From: Bruce Jarnot <jarnotb@api.org>
Sent: Thursday, January 10, 2008 5:31 PM (GMT)
To: BenzConsort-TC@listserve.api.org; BenzConsort-OC@listserve.api.org
Subject: BHRC-TC, -OC... Case Control Progress Report: Jul-Dec2007
Attach: CC Progress Report Jan-2008.pdf

BHRC Technical and Oversight Committee members –

Dr. Wong submitted a semi-annual Progress Report for the Case Control Study earlier today, covering the 6-month period July – December 2007. A pdf copy of this nine-page report (CC Progress Report Jan-2008.pdf) is attached for your review.

Best Regards – Bruce.

From: ottowong@aol.com [mailto:ottowong@aol.com]
Sent: Thursday, January 10, 2008 3:30 AM
To: Bruce Jarnot
Subject: Progress report

Dear Bruce:

Attached is the progress report for the case-control study dated 10 January 2008. If you have any questions, please let me know.

Looking forward to seeing you in our beautiful City by the Bay.

Regards,

Otto

****

Otto Wong, Sc.D., F.A.C.E.
Fellow, American College of Epidemiology
Chief Epidemiologist, Applied Health Sciences, San Mateo, California, USA
Adjunct Professor of Epidemiology, University of North Texas, Fort Worth, Texas, USA
Adjunct Professor of Community Medicine, Chinese University of Hong Kong
Honorary Professor of Community Medicine, University of Hong Kong
Visiting Professor of Occupational Epidemiology, Fudan University, Shanghai
Associate Editor, Annals of Epidemiology
Editorial Board, Regulatory Toxicology and Pharmacology

Tel: 1 (650) 347-7898
Email: OttoWong@aol.com

Case-control Study of
Acute Myeloid Leukemia and Lymphoid Neoplasm
Shanghai Health Study

Progress Report
Activities through December 2007

Prepared by:
Otto Wong, Sc.D.
Applied Health Sciences, Inc.
Post Office Box 2078
San Mateo, CA 94401
USA

Tel: 1 (650) 347-7898 USA
ottowong@aol.com
ottowong@fudan.edu.cn

With contributions from:
Fu Hua, MB, Ph.D.
and
Wang Yiying, MB
Fudan University School of Public Health
Shanghai, China

10 January 2008

Preliminary Dataset

According to the previous timeline, a complete preliminary dataset for the case-control study was to be provided to us in September 2007. The purpose of the preliminary dataset is for us to gain familiarities of the data, which will be helpful in the development of analysis plans. We received a preliminary dataset on 31 October 2007 from Gail Jorgensen, who is responsible for extracting the case-control study data from the underlying JCML database at the University of Colorado. The preliminary dataset that we received consisted of 3 separate data files:

- GENERAL: contains basic demographic information, exposure status, diagnosis, etc.
- ORIGQUEST: contains data from original questionnaires, and
- TRANSQUEST: contains "translated" data from questionnaires in SAS usable format.

These files contain information of 3516 patients (AML and NHL patients and their controls). The information in the preliminary dataset, however, is incomplete at this point. In particular, work history and exposure assessment information is not included in any of the files. Year of birth is not available in the dataset. Answers to some questions of lifestyle are incomplete. In addition, fields with Chinese characters are not recognized.

The files contain some obvious errors, such as out-of-range values. These out-of-range errors can be identified easily and corrected. Some examples of errors found are listed below:

- Age (year): 13, 17, 939, 979 (minimum age = 18 per study protocol)
- Height (cm): 1.72, 56, 1969
- Weight (kg): 4, 570
- Age started smoking (year): 1, 2

In spite of the incompleteness and the obvious errors, the preliminary dataset did provide us with the opportunity to get familiar with the structure of the data and the frequencies and ranges of some variables. It will also allow us to estimate the magnitude of missing values of some variables. The information will be helpful to us in developing analytical plans.

The data files are in SAS usable format, which is extremely convenient and helpful, since we will use SAS statistical programs for data management and data analysis. The files in the preliminary dataset are well organized and we did

not have any problems in merging the files through the use of the unique personal study ID of each patient.

Although based on preliminary data, some interesting information of the 3516 patients can be derived from the preliminary dataset:

- 676 AML patients, 496 NHL patients, and 2344 controls
- 58% males and 42% females
- total number of hospitals that contributed patients = 30
- number of hospitals that contributed more than 200 patients each = 7
- number of hospitals that contributed less than 10 patients each = 7
- number of diagnostic subtypes with more than 100 patients = 4
- 88% interviews at bedside, 12% in hospital interview rooms
- interviewees: self = 83%, spouse or child = 2%, others = 15%
- 11% patients with a history of blood transfusion
- 36% patients with a history of cigarette smoking
- 18% patients with a history of alcohol consumption
- 35% patients with a history of using hair dye
- 20% patients with a history of living within 100 meters of power lines
- 43% patients with a history of living on a farm
- 23% patients with a history of raising crops
- 9% patients with a history of raising livestock
- average age = 52.2 years
- average height = 166.5 cm
- average weight = 63.2 kg

Although there are 33 hospitals participating in the Shanghai Health Study, according to the preliminary dataset, only 30 hospitals actually contributed patients to the case-control study. Several hospitals contributed only a few patients over the last 5 years.

The preliminary dataset has identified the variables or groups of patients with adequate frequency for statistical analysis; such as certain diagnostic subtypes, cigarette smoking, use of hair dye, living on a farm, body mass index, raising crops, living within 100 meters of power lines, etc. On the other hand, the dataset has also tentatively identified some variables that, most likely, will not be adequate for statistical analysis; such as furniture making, wood working, motor engine repair (with less than 5 patients who reported "yes" to the corresponding question).

We had several discussions with Gail Jorgensen regarding the preliminary dataset. She was aware of the incompleteness of the data. She is now preparing an updated and more complete version of the data for us. In our discussions, Gail Jorgensen has clarified some of our questions regarding the data. However,

there are areas in the data that she is not familiar with. We also discussed with Gail Jorgensen the procedure of correcting errors identified. We need to identify or create a mechanism for data clarification, correction and documentation.

Completeness of Case Ascertainment

As noted in previous progress reports, case ascertainment of AML and NHL (lymphoid neoplasms) patients might not be complete at some participating hospitals (NHL patients in particular). This suspicion was based on the observation that some hospitals had not referred any patients to JCML for months and no interviews were conducted at these hospitals, although the lack of eligible patients might very well be the reason for some small hospitals. Furthermore, some degree of incomplete case ascertainment is almost always present in most epidemiologic studies. Nevertheless, the magnitude and sources of the incompleteness should be investigated.

A conference call was held in August 2007 (Drs. Jerry Rice, Harvey Checkoway, Howard Rockette, Bruce Jarnot and myself). Approaches to the investigation of completeness of case ascertainment in the case-control study using independent data sources were discussed. Dr. Checkoway suggested that we used the data maintained at the Shanghai Tumor Registry to cross-check the number of cases in our case-control study. Dr. Rockette suggested that samples of patients in the Shanghai Tumor Registry could be used to match against those in our study. The use of patient data maintained at participating hospitals was also discussed.

In my subsequent trips to Shanghai, I worked with Professor Fu's group to explore the potential of these independent data sources. Professor Fu, Wang Yiying (a graduate assistant) and I met with Dr. Zheng Yin (郑莹) and Dr. Wu Chunxiao (吴春晓) at the Shanghai Municipal Center for Disease Control and Prevention (CDCP). Dr. Zheng gave us some background information of the Shanghai Tumor Registry and the organizational structure. The Shanghai Tumor Registry (STR) has been in operation for a long time. Before 2002, it was part of the Shanghai Tumor Institute. In 2002, it became part of the Shanghai Municipal Center for Disease Control and Prevention. It is now in the Department of Cancer Control and Prevention within CDCP. Dr. Wu works in the Department, who is responsible for generating statistics and patient listings. Dr. Zheng was a graduate student under Professor Fu Hua at Fudan University. She is very familiar with epidemiologic research. She spent 6 months at the California Department of Health a couple of years ago. She told us that the Department collaborated with international groups, including the University of Washington and Vanderbilt University in the United States.

The STR is a population-based registry – all newly diagnosed cases of malignant neoplasms and benign tumors of the CNS among Shanghai residents are reportable to the registry by law. All Class 2 or 3 hospitals in Municipal Shanghai participate in the surveillance program. (Hospitals in China are classified into three categories – Class 3 being the highest in medical care delivery and Class 1 hospitals are usually small community hospitals.) All 33 participating hospitals in the SHS are Class 2 or 3 categories.

Dr. Zheng pointed out a number of problems with the registry, which may complicate a comparison between the registry data and SHS data. First, some of the hospitals under-report cases, especially some of the Class 3 hospitals. Under-reporting by small hospitals in the rural areas of Shanghai is also common. She was not able to quantify the magnitude of under-reporting. As an indication, however, she stated that about 91% of cases in the STR database were reported by hospitals and the remaining 9% were ascertained through death certificates (i.e., not reported by hospitals). Second, only Shanghai residents are included in the reporting system. For hospitals with a good reputation, non-resident patients can make up a significant portion of the patient population of the hospital, which can be as high as 50 or 60% in some hospitals. SHS includes all patients admitted to the participating hospitals, regardless of their legal residence status. According to the Shanghai Municipal Statistics Bureau, the total population of Shanghai in 2005 was 19.2 million. However, only 13.4 millions (69.8%) have legal permanent Shanghai residence status (Shanghai *"hukou"* 上海户口). Nonresident patients at Shanghai hospitals include not only the remaining 5.8 millions (30.2%) who live in Shanghai without Shanghai *hukou* but also residents of nearby areas in adjacent provinces, such as Jiangsu (江苏), Zhejiang (浙江) and Anhui (安徽). Third, the STR includes only incidence (newly diagnosed) cases, whereas SHS includes both incidence and prevalence cases. Fourth, according to Dr. Zheng, national IDs are recorded for approximately 80% patients in the registry. Since Chinese names are not very specific (due to common family names and popular given names) and, therefore, not good identifiers, the lack of national IDs will make a cross-check difficult. A possible approach is to use both name and birth date.

Dr. Zheng stated that she would be glad to generate the number of patients by diagnostic category by hospital by time period for us. However, because of patient confidentiality, if we need any personal information of individual patients, we will need to file an application.

We subsequently obtained some overall statistics of cancer patients registered at the STR. However, the statistics were not useful in assessing the completeness of case ascertainment in the case-control study due to the following reasons:

- The case-control study is hospital-based, whereas the STR is population-based.
- Only legal Shanghai residents are included in the STR, whereas all patients from the participating hospitals are potentially eligible for the case-control study regardless of residence status.
- There is a difference between the time of diagnosis at the hospital (maintained at STR) and that of the JCML diagnosis.
- There can be a difference between the original diagnosis at the hospital and the final diagnosis at JCML.

It appeared that the only viable approach using the STR data is to generate samples of patients registered at the STR and compare them to the patients in the case-control study. This approach involves an official application to the CDCP and will likely be time-consuming.

We have also selected three participating hospitals (Tumor Hospital 肿瘤医院, Huadong Hospital 华东医院 and Renji Hospital 仁济医院) to explore the use of hospital patient records as an independent data source. We have had some initial discussions with hospital personnel. Based on some very incomplete data, we identified some potentially eligible hospital patients who are not in our study. However, we were not able to determine the reason for their absence or non-participation in the study. Again, the only viable approach appeared to be comparing samples of hospital patients to patients in the case-control study. This approach requires an approval from hospital administration and will likely be time-consuming.

The completeness of case ascertainment (or completeness of patient referral by participating hospitals) is basically a data quality issue. Logically, the investigation should be part of the overall QA/QC effort. A meeting between the PIs and selected SRP members to discuss this issue further (during the upcoming annual meeting in San Francisco) is recommended.

Exposure Assessment

Under the direction of Dr. Thomas Armstrong, the exposure assessment (EA) team continues to develop exposure data/estimates using the "sector analysis" approach. Two Scientific Review Panel members, Dr. John Cherrie and Dr. Robert Herrick, are working closely with Dr. Armstrong and the EA team. We have several discussions with Dr. Armstrong regarding the status of exposure assessment in the case-control study. An EA workshop has been scheduled to be held in Shanghai from 20-22 February, 2008. Both US and Chinese EA team members will be present at the workshop. I will attend the EA workshop in Shanghai prior to the annual meeting in San Francisco.

An English manuscript of benzene exposure in the paint industry based on the Chinese medical literature entitled "Benzene exposure potential in industries using or manufacturing paint in China: a literature review, 1956-2005" has been completed and submitted to an international journal for publication. The manuscript is currently under journal peer review.

QA/QC

The team at Fudan University in Shanghai continues to carry out routine QA/QC of questionnaire interviews and data entry. Pan Miaoting, who replaced Pei Xiaodan in February 2007, is now the person responsible for QA/QC. I continue to receive monthly QA/QC reports from Pan Miaoting. These monthly reports include accuracy checks of the following items: primary and secondary questionnaires, data entry and initial exposure assessment. According to these monthly reports, QA/QC of questionnaires taken in the second half of 2007 indicated that approximately 96%-97% of primary questionnaires sampled for review were "qualified" (i.e., no problems were found). The data entry errors during the same time period ranged from 1% to 2%.

During my trips to Shanghai in August and December 2007, I conducted *ad hoc* QA checks of samples of questionnaires taken in the recent months. Only a few minor mistakes were found in these questionnaires (with literally thousands of items/entries) per QA check. These results reinforce/validate the routine QA/QC checks by Pan Miaoting.

In addition to the QA/QC activities described above, I also participated in the monthly QA/QC conference calls, which provided updated information of the progress of QA/QC work contracted to PharmOlam.

Epidemiology Review of Risk Factors of AML and NHL

We continue identifying and reviewing epidemiologic studies of AML and NHL in relation to occupational and environmental risk factors. We are particularly interested in studies that provide information on subtypes of AML and NHL. We have collected the abstracts of more than 300 potentially relevant papers. One interesting risk factor reported in recent publications is the body mass index (BMI). For example, Ross et al. (*Cancer Epidemiol Biomarkers Prev* 2004;13:1810-1813) studied the relationship between anthropometric parameters and the risk of overall leukemia, AML and CLL in more than 41,000 Iowa women. The most important parameters were weight and BMI. The results of the Ross et al. study are summarized below:

Risk of AML in relation to weight and body mass index (BMI) in 41,368 Iowa women (derived from Ross et al., 2004)

Anthropometric parameter	Number of cases	Relative Risk (95% confidence interval)	p-value for upward trend
Weight (pounds)			
<137	13	1.0 (reference)	
138-160	26	1.8 (0.9-3.6)	0.01
>160	33	2.3 (1.2-4.4)	
BMI			
18.5-24.9	16	1.0 (reference)	
25.0-29.9	30	1.9 (1.0-3.4)	0.006
>30.0	26	2.4 (1.3-4.5)	

The finding of Ross et al. (2004) is supported by other similar studies. In a large-scale cohort study of 4.5 million US veterans followed for 27 years, Samanic et al. (*Cancer Causes & Control* 2004;15:35-43) reported AML relative risks (RRs) of 1.59 (95% CI: 1.33-1.90) for obese white men and 2.64 (95% CI: 1.80-3.85) for obese black men. From Canada, Kasim et al. (*Cancer Causes & Control* 2005;16:489-500) reported a case-control study of adult leukemia and lifestyle factors. For BMI>30, the AML RR was 1.6 (95% CI: 1.2-2.2). Furthermore, a significant upward trend of AML in relation to BMI was found (p=0.005). We are now in the process of conducting a literature search and review of Chinese medical articles on the same topic.

The preliminary dataset indicated that information on both height and weight was available for almost all study subjects. Therefore, BMI can be calculated. The table below is based on the preliminary dataset.

BMI	Female	Male	Total
Average	22.6	22.9	22.8
Range	13.7-57.5	13.7-50.7	13.7-57.5
<18 (underweight)	6.9%	5.9%	6.3%
18-25 (desirable)	72.3%	70.0%	71.0%
25-30 (overweight)	17.5%	21.5%	19.8%
>30 (obese)	3.4%	2.6%	2.9%

The average BMI of the patients in the preliminary dataset was 22.8, with a wide range of (13.7-57.5). According to the preliminary dataset, the "overweight" and "obese" categories accounted for 22.7%. These categories are based on the classification that is commonly used in western countries. For Asian populations, the categories are usually based on lower BMI. Below are two

commonly used classifications in China and the corresponding distributions of patients in the preliminary dataset:

Classification	Underweight	Desirable	Overweight	Obese
2002 CNNS*	BMI <18.5	BMI 18.5-24.0	BMI 24.0-28.0	BMI >28.0
% patients	9.2%	58.3%	25.5%	7.0%
Other surveys	BMI <18.5	BMI 18.5-22.9	BMI 23.0-26.9	BMI >27.0
% patients	9.2%	46.5%	33.4%	10.9%

* 2002 Chinese National Nutrition Survey

Using the Chinese classifications, there are more patients in the "overweight" and "obese" categories in our study, which will make the analysis statistically more meaningful. We will do some additional research to determine which system is the most widely accepted in public health research in China. We will work closely with Professor Fu on this issue.

Project Management and Timeline

Two trips to Shanghai were made in the second half of 2007. Professor Fu and his staff remain as my primary contact and major support in Shanghai for the case-control study. Unfortunately, the contract for providing support work in the case-control study between Professor Fu's group at the Fudan University School of Public Health and API has not yet been renewed. We are grateful to Professor Fu's group, who continue to provide support to us without a contract.

We will make every attempt to maintain the timeline of the case-control study specified in our last progress report (also in our contract amendment). The projected timeline depends heavily on the timing when complete demographic, diagnostic, exposure and lifestyle data (both preliminary dataset and final complete dataset) are provided to us. As discussed above in this progress report, there was a delay of one month in receiving the preliminary dataset. In addition to the delay, the preliminary dataset was incomplete. The lack of work history and exposure assessment information in the preliminary dataset will delay our effort in analysis planning. Any further delay in receiving additional updated preliminary datasets, the final dataset or any unanticipated additional preparation or processing of the final dataset will have a significant impact on the projected timeline.