Reliability of Obituaries as a Data Source in Epidemiologic Studies: Agreement in Age, Residence and Occupation

Donna B. Gerstle¹, MA, MS, JD, Kristen B. White², MPH, CPH, Alfred M. Levine³, PhD.

¹Donna B. Gerstle, MA, MS, JD. Executive Director and Principle Investigator for the Staten Island Breast Cancer Research Initiative at the College of Staten Island, Center for Environmental Science. Email: Donna.Gerstle@csi.cuny.edu

²Kristen White, MPH, CPH. Epidemiologic Research Associate at the College of Staten Island, Center for Environmental Science. Email: Kristen.White@csi.cuny.edu

³Alfred Levine, PhD. Professor of Engineering Science and Physics at the College of Staten Island, Center for Environmental Science. Email: Alfred.Levine@csi.cuny.edu

Accepted 21 April, 2013

Abstract: Newspaper obituaries are an easily accessible and cost-effective data source. Although obituaries offset many of the challenges involved with the cohort study designs, their use in public health research is limited and their reliability remains untested. The current study investigates the reliability of life history data obtained from obituaries for 620 lung cancer cases and controls between the years 1980-1990. New York City Department of Health & Mental Hygiene death certificates, in-person/surrogate interviews, and obituary notices from the Staten Island Advance were obtained and compared on the following life history variables: Age, Length of Residence on Staten Island, Major Job Title, and Occupational Risk Exposure. Obituaries and in-person/surrogate interviews were in excellent overall percent agreement on 'Age' [97.3% (+/-1yr)], 'Length of Residence on Staten Island' [76.1% (+/-1 year); 81.1% (+/-2 years); 85.3% (+/-3 years); 87.6% (+/-4 years); 91.9% (+/-5 years)], ‘Major Job Title’ [Kappa=0.87; 95% CI [0.83, 0.91]; p<0.0001], and ‘Occupational Risk Exposure’ [Kappa=0.51; 95% CI [0.47, 0.55]; p<0.0001]. We show that obituary notices are comparable to in-person and surrogate interviews on a number of personal and exposure history variables. These highly correlated results indicate that newspaper obituaries are a reliable source of life history information and could be utilized more commonly in epidemiologic research. We demonstrate that when detailed life history data obtained in an obituary is coupled with death certificate information, researchers can conduct original and progressive epidemiologic studies on chronic disease.

Keyword: obituaries, cancer, Staten Island, epidemiologic, Residence and Occupation.

Introduction

Background

Researchers at the Center for Environmental Science at the College of Staten Island, a City University of New York, are currently conducting a large case-control study on cancer mortality for the calendar years 1980-2006. This ongoing study utilizes epidemiologic data provided by the New York City Department of Health & Mental Hygiene (NYCDOHMH), in-person/surrogate interviews, and local newspaper obituaries collected from the Staten Island Advance along the course of the study. An innovative method of obtaining individual residence history was developed by combining NYCDOHMH data with data obtained in newspaper obituaries. Although the obituary notice has been used as a corroborating data source in several previous studies (Liebling et al., 1984; Adesina et al., 1991; Camilli et al., 1991; Heilbrun et al., 1991; Klein et al., 1995), the reliability of obituary data has not been closely examined. In contrast, the use of in-person and surrogate interviews has been established as a reliable source of personal history information for specific variables (Prochazka et al., 2008).

In order to validate the use of obituaries as an alternative reliable data source, the purpose of the current study is to examine agreement between life history data acquired through newspaper obituaries and in-person/surrogate interviews occurring between the years 1980-1990. Agreement was investigated between the two data sources by 'Age,' 'Length of Residence on Staten Island,' 'Major Job Title,' and 'Occupational Risk Exposure.'

Staten Island Advance

The Staten Island Advance is a daily newspaper published in the New York City borough of Staten Island (Richmond County). The Advance is the only major newspaper published in the borough and covers local news and events of interest to the community. In addition to publishing brief death notices with funeral and contact information, the Advance also publishes long narratives including information about residence history, occupational history, socioeconomic variables, military record, and interests or hobbies of the decedent.

Methods

The current study was conducted within a larger case-control study of cancer mortality on Staten Island during the period 1980-1990. Death certificate data for cases and controls were obtained with permission through the NYCDOHMH. In addition to death certificates, in-person/surrogate interviews and newspaper obituaries were located for 620 (265-cases; 355-controls) deceased Staten Island residents from the years 1980-1990. All in-person and surrogate interviews were conducted during the five-year period 1987-1991 and each were completed within ten years of the occurring death. Surrogate interviews were performed when in-person interviews were not available or appropriate and newspaper obituaries were extracted from the Staten Island Advance. Cases were identified by the NYCDOHMH who had histologically confirmed primary lung cancer. Controls were also identified from NYCDOHMH data, but were classified as those individuals who had died from causes other than malignant neoplasms of the lung.

The protocol for the current study obtained approval from the College of Staten Island's Institutional Review Board and
Human Subjects Committee. For in-person and surrogate interviews, informed consent papers were given to each interviewee and signed before an interview was conducted.

In all of the data reported in the study, a unique match between obituary and in-person/surrogate interview was obtained on as many informative categories as possible. Those individuals without a death certificate record from the NYCDOHMH, in-person/surrogate interview, and printed obituary or who we were unable to match, were excluded from analysis.

Age, Residence, and Occupation.

Newspaper obituaries for 620 Staten Island residents were matched to both in-person/surrogate interviews and NYCDOHMH data on life history variables: 'Age', 'Length of Residence on Staten Island', 'Major Job Title', and 'Occupational Risk Exposure'. The descriptive variable, 'Age', was matched between newspaper obituaries and in-person/surrogate interviews with the agreement criteria of +/- one year. Obituary and interview 'Age' was also matched to the NYCDOHMH death certificate value, which was considered to be the most reliable standard.

'Length of Residence on Staten Island' was defined as the total number of years individuals had resided in the New York City borough. The percent of exact agreement between obituaries and interviews was determined first and then in order to investigate the difference when an exact match was not obtained, we used the agreement criteria of +/- one, two, three, four, and five years between obituaries and interviews. The duration of residence at each address among individuals who had lived at more than one address on the island was not included in this analysis. We were interested only in the total aggregated years that an individual spent on Staten Island.

Occupation history was obtained in the form of 'Major Job Title' and 'Occupational Risk Exposure' for both interviews and matching subsequent obituaries. In-person/surrogate interview and obituary 'Major Job Title' were considered in agreement if both were classified in the same primary industrial category (i.e. finance, construction, etc.) and if the majority of the daily functions of both jobs were congruent (i.e. treasurer versus bank officer).

Risk exposure for occupation was based on the chemical classification system established by the International Agency for Research on Cancer (IARC), which consists of four major divisions of agents, mixtures, and exposures: Group 1 (known human carcinogen); Group 2A (probable human carcinogen); Group 2B (possible human carcinogen); Group 3 (not classifiable as a human carcinogen); and Group 4 (probably not a human carcinogen) (IARC, 2011; Siemiatycki et al., 2004). Risk was assigned to 'Major Job Title' based on a list compiled by Siemiatycki et al. (2004) that allocated IARC classified carcinogens to specific industrial divisions. The risk was ranked 'high' (≥1 exposure in IARC Group 1), 'medium' (≥1 exposure in IARC Group 2A or 2B), or 'low' (≥1 exposure in IARC Group 3 or 4) for 'Major Job Titles' in both obituaries and interviews. In order to minimize misclassification bias, job titles were randomized before risk exposure was individually assigned.

Staten Island Advance

The Staten Island Advance obtained personal history information through a comprehensive and thorough surrogate interview process. The interviewee, most commonly a surviving spouse or family member, was contacted by a Staten Island Advance reporter within twenty-four to forty-eight hours of the individual’s death. The interview was conducted by a skilled reporter and was based upon a pre-established, systematic questionnaire. The substantive portion of the questionnaire used by the Staten Island Advance for the use of obituary reports has remained unaltered for decades.

Statistical Methods

Overall percent agreement between obituaries and interviews were calculated for each major variable and linear regression analysis was conducted to obtain statistical associations, including the correlation coefficient (Pearson’s r) and the coefficient of determination ($r^2$). Kappa statistics, along with 95% confidence intervals, were computed to assess inter-observer variation and reliability. Landis & Koch (1977) defined the interpretation of the kappa statistic as: <0 as poor agreement, 0.0-0.20 as slight agreement, 0.21-0.40 as fair agreement, 0.41-0.60 as moderate agreement, 0.61-0.80 as substantial agreement, and 0.81-1.0 as almost perfect agreement. All statistical analyses were performed in SPSS and Microsoft Excel with a significance of p<0.05.

Results

Newspaper obituaries, in-person/surrogate interviews, and death certificates were obtained for 620 lung cancer cases (N=265) and controls (N=355) from 1980-1990. The Staten Island Advance reported on approximately 86.6% of all deaths on Staten Island, thereby representing a substantial majority of the study population. Cases were younger than the controls in both obituaries and in-person/surrogate interviews with a mean age of 67 years compared to a mean age of 70 years. The mean length of residence years spent on Staten Island varied between obituaries and in-person/surrogate interviews for controls and cases, which was 50 years compared to 68 years respectively. Among the controls, average 'Length of Residence on Staten Island' was reported as 45 years for both obituaries and interviews.

We examined the time between an individual’s death and the occurrence of a surrogate interview and found that 66.1% of all surrogate interviews were conducted within five years and that 99.5% of all surrogate interviews took place in less than ten years of the individual’s date of death.

Newspaper obituaries and in-person/surrogate interviewwerenon excellent overall percent agreement on 'Age' [97.3% (+/-1yr)] Table 1. The relationship between reported 'Age' in newspaper obituaries and in-person/surrogate interviews were significantly correlated (r²=0.9941; p<0.0001) Fig. 1.Spearman’s ranks were significantly correlated (rₛ=1.0; p<0.0001) Table 3. High
overall percent agreement was also observed between obituaries and NYCDOHMH death certificate data on reported 'Age' [96.5% (+/-1yr)] Table 1.The relationship between reported 'Age' in obituaries and NYCDOHMH death certificate data were significantly correlated ($r^2=0.959; p<0.0001$) Fig. 2. Spearman’s ranks were also significantly correlated ($r_s=0.98; p<0.001$) Table 3. In-person/surrogate interviews and NYCDOHMH death certificate data were in excellent overall percent agreement on 'Age' [95.0% (+/-1yr)] Table 1. The relationship between reported 'Age' in in-person/surrogate interviews and NYCDOHMH death certificate data were significantly correlated ($r^2=0.9571; p<0.0001$) Fig. 3. Spearman’s ranks were also significantly correlated ($r_s=0.98$) Table 3.

Seventy-six percent of the compared obituaries and interviews on recorded 'Length of Residence on Staten Island' were an exact match. Total overall agreement on 'Length of Residence on Staten Island' between newspaper obituaries and in-person/surrogate interviews was outstanding [76.1% (+/-1 year); 81.1% (+/-2 years); 85.3% (+/-3 years); 87.6% (+/-4 years); 91.9% (+/-5 years)] Table 2. We examined the 8.1% of remaining, non-matching obituaries and interviews (+/- 5 years) and found no evidence of systematic bias present or any apparent reason for the discrepancy shown. Approximately 3% of the 620 lung cancer cases and controls did not agree within the matching criteria of +/- 10 years. The relationship between reported 'Length of Residence on Staten Island' between obituaries and in-person/surrogate interviews were significantly correlated ($r^2=0.9704; p<0.0001$) Fig. 3. Spearman’s ranks were also significantly correlated ($r_s=0.97; p<0.0001$) Table 3.

Regarding occupational variables, the inter-rater reliability was very good for matching on 'Major Job Title' between obituaries and interviews [Kappa=0.87; 95% CI (0.83, 0.91); p<0.0001] Table 4 and moderate agreement was seen matching on 'Occupational Risk Exposure' between obituaries and interviews [Kappa=0.51; 95% CI (0.47, 0.55); p<0.0001] Table 4.

**Discussion**

Data obtained from routine surveillance efforts and other forms of secondary data collection have been commonly used in epidemiology for several years (Stang et al., 1999). In cancer studies, traditional methods of data collection commonly include, but are not limited to: health departments, medical records, patient/surrogate interviews, death certificates, state and national cancer registries, Medicare claims data, and the Surveillance, Epidemiology, and End Results (SEER) program (Stang et al., 1999). The primary intent of the current study was to evaluate the role of local newspaper obituaries as a reliable source of life history data in cancer studies. We found obituaries and interviews to be highly correlated in 'Age,' 'Length of Residence on Staten Island,' and 'Major Job Title' for both cases and controls. The significant correlations seen among all primary variables indicate that obituaries are an alternative data source for life history information.

The results from our study have the potential to greatly influence cancer research methodologies in a variety of ways. Cohort studies are the most valid observational study design in epidemiologic research (Gordis, 2009). Although ideal in deducing causal inferences, cohort studies have inherent complications that include: extensive amount of study time required, high associated cost, potential loss-to-follow up of study participants, and difficulty obtaining a large sample size (Gordis, 2009). Newspaper obituaries are a widely available research tool that eliminates several common limitations of the cohort design.

First, obituary notices offer a way to identify epidemiologic data on cases and controls quickly, with little associated cost. Large cancer databases, although providing a substantial research population, can be costly and may not be the most feasible option for many investigators. In contrast, obituaries are made available to the public through free of charge. Even as technology continues to advance, obituaries remain a constant and their availability through the internet in online newspapers has only enhanced the accessibility of these documents.

As well as cost, obituaries allow researchers the ability to generate large datasets in less time than conventional methods. Obtaining large cancer datasets can take several months or more to receive, creating an inconvenient waiting period for investigators. This waiting period can be particularly damaging in situations requiring an urgent public health intervention. Utilizing existing life history data from obituaries helps to offset limitations of the cohort study in both speed and cost-effectiveness.

In addition to the benefits described above, we speculate that not only does the use of life history data provided in newspaper obituaries decrease both cost and time for investigators, but that they challenge the traditional epidemiologic use of in-person/surrogate interviews. To the best of our knowledge, obituaries have never been used in epidemiologic cancer research to establish residence or exposure histories. Previously, the only method used in the field of epidemiology for obtaining past information on a deceased individual was to conduct a surrogate interview. Surrogate interview questionnaires are frequently subjective in content, can be inconsistently collected, and vary substantially from study to study. We also know that traditional surrogate interviews, as well as in-person interviews, can be subject to forms of recall bias. In the current study, the time between the individual’s date of death and the surrogate interview process varied, but all were conducted within a ten-year span. The longer the period between the death and the interview becomes, the more the information obtained in the interview becomes subject to recall bias. We recognize that similar effects of recall bias subjectivity exist in the reported obituary notices; however, obituary interviews are conducted within a very short period (twenty-four to forty-eight hours) by a highly trained, paid reporter and are based upon a consistent, systematic document that has remained unaltered for decades, as is the case with the Staten Island Advance.

**Limitations**
Although our study indicates a number of strengths associated with use of obituary data, inherent limitations exist. The first limitation is that the Staten Island *Advance* represents a very specific population and practices superior reporting common to a small community setting. The quality, degree, and sampling of deceased individuals covered by the Staten Island newspaper may not be representative of other populations. In addition to coverage area, the questionnaires used by reporters of the Staten Island *Advance* may vary depending on location and may provide more or less personal history information.

**Conclusion**

Our current results indicate that obituary notices are comparable to in-person/surrogate interviews on a number of personal and exposure history variables. This finding affirms the reliability of the life history data obtained in newspaper obituaries and the subsequent use of this information in cancer research efforts. We conclude that personal history information acquired from newspaper obituaries is a reliable source of epidemiologic data in cancer research. We determined that when detailed life history information obtained in an obituary is coupled with NYCDOHMH death certificate data, researchers can conduct original and progressive epidemiologic studies. Moreover, we hypothesize that future research into the use of epidemiologic data obtained through obituaries may prove a superior method of data collection compared to the traditional in-person/surrogate interview process.

**References**


Table 1. Overall Percent Agreement on ‘Age’ between Newspaper Obituaries, In-person/Surrogate Interviews, and NYCDOHMH Death Certificates.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Total Percent Agreement (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obituary v. In-person/Surrogate Interview</td>
<td></td>
</tr>
<tr>
<td>+/- 1 year</td>
<td>97.3</td>
</tr>
<tr>
<td>Obituary v. NYCDOHMH Death Certificate</td>
<td></td>
</tr>
<tr>
<td>+/- 1 year</td>
<td>96.5</td>
</tr>
<tr>
<td>In-person/Surrogate Interview v. NYCDOHMH Death Certificate</td>
<td>95</td>
</tr>
<tr>
<td>Total</td>
<td>N=620</td>
</tr>
</tbody>
</table>

Table 2. Relationship between ‘Length of Residence on Staten Island’ between Newspaper Obituaries and In-person/Surrogate Interviews.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Total Percent Agreement (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obituary v. In-person/Surrogate Interview</td>
<td></td>
</tr>
<tr>
<td>+/- 1 year</td>
<td>76.1</td>
</tr>
<tr>
<td>+/- 2 year</td>
<td>81.1</td>
</tr>
<tr>
<td>+/- 3 year</td>
<td>85.3</td>
</tr>
<tr>
<td>+/- 4 year</td>
<td>87.6</td>
</tr>
<tr>
<td>+/- 5 year</td>
<td>91.9</td>
</tr>
<tr>
<td>Total</td>
<td>N=620</td>
</tr>
</tbody>
</table>

Table 3. Relationship of Obituaries, Interviews, and NYCDOHMH Death Certificates on ‘Age’ and ‘Length of Residence on Staten Island.’

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Overall Percent Agreement (%)</th>
<th>R</th>
<th>r²</th>
<th>Spearman Rho (r₂)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (+/- 1 year)</td>
<td>97.3</td>
<td>1</td>
<td>0.99</td>
<td>1</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Obituary v. In-person/Surrogate Interview</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obituary v. NYCDOHMH Death Certificate</td>
<td>96.5</td>
<td>0.98</td>
<td>0.96</td>
<td>0.98</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>In-person/Surrogate Interview v. NYCDOHMH Death Certificate</td>
<td>95</td>
<td>0.98</td>
<td>0.96</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td>Length of Residence on Staten Island</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Obituary v. In-person/Surrogate Interview</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Relationship between Obituaries and In-person/Surrogate Interviews for Occupational Variables.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Total Percent Agreement (%)</th>
<th>Kappa Statistic (K)</th>
<th>95% C.L</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major Job Title</td>
<td>94.7</td>
<td>0.87</td>
<td>(0.83, 0.91)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Occupational Risk Exposure</td>
<td>68</td>
<td>0.51</td>
<td>(0.47, 0.55)</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>
**Figure 1.** Relationship between reported 'Age' in Newspaper Obituaries and In-person/Surrogate Interviews.

![Graph showing the relationship between Newspaper Obituary 'Age' and In-person/Surrogate Interview 'Age'. The equation is y = 1.000x + 0.177 and R² = 0.994.]

**Figure 2.** Relationship between reported 'Age' in Newspaper Obituaries and NYCDOHMH death certificates.

![Graph showing the relationship between Newspaper Obituary 'Age' and NYCDOHMH 'Age'. The equation is y = 0.986x + 0.925 and R² = 0.959.]

**Figure 3.** Relationship between reported 'Age' in In-person/Surrogate Interviews and NYCDOHMH death certificates.

**Figure 4.** Relationship between reported 'Length of Residence on Staten Island' in Newspaper Obituaries and In-person/Surrogate Interviews.