

Agreement Between Self-Reported and Confirmed Cancer Diagnoses in New York City Firefighters and EMS Workers, 2001–2011

RACHEL ZEIG-OWENS, DrPH^{a,b}
ANDREA KABLANIAN, MPH^b
MAYRIS P. WEBBER, DrPH^{b,c}
YANG LIU, MS^{b,c}
EDWARD MAYERSON, MA^b
THERESA SCHWARTZ, MS^{b,c}
NADIA JABER, RPA-C^b
KERRY J. KELLY, MD^b
DAVID J. PREZANT, MD^{b,d}

ABSTRACT

Objectives. Because of the delay in availability of cancer diagnoses from state cancer registries, self-reported diagnoses may be valuable in assessing the current cancer burden in many populations. We evaluated agreement between self-reported cancer diagnoses and state cancer registry-confirmed diagnoses among 21,437 firefighters and emergency medical service workers from the Fire Department of the City of New York. We also investigated the association between World Trade Center (WTC) exposure and other characteristics in relation to accurate reporting of cancer diagnoses.

Methods. Participants self-reported cancer status in questionnaires from October 2, 2001, to December 31, 2011. We obtained data on confirmed cancer diagnoses from nine state cancer registries, which we used as our gold standard. We calculated sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), comparing self-reported cancer diagnoses with confirmed cancer diagnoses. We used multivariable logistic regression models to assess the association between WTC exposure and correct self-report of cancer status, false-positive cancer reports, and false-negative cancer reports.

Results. Sensitivity and specificity for all cancers combined were 90.3% and 98.7%, respectively. Specificities and NPVs remained high in different cancer types, while sensitivities and PPVs varied considerably. WTC exposure was not associated with accurate reporting.

Conclusion. We found high specificities, NPVs, and general concordance between self-reported cancer diagnoses and registry-confirmed diagnoses. Given the low population prevalence of cancer, self-reported cancer diagnoses may be useful for determining non-cancer cases. Because of the low sensitivities and PPVs for some individual cancers, however, case confirmation with state cancer registries or medical records remains critically important.

^aMontefiore Medical Center, Department of Medicine, Bronx, NY

^bFire Department of the City of New York, Bureau of Health Services, Brooklyn, NY

^cMontefiore Medical Center and Albert Einstein College of Medicine, Department of Epidemiology and Population Health, Division of Epidemiology, Bronx, NY

^dMontefiore Medical Center and Albert Einstein College of Medicine, Department of Pulmonary Medicine, Bronx, NY

Address correspondence to: Mayris P. Webber, DrPH, Fire Department of the City of New York, 9 Metrotech Center 5E63K, Brooklyn, NY 11201; tel. 718-999-2665; fax 718-999-0681; e-mail <webberm@fdny.nyc.gov>.

©2016 Association of Schools and Programs of Public Health

Many epidemiologic studies rely upon participant-reported survey data to estimate the prevalence of common diseases and conditions among different communities and populations.^{1,2} The Fire Department of the City of New York (FDNY) previously reported on the agreement between medical records and self-reported physician-diagnosed lower and upper airways disease.^{3,4} Others assessed the validity of self-reported physician cancer diagnoses by reviewing medical records and health registries in specific cohorts, such as an elderly French population and female California teachers, and national population-based studies from the United States and Australia. The sensitivity of self-report for cancer has been shown to vary widely depending on the population surveyed and the cancer site. For example, sensitivities ranged from 15.5% to 90.0% for self-reported prostate cancer, and from 36.9% to 85.3% for self-reported melanoma.⁵⁻⁸ However, the findings from these studies may not be generalizable to other populations, possibly including FDNY firefighters and emergency medical service (EMS) workers, given the recent focus on World Trade Center (WTC)-related illnesses, including cancer.

Our group previously reported the cancer incidence among WTC-exposed and WTC-unexposed FDNY firefighters using state cancer registry and medical record-confirmed cancer cases.⁹ Confirmation of cancer cases can take several years; for example, state cancer registry data are not considered complete until about 2.5 years after the diagnosis year.¹⁰ In addition to the time delay, certain cancer types (e.g., hematologic cancers) are more likely to be diagnosed in doctors' offices rather than in hospitals, leading to longer lag times or even absent reporting to state registries.⁵ Given these understandable issues, the ability to use state cancer registry data to confirm recent cancer trends is limited. Furthermore, confirmation using state cancer registry data or medical records can be costly.

As such, it could be valuable to use self-reported cancer diagnoses from questionnaires, if validated, to assess the current cancer burden in the FDNY population and to provide an early warning of emerging cancer clusters. To our knowledge, no study has assessed the validity of self-reported cancer diagnoses in WTC-exposed individuals and examined whether they were more or less likely than WTC-unexposed individuals to self-report a cancer diagnosis.

This study (1) evaluated the agreement between self-reported cancer diagnoses and state cancer registry-confirmed cancer diagnoses, using the latter as the gold standard among FDNY firefighters and EMS workers; and (2) investigated the association between

WTC exposure and other characteristics in relation to accurate reporting of cancer diagnoses.

METHODS

Study population

The source population consisted of 22,982 FDNY firefighters and EMS workers who provided informed consent to participate in research and completed a medical monitoring questionnaire on or after October 2, 2001, the date self-reported cancer questions were added to the medical monitoring questionnaire. We excluded individuals who did not complete a questionnaire by December 31, 2011 ($n=1,296$), because, at the time of this analysis, cancer cases from state cancer registries were considered complete only as of this date. Furthermore, we restricted the population to people living in one of the nine states where FDNY conducts state cancer registry linkages, thereby excluding an additional 249 individuals. The final study population consisted of 21,437 participants.

Data sources

FDNY Bureau of Health Services (FDNY-BHS) schedules monitoring evaluations every 12 to 18 months for active and WTC-exposed retired responders. These evaluations include self-administered medical monitoring questionnaires. We captured data on participants' self-reported cancer status, highest education level, smoking status (ever/never), and WTC exposure status from their questionnaire responses. We classified individuals as WTC exposed if they reported working at the WTC site for at least one day prior to the FDNY closing of the site on July 25, 2002. We obtained participant demographic information, retirement status, and work assignment (firefighter or EMS worker) from the FDNY employee database.

Self-reported cancer

We defined self-reported cancer for all cancers combined by a positive response to the question, "Has your doctor ever told you that you have or had cancer?" on the medical monitoring questionnaire; the exact wording of the question has been modified slightly over time. Participants who affirmatively answered this question were then asked to identify the type or location of their cancer. We defined self-report of cancer type by the response to this follow-up question: "Identify organ(s) with current or cured cancer. Do not identify organs with benign, non-cancerous growths." Participants were presented with the following list of cancer types to choose from: bladder/kidney, bone/sarcoma, brain/central nervous system (CNS), breast,

colorectal, esophageal/gastric, hematologic (excluding lymphoma), lung (including mesothelioma), lymphoma, melanoma, oral/nasal/throat, ovarian/uterine/cervical, prostate, testicular, thyroid, and other. If an individual completed more than one questionnaire during the study period (October 2, 2001, to December 31, 2011), we used information from the most recent questionnaire.

Confirmed cancer

For the primary analyses, we obtained confirmed cancer diagnoses and diagnosis dates through linkages using complete social security numbers as well as name, date of birth, and sex for all participants to the following state cancer registries: Arizona, Connecticut, Florida, New Jersey, New York, North Carolina, Pennsylvania, South Carolina, and Virginia. These state cancer registries collect information on all cases of cancer diagnosed among their residents. To be consistent with state cancer registry reporting, we excluded non-melanoma skin cancers and in situ cancers, with the exception of in situ bladder cancers. As noted previously, cancers with diagnosis dates after December 31, 2011, were excluded. We considered participants to have confirmed cancer at the time of the questionnaire if the diagnosis date was the same as or prior to the date of their medical monitoring questionnaire, regardless of the questionnaire response regarding cancer type or site.

Statistical analysis

General population characteristics were represented as proportions by work assignment (firefighter or EMS worker) and combined. We calculated sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) comparing self-reported cancer diagnoses with confirmed cancer diagnoses. We defined sensitivity as the proportion of participants with a confirmed cancer diagnosis who correctly self-reported a cancer diagnosis in their medical monitoring questionnaire. We defined specificity as the proportion of participants without a confirmed cancer diagnosis who self-reported no cancer diagnosis. We defined PPV as the proportion of participants who self-reported a cancer diagnosis and had a confirmed cancer diagnosis. We defined NPV as the proportion of participants who self-reported no cancer diagnosis and did not have a confirmed cancer diagnosis. These analyses were repeated for all cancer types with at least five confirmed cancer diagnoses in our cohort.

We used multivariable logistic regression models to assess the association between WTC exposure and a correct self-report of cancer status, controlling for

potential confounders including age (as a continuous variable), smoking status at time of questionnaire (ever/never), education level, sex, and race, if associated with the outcome in univariable models. Participants missing WTC exposure information ($n=88$) were excluded from these models. Additionally, we assessed the association between WTC exposure and a false-positive cancer report (i.e., a self-report of cancer and no confirmed cancer diagnosis) and the association between WTC exposure and a false-negative cancer report (i.e., a self-report of no cancer and a confirmed cancer diagnosis). We repeated all analyses separately for firefighters and EMS workers to account for possible underlying differences in population characteristics. We calculated odds ratios (ORs) and 95% confidence intervals (CIs) from the models.

In addition to cancer cases confirmed using state cancer registry data, the FDNY database also includes cancer cases reported to FDNY and confirmed using medical records. Specifically, confirmation was completed by a trained clinician (Nadia Jaber) who contacted participants who reported a cancer not identified through matches to state cancer registries and requested documentation. For these cases, we required a pathology report or detailed treating physician notes/evaluations (e.g., operative reports, oncology notes with diagnosis/treatment, formal consultations from related specialists, and/or physical findings consistent with oncologic treatments/modalities) to confirm cases. After confirmation, we conducted a secondary analysis assessing agreement by including additional cancer cases confirmed using medical records, but not reported in matches to state cancer registries. We calculated sensitivity, specificity, PPV, and NPV using the expanded confirmed cancer diagnoses. We conducted analyses for all cancers combined and for cancer types using SAS[®] version 9.4.¹¹

RESULTS

Of 21,437 study participants, 17,279 (80.6%) were firefighters and 4,158 (19.4%) were EMS workers. The mean age at the time of the survey was 43.9 years (standard deviation ± 11.5). WTC exposure varied between firefighters and EMS workers in our study, with 70.9% of firefighters but only 48.0% of EMS workers having worked at the WTC site (Table 1).

During the study period, 890 individuals (4.2%) self-reported one or more cancer diagnoses, but only 681 (3.2%) had at least one confirmed cancer case at the time of their questionnaire. Among these 681 individuals, 615 self-reported having cancer, yielding a sensitivity of 90.3% for all cancers combined (Table 2).

Table 1. Selected characteristics of the Fire Department of the City of New York study population, including World Trade Center exposed and unexposed workers, October 2, 2001, to December 31, 2011

Characteristics	Total N (percent) ^a	Firefighters N (percent) ^a	Emergency medical service workers N (percent) ^a
Total	21,437 (100.0)	17,279 (80.6)	4,158 (19.4)
World Trade Center exposure			
Exposed	14,251 (66.5)	12,255 (70.9)	1,996 (48.0)
Unexposed	7,098 (33.1)	4,979 (28.8)	2,119 (51.0)
Missing	88 (0.4)	45 (0.3)	43 (1.0)
Age at questionnaire (mean years, SD)	43.9 (11.5)	44.9 (11.4)	39.7 (10.9)
Race			
White	17,500 (81.6)	15,572 (90.1)	1,928 (46.4)
Nonwhite	3,937 (18.4)	1,707 (9.9)	2,230 (53.6)
Sex			
Male	20,282 (94.6)	17,216 (99.6)	3,066 (73.7)
Female	1,155 (5.4)	63 (0.4)	1,092 (26.3)
Work status			
Active	15,233 (71.1)	11,589 (67.1)	3,644 (87.6)
Retired	6,204 (28.9)	5,690 (32.9)	514 (12.4)
Smoking status ^b			
Ever	8,394 (39.2)	6,381 (36.9)	2,013 (48.4)
Never	13,043 (60.8)	10,898 (63.1)	2,145 (51.6)
Highest education ^c			
High school	4,864 (22.7)	3,491 (20.2)	1,373 (33.0)
College	16,573 (77.3)	13,788 (79.8)	2,785 (67.1)

^aPercentages may not sum to 100 due to rounding.

^bA total of 52 individuals with missing smoking status were defined as "never smokers."

^cA total of 1,011 individuals with missing education information were defined as "high school."

SD = standard deviation

Of the 20,756 participants without confirmed cancer cases, 20,481 self-reported not having cancer, yielding a specificity of 98.7%. The PPV and NPV were 69.1% and 99.7%, respectively. Overall, 21,096 individuals (98.4%) correctly self-reported their cancer status, 275 (1.3%) had a false-positive report of cancer, and 66 (0.3%) had a false-negative report of cancer.

Although specificities and NPVs remained similarly high across all cancer types, sensitivities and PPVs varied by cancer diagnosis. Sensitivities ranged from 33.3% for brain/CNS cancer to 100.0% for esophageal/gastric cancer. PPVs ranged from 16.3% for bone cancer/sarcoma to 94.2% for prostate cancer (Table 2).

Final logistic regression models used the following outcomes: correct self-report of cancer status, false-positive report of cancer, and false-negative report of cancer among all cancers combined (Table 3). WTC exposure was not significantly associated with agreement between self-reported cancer and confirmed cancer (OR=0.85, 95% CI 0.59, 1.23). Older individuals and ever-smoking individuals were more likely than never smokers to self-report a cancer that, after review, was not confirmed. Race, sex, and active work status were also associated with agreement between

self-reported cases and confirmed cases. When analyses were repeated separately for firefighters and EMS workers, results were similar.

The secondary analysis added 70 additional people with at least one confirmed cancer diagnosis. Although the sensitivity, specificity, and NPV remained similar to those in the primary analysis, the PPV increased to 75.1% from 69.1%. Results for types of cancer were similar to those in the primary analysis. Melanoma skin cancer and hematologic cancers had the greatest number of cases confirmed using only medical records. As a result, the PPV was higher than in the primary analysis: 35.4% vs. 26.3% for melanoma skin cancer and 70.2% vs. 51.1% for hematologic cancers.

DISCUSSION

The results from this study demonstrate adequate agreement between self-reported and confirmed cancer diagnoses. Furthermore, the specificities and NPVs for self-reporting all cancers combined, as well as for specific cancer types, indicated that cancer self-reporting may be most useful for identifying individuals who do not have cancer when resources are limited and

medical records are unavailable. However, we note that more than 30% of the self-reported cancer cases were not confirmed (i.e., false positive). Therefore, caution in relying too heavily on self-report is appropriate, as we found that PPVs for self-reported cancers varied considerably, depending on the specific cancer type of interest; as a result, it may not be a useful sentinel indicator of emerging disease.

Because we did not find an association between WTC exposure and accurate reporting of cancer diagnoses, these findings are likely to be generalizable to other populations. We did observe, however, that older individuals and those with a history of smoking were more likely to report cancer diagnoses that were not confirmed either by registry matches or medical records. These individuals may be exposed to higher rates of cancer screening due to a perceived or increased cancer risk; as a result, they may have stress-related concerns or have experienced suspicious/abnormal screenings that were ultimately benign, thereby increasing their likelihood of providing an unconfirmed positive self-report of cancer.

Similar to studies on non-WTC-exposed populations, we found inconsistencies by specific cancer types.⁵⁻⁸ One possible explanation is that participants may have difficulty differentiating between primary and

metastatic cancer sites. For example, after reviewing the medical records of individuals with confirmed cancer, we found that some respondents self-reported bone cancer but actually had a different type of confirmed primary cancer that had metastasized to the bone. Therefore, for some cancer types, this confusion between primary and metastatic sites was a potential contributor to a low PPV.

Melanoma skin cancer and hematologic cancers had the greatest number of cases that were classified as false-positive results in our primary analysis, but were confirmed as a cancer diagnosis in the secondary analysis using medical records to supplement state registry data. We believe this situation may be related to national trends that suggest these diagnoses are currently more likely to be made in doctors' offices rather than in hospitals, leading to longer lag times or even absent reporting to state registries.⁵ Therefore, depending on the type of cancer being studied, investigators should not rely exclusively on cancer registry data as the gold standard to detect or confirm cases.

Strengths and limitations

The primary strength of our study was our comparison of self-reported cancer diagnoses against state registry and medical record-confirmed cancer diagnoses. These

Table 2. Sensitivity, specificity, positive predictive value, and negative predictive value of self-reported cancer diagnoses in relation to confirmed cancer diagnoses in World Trade Center exposed and unexposed workers, United States, October 2, 2001, to December 31, 2011

Cancer type	Number of confirmed cases	Sensitivity	Specificity	Positive predictive value	Negative predictive value	Correct self-report
All cancers combined ^a	681	90.3	98.7	69.1	99.7	98.4
Prostate ^b	241	94.6	99.9	94.2	99.9	99.9
Melanoma	67	82.1	99.3	26.3	99.9	99.2
Lymphoma	58	79.3	99.9	73.0	99.9	99.9
Bladder/kidney	58	87.9	100.0	85.0	100.0	99.9
Testicular	50	92.0	100.0	82.1	100.0	99.9
Colorectal	45	80.0	99.9	58.1	100.0	99.8
Hematologic ^c	37	64.9	99.9	51.1	99.9	99.8
Thyroid	40	92.5	100.0	82.2	100.0	99.9
Oral/nasal/throat	29	89.7	99.9	56.5	100.0	99.9
Lung ^d	24	95.8	99.9	57.5	100.0	99.9
Bone/sarcoma	13	61.5	99.8	16.3	100.0	99.8
Esophageal/gastric	12	100.0	100.0	63.2	100.0	100.0
Ovarian/uterine/cervical ^e	6	66.7	99.1	28.6	99.8	99.0
Brain/central nervous system	6	33.3	100.0	28.6	100.0	100.0

^aIncludes people with one or more cancer diagnoses

^bRestricted to males only

^cIncludes leukemia (n=21), multiple myeloma (n=10), and other hematopoietic (n=6) cancers

^dIncludes mesothelioma (n=1)

^eRestricted to females only

Table 3. Correct, false-positive, and false-negative self-reports of cancer status among all cancers combined in World Trade Center exposed and unexposed workers, October 2, 2001, to December 31, 2011

Variable	Correct self-report ^a		False-positive self-report ^a		False-negative self-report ^a	
	OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value
World Trade Center exposure						
Exposed	0.85 (0.59, 1.23)	0.393	1.28 (0.84, 1.95)	0.248	0.90 (0.44, 1.84)	0.776
Unexposed	Ref.		Ref.		Ref.	
Race						
Nonwhite	2.30 (1.48, 3.55)	<0.001	0.35 (0.21, 0.59)	<0.001	NA	
White	Ref.		Ref.			
Age at questionnaire (in years)	0.94 (0.93, 0.95)	<0.001	1.06 (1.05, 1.08)	<0.001	1.07 (1.05, 1.10)	<0.001
Sex						
Male	2.85 (1.78, 4.56)	<0.001	0.28 (0.17, 0.46)	<0.001	NA	
Female	Ref.		Ref.			
Work status						
Active	1.49 (1.12, 1.99)	0.007	0.60 (0.43, 0.83)	0.002	NA	
Retired	Ref.		Ref.			
Smoking status						
Ever	0.78 (0.62, 0.97)	0.028	1.33 (1.04, 1.70)	0.025	NA	
Never	Ref.		Ref.			

^aNon-significant variables in univariable analyses were not included in the final model. Work assignment and education were not significant in any of the models.

OR = odds ratio

CI = confidence interval

Ref. = reference group

NA = not applicable

state cancer registries were considered to be at least 90% complete during the study period,^{12,13} giving us further confidence in case classifications. However, because the state cancer registries are not 100% complete, and we were unable to access medical records outside of FDNY, there was a potential for misclassification. This limitation was likely to be minimal and limited to those who did not self-report a cancer diagnosis, as in our secondary analyses we included cases confirmed using medical records provided to FDNY after a self-report of cancer on the medical monitoring questionnaire. Our findings from the primary analysis found that sensitivity and specificity for all cancers remained relatively high even when cases were limited to those obtained from state cancer registries. Another limitation was that we were unable to assess agreement between self-reported and in situ cancers, because cancer registries do not report in situ cases.

CONCLUSION

This study demonstrated the general usefulness of self-reported cancer when time or resources are limited, and contrasted our previous findings where agreement

between self-reported diagnoses and medical records of lower and upper respiratory conditions was lower. However, as demonstrated by the low PPVs for some cancers, especially for brain, bone, melanoma, and hematologic cancers, it remains necessary to confirm all self-reported cancer cases with either state cancer registry data or medical records.

The authors thank the New York State Cancer Registry, the New York State Department of Health, and the New Jersey State Cancer Registry (NJSCR) for their comments while writing the manuscript. NJSCR is supported by the Centers for Disease Control and Prevention under cooperative agreement #5U58DP003931-02; the National Cancer Institute's Surveillance, Epidemiology, and End Results Program under contract #HHSN 261201300021I NCI Control No. N01PC-2013-00021; and the State of New Jersey.

This study was approved by the Institutional Review Board at Montefiore Medical Center and by the institutional review boards of the state cancer registries when required.

REFERENCES

- Centers for Disease Control and Prevention (US). National Health Interview Survey: questionnaires, datasets, and related documentation, 1997 to the present [cited 2014 Nov 11]. Available from: http://www.cdc.gov/nchs/nhis/quest_data_related_1997_forward.htm

2. The City of New York. WTC health registry [cited 2014 Nov 1]. Available from: <http://www.nyc.gov/html/doh/wtc/html/registry/registry.shtml>
3. Weakley J, Webber MP, Ye F, Zeig-Owens R, Cohen HW, Hall CB, et al. Agreement between obstructive airways disease diagnoses from self-report questionnaires and medical records. *Prev Med* 2013;57:38-42.
4. Weakley J, Webber MP, Ye F, Zeig-Owens R, Cohen HW, Hall CB, et al. Agreement between upper respiratory diagnoses from self-report questionnaires and medical records in an occupational health setting. *Am J Ind Med* 2014;57:1181-7.
5. 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;147:556-62.
6. Berthier F, Grosclaude P, Bocquet H, Faliu B, Cayla F, Machelard-Roumagnac M. Prevalence of cancer in the elderly: discrepancies between self-reported and registry data. *Br J Cancer* 1997;75:445-7.
7. 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;157:539-45.
8. 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;38:35-8.
9. Zeig-Owens R, Webber MP, Hall CB, Schwartz T, Jaber N, Weakley J, et al. Early assessment of cancer outcomes in New York City firefighters after the 9/11 attacks: an observational cohort study. *Lancet* 2011;378:898-905.
10. National Cancer Institute (US). SEER Cancer Statistics Review (CSR) 1975–2011. 2014 [cited 2014 Nov 1]. Available from: http://seer.cancer.gov/csr/1975_2011
11. SAS Institute, Inc. SAS®: Version 9.4. Cary (NC); SAS Institute, Inc.; 2013.
12. New York State Department of Health. New York State cancer registry [cited 2014 Nov 1]. Available from: <http://www.health.ny.gov/statistics/cancer/registry/about.htm>
13. North American Association of Central Cancer Registries. Certification levels [cited 2014 Nov 1]. Available from: <http://www.naacr.org/certification/certificationlevels.aspx>