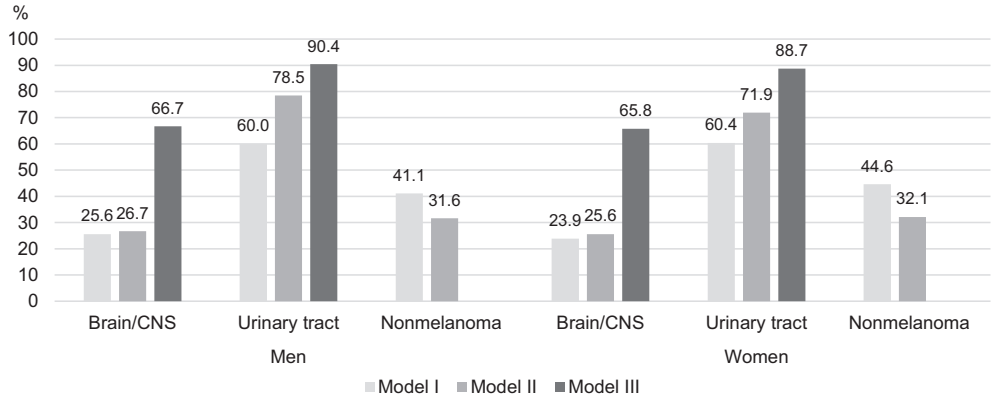

FIGURE 1 Brain/CNS: *Model I (full diagnostic definition)*: C70-C72, C75.1-C75.3, D32-D33, D35.2-D35.4, D42-D43, D44.3-D44.5; *Model II*: C70-C72, D32-D33, D42-D43; *Model III*: C70-C72. Urinary tract (D-diagnoses were restricted to morphology codes 812-813): *Model I (full diagnostic definition)*: C64-C68, D09.0-D09.1, D09.5-D09.6, D30.1-D30.9, D41.1-D41.9; *Model II*: C64-C68, D41.1-D41.9; *Model III*: C64-C68. Nonmelanoma: *Model I (full diagnostic definition)*: C44 (% morphology code 809); *Model II*: C44 (+ morphology code 809). CNS, central nervous system.

TABLE 4 Multiple logistic regression models for sensitivity of self-reported cancer compared to registry-documented site-specific cancer diagnoses among men.

Men		Prostate (n = 1906) AOR[a] (95% CI)	Urinary tract (n = 867)	Melanoma skin (n = 582)	Nonmelanoma skin (n = 579)	Colon (n = 560)
Age at survey participation		0.92 (0.90;0.94)	0.99 (0.97;1.00)	0.94 (0.91;0.96)	0.97 (0.96;0.99)	0.95 (0.91;0.99)
Educational level[b]	Basic school/no information	0.58 (0.40;0.85)	0.83 (0.57;1.22)	0.78 (0.40;1.52)	1.05 (0.67;1.65)	0.73 (0.29;1.85)
	Secondary/vocational education	0.54 (0.38;0.76)	0.89 (0.64;1.24)	0.58 (0.34;0.98)	1.02 (0.68;1.55)	0.85 (0.36;2.00)
Marital status[c]	Not married[d]	0.81 (0.60;1.11)	0.94 (0.70;1.27)	1.41 (0.80;2.47)	0.99 (0.68;1.43)	1.06 (0.48;2.32)
Time between diagnosis and self-report		1.06 (1.02;1.10)	0.98 (0.96;1.00)	1.00 (0.98;1.03)	0.99 (0.96;1.02)	0.99 (0.94;1.03)

Note: Age ≥ 25 years. Adjusted odds ratios (AORs)[a] and 95% confidence intervals (CIs).
Abbreviations: AOR, adjusted odds ratio; CI, confidence interval; CNS, central nervous system; y, years.
[a]Adjusted for all variables in model.
[b]Ref.: Higher education.
[c]Ref.: Married.
[d]Divorced, widowed, or unmarried.

sensitivity. This tendency was also seen for some of the cancer diagnoses for increasing time between diagnosis and self-report and lower educational level.

There are several possible explanations for the demonstrated variations in the sensitivity across site-specific cancer diagnoses. Indeed, it has been suggested that diagnoses with very clear and well-defined diagnostic definitions, such as breast cancer or prostate cancer, are more likely to be accurately reported than cancer diagnoses with more ambiguous diagnostic procedures.[14,16]

In the present study, cancer diagnoses defined by only ICD-10 C-diagnoses exhibited the highest sensitivity (eg, 96.7% for testicular cancer (C62) among men and 98.9% for breast cancer (C50) among women). In contrast, the lowest sensitivity was found among cancer diagnoses that, in addition to C-diagnoses, also included D-diagnoses and/or were further specified by morphology codes. Accordingly, the lowest sensitivity among both men and women was found for cancer in brain/CNS (25.6% and 23.9%, respectively), nonmelanoma skin cancer (41.1% and 44.6%, respectively), and urinary tract cancer (60.0% and 60.4%, respectively).

ICD-10 codes used to define cancer in brain/CNS and urinary tract cancer included cases of both confirmed benign and potentially benign tumors, thus constituting a rather heterogeneous group. However, treatment for benign tumors in the brain/CNS is similar to that of malign tumors at this site, as benign tumors may exhibit a behavior leading to death of the patient.[26] Also, definitions and coding practices for urinary tract cancer have varied over time as to whether in situ lesions are reported as invasive tumors or not. Further, treatment for benign tumors is similar to those of malign tumors at this site.[27] Therefore, respondents may correctly reply in the questionnaire that they do *not* have a cancer history, but still be registered in the DCR with cancer brain/CNS or urinary tract cancer according to the diagnostic definitions.

TABLE 5 Multiple logistic regression models for sensitivity of self-reported cancer compared to registry-documented site-specific cancer diagnoses among women.

Women		Breast (n = 3078) AOR[a] (95% CI)	Melanoma (n = 860)	Colon (n = 552)	Nonmelanoma (n = 461)	Uterus (n = 418)	Brain and CNS (n = 416)
Age at survey participation		0.94 (0.91;0.98)	0.97 (0.95;0.98)	1.00 (0.96;1.03)	0.99 (0.98;1.01)	0.96 (0.92;1.00)	1.02 (1.00;1.04)
Educational level[b]	Basic school/no information	0.22 (0.06;0.75)	1.06 (0.59;1.90)	0.80 (0.29;2.19)	1.16 (0.71;1.89)	0.20 (0.06;0.68)	0.83 (0.45;1.52)
	Secondary/vocational education	0.27 (0.08;0.94)	0.86 (0.54;1.38)	0.89 (0.32;2.48)	1.39 (0.88;2.20)	0.22 (0.06;0.76)	0.77 (0.45;1.31)
Marital status[c]	Not married[d]	1.72 (0.81;3.67)	0.99 (0.64;1.51)	0.73 (0.33;1.62)	0.97 (0.65;1.43)	1.03 (0.52;2.06)	1.53 (0.96;2.44)
Time between diagnosis and self-report		0.99 (0.95;1.02)	0.99 (0.97;1.01)	0.98 (0.94;1.02)	1.00 (0.96;1.03)	0.97 (0.94;1.01)	0.96 (0.93;0.99)

Note: Age ≥ 25 years. Adjusted odds ratios (AORs)[a] and 95% confidence intervals (CIs).
Abbreviations: AOR, adjusted odds ratio; CI, confidence interval; CNS, central nervous system; y, years.
[a]Adjusted for all variables in model.
[b]Ref.: Higher education.
[c]Ref.: Married.
[d]Divorced, widowed, or unmarried.

Sensitivity increased markedly when restricting the diagnostic definition to only malignant neoplasms for cancer in brain/CNS (from 25.6% to 66.7% among men and from 23.9% to 65.8% among women) and for urinary tract cancer (from 60.0% to 90.4% among men and from 60.4% to 88.7% among women). Accordingly, these analyses emphasize the impact of *which* diagnoses that are included in the diagnostic definitions. This is especially relevant for site-specific cancer diagnoses not only defined by C-diagnoses, that is, diagnoses where there is uncertainty as to whether certain benign conditions are considered cancer.

As expected, the sensitivity for nonmelanoma skin cancer was further decreased by inclusion of the morphology code for basal cell carcinoma ("809"), from 41.1% to 31.6% among men and from 44.6% to 32.1% among women. Given that the survival rate of basal cell carcinoma remains very high after both 3 years (100%), 5 years (99.8%), and 10 years (98.5%),[28] respondents registered with this cancer diagnosis in registries may be inclined to not consider themselves as individuals with a cancer history. Accordingly, the cancer diagnosis may be perceived as "as little severe" that it seems irrelevant to report in the survey questionnaire. Squamous cell carcinoma is the second most frequent cutaneous malignancy after basal cell carcinoma and also exhibits a relatively high survival rate (ie, 94.2%, 90.1%, and 86.0% after 3, 5, and 10 years, respectively).[28] Thus, inclusion of squamous cell carcinoma in the analyses of nonmelanoma may then, at least partly, explain the low sensitivity.

Other factors that may, at least partly, explain the relatively low sensitivity demonstrated for cancer in brain/CNS, urinary tract cancer, and nonmelanoma skin cancer include the applied communication approaches by the responsible health professionals. Thus, if health professionals communicate to the patient that the identified tumor in for example, brain/CNS or urinary tract was *not* malign, patients are perhaps likely to be inclined to stick to this key message even though the doctor may also inform about for example, future medical or surgical treatment, an increased mortality risk, and so forth. In relation to nonmelanoma skin cancer, it is likely that diagnoses are provided by the health professionals who may emphasize in their communication with the patient that the diagnosis is "as little malignant as almost benign." This is even more likely in cases of basal cell carcinoma given the very high survival rate. This may affect the diagnostic perception among the patient, which would then be reflected in survey response.

In general, the demonstrated sensitivity values in the present study are higher than or comparable to those from other studies.[7,9-16] But in line with our results, other studies have also documented a relatively high sensitivity for, for example, breast cancer (81.2%-96.4%)[7,10-16] and prostate cancer (67%-90%)[10,12-14] and a relatively low sensitivity for, for example, nonmelanoma skin cancer (45.0%-53.6%).[10,16] However, the sensitivity for cervical cancer was substantially higher in our study (84.2%) compared to that of other studies (13.2%-52.1%).[7,10,13,15,16] This may be cause by a high societal awareness of this cancer diagnosis in Denmark following the implementation of a national screening program for cervical cancer as well as a free HPV vaccination at age 12 to 17 years.[29]

Excellently high NPV >99.6% were demonstrated in the present study across all included site-specific cancer diagnoses. That is, respondents who in their survey response indicated no history of cancer of any site were very likely *not* to be registered in DCR with a cancer history. This finding is in line with those from other studies, demonstrating NPV of >98% for, for example, breast, prostate, and cervical cancer.[14,17]

For some of the included site-specific cancer diagnoses and in line with other studies, we found higher age,[9,10,13,14,16] lower educational level,[7,13] and longer time between cancer diagnosis and self-report[9,13,14] to be associated with lower sensitivity. However, some studies did not find an association with educational level[10,11,14] and

time from diagnosis.[16] Several factors have been proposed as explanatory variables for the accuracy self-reported diseases. For example, it has been suggested that emotional events, for example, those considered as threatening such as severe disease diagnosis, are associated with enhanced temporal memory, that is, recall.[30] While studies have found that cognitive and memory decline as a natural result of aging,[31] a cancer diagnosis itself or the treatment received may also impact self-report accuracy,[32,33] which may indeed be the case for cancer in brain/CNS. Higher educational level has been found to be associated with more favorable health behaviors and better health literacy, which may also cause these individuals to better gain access to and understand health information compared to those with lower educational level.[34] However, according to Schrijvers and colleagues,[8] underreporting of cancer may be due to not only respondents' misunderstanding of the term "malignant disease or cancer" but may also be caused by cancer being regarded a taboo subject in some segments of society, for example, older and less educated individuals. Also, doctors may differentiate their information about a cancer diagnosis, for example, according to educational level, that is, reflecting an overall social inequality with those higher educated finding it less difficult to navigate the health system and understand information about for example, diagnoses.

A major strength of the present study is that it was conducted in a large representative sample of adults in the general Danish population, enabling us to compare on an individual level self-reported data on cancer history with cancer registry data. Moreover, it was possible to adjust for other sociodemographic variables in our analyses. Also, the DCR has a very high quality, coverage and completeness, which strengthens the validity of our results.[5] A potential limitation of using registry data as a criterion standard is the risk of inconsistent coding practices and misclassification of diagnoses, the latter which will, however, most often be nondifferential.[35] Another limitation is the risk of left truncation as some individuals may have been diagnosed with one of the included site-specific cancer diagnoses prior to 1978 and (correctly) self-reported a cancer history in the questionnaire. These cancer cases would then not be included in our registry data, that is, be misclassified with no cancer history. However, we consider the risk of this type of bias to be negligible as it is rather unlikely that individuals were registered with incident cancer prior to 1978, participated and reported a cancer history in the survey in 2017 as cancer incidence is very low at younger ages. Also, truncation bias would neither affect sensitivity nor NPV (but might affect specificity and PPV, which were not included in the present study). In the present study, marital status was dichotomized as "married" vs "not married," as registry data on cohabitation status were not available. Thus, individuals not married but cohabiting were categorized as "not married" although they may to a large degree share characteristics of those married, for example, in terms of mental and practical support in times of disease. Accordingly, our dichotomization may have resulted in marital status not being an explanatory variable for cancer sensitivity. In studies examining the agreement between self-reported diseases and registry data, measures of specificity, PPV, total agreement, or kappa are often included, too. However, because no questions on site-specific cancer diagnoses were included in the questionnaire, it was not possible to calculate these outcomes in the present study. Thus, it would also have been interesting to include such questions on site-specific cancer diagnoses in the questionnaire to calculate other agreement measures and explore the impact on sensitivity and NPV.

In conclusion, results from the present study revealed large variations in the sensitivity of self-reported cancer history in a nationally representative health survey, whereas NPV was consistently high across included site-specific cancer diagnoses. Thus, only for some cancer diagnoses, self-reported data could be considered a reasonable proxy for cancer registry data. Moreover, for several cancer diagnoses, sensitivity decreased with increasing age and lower educational level, whereas conflicting results were observed for time from diagnosis to self-report. The relatively low sensitivity for cancer in brain/CNS and urinary tract cancer may be explained by for example, applied diagnostic definitions, whereas nonmelanoma skin cancer is likely to be perceived as an almost benign condition, especially basal cell carcinoma, which may have caused the sensitivity to decrease. The results have implications for studies that depend on self-reported cancer history, as these data may deviate markedly from those from registries. Moreover, such studies may consider to only include questions on site-specific cancer diagnoses that exhibit high agreement with criterion standard data.

AUTHOR CONTRIBUTIONS

Heidi Amalie Rosendahl Jensen: Conceptualization, Methodology, Writing-Original draft preparation, Writing-Reviewing and Editing; **Trine Allerslev Horsbøl:** Methodology, Writing-Reviewing and Editing; **Lau Caspar Thygesen:** Methodology, Writing-Reviewing and Editing; **Michael Davidsen:** Methodology, Writing-Reviewing and Editing; **Anne Illemann Christensen:** Methodology, Writing-Reviewing and Editing; **Ola Ekholm:** Conceptualization, Methodology, Formal Analysis, Software, Writing-Reviewing and Editing. The work reported in the paper has been performed by the authors, unless clearly specified in the text.

FUNDING INFORMATION

No funding was received to conduct the present study. The Danish National Health Survey was funded by The Capital Region, Region Zealand, The South Denmark Region, The Central Denmark Region, The North Denmark Region, Ministry of Health and the National Institute of Public Health, University of Southern Denmark.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data used in the present study were stored at Statistics Denmark on secure logged servers and was pseudo-anonymized before the authors were granted access. Further information is available from the corresponding author and with the permission of Statistics Denmark.

ETHICS STATEMENT

Ethical approval for the study was provided by the legal department at the University of Southern Denmark. Moreover, according to

Danish law, individual-level linkage of data from for example, surveys and administrative registries is allowed without further consent when it is for research purposes and when thoroughly ensuring that results are presented in an anonymized way.

ORCID

Heidi Amalie Rosendahl Jensen https://orcid.org/0000-0001-7171-7429

REFERENCES

1. Dagenais GR, Leong DP, Rangarajan S, et al. Variations in common diseases, hospital admissions, and deaths in middle-aged adults in 21 countries from five continents (PURE): a prospective cohort study. *Lancet*. 2020;395:785-794.
2. The Cause of Death Register 2020. Danish Health Data Authority, 2022. Danish: E-Sundhed. Dødsårsagsregistret 2020. Sundhedsdatastyrelsen, 2022. https://www.esundhed.dk/Emner/Hvad-doer-vi-af/Doedsaarsager. Accessed February 23, 2023.
3. Danish Cancer Society. Cancer in Denmark 2022. *Danish: Kræft i Danmark 2022*. Copenhagen: Danish Cancer Society; 2022.
4. World Health Organization. *International Agency for Research on Cancer*. Global Cancer Observatory. https://www-dep.iarc.fr/. Accessed February 13, 2023.
5. Gjerstorff ML. The Danish Cancer Registry. *Scand J Public Health*. 2011;39:42-45.
6. Jensen HAR, Lau CJ, Davidsen M, Ekholm O, Christensen AI. Agreement between self-reported diseases from health surveys and national health registry data: a Danish nationwide study. *J Epidemiol Community Health*. 2022;77:116-122.
7. Navarro C, Chirlaque MD, Tormo MJ, et al. Validity of self reported diagnoses of cancer in a major Spanish prospective cohort study. *J Epidemiol Community Health*. 2006;60:593-599.
8. Schrijvers CT, Stronks K, van de Mheen DH, Coebergh JW, Mackenbach JP. Validation of cancer prevalence data from a postal survey by comparison with cancer registry records. *Am J Epidemiol*. 1994;139:408-414.
9. Desai MM, Bruce ML, Desai RA, Druss BG. Validity of self-reported cancer history: a comparison of health interview data and cancer registry records. *Am J Epidemiol*. 2001;153:299-306.
10. Manjer J, Merlo J, Berglund G. Validity of self-reported information on cancer: determinants of under- and over-reporting. *Eur J Epidemiol*. 2004;19:239-247.
11. Klein BEK, Lee KE, Moss SE, Trentham-Dietz A, Klein R. Self- and registry-reported cancer in a population-based longitudinal study. *WMJ*. 2010;109:261-266.
12. Bergmann MM, Calle EE, Mervis CA, Miracle-McMahill HL, Thun MJ, Heath CW. Validity of self-reported cancers in a prospective cohort study in comparison with data from state cancer registries. *Am J Epidemiol*. 1998;147:556-562.
13. Cho S, Shin A, Song D, et al. Validity of self-reported cancer history in the health examinees (HEXA) study: a comparison of self-report and cancer registry records. *Cancer Epidemiol*. 2017;50:16-21.
14. Loh V, Harding J, Koshkina V, Barr E, Shaw J, Magliano D. The validity of self-reported cancer in an Australian population study. *Aust N Z J Public Health*. 2014;38:35-38.
15. Stavrou E, Vajdic CM, Loxton D, Pearson SA. The validity of self-reported cancer diagnoses and factors associated with accurate reporting in a cohort of older Australian women. *Cancer Epidemiol*. 2011;35:e75-e80.
16. Parikh-Patel A, Allen M, Wright WE, California Teachers Study Steering Committee. Validation of self-reported cancers in the California Teachers Study. *Am J Epidemiol*. 2003;157:539-545.
17. Takamatsu K, Ideno Y, Kikuchi M, et al. Validity of self-reported diagnoses of gynaecological and breast cancers in a prospective cohort study: the Japan Nurses' Health Study. *BMJ Open*. 2021;11:e045491.
18. Colditz GA, Martin P, Stampfer MJ, et al. Validation of questionnaire information on risk factors and disease outcomes in a prospective cohort study of women. *Am J Epidemiol*. 1986;123:894-900.
19. Christensen AI, Lau CJ, Kristensen PL, et al. The Danish National Health Survey: study design, response rate and respondent characteristics in 2010, 2013 and 2017. *Scand J Public Health*. 2022;50:180-188.
20. Pedersen CB. The Danish Civil Registration System. *Scand J Public Health*. 2011;39:22-25.
21. Jensen VM, Rasmussen AW. Danish Education Registers. *Scand J Public Health*. 2011;39:91-94.
22. Lynge E, Sandegaard JL, Rebolj M. The Danish National Patient Register. *Scand J Public Health*. 2011;39:30-33.
23. Bjerregaard B, Larsen OB. The Danish pathology register. *Scand J Public Health*. 2011;39:72-74.
24. The Danish Health Data Authority. *Applied Grouping in The Danish Cancer Registry*. Danish: Grupperinger anvendt i Cancerregisteret. 2022.
25. Okura Y, Urban LH, Mahoney DW, Jacobsen SJ, Rodeheffer RJ. 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;57:1096-1103.
26. Bray F, Engholm G, Hakulinen T, et al. Trends in survival of patients diagnosed with cancers of the brain and nervous system, thyroid, eye, bone, and soft tissues in the Nordic countries 1964-2003 followed up until the end of 2006. *Acta Oncol*. 2010;49:673-693.
27. Engholm G, Hakulinen T, Gislum M, et al. Trends in the survival of patients diagnosed with kidney or urinary bladder cancer in the Nordic countries 1964-2003 followed up to the end of 2006. *Acta Oncol*. 2010;49:655-664.
28. Rubio-Casadevall J, Hernandez-Pujol AM, Ferreira-Santos MC, et al. Trends in incidence and survival analysis in non-melanoma skin cancer from 1994 to 2012 in Girona, Spain: a population-based study. *Cancer Epidemiol*. 2016;45:6-10.
29. Danish Cancer Society. *Screening for Cervical Cancer*. https://www.cancer.dk/international/english/screening-cervical-cancer-english/. Accessed March 10, 2023.
30. Cocenas-Silva R, Bueno JLO, Droit-Volet S. Temporal memory of emotional experience. *Mem Cognit*. 2012;40:161-167.
31. Deary IJ, Wright AF, Harris SE, Whalley LJ, Starr JM. Searching for genetic influences on normal cognitive ageing. *Trends Cogn Sci*. 2004;8:178-184.
32. Boykoff N, Moieni M, Subramanian SK. Confronting chemobrain: an in-depth look at survivors' reports of impact on work, social networks, and health care response. *J Cancer Surviv*. 2009;3:223-232.
33. Bray VJ, Dhillon HM, Vardy JL. Systematic review of self-reported cognitive function in cancer patients following chemotherapy treatment. *J Cancer Surviv*. 2018;12:537-559.
34. Friis K, Lasgaard M, Rowlands G, Osborne RH, Maindal HT. Health literacy mediates the relationship between educational attainment and health behavior: a Danish population-based study. *J Health Commun*. 2016;21:54-60.
35. Thygesen LC, Ersboll AK. When the entire population is the sample: strengths and limitations in register-based epidemiology. *Eur J Epidemiol*. 2014;29:551-558.

How to cite this article: Jensen HAR, Horsbøl TA, Thygesen LC, Davidsen M, Christensen AI, Ekholm O. Variations in the agreement of self-reported cancer: A Danish nationwide study. *Int J Cancer*. 2024;154(2):217-225. doi:10.1002/ijc.34692