COMPARISON BETWEEN RESPONDENTS AND NON-RESPONDENTS IN A NESTED CASE-CONTROL STUDY OF BRAIN TUMORS

by

Hui Xu

B.Med., The Suzhou University, China, 1998

Submitted to the Graduate Faculty of

Department of Biostatistics

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2009

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This thesis was presented

by

Hui Xu

It was defended on

December 10, 2009

and approved by

Gary M. Marsh, Ph.D, Professor
Department of Biostatistics, Graduate School of Public Health
University of Pittsburgh

Jeanine Buchanich, Ph.D, Research Assistant Professor
Department of Biostatistics, Graduate School of Public Health
University of Pittsburgh

Marike Vuga, Ph.D, Research Assistant Professor
Department of Epidemiology, Graduate School of Public Health
University of Pittsburgh

Thesis Advisor: Ada O. Youk, Ph.D, Research Assistant Professor
Department of Biostatistics, Graduate School of Public Health
University of Pittsburgh

COMPARISON BETWEEN RESPONDENTS AND NON-RESPONDENTS IN A NESTED CASE-CONTROL STUDY OF BRAIN TUMORS

Hui Xu, M. S.

University of Pittsburgh, 2009

The purpose of this study was to identify characteristics of non-respondents and estimate the potential non-response bias by comparing the respondents with non-respondents from a nested case-control study of brain tumors. The nested case-control study was conducted in eight Pratt & Whitney plants in Connecticut. Information about demographic and some work related variables of 239 cases and 116 controls who responded to an interview, as well as 483 cases and 604 controls who did not respond were obtained from the plant records. Pearson's chi-square test was used to test whether these common known variables were differently distributed between respondents and non-respondents by case-control status. There were no differences detected between the respondents and the non-respondents in the control group. However, significant distribution differences were identified between the case respondents and the case non-respondents with respect to the variables: age at hire, age at termination, and duration of time worked. Multivariate logistic regression was conducted to specify which variables were significantly associated with non-response. The probability of being a non-respondent in the case group was significantly associated with age at hire and age at termination. Furthermore, case-control status, age at hire, and duration of time worked were significant predictors for being a non-respondent in the whole dataset. In addition, the non-response biases in brain tumor risk associated with age at hire and age at termination were calculated by comparing risk among respondents and all subjects. The bias varied from -9% to 43%, indicating that difference

between the respondents and the non-respondents may result in a large bias in the risk estimate for brain tumors in the nested case-control study. Our study has great public health relevance because survey data with low response rate could undermine the results of a case-control study of some exposure of interest and a specific disease, or worse lead to erroneous conclusions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

Case-control studies are faced with increasing challenges in obtaining a sufficiently high response rate, because fewer and fewer people are willing to participate in research studies. In our nested case-control study of brain tumors, a plan to increase the response rate was implemented, including at least sending letters and consent forms two times for non-responders and the use of selective financial incentives. The study still suffered from low response rate. In this study, we investigated the feasibility of using the respondent data only to represent the overall data by comparing the respondents and non-respondents of the nested case-control study of brain tumors and estimating potential non-response bias.

1.1 REVIEW AND PERSPECTIVE OF PRATT&WHITNEY STUDY

Pratt & Whitney (P&W), one of the largest aircraft engine manufacturers in the world, was founded in Hartford, Connecticut (CT), in 1925. Since then, P&W has established up to eight plants in North Haven (NH), East Hartford, Middletown, Rocky Hill, Southington-Aircraft Road, Southington-Newell Street, Cheshire and Manchester Foundry in CT successively. Most of these plants operated from the 1960s to the 1990s. However, only three of these plants in East Hartford, Middletown, and Cheshire are still operating now. The others were closed from 1988 to 2002.

The CT Department of Public Health (CTDPH) started an investigation in May 2000 as a response to a suspected cluster of brain tumors at the P&W engine manufacturing plant in NH, CT. Around one year later, 14 cases of primary malignant brain tumor were identified by CTDPH. All of these cases were verified by the CT Tumor Registry Program; all occurred in white male workers. Twelve of these cases were the same aggressive type: glioblastoma multiforme. CTDPH worked further to conduct a preliminary comparative cancer incidence analysis, which indicated that 14 cases may be excessive, so the CTDPH suggested a carefully designed epidemiologic study should be performed by an independent research group.

In August 2001, the National Institute for Occupational Safety and Health and the CTDPH recommended two experts (Dr. Marsh, a biostatistician of the University of Pittsburgh, Department of Biostatistics (UPitt), and Dr. Esmen, an expert in assessing and reconstructing workplace exposures at University of Illinois at Chicago (UIC)) to determine the feasibility of conducting a formal epidemiological study of the suspected excessive brain tumors. Drs. Marsh and Esmen began the feasibility study in September, 2001. They met with P&W administrative, medical and environmental staff at the East Hartford office complex and tried to assess the feasibility of reconstructing the NH workforce and plant processes. The feasibility study used the related information provided by P&W from the plants in NH, East Hartford, Middletown, Rocky Hill, Southington-Aircraft Road and Southington-Newell Street. The feasibility study results confirmed that there were sufficient data to ensure a formal epidemiological investigation and recommended that two other P&W central CT plants at Cheshire and Manchester Foundry should also be included in the study to increase the study power. Finally, all eight plants in CT were included in P&W study. Thus, a study initially estimated to cover about 100,000 workers increased to a $12-million, seven-year investigation of about 224,000 workers during the years

from 1952 to 2001, which is one of the largest workplace health studies in America ever conducted up until now.

This multipart epidemiology study was started in July 2002. The Department of Biostatistics at UPitt is responsible for the epidemiology and biostatistical component of the study, which involves mainly four parts: historical cohort mortality, cancer incidence studies, a nested case–control study of malignant and benign brain tumor and other central nervous system (CNS) cancer sites, and a genetics study based on tissue samples obtained from cases of brain tumor. The Division of Environmental and Occupational Health Sciences at UIC took charge of the companion exposure assessment project, by which investigators tried to characterize the historical work practices and exposures that occurred in each P&W study plant. At the end of the study, UPitt will determine the association between the past working environment and mortality and incidence from CNS by linking exposure information and work history. To date, the total study cohort includes 223,894 subjects during the study period of 1952 to 2001. The investigators have reported that at the total cohort level, the total and cause-specific mortality including CNS neoplasms at eight P&W study plants, were not statistically significant elevated compared with the general populations of the US and state of CT[1,2].

The third part of the P&W study is a nested case-control study of brain tumors. This exploratory study aims to investigate whether the suspected excessive brain tumors are associated with exposures obtained by the workers in our study.

1.2 LITERATURE REVIEW ABOUT CASE-CONTROL STUDY

A case-control study is defined as an investigation to compare the frequency of the exposure of interest between an outcome group and a control group to determine whether an association exists between outcome and exposure. Case-control studies are often considered to be an alternative to the cohort study. The case-control study can be distinguished from a cohort study by three key features: sampling by outcome as opposed to sampling by exposure; investigative movement from outcome to exposure as opposed to from exposure to outcome; and always being retrospective as opposed to being prospective or retrospective[3]. A case-control study is always retrospective, because it begins with an outcome and then traces back to exposures of interest. When the participants are recruited in their corresponding groups, the outcome of each participant is already known by the researcher. Because case-control studies are used to estimate the strength of the association between an exposure and the outcome, odds ratios (OR) are calculated. An OR is the ratio of the odds of an exposure in the case group to the odds of an exposure in the control group. Associated confidence intervals are also calculated for OR. If confidence interval includes 1.0, the relationship between the exposure and outcome may have been generated only due to chance and the observed relationship is not statistically significant with respect to the corresponding significance level. In the design of a case-control study, we can increase power by selecting more than one control per case. However, the number of controls should be no more than four in general, because there is not much further gain of power above four controls per case.

Case-control studies have the following advantages compared to other study designs: 1) they are comparatively inexpensive and more efficient, 2) they make it feasible for rare diseases or those with long latency between exposures and outcome, 3) they can be used to study several

potential exposures at the same time, 4) they are perhaps the only ethical way to investigate the exposure if it is known that the specific exposure is harmful to people, 5) they permit the evaluation and control of confounding factors and 6) they permit the investigation of potential interaction between two or more factors. However, case-control studies also have their disadvantages: 1) information on exposure mainly relies on memory or records, which can be hard or even impossible to confirm, 2) selection of appropriate control group is sometimes difficult, 3) they are subject to many types of potential bias, including recall bias, non-response bias and information bias, 4) researchers are unable to infer causality and determine the temporal relationship between the outcome and exposure and 5) they usually cannot give information on disease incidence rate[3,4]. The practical benefits such as the feasibility to investigate rare diseases or produce rapid result without being too expensive, may outweigh the disadvantages of case-control studies. In addition, for the rare diseases, this study design, because of its efficiency, is also ideal for preliminary investigation of a suspected risk factor for a common condition. The conclusions of the case-control study for a preliminary investigation can be further used to justify a more time-consuming and costly cohort study later.

The nested case-control study, which is alternatively called the case-control in a cohort study, is a variant study design of the classic case-control study. In a nested case-control study, cases for an outcome that occur in a predefined cohort are identified and a specified number of controls are selected for each case among those in the cohort who have remained free of the outcome by the time the outcome occurs in the case. The essence of a case-control study is that cases emerge in a source population and controls are a representative sample of the exact same source population[5]. It was noticed as early as 1959 that the challenge to make certain that both cases and controls are a representative sample of the same source population is a major weakness

inherent to the design of case-control studies[6]. However, in a nested case-control study, the cases arise from a predefined population and the controls are sampled from the exact same population.

The main difference between a nested case-control and a case-control study is that the cases and controls are sampled from a source population with unknown size in a case-control study, whereas a nested case-control study is conducted within a predefined source population with known sample size. In a nested case-control study, cases can act as both controls and cases, because a cohort member who acts as a control at one time may turn into a case later. In this study, a cohort member can also be chosen to be a control for different cases. But if cases are excluded as controls, the usual case-control analyses can still be performed by calculating OR. The only disadvantages to nested case-control studies are the reduced power and precision because of control sampling, and the possibility of flaws in the sampling design or its implementation[7]. However, any fundamental problem with nested case-control studies must also be a problem for conventional case-control studies. For many research questions, the relatively minor loss in statistical efficiency for this type of observational study is offset by impressive reductions in the number of study subjects and in the costs and efforts of data collection and analysis[8].

1.3 STATISTICAL CONSIDERATIONS

In most epidemiologic studies which need to collect information through direct contact with eligible subjects or their next of kin, there is always a certain proportion of the target population who refuses or is not able to participate. The failure to obtain information from a designated individual for any reason is often called a non-response and the proportion of such sample

individuals is called the non-response rate. Four common reasons of non-response are: not-at-homes, not found, refusals, and unable to answer[4]. Not-at-homes are people who live in the location but are away such as for travel or work when the interviewer tried to reach them. Not founds are people who are dead or have moved to other place without a forwarding address. Refusals are people who are contacted but refuse to participate in the study. Refusals may result from not being interested in the study objectives, fear of invasion of privacy or other reasons. Finally, unable to answer includes persons who are physical or mental incapable to provide response[9]. Among the non-respondents in our study, there are 63 cases who have not been contacted for an interview due to the failure to locate a contact and 118 controls who have not been contacted for an interview because the survey period was over.

Case-control studies are faced with increasing challenges in obtaining a sufficiently high response rate, because fewer and fewer people are willing to participate in research studies. The Centers for Disease Control and Prevention (CDC) considers a response rate of 70 percent to be a minimum threshold to reduce non-response bias, which is defined as a systematic difference in exposures and other factors between responding and non-responding groups[10]. There are two types of non-response bias: in the estimate of the measure of association between outcome and exposure, and in the measure of outcome or exposure[11]. Although cases in case-control studies are often more likely to respond than individuals who are randomly selected as the controls (because the cases usually tend to have more interest in the study), different rates of non-response between the case and control group does not introduce non-response bias by itself. If the exposure rates were equal between the response and non-response case and the exposure rates were also equal between the response and non-response control, then the OR calculated by only using the respondents data is not different from the OR calculated by using the whole set of

eligible subjects. Non-response bias may become an issue when response cases, controls, or both differ from non-respondent with respect to the distribution of outcome risk factors[12]. When patterns of non-response in the case group differ from those in the control group, biased risk estimates may occur. In general, higher rates of non-response increase the likelihood of non-response bias and errors in risk estimates.

In previous studies, many methods have been applied to compare the characteristics between the respondents and the non-respondents in a case-control study. One approach is to compare characteristics of common variables between the respondents and the non-respondents for the case group and the control group, respectively. Homogeneity is tested with respect to these common variables by using the Pearson chi-square test[13, 18, 19]. Another approach is using logistic regression to specify which variables are statistically significant associated with being non-respondent for the case group and the control group, respectively[10, 18, 19].

2.0 METHODS

2.1 CASES AND CONTROLS SELECTION CREITERIA AND RECURITMENT PROCEDURE

All living and deceased cases of primary malignant and benign brain tumors identified in the predefined cohort (the whole population of P&W study plants from 1952 to 2001) were included as the cases in our study. For each brain tumor case, one control was matched with event age of the case (exact age at diagnosis time), gender, and year of birth ($\pm$1 year) from everyone in the cohort who had not developed a brain tumor. All participants are defined as respondents, and all selected eligible subjects who did not participate for any reason are defined as non-respondents.

After receiving the signed informed consent, willing participants were contacted by a professional interviewer for a 15-20 minute telephone interview to collect the following information: cigarette smoking, occupational history, hobbies, and personal medical history. The information about demographics and work history: race, gender, date of birth, age at hire, age at termination, and the duration of time worked in the P&W plants for all eligible subjects was collected from the plant records. There were 63 cases and 118 controls never contacted for an interview. The reason for never being contacted for interview in the case group was the failure to contact them. For the controls, they were not contacted for interview because the survey period was over. The final study pool consisted of 722 cases and 722 controls.

2.2 DATA DESCRIPTION

2.2.1 Response rate

In the original dataset, there were 1444 observations composed of 722 cases and 722 controls. Two non-respondent subjects were selected as control twice for different cases, so these subjects are only used once in the analysis giving a total of 1442 observations which includes 722 cases and 720 controls. There were 256 variables recorded such as race, gender, age at hire date, duration of time worked, cigarette smoking, occupational history, hobbies and personal medical history. The overall number of respondents was 355 (24.6%) and the overall number of non-respondents was 1087 (75.4%). In the case group, the response rate and the non-response rate was 33.1% and 66.9%, respectively. And in the control group, the response rate and the non-response rate was 16.1% and 83.9%, respectively (table 1). There were 63 cases and 118 controls who were not contacted for an interview during the survey period. They were included in the non-respondents, so the response rates were lowered as a result in both the case and control group.

Table 1. Overall description of the data set

	Response		Non-response		Total	
	No.	%	No.	%	No.	%
Case	239	33.1	483	66.9	722	50.1
Control	116	16.1	604	83.9	720	49.9
Total	355	24.6	1087	75.4	1442	100.0

2.2.2 Non-response reasons

Table 2. Summarization for the non-respondents

	Non-response							
	Non-response to the contact		Refusal to do the interview		Never contacted for interview		Total	
	No.	%	No.	%	No.	%	No.	%
Case	290	60.0	130	26.9	63	13.1	483	44.4
Control	398	65.9	88	14.6	118	19.5	604	55.6
Total	688	63.3	218	20.1	181	16.6	1087	100.0

As shown in table 2, there were three subgroups of non-respondents: non-response to the contact, refusal to do the interview, and never contacted for an interview in both the cases and controls. Overall, non-response to the contact was the majority for both the case non-respondents (60.0%) and the control non-respondents (65.9%). The proportion of refusal to do the interview was much larger in the case group than that of the control group (26.9% versus 14.6%). In contrast, the proportion of never contacted for interview in the case group was lower than that of the control group (13.1% versus 19.5%).

The detailed reasons for refusal to do the interview are summarized in table 3 for the case group and the control group separately. Being indifferent, not interested or other reasons were the two most important refusal reasons in both groups, which were 23.9% and 16.9% versus 30.7% and 25.0% in the case group and the control group, respectively. In contrast, the reason that their loved one would not want their information shared stood for the least important reason in the both groups. There were two additional reasons for refusal in the case group: getting advice from lawyer and believing that the brain tumor was not work related.

Table 3. Reasons for being refusal of the interview in the case group and the control group

Reasons for being refusal of the interview	Case		Control	
	No.	%	No.	%
Advice from lawyer	3	2.3	0	0
Belief that the brain tumor was not work related	9	6.9	0	0
Belief that their loved one would not want their information shared	2	1.5	2	2.3
Elderly	9	6.9	8	9.1
Indifference, or not interested	31	23.9	27	30.7
Not specified	10	7.7	5	5.7
Other	22	16.9	22	25.0
Refused to give reasons	17	13.1	11	12.5
Short employment with P & W	11	8.5	11	12.5
Still grieving	16	12.3	2	2.3
Total	130	100.0	88	100.0

2.2.3 Common known variables in the respondents and non-respondents

The common known variables: race, gender, age at hire, age at termination, duration of time worked, and year of hire are summarized for all data according to case-control status.

2.2.3.1 Race

As shown in figure 1, in the case group race was recorded as white (42.1%), missing (54.4%), and other (3.5%). The missing values accounted for 47.3% (113/239) in the case response group and 58.0% (280/483) in the case non-response group. In comparison, the problem of missing racial data was even worse in the control group, where the missing race accounted for 66.9%

overall (66.4% (77/116) in the control respondents and 67.1% (405/604) in the control non-respondents). Due to the large amount of missing racial data, we were not be able to use race in the subsequent analyses.

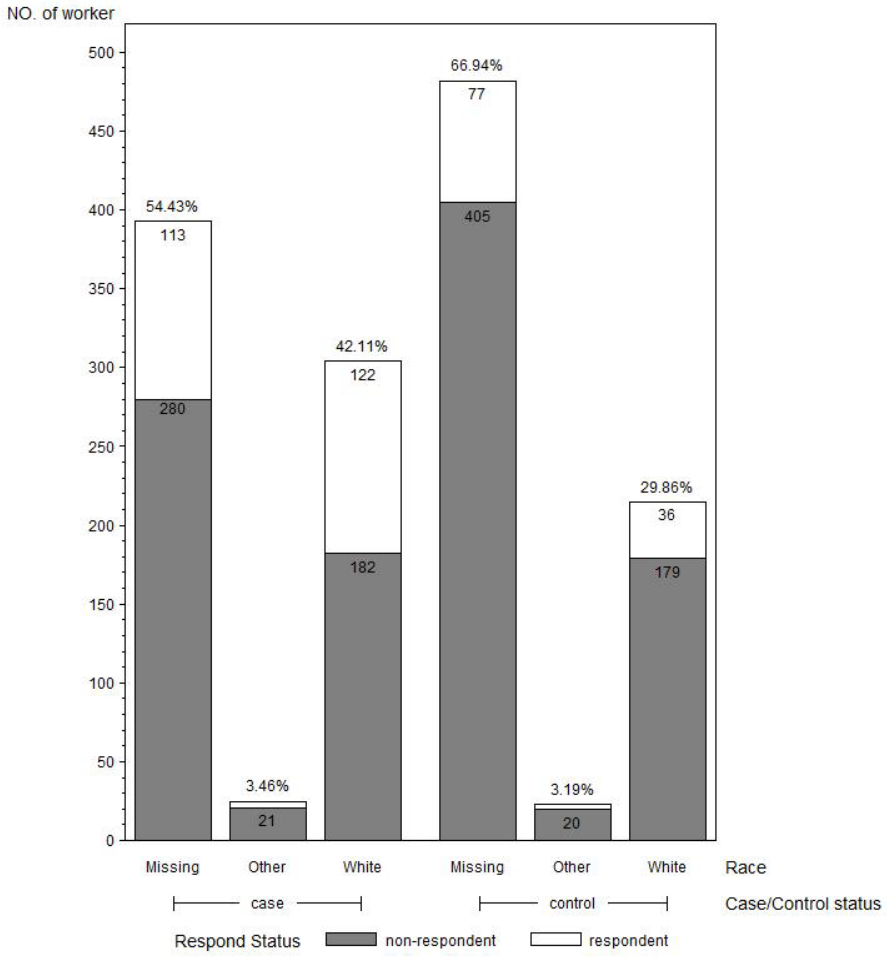

Figure 1. Race distribution

2.2.3.2 Gender

As shown in figure 2, male was the major gender in the both groups. The overall proportions of male were both around 81% in the case group and the control group. The proportions of male responding and non-responding were also similar in the case group and the control group, which

was 84.5% (202/239) versus 79.5% (384/483) and 85.3% (99/116) versus 80.5% (486/604), respectively. This indicates that the non-response rate may not be associated with gender in both groups.

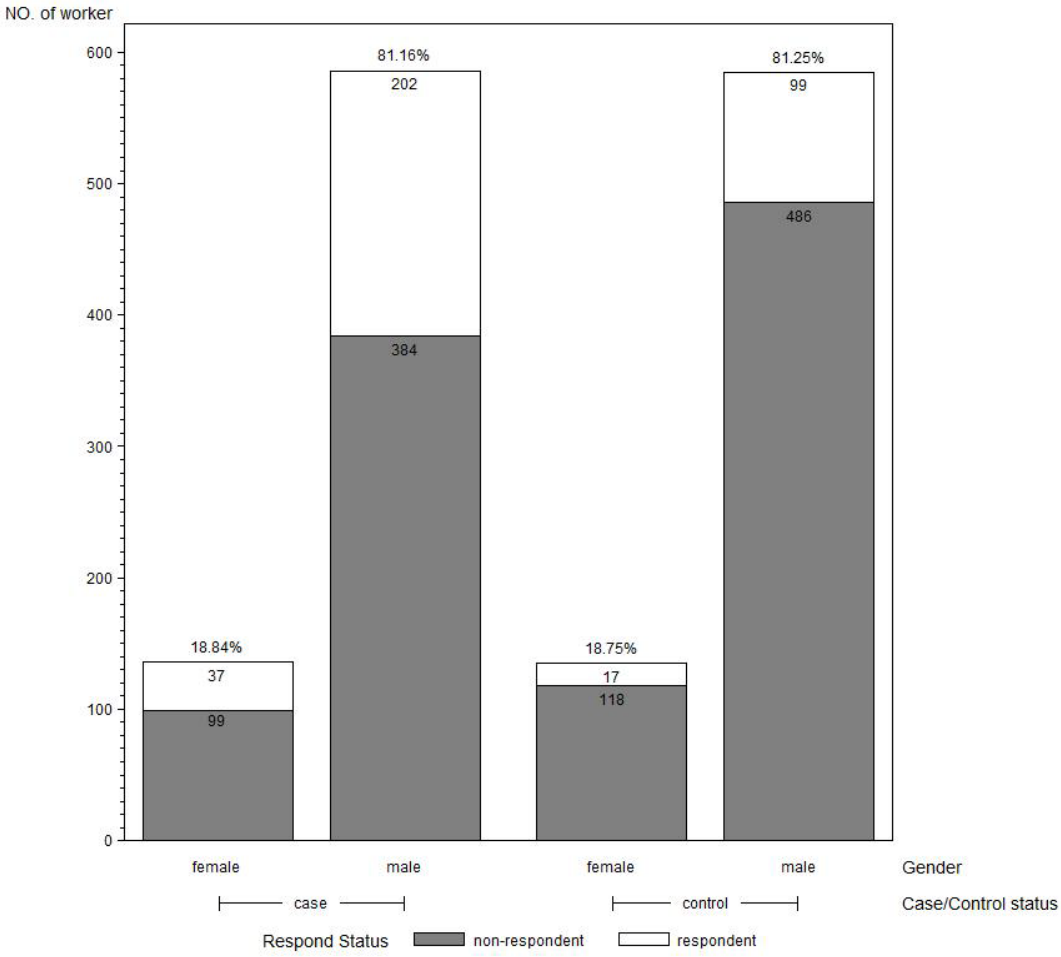

Figure 2. Gender distribution

2.2.3.3 Age at hire

As shown in table 4, the overall mean ages at hire for the case group and the control group were both around 28. In both groups, the mean ages at hire for the respondents were relatively younger than that of the non-respondents: 27.7 versus 29.1 in the case group and 26.9 versus 28.4 in the

control group. When age at hire is categorized, the distribution is shown in figure 3 according to respondent status for the case group and the control group.

Table 4. Summarization of age at hire in year

Age at hire		N	Mean	Std. Dev	Minimum	Maximum
Case	Response	239	27.7	9.3	16	58
	Non-response	483	29.1	9.3	17	60
	All	722	28.6	9.3	16	60
Control	Response	116	26.9	8.6	18	53
	Non-response	604	28.4	9.1	14	60
	All	720	28.1	9.0	14	60

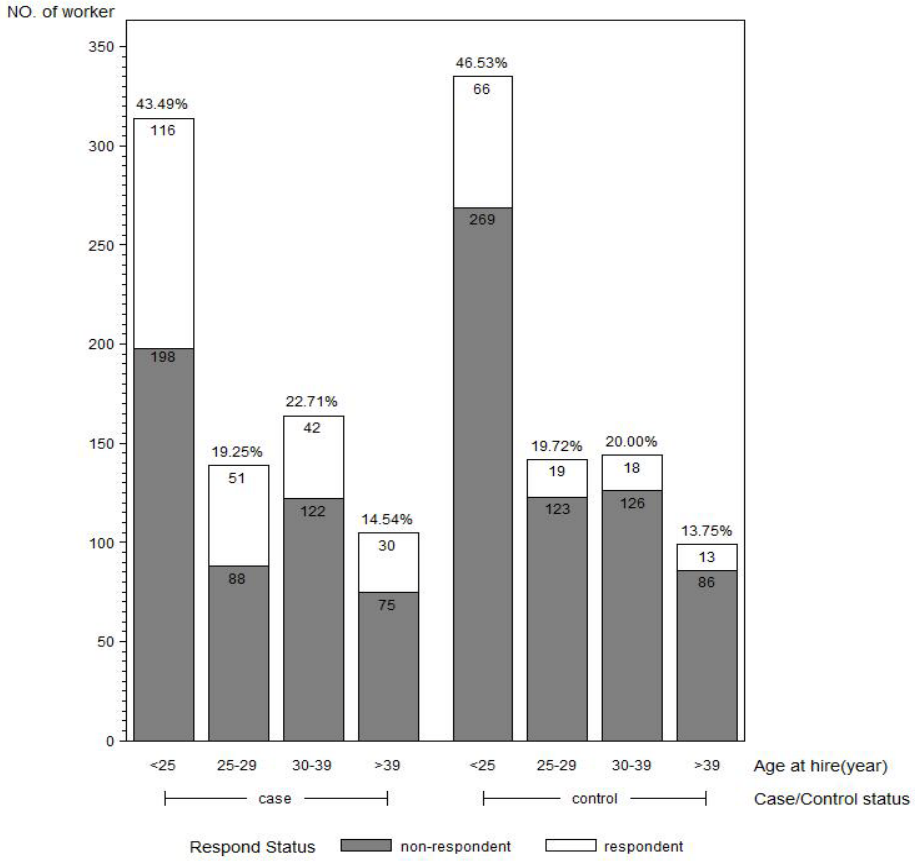

Figure 3. Age at hire distribution

As shown in figure 3, the values of age at hire were divided into four categories: less than 25 years of age, 25 to 29 years of age, 30 to 39 years of age, and older than 39 years of age. In both the case group and the control group, most of the respondents and the non-respondents were hired at less than 25 years of age. The proportions of respondents less than 25 years of age were higher than those of non-respondents both in the case group (48.5% (116/239) versus 41.0% (198/483)) and the control group (56.9% (65/116) versus 44.5% (269/604)). In contrast, the proportions of respondents at age of 30 to 39 or older than 39 years of age were lower than those of non-respondents both in the case group (17.6% (42/239) versus 25.3% (122/483) and 12.6% (30/239) versus 15.5% (75/483)) and in the control group (15.5% (18/116) versus 20.9% (126/604) and 11.2% (13/116) versus 14.2% (86/604)).

2.2.3.4 Age at termination

Table 5. Summarization of age at termination in year

Age at terminating		N	Mean	Std. Dev	Minimum	Maximum
Case	Response	239	40.8	16.3	18	69
	Non-response	483	38.2	15.2	18	69
	All	722	39.1	15.6	18	69
Control	Response	116	38.4	15.8	18	66
	Non-response	604	37.6	15.1	18	69
	All	720	37.7	15.2	18	69

As shown in table 5, the overall mean age at termination for the case group and the control group were not very different (39.1 and 37.7, respectively). In both groups, all the minimum ages at termination by the respondents and the non-respondents were 18. In both groups, the mean ages at termination for the respondents were a little older than those of the non-respondents: 40.8

versus 38.2 in the case group and 38.4 versus 37.6 in the control group. When age at termination is categorized, the distribution is shown in figure 4 according to respondent status for the case group and the control group.

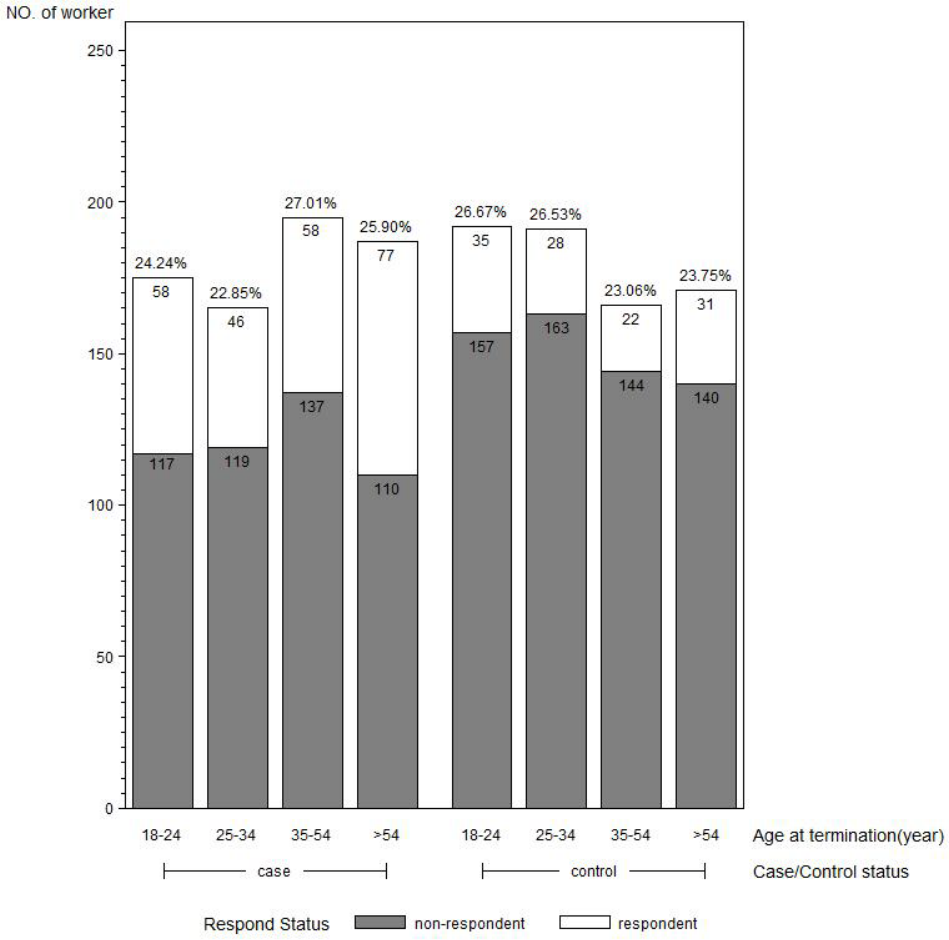

Figure 4. Age at termination distribution

As shown in figure 4, the values of age at termination were divided into four categories: 18 to 24 years of age, 25 to 34 years of age, 35 to 54 years of age, and older than 54 years of age. In the case group, the proportions of people who terminated their work at P&W at age of 18 to 24 were both around 24% for the respondents and non-respondents. However, the proportion of

respondents terminated their work when they were older than 54 was much larger than that of the non-respondents (32.2% (77/239) versus 22.8% (110/483)). In the control group, the proportions of four respondent categories were not very different than those of the non-respondent categories.

2.2.3.5 Duration of time worked

Table 6. Summarization of duration of time worked in years

Duration of time worked		N	Mean	Std. Dev	Minimum	Maximum
Case	Response	239	12.6	13.6	<1	43
	Non-response	483	8.8	11.7	<1	41
	All	722	10.0	12.5	<1	43
Control	Response	116	11.1	13.4	<1	37
	Non-response	604	8.8	12.1	<1	45
	All	720	9.1	12.3	<1	45

As shown in table 6, the overall mean durations of time worked for the case group and the control group were not very different (10.0 and 9.1 years, respectively). In both groups, the minimum durations of time worked by the respondents and the non-respondents were all less than one year. In both groups, the mean durations of time worked by the respondents were relatively longer than those of the non-respondents: 12.6 versus 8.8 years in the case group and 11.1 versus 8.8 years in the control group. When duration of time worked was categorized, the distributions in the case group and the control group were shown in figure 5 according to respondent status.

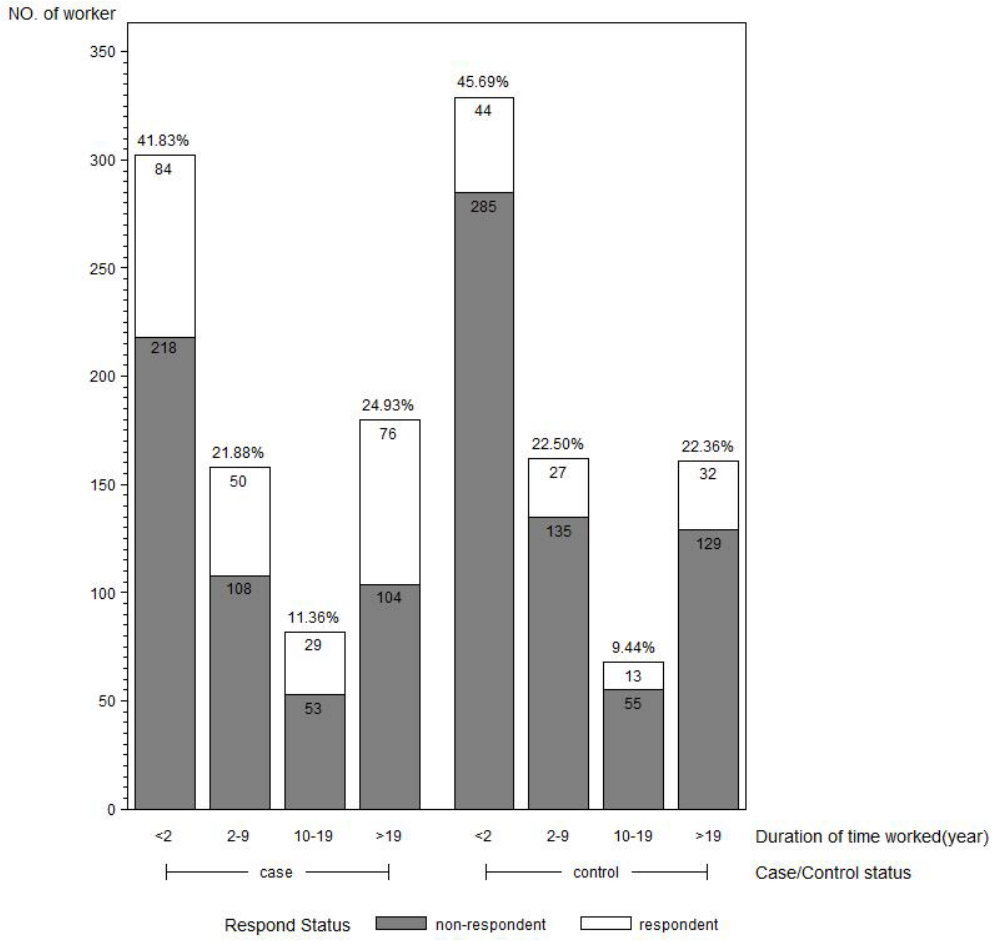

Figure 5. Duration of time worked distribution

As shown in figure 5, the values of duration of time worked were divided into four categories: shorter than two years, two to nine years, 10 to 19 years and longer than 19 years. In both the case group and the control group, the largest proportions were working less than two years, which was 41.8% in the case group and 45.7% in the control group. In the case group, the proportions of responding people who worked shorter than two years (35.2% (84/239)) or longer than 19 years (31.8% (76/239)), were higher than those of the other two categories. About 45% (218/483) of the non-respondent cases worked less than two years, and this proportion was much

higher than the proportions of the other three categories. In the control group, the proportions of less than two years duration of time worked, which were 37.9% (44/116) for the respondents and 47.2% (285/604) for the non-respondents, were obviously higher than proportions of the other three categories in the corresponding respond status.

2.2.3.6 Year of hire

Table 7. Summarization of the year of hire

Year of hire		N	Mean	Std. Dev	Minimum	Maximum
Case	Response	239	1961	9.9	1939	1990
	Non-response	483	1960	9.4	1937	2001
	All	722	1960	9.6	1937	2001
Control	Response	116	1959	8.1	1937	1980
	Non-response	604	1960	9.3	1935	1999
	All	720	1960	9.2	1935	1999

As shown in table 7, the overall mean year of hire was the year of 1960 both in the case group and the control group. The mean year of hire was very similar in the respondents and the non-respondents in the case group (1961 versus 1960) and in the control group (1959 versus 1960). In both groups, the maximum years of hire by the respondents were much earlier than those of the non-respondents, which was 1990 versus 2001 in the case group and 1980 versus 1999 in the control group. When year of hire was categorized, the distributions in the case group and the control group were shown in figure 6 according to response status.

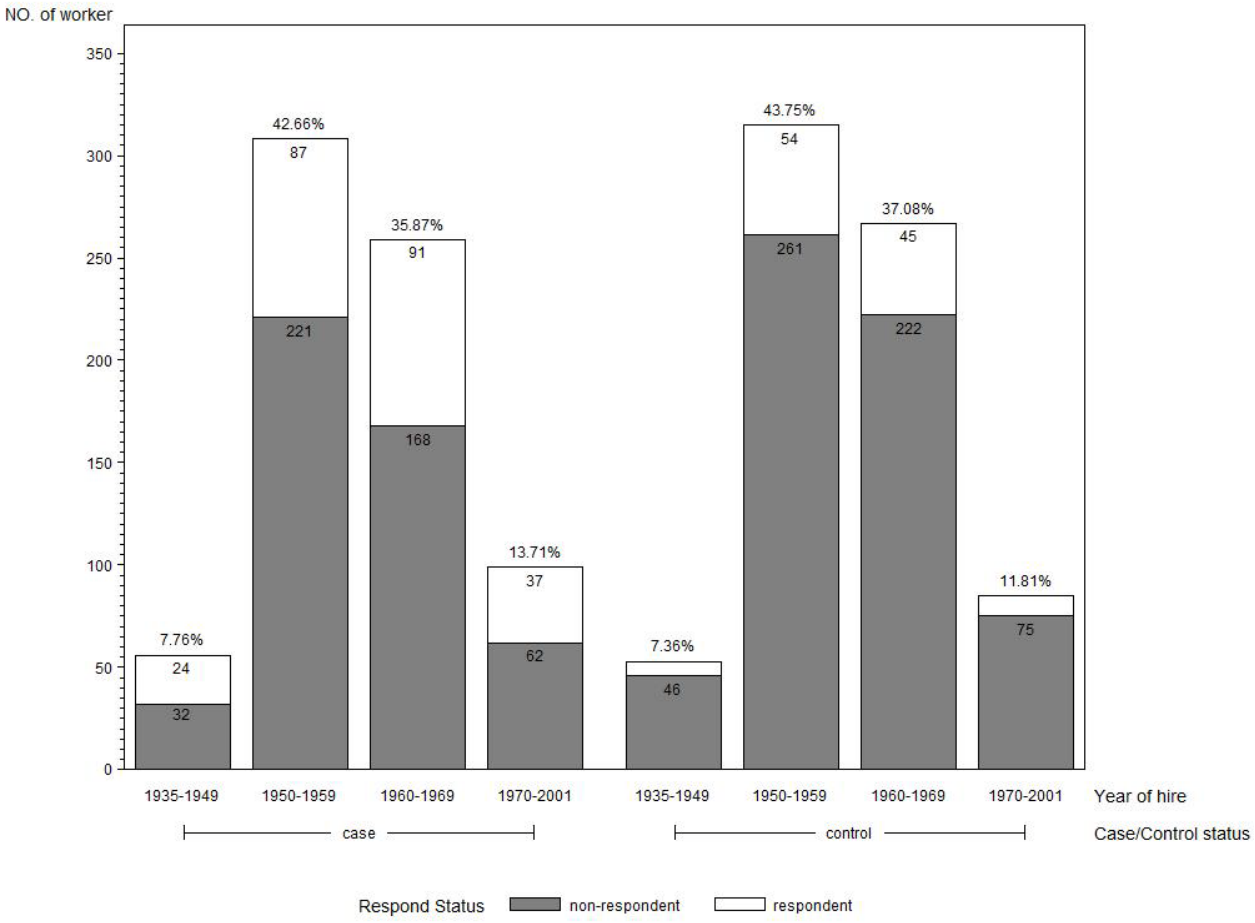

Figure 6. Year of hire distribution

As shown in figure 6, the values of year of hire were divided into four categories: from the year of 1935 to 1949, from the year of 1950 to 1959, from the year of 1960 to 1969, and from the year of 1970 to 2001. In both the case group and control group, the largest proportions were being hired between the year of 1950 to 1959, which was 42.7% in the case group and 43.8% in the control group. The least proportions were being hired between the year of 1935 to 1949, which were around 7% in both groups. The majority of the respondents and non-respondents

were hired between the year periods of 1950 to 1959 and 1960 to 1969 in both the case group and control group.

2.2.4 Survey data

From the survey data, the variables related to cigarette smoking, occupational history, hobbies, and personal medical history were summarized for the respondent cases and controls only.

2.2.4.1 Cigarette smoking

Table 8. Summarization of cigarette smoking

Respondent	Current smoker		Former smoker		Never smoker		Unknown		Total	
	No.	%	No.	%	No.	%	No.	%	No.	%
Case	25	10.5	107	44.8	91	38.1	16	6.7	239	67.3
Control	7	6.0	55	47.4	46	39.7	8	6.9	116	32.7
Total	32	9.0	162	45.6	137	38.6	24	6.8	355	100.0

As shown in table 8, the majority of the cases and controls were people who either smoked in the past (44.8% in the case group and 47.4% in the control group) or never smoked (38.1% in the case group and 39.7% in the control group). The proportion of the current smoker was larger in the case group than that of the control group (10.5% versus 6.0%). The proportions of people, whose cigarette smoking history was unknown, were very similar in the both groups, which was 6.7% in the case group and 6.9% in the control group.

2.2.4.2 Occupational history

As shown in table 9, the mean numbers of jobs before employment at P&W, they had were not very different (2.2 for the cases and 2.9 for the controls). These jobs included painter, soldier, clerk, engineer, teacher, and cook. However, the mean time worked was much longer in the controls (14.9 years) than that of the cases (5.2 years).

After the workers left P&W, the mean number of jobs outside of P&W was relatively more in the controls (3.0) than in the cases (1.9). The places they worked involved post offices, restaurants, traveler's insurance companies, power companies, and hospitals. Similarly, the mean time worked was much longer in the controls (15.4 years) than that of the cases (3.3 years).

Table 9. Summarization of occupational history

Respondents	Before employment at P&W		After employment at P&W	
	Mean No. of jobs	Mean years worked	Mean No. of jobs	Mean years worked
Case	2.2	5.2	1.9	3.3
Control	2.9	14.9	3.0	15.4
Total	2.4	8.1	2.4	6.8

2.2.4.3 Hobbies

For contact sports, 198 (82.9%) cases and 82 (70.7%) controls never took part in any contact sports such as boxing, wrestling and football. In the case group, only 38 (15.9%) people ever participated in at least one of the above sports. 25 (65.8%) of them played football and 7 (18.4%) of them took part in two or more of these contact sports. In the control group, 32 (27.6%) people participated in at least one of those sports and the majority (78.1%) of them played football, too (table 10).

As for workshops, 69.5% of the respondents in the case group never participated in any workshop and 29.3% of the case respondents did. In the control group, this participating rate (55.2%) is much lower (table 11). In the both cases and controls, most workshops involved are woodworking or house repairs. The time period they participated in these workshops varied from 1 year to 62 years in the both groups.

Table 10. Summarization of contact sports

Respondent	No		Yes		Don't know		Missing		Total	
	No.	%	No.	%	No.	%	No.	%	No.	%
Case	198	82.9	38	15.9	2	0.8	1	0.4	239	67.3
Control	82	70.7	32	27.6	0	0	2	1.7	116	32.7
Total	280	78.9	70	19.7	2	0.6	3	0.8	355	100.0

Table 11. Summarization of workshops

Respondent	No		Yes		Don't know		Missing		Total	
	No.	%	No.	%	No.	%	No.	%	No.	%
Case	166	69.5	70	29.3	1	0.4	2	0.8	239	67.3
Control	64	55.2	52	44.8	0	0	0	0	116	32.7
Total	230	64.7	122	34.4	1	0.3	2	0.6	355	100.0

2.2.4.4 Personal medical history

As shown in tables 12 and 13, excluding the year they were diagnosed with brain cancer, 86.1% of cases and 87.0% of controls never saw a doctor or went to a hospital because of head injury; and only 27 (11.3%) cases and 13 (11.2%) controls ever did so. The mean ages for those who ever saw a doctor because of head injury was 31.2 for cases and 35.6 for controls. The most common cause for head injury was because of falling for the cases and was not specified for the controls. Excluding the year they were diagnosed with brain cancer, 220 (92.1%) cases and 114

(98.3%) controls never had any operations on the brain; only 17 (7.1%) cases and 2 (1.7%) controls ever had. One of these 17 cases had a brain operation twice at the age of 29 and 39, respectively. The mean age when these 17 cases had a brain operation was 48. Around 81% (193) cases and 109 (93.9%) controls once had dental X-rays; a much larger proportion of cases (13.8%) than that of controls (5.2%) did not know if they had any dental X-rays. Except for dental X-rays, a much larger proportion of controls (70.7%) than that of cases (51.9%) never had X-rays to any part of his/her head or neck. Almost 63% of cases had CAT or CT scans to some part of his/her head or neck and 20% cases never had. However, 74.1% of controls never had this kind of CAT or CT scans and only 18.1% of controls did. Around 53% of cases had MRI/NMR/MR scans to some part of the head or neck and 26.1% of cases did not. For controls, the majority (77.6%) never had the MRI/NMR/MR scans to any part of head or neck. A much larger proportion of controls who never had radioisotope or PET scans or an angiogram to some part of the head or neck than cases who never had radioisotope/PET scans/angiogram to the head or neck (88.8% versus 58.1% and 84.5% versus 65.3%).

Table 12. Summarization of personal medical history in the case group (total=239)

Related questions	No		Yes		Don't know		Missing	
	No.	%	No.	%	No.	%	No.	%
Ever saw a doctor or went to hospital because of a head injury*	206	86.1	27	11.3	3	1.3	3	1.3
Ever had operations on the brain*	220	92.1	17	7.1	0	0	2	0.8
Ever had dental X-rays	12	5.0	193	80.8	33	13.8	1	0.4
Ever had a full mouth or Panorex X-ray exam	108	45.2	33	13.8	96	40.2	2	0.8

Table 12. continued

Ever had X-rays to head or neck[#]	124	51.9	74	31.0	37	15.5	4	1.6
Ever had CAT or CT scans to head or neck[#]	48	20.1	150	62.8	38	15.9	3	1.2
Ever had MRI/NMR/MR scans to head or neck[#]	61	25.5	127	53.1	46	19.3	5	2.1
Ever had radioisotope or PET scans to head or neck[#]	139	58.1	27	11.3	70	29.3	3	1.3
Ever had an angiogram to head or neck[#]	156	65.3	26	10.9	55	23.0	2	0.8

***excluding the year of being diagnosed with brain cancer**

#excluding the dental X-rays

Table 13. Summarization of personal medical history in the control group (total=116)

Related questions	No		Yes		Don't know		Missing	
	No.	%	No.	%	No.	%	No.	%
Ever saw a doctor or went to hospital because of a head injury*	101	87.0	13	11.2	1	0.9	1	0.9
Ever had operations on the brain*	114	98.3	2	1.7	0	0	0	0
Ever had dental X-rays	1	0.9	109	93.9	6	5.2	0	0
Ever had a full mouth or Panorex X-ray exam	58	50.0	40	34.5	18	15.5	0	0
Ever had X-rays to head or neck[#]	82	70.7	26	22.4	8	6.9	0	0
Ever had CAT or CT scans to head or neck[#]	86	74.1	21	18.1	9	7.8	0	0
Ever had MRI/NMR/MR scans to head or neck[#]	90	77.6	18	15.5	8	6.9	0	0
Ever had radioisotope or PET scans to head or neck[#]	103	88.8	5	4.3	8	6.9	0	0

Table 13. continued

Ever had an angiogram to head or neck#	98	84.5	8	6.9	10	8.6	0	0

***excluding the year of being diagnosed with brain cancer**

#excluding the dental X-rays

2.3 STATISTICAL METHODS

Even though a plan to increase the response rate was implemented, including at least sending letters and consent forms two times for non-responders and the use of selective financial incentives. The study still suffered from low response rate. The overall response rate of our data was 24.6%. The overall non-response rate was 75.4%, which can be broken down to 63.3% non-response to the contact, 20.1% refusal to do the interview, and 16.6% never contacted for the interview.

Based on previous studies, age, gender, race, educational level, marital status and soci-economic status are considered as being associated with non-response[10,14,15]. However, in our study there are only six common known variables: race, gender, age at hire, age at termination, duration of time worked, and year of hire between the respondents and the non-respondents. As for race, because the majority of the values are missing, which is 54.4% in the case group and 67.0% in the control group; it was not be used in the comparison. The following methods were conducted to compare the respondents and the non-respondents by the case-control status and to estimate non-response bias in our study.

2.3.1 Comparison of common known variables among the non-respondent subgroups according to the case-control status

The five common known variables: gender, age at hire, age at termination, year of hire, and duration of time worked in the P&W plants were compared in the non-respondent subgroups according to case-control status. Age at hire, age at termination, year of hire, and duration of time worked were treated as categorical variables by dividing values into different categories. Pearson's chi-square test or Fisher's exact test (if the minimum expected value is below 5) was used to test for homogeneity. A P value of 0.05 or less was interpreted as indication for a difference in the distribution of the characteristic among the subgroups.

2.3.2 Comparison of common known variables between the respondents and the non-respondents according to the case-control status

Because response rates were relatively low and many exposures of interest may vary by some key characteristics such as gender, age at hire, and duration of time worked, we compared the common known variables between the respondents and the non-respondents in the case group and the control group separately to evaluate the potential non-response bias. Pearson's chi-square test was used to test for homogeneity. A chi-square probability of 0.05 or less was interpreted as that the related characteristics are unequally distributed between the respondents and the non-respondents in the corresponding group. The presence of differences indicates non-response bias and that caution is necessary in making inferences.

2.3.3 Estimating adjusted odds ratio of being a non-respondent using logistic regression models

All of the common known variables (gender, age at hire, age at termination, year of hire, and duration of time worked) are suspected to be important predictors of a non-response and were therefore included in the analysis for the case group and control group, respectively. First, univariate logistical regression models were fit to identify the important univariate predictors ($p<0.25$). Second, backward selection multivariate logistic regression was used to identify the most significant variables using significance level of 0.05 for keeping a variable in the model. Third, interactions were assessed. Model diagnostics regarding adequacy, validity, and stability were checked in the final models. A logistic regression model combining the case group and control group was fit to determine whether case-control status was a strong factor for being a non-respondent.

The SAS procedure PROC LOGISTIC in SAS (9.2) was used for the logistic regression analysis to estimate odds ratios and 95% confidence intervals. The likelihood ratio test was used for significant testing for all variables. A two-sided P value of 0.05 was used as the alpha level for testing of statistical significance.

2.3.4 Evaluation of the potential for bias in the estimate of brain tumor risk associated with all common known variables

Finally, we evaluated the potential for bias in the estimate of brain tumor risk associated with all common known variables by comparing the OR based on the respondents only with that based on all subjects. Bias was calculated using the formula: Bias (%) = [(OR in respondents - OR in

the all subjects)/OR in the all subjects] $\times 100\%$[19]. Biases based on each univariate logistic regression model and final multivariate logistic regression model were calculated.

3.0 RESULTS

3.1 COMPARISON AMONG THE NON-RESPONDENT SUBGROUPS BY CASE-CONTROL STATUS

Table 14. Comparison of common known variables among the non-respondent subgroups by case-control status

variable	Case non-respondents						Control non-respondents					
	never contacted for interview		non-response to the contact		refusal to do the interview		never contacted for interview		non-response to the contact		refusal to do the interview	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Gender												
Female	17	17.2	58	58.6	24	24.2	26	22.0	77	65.3	15	12.7
Male	46	12.0	232	60.4	106	27.6	92	18.9	321	66.1	73	15.0
P value	0.37*						0.66*					
Age at hire (years)												
<25	14	7.1	136	48.7	48	24.2	55	20.5	172	63.9	42	15.6
25-29	10	11.3	51	58.0	27	30.7	28	22.8	77	62.6	18	14.6
30-39	23	18.9	63	51.6	36	29.5	19	15.1	91	72.2	16	12.7
>39	16	21.3	40	53.3	19	25.4	16	18.6	58	67.4	12	14.0
P value	<0.01*						0.71*					
Age at termination (years)												

Table 14. continued

18-24	9	7.7	82	70.1	26	25.2	37	23.6	96	61.1	24	15.3
25-34	10	8.4	67	56.3	42	35.3	29	17.8	110	67.5	24	14.7
35-54	26	19.0	74	54.0	37	27.0	30	20.8	95	66.0	19	13.2
>54	18	16.4	67	60.9	25	22.7	22	15.7	97	69.3	21	15.0
P value	0.01*						0.70*					
Duration of time worked (years)												
<2	31	14.2	126	57.8	61	28.0	58	20.3	168	65.3	41	14.4
2-9	9	8.3	68	63.0	31	28.7	26	19.3	88	65.2	21	15.5
10-19	7	13.2	29	54.7	17	32.1	10	18.2	41	74.5	4	7.3
>19	16	15.4	67	64.2	21	20.2	24	18.6	83	64.3	22	17.1
P value	0.43*						0.74*					
Year of hire												
1935-1949	6	18.8	17	53.1	9	28.1	9	19.6	29	63.0	8	17.4
1950-1959	34	15.4	114	51.6	73	33.0	38	14.6	181	69.3	42	16.1
1960-1969	18	10.7	110	65.5	40	23.8	51	23.0	138	62.2	33	14.8
1970-2001	5	8.1	49	79.0	8	12.9	20	26.7	50	66.6	5	6.7
P value	<0.01#						0.07*					

*** Pearson's chi-square test**

 # Fisher's exact test

Table 14 shows the results of Pearson's chi-square test or Fisher's exact test for differences among the case non-respondent subgroups and the control non-respondent subgroups. There were no differences detected in the control non-respondent subgroups. However, significant differences were identified in the case non-respondent subgroups with respect to the variables: age at hire ($p<0.01$), age at termination ($p=0.01$), and year of hire ($p<0.01$).

3.2 COMPARISON BETWEEN THE RESPONDENTS AND NON-RESPONDENTS BY CASE-CONTROL STATUS

Table 15. Comparison of common known variables between the respondents and the non-respondents, by case-control status

variable	Cases				Controls			
	Respondents		Non-respondents		Respondents		Non-respondents	
	(n=239)		(n=483)		(n=116)		(n=604)	
	No.	%	No.	%	No.	%	No.	%
Gender								
Male	202	34.5	384	65.5	99	16.9	486	83.1
Female	37	27.2	99	72.8	17	12.6	118	87.4
P value	0.10				0.22			
Age at hire (years)								
<25	116	36.9	198	63.1	66	19.7	269	80.3
25-29	51	36.7	88	63.3	19	13.4	123	86.6
30-39	42	25.6	122	74.4	18	12.5	126	87.5
>39	30	28.6	75	71.4	13	13.1	86	86.9
P value	0.05				0.11			
Year of hire								
1935-1949	24	42.9	32	57.1	7	13.2	46	86.8
1950-1959	87	28.2	221	71.8	54	17.1	261	82.9
1960-1969	91	35.1	168	64.9	45	16.8	222	83.2
1970-2001	37	37.4	62	62.6	10	11.8	75	88.2
P value	0.07				0.60			
Age at termination (years)								
18-24	58	33.1	117	66.9	35	18.2	157	81.8
25-34	46	27.9	119	72.1	28	14.7	163	85.3

Table 15. continued

35-54	58	29.9	136	70.1	22	13.2	144	86.8
>54	77	41.0	111	59.0	31	18.1	140	81.9
P value	0.04				0.48			
Duration of time worked (years)								
<2	84	27.8	218	72.2	44	13.4	285	86.6
2-9	50	31.6	108	68.4	27	16.7	135	83.3
10-19	29	35.4	53	64.6	13	19.1	55	80.9
>19	76	42.2	104	57.8	32	19.9	129	80.1
P value	0.01				0.26			

Table 15 shows the results of Pearson's chi-square test for different characteristic distribution between the respondents and non-respondents for the case group and the control group separately. There were no significant differences detected between the respondents and the non-respondents in the control group. However, significant distribution differences were identified between the case respondents and the case non-respondents regarding the variables: age at hire ($p=0.05$), age at termination ($p=0.04$), and duration of time worked ($p=0.01$). These differences indicate the case respondents may not represent the overall case data well.

3.3 LOGISTIC REGRESSION MODELS TO ESTIMATE ADJUSTED ODDS RATIO OF BEING A NON-RESPONDENT

Table 16. Comparison of common known variables between the respondents and the non-respondents, by case-control status using univariate logistic regression

Variable	Cases				Controls			
	No. of respondents	No. of non-respondents	OR	95% CI*	No. of respondents	No. of non-respondents	OR	95% CI
Gender								
Female	37	99	1.0		17	118	1.0	
Male	202	384	0.7	0.5-1.1	99	486	0.7	0.4-1.2
P value			0.10				0.21	
Age at hire (years)								
<25	116	198	1.0		66	269	1.0	
25-29	51	88	1.0	0.7-1.5	19	123	1.6	0.9-2.8
30-39	42	122	1.7	1.1-2.6	18	126	1.7	1.0-3.0
>39	30	75	1.5	0.9-2.4	13	86	1.6	0.9-3.1
P value			0.04				0.11	
Age at termination(years)								
18-24	58	117	1.0		35	157	1.0	
25-34	46	119	1.3	0.8-2.0	28	162	1.3	0.8-2.2
35-54	58	136	1.2	0.8-1.8	20	143	1.5	0.8-2.6
>54	77	111	0.7	0.5-1.1	33	142	1.0	0.6-1.7
P value			0.04				0.49	
Year of hire								
1935-1949	24	32	1.0		7	46	1.0	
1950-1959	87	221	1.9	1.1-3.4	54	261	0.7	0.3-1.7
1960-1969	91	168	1.4	0.8-2.5	45	222	0.8	0.3-1.8
1970-2001	37	62	1.3	0.6-2.5	10	75	1.1	0.4-3.2

Table 16. continued

P value			0.07				0.60	
Duration of time worked (years)								
<2	84	218	1.0		44	285	1.0	
2-9	50	108	0.8	0.5-1.3	27	135	0.8	0.5-1.3
10-19	29	52	0.7	0.4-1.2	13	53	0.7	0.3-1.3
>19	76	105	0.5	0.4-0.8	32	131	0.6	0.4-1.0
P value			0.01				0.26	

*** CI: confidence interval**

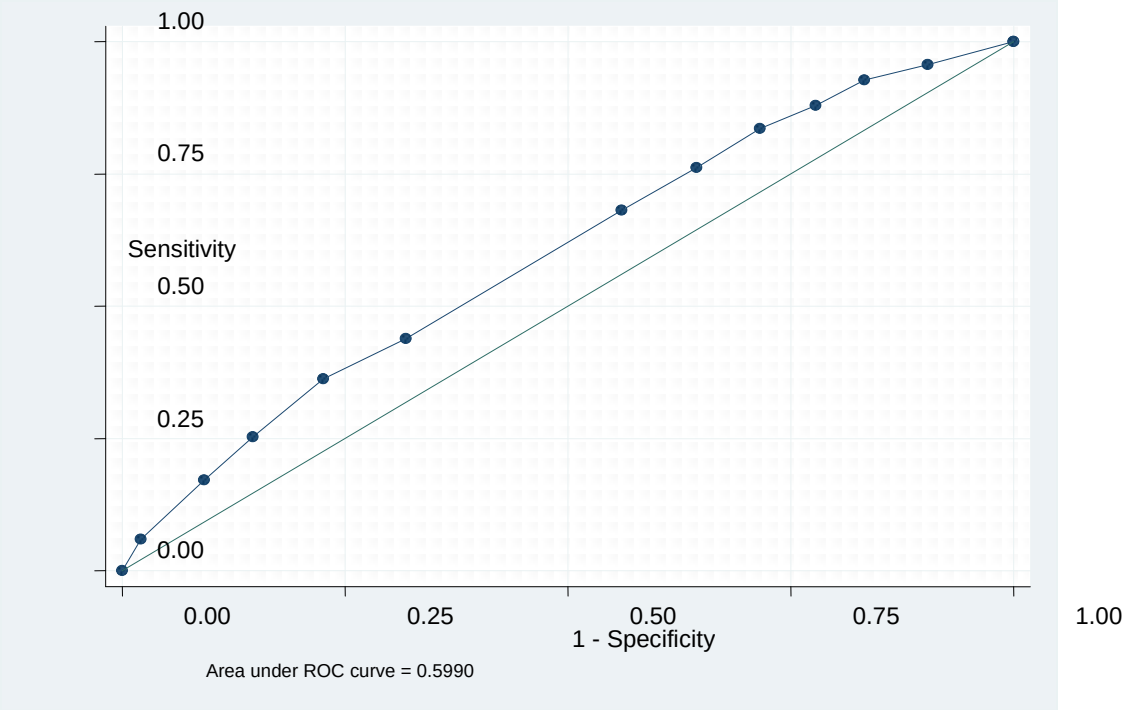

Figure 7. ROC curve for evaluating the correctly classified proportion in the case group

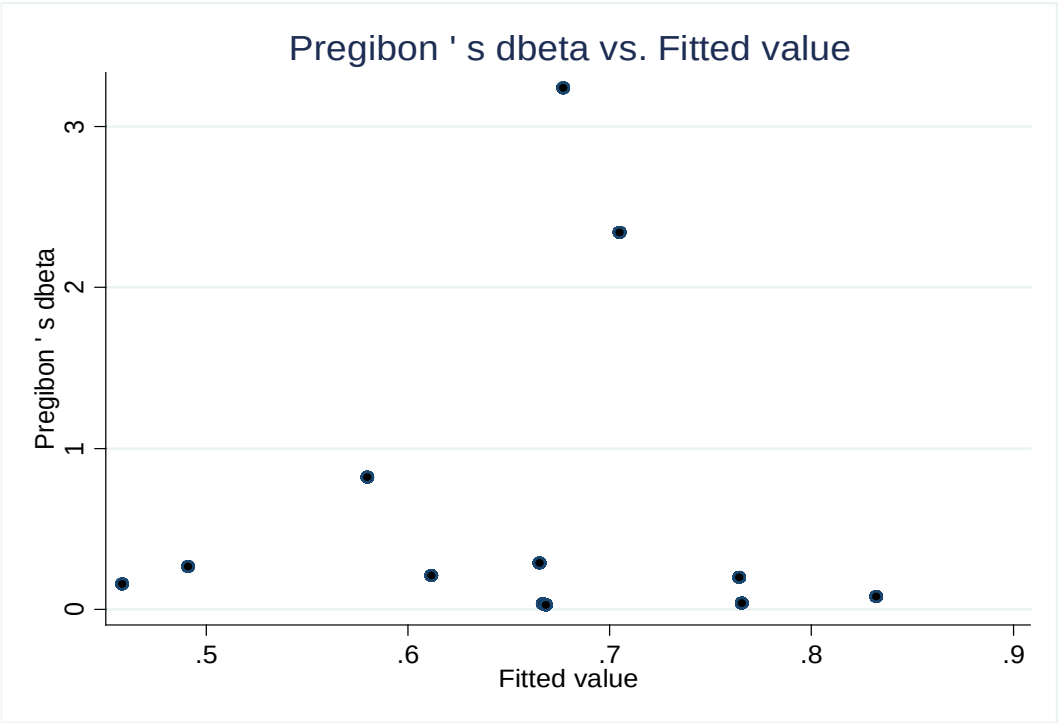

Figure 8. Pregibon's dbeta plot for evaluating influential data points in the case group

Table 17. Odds ratios of being a non-respondent estimated in multivariate logistic regression model for the case group

Variable	Cases			
	No. of respondents	No. of non-respondents	OR	95% CI
Age at hire (years)				
<25	116	198	1.0	
25-29	51	88	1.1	0.7-1.9
30-39	42	122	2.4	1.4-3.9
>39	30	75	2.3	1.3-4.2
P value			<0.01	
Age at termination(years)				
18-24	58	117	1.0	
25-34	46	119	1.0	0.6-1.8
35-54	58	136	0.7	0.4-1.2
>54	77	111	0.4	0.2-0.7
P value			<0.01	

Table 16 shows the results of comparison of common known variables between the respondents and the non-respondents using univariate logistic regression for the case group and control group separately. The variables with $p<0.25$ were kept for fitting multivariate logistic regression model using backward selection. Among controls, only variables: gender and age at hire were kept. However, neither gender nor age at hire was selected in the backward selection multivariate logistic regression model, so the probability of being a non-respondent is not significantly associated with any variable. Among cases, all variables were kept for the backward selection multivariate logistic regression model. The main effect model built for the case group included

age at hire (p<0.01) and age at termination (p<0.01) using backward selection method. The interaction between age at hire and age at termination was not significant (p=0.83). The final model to estimate the probability of being a non-respondent in the case group resulted in:

$$\log it(\hat{p}) = 0.70 + 0.13 \times age\_hirecat2 + 0.86 \times age\_hirecat3 + 0.85 \times age\_hirecat4$$
$$+ 0.04 \times age\_termcat2 - 0.38 \times age\_termcat3 - 0.87 \times age\_termcat4$$

The final model was checked for adequacy, validity, and stability. 1) Hosmer and Lemeshow goodness of fit test was not significant with x^2=1.21 p=0.99 with 7 degrees of freedom. 2) As shown in figure 7, receiver operating curve (ROC) was plotted to show the model has 59.9% (area under the curve) correctly classified proportion (The points on the curve are generated using each possible outcome of the diagnosis test as a classification cutpoint and computing the corresponding sensitivity and 1-specificity). 3) Pregibon's dbeta versus fitted value was plotted to check the model stability (figure 8). There were two points with Pregibon's dbeta greater than one. The model was refitted without these potential influence points. The values of parameters changed a little, but the significance of each parameter did not change. So, this final model is adequate, valid, and stable. As shown in table 17, among cases the odds of being a non-respondent increased significantly as the age at hire became older adjusted for the other covariate: age at termination. The odds decreased significantly with increasing age at termination adjusted for the covariate: age at hire.

Table 18. Comparison between the respondents and the non-respondents over all the dataset using univariate logistic regression

Variable	No. of respondents	No. of non-respondents	OR	95% CI
Gender				
Female	54	217	1.0	
Male	301	870	0.7	0.5-1.0

Table 18. continued

P value			0.04	
Case/control status				
case	239	483	1.0	
control	116	604	2.6	2.0-3.3
P value			<0.01	
Age at hire (years)				
<25	182	467	1.0	
25-29	70	211	1.2	0.9-1.6
30-39	60	248	1.6	1.2-2.2
>39	43	161	1.5	1.0-2.1
P value			0.02	
Age at termination(years)				
18-24	93	274	1.0	
25-34	74	282	1.3	0.9-1.8
35-54	80	281	1.2	0.8-1.7
>54	108	250	0.8	0.6-1.1
P value			0.02	
Year of hire				
1935-1949	31	8	1.0	
1950-1959	141	482	1.4	0.9-2.1
1960-1969	136	390	1.1	0.7-1.8
1970-2001	47	137	1.2	0.7-2.0
P value			0.44	
Duration of time worked (years)				
<2	128	503	1.0	
2-9	77	243	0.8	0.6-1.1
10-19	42	108	0.7	0.4-1.0
>19	108	233	0.5	0.4-0.7
P value			<0.01	

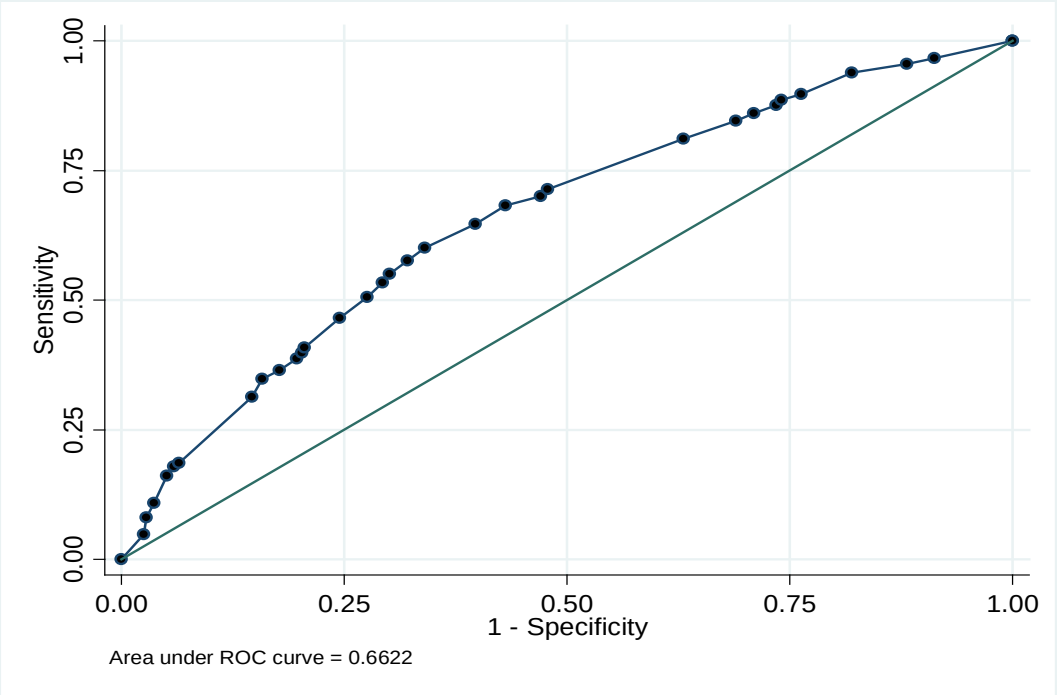

Figure 9. ROC curve for evaluating the correctly classified proportion in the whole dataset

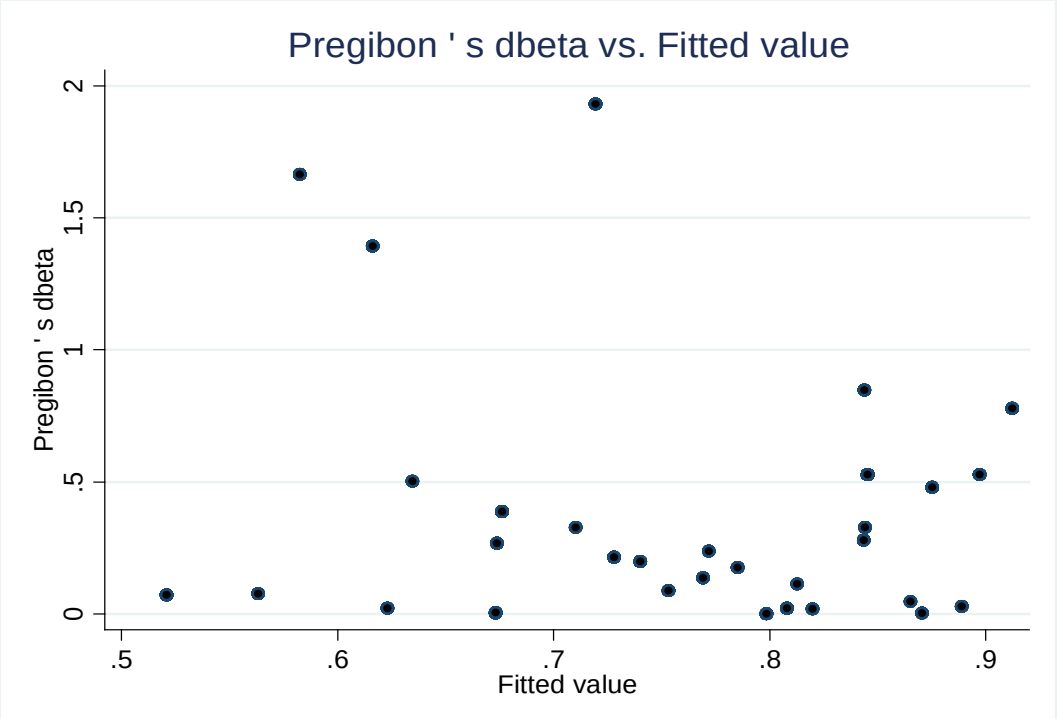

Figure 10. Pregibon's dbeta plot for evaluating influential data points in the whole dataset

Table 19. Odds ratios of being a non-respondent estimated in multivariate logistic regression model built for the overall dataset

Variable	No. of respondents	No. of non-respondents	OR	95% CI
Case/control status				
case	239	483	1.0	
control	116	604	2.6	2.0-3.4
P value			<0.01	
Age at hire (years)				
<25	182	467	1.0	
25-29	70	211	1.3	0.9-1.8
30-39	60	248	1.9	1.3-2.7
>39	43	161	1.6	1.1-2.4
P value			<0.01	
Duration of time worked (years)				
<2	128	503	1.0	
2-9	77	243	0.8	0.6-1.1
10-19	42	108	0.6	0.4-1.0
>19	108	233	0.5	0.4-0.7
P value			<0.01	

Table 18 shows the results of comparison between the respondents and the non-respondents using univariate logistic regression with case-control status also as a predictor variable. The variables with p <0.25 (case-control status, gender, age at hire, age at termination, and duration of time worked) were kept for building multivariate logistic regression model using backward

selection. The main effects model includes case-control status (p<0.01), age at hire (p<0.01), and duration of time worked (p <0.01). Interaction terms among these three variables were checked. However, none of them was kept in the final model. The final model to estimate the odds ratios of being a non-respondent in the overall dataset by using backward selection resulted in:

$$\log it(\hat{p}) = 0.74 + 0.96 \times control + 0.25 \times age\_hirecat2 + 0.64 \times age\_hirecat3 + 0.47 \times age\_hirecat4 - 0.26 \times durationcat2 - 0.48 \times durationcat3 - 0.65 \times durationcat4$$

The final model was checked for adequacy, validity, and stability. 1) Hosmer and Lemeshow goodness of fit test was not significant with x^2=5.67 p=0.68 with 8 degrees of freedom. 2) ROC was plotted to show the model has 66.2% correctly classified proportion (figure 9). 3) Pregibon's dbeta versus fitted value was plotted to check the model stability (figure 10). Because there were three points with Pregibon's dbeta grater than one, the model was refitted without these potential influence points. The values of parameters changed a little, but the significance of each parameter did not change. So, we concluded that the final model for the overall dataset is adequate, valid, and stable. As shown in table 19, the odds of being a non-respondent is 2.6 times for controls compared to cases adjusted for the covariates: age at hire and duration of time worked. The odds increased significantly as the age at hire became older adjusted for the other covariates. The odds decreased significantly with increasing duration of time worked adjusted for the other covariates.

3.4 THE NON-RESPONSE BIASES IN THE ESTIMATE OF BRAIN TUMOR RISK ASSOCIATED WITH THE COMMON KNOWN VARIABLES

Table 20. Non-response bias in the estimate of brain tumor risk associated with each variable

Variable	Respondents only				All subjects				Bias(%)
	No. of cases	No. of controls	OR	95% CI	No. of cases	No. of controls	OR	95% CI	
Gender									
Female	37	17	1.0		136	135	1.0		
Male	202	99	0.94	0.50-1.75	586	585	0.99	0.76-1.30	-5
Age at hire (years)									
<25	116	66	1.0		314	335	1.0		
25-29	51	19	1.53	0.83-2.80	139	142	1.04	0.79-1.38	47
30-39	42	18	1.33	0.71-2.49	164	144	1.22	0.93-1.59	9
>39	30	13	1.31	0.64-2.69	105	99	1.13	0.83-1.55	16
Age at termination (years)									
18-24	58	35	1.0		175	192	1.0		
25-34	46	28	0.99	0.53-1.86	165	191	0.95	0.71-1.27	4
35-54	58	22	1.59	0.83-3.03	195	166	1.29	0.96-1.72	23
>54	77	31	1.50	0.83-2.71	187	171	1.20	0.90-1.61	25
Year of hire									
1935-1949	24	7	1.0		56	53	1.0		
1950-1959	87	54	0.47	0.19-1.17	308	315	0.93	0.62-1.39	-49
1960-1969	91	45	0.59	0.24-1.47	259	267	0.92	0.61-1.39	-36
1970-2001	37	10	1.08	0.36-3.22	99	85	1.10	0.69-1.77	-2
Duration of time worked (years)									
<2	84	44	1.0		302	329	1.0		

Table 20. continued

2-9	50	27	0.97	0.54-1.76	158	162	1.06	0.81-1.39	-8
10-19	29	13	1.17	0.55-2.47	82	68	1.31	0.92-1.88	-11
>19	76	32	1.24	0.72-2.16	180	161	1.22	0.94-1.59	2

Table 21. Non-response bias in the estimate of brain tumor risk associated with one variable adjusted for the covariate

Variable	Respondents only				All subjects				Bias(%)
	No. of cases	No. of controls	OR	95% CI	No. of cases	No. of controls	OR	95% CI	
Age at hire (years)									
<25	116	66	1.0		314	335	1.0		
25-29	51	19	1.50	0.75-3.00	139	142	1.05	0.75-1.46	43
30-39	42	18	1.13	0.55-2.34	164	144	1.10	0.79-1.53	3
>39	30	13	1.02	0.45-2.31	105	99	0.96	0.66-1.41	6
Age at termination (years)									
18-24	58	35	1.0		175	192	1.0		
25-34	46	28	0.83	0.40-1.68	165	191	0.91	0.64-1.29	-9
35-54	58	22	1.40	0.64-3.04	195	166	1.25	0.86-1.82	12
>54	77	31	1.35	0.68-2.70	187	171	1.17	0.82-1.67	15

We evaluated the association between gender, age at hire, age at termination, year of hire, and duration of time worked and risk of brain tumors among respondents only and among all eligible subjects (table 20). When results were based on respondents only, the OR for male versus female was underestimated by 5% as compared with OR for gender among all subjects. The bias related to age at hire was 47%, 9%, and 16% for comparisons between the three older age at hire categories, respectively, as compared with the youngest age at hire category. The bias related to age at termination was 4%, 23%, and 25% for comparisons between the three older age at

termination categories, respectively, as compared with the youngest age at termination category. The bias related to year of hire was -49%, -36%, and -2% for comparisons between the three later year of hire categories, respectively, as compared with the earliest year of hire category. The bias related to duration of time worked was -8%, -11%, and 2% for comparisons between the three longer duration of time worked categories, respectively, as compared with the shortest duration of time worked category.

Because age at hire and age at termination were significantly associated with being a non-respondent in the case group, the association between age at hire and age at termination and risk of brain tumors among respondents only and among all subjects were also calculated (table 21). When results were based on respondents only, the OR for age at hire between 25 to 29 versus age at hire less than 25, age at hire between 30 to 39 versus age at hire less than 25, and age at hire greater than 39 versus age at hire less than 25 controlling for the covariate of age at termination, were overestimated by 43%, 3% and 6% as compared with the respective OR for age at hire among all subjects. The bias related to age at termination was -9%, 12%, and 15% for comparisons between the three older ages at termination categories, respectively, as compared with the youngest age at termination category controlling for the covariate of age at hire.

4.0 DISCUSSION

Response rates in observational epidemiologic studies, especially population-based case-control studies, have been decreasing for around 20 years despite increasing efforts to enhance it[20]. The likelihood of non-response bias increases as a result of low response rates. In our study, the response rate was 33.1% among cases and 16.1% among controls. Therefore, we were concerned there may be significant differences regarding exposure variables between respondents and non-respondents and it may not be feasible to generalize the results from the respondent data to the overall predefined cohort directly.

Even though there were significant differences detected in the case non-respondent subgroups with respect to the variables: age at hire, age at termination, and year of hire, we ignored these differences and combined the non-respondent subgroups according to case-control status. This was done because the key objective of this study was to determine whether the respondent data we did obtain was representative for the overall data. Several previous studies conducted similar comparisons between respondents and non-respondents by combining all the subgroups of non-respondents[18,19].

Among cases, the distribution of age at hire, age at termination, and year of hire differed significantly between respondents and non-respondents. The probability of being a non-respondent in the case group was significantly associated with age at hire and age at termination. Furthermore, case-control status, age at hire, and duration of time worked were all significant

predictors for being a non-respondent in the overall dataset. The odds of being a non-respondent is much higher for controls (almost three times) compared to cases adjusted for other covariates, being consistent with several previous studies.

Although the response rate was much lower in the control group than that of the case group, our results indicate there was no significant difference between the control non-respondents and control respondents. Our results also show that there was no significant predictor regarding being a non-respondent in the control group. These results demonstrate that the lower response rate does not introduce non-response bias by itself.

As for the bias estimation, biases varied from -49% to 47% in univariate logistic regression and varied from -9% to 43% in multivariate logistic regression model. In univariate logistic regression, large biases were detected in variables: age at hire, age at termination, and year of hire. In multivariate logistic regression model, large bias (43%) related to age at hire was identified at hire age of 25 to 29 compared to hire age less than 25, with the same age category of termination. However, the results show acceptable biases for some variable subgroups such as age at termination between 25 to 34, year of hire between 1970 to 2001, duration of time worked between 2 years to 9 years and longer than 19 years in univariate logistic regression models and age at hire between 30 to 39 and older than 39 in multivariate logistic regression model. These subgroup data may be used to represent the corresponding overall sub dataset. The directions of non-response biases for some variable subgroups were not consistent such as: duration of time worked in univariate logistic regression and age at termination in multivariate logistic regression model. Further investigation should be conducted in this part. Although the common known variables we obtained between respondents and non-respondents were limited, due to the large

non-response bias in the OR estimate, special caution should be put in generalizing the results of the further nested case-control study to the overall predefined cohort.

5.0 CONCLUSION

We found there were significant differences between the responding and non-responding cases regarding the variables: age at hire, age at termination, and duration of time worked. The probability of being a non-respondent in the case group was significantly associated with age at hire and age at termination. Furthermore, case-control status, age at hire, and duration of time worked were significant predictors for being a non-respondent in the whole dataset. The biases in the risk estimate for brain tumors associated with age at hire and age at termination were large, which varied from -9% to 43%. Even though there is not enough evidence to use the respondent data we obtained to represent the overall dataset, it could be reasonable to use some subgroups of variables to represent the corresponding overall sub datasets. This study supports the understanding that non-response is one of the most important potential sources of bias in population-based case-control studies where should be considered and discussed. Special caution should be used in inferring the relationship between brain tumors and exposures of interest in the overall dataset based on the respondents only.

BIBLIOGRAPHY

1. Marsh G, Buchanich J, Youk A, et al. Long-term health experience of jet engine manufacturing workers: I. Mortality from central nervous system neoplasms. J Occup Environ Med. 2008; 50:1099 – 1116.

2. Marsh G, Buchanich J, Youk A, et al. Long-term health experience of jet engine manufacturing workers: II. Total and cause-specific mortality excluding central nervous system neoplasms. J Occup Environ Med. 2008; 50:1117–1129.

3. Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research: principles and quantitative methods. Belmont: Wadsworth, 1982.

4. Glenn D. Sampling Issues: Nonresponse. University of Florida. PEOD-9. November 1992.

5. Rothman KJ, Greenland S: Modern epidemiology. Second edition. Philadelphia, Lincot-Raven Publishers; 1998.

6. Mantel N, Haenszel W: Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 1959, 22(4):719-748.

7. Wacholder, Sholom. Bias in Full Cohort and Nested Case-Control Studies? Epidemiology: May 2009 - Volume 20 - Issue 3 - pp 339-340

8. Ernster V. L. Nested Case-Control Studies. Prev Med. 1994 Sep; 23(5):587-90.

9. Glenn I. Sampling Issues: Nonresponse. PEOD9. 2009.

10. Fatma S, Harry H. North Carolina PRAMS non-response analysis, 2003-2004. SCHS study NO.154-october 2007.

11. Austin M, Criqui MH, Barret-Connor E, Holdbrook MJ. The effect of response bias on the odds ratio. Am J Epidemiol 1981; 114:137–43.

12. Rothman KJ, Greenland S. Precision and validity in epidemiologic studies. In: Rothman KJ, Greenland S, eds. Modern epidemiology. 2nd ed. Philadelphia, PA: LippincottWilliams & Wilkins, 1998:115–34.

13. Madigan MP, Troisi R, Potischman N, et al. Characteristics of respondents and non-respondents from a case-control study of breast cancer in younger women. Int J Epidemiol 2000; 29:793–8.

14. Boshuizen HC, Viet AL, Picavet HS, Botterweck A, van Loon AJ. Nonresponse in a survey of cardiovascular risk factors in the Dutch population: determinants and resulting biases. Public Health 2006; 120(4): 297–308.

15. Gundgaard J, Ekholm O, Hansen EH. The effect of non-response on estimates of health care utilisation: linking health surveys and registers. Eur J Public Health 2008; 18(2): 189–194.

16. Pope D, Croft P. Surveys using general practice registers: who are the non-responders? J Public Health Med 1996; 18:6–12.

17. Etter JF, Perneger TV. Analysis of non-response bias in a mailed health survey. J Clin Epidemiol 1997; 50:1123–8.

18. Min S, Wendy C, Lan H, et al. Census and Geographic Differences between Respondents and Nonrespondents in a Case-Control Study of Non-Hodgkin Lymphoma. Am J of Epidemiol. 2008; 167:350-361.

19. Lorenzo R, Paolo B, Franco M. Analysis of nonresponse bias in a population-based case–control study on lung cancer. Journal of clinical epidemiology 55(2002) 1033-1040.

20. Morton LM, Cahill J, Hartge P. Reporting participation in epidemiologic studies: a survey of practice. Am J Epidemiol 2006; 163:197–203.