Progress Report
AML and NHL Case-control Study
Activities through August 8, 2006

Prepared by:
Otto Wong, Sc.D.
Applied Health Sciences, Inc.
181 Second Avenue, Suite 628
San Mateo, CA 94401
USA

Tel: 1 (650) 347-7898 USA
Fax: 1 (650) 344-6887 USA
ottowong@aol.com
ottowong@shmu.edu.cn

Revised: August 9, 2006

♦♦♦♦♦♦♦♦

Patient Enrollment

- The number of participating hospitals has increased to 33 (from 29 in 2004). The majority of participating hospitals are municipal hospitals (17), followed closely by district hospitals (13). Of the three remaining hospitals, two are occupational disease hospitals and one is a private hospital.

- As of June 2006, approximately 8300 patients have been interviewed. These patients included both cases and controls in the case-control study of AML and NHL as well as subjects in the disease progression study. Among the cumulative number of 8300 patients, approximately 5700 questionnaires were entered and verified.

- In terms of the case-control study, the case enrollment rate for NHL patients during the last 6 months appeared to be quite encouraging. According to the numbers provided by Dr. Richard Irons, the cumulative numbers of AML and NHL patients (ICD 9) in the study as of May 1, 2006 are as follows:

    - 353 acute myeloid leukemia patients
    - 336 "NHL" patients

- According to Dr. Richard Irons' latest projection, we will likely end up with the following number of cases by the end of the year (end of data collection according to the current study plan):

    - 424 acute myeloid leukemia cases
    - 520 lymphoid cancer (WHO definition) cases
    - 433 "NHL" cases (part of lymphoid cancer above)

- In general, compared to previous years, case identification by the Clinical Coordinators at participating hospitals during the last six months has been relatively timely and consistent. This is particularly true for AML patients. Previously, by comparison, the enrollment of NHL patients fluctuated from time to time. Most of the NHL patients come from the Shanghai Tumor Hospital. In 2005 the team at Fudan made periodic reminders to the Clinical Coordinators at the Shanghai Tumor Hospital for timely submission of NHL patients. These periodic reminders seemed to have worked well. Currently, the enrollment of NHL patients seems to be both timely and consistent.

Power Calculation

- According to the latest projection, assuming we will end case accrual by the end of 2006, we will likely have approximately 420 cases of AML or NHL, slightly below the original target of 500 cases. The following table shows the smallest risk ratios that the overall analysis of the study (420 cases) will be able to detect. As can be seen, the smallest detectable risk ratios depend on the exposure rate in the controls. Currently, I do not have a good estimate for that, but it should be around 2% or 3%.

    Control-to-case ratio = 2:1
    Alpha = 0.05
    Power = 0.80
    One-sided test

Exposure rate in controls	Minimum detectable risk ratios for 420 cases
0.05	1.82
0.04	1.93
0.03	2.10
0.02	2.40
0.01	3.17

- In addition to the overall analysis, I anticipate that there will be quite a few stratified analyses (such as major diagnostic subtypes of AML and NHL, stratifications by major industry, occupation and/or exposure category). Such stratified analyses will be based on subsets of the overall data. The table below shows the smallest risk ratios that subsets of certain sizes can detect.

  Control-to-case ratio = 2:1
  Alpha = 0.05
  Power = 0.80
  One-sided test

Exposure rate in controls	Minimum detectable risk ratios						
	25 cases	50 cases	100 cases	200 cases	300 cases	400 cases	500 cases*
0.05	6.72	4.34	3.04	2.29	2.01	1.85	1.74
0.04	7.72	4.87	3.34	2.47	2.14	1.96	1.84
0.03	9.34	5.74	3.82	2.76	2.35	2.13	1.99
0.02	12.54	7.42	4.74	3.27	2.73	2.44	2.25
0.01	21.96	12.25	7.28	4.67	3.74	3.24	2.95

* Original target sample size.

Selection of Controls

- The selection of controls for both AML and NHL patients seemed to be working properly. Matching criteria have been followed closely. During January-June 2006, there were only a few minor isolated problems.

- At a couple of small hospitals with limited number of patients, finding appropriate controls in the right age range, especially when the cases were either very young or very old, remained to be difficult. In 2005, there was some confusion as to whether the age matching requirement was $\pm$5 or $\pm$10 years. This point was subsequently clarified: $\pm$5 years should be attempted first and, if none available, $\pm$10 years will be accepted. A review of the recently selected controls confirms that this matching criterion is now clearly understood by the Clinical Coordinators at participating hospitals in Shanghai.

- The participation rate among eligible cases or controls has been extremely high. During the last six months, only 3 to 5 patients (mostly controls) per month refused to participate in the study.

- We have been able to select two controls for every case.

- We continue to monitor the distribution of controls by diagnostic category to ensure that there will be no over-representation of any diagnostic category, which in turn may be associated with certain socio-economic class, occupation or lifestyle. The table below shows the distribution of controls by major diagnostic category, which seems to be in agreement with that of the general population of patients in Shanghai.

Diagnostic category	Percent
Cancers	11.26%
Nutritional, metabolic & immune system diseases	12.25%
Neurological system diseases	8.42%
Circulatory system diseases	10.40%
Respiratory system diseases	9.16%
Digestive system diseases	14.48%
Urinary system diseases	13.49%
Musculoskeletal diseases	7.8%
Blood & hematopoietic diseases	0
Mental diseases	0

Please note that, as specified in the protocol, patients with "blood & hematopoietic diseases" or "mental diseases" are not eligible.

Exposure Assessment

- Dr. Richard Irons will present an updated progress report of exposure assessment at the upcoming August 1-2 meeting in Houston. At the moment, the updated information is not available to us to be included in this report.

- For reference, however, we have reproduced some previous exposure assessment information (as of January 2006) below (in blue). The information should still provide a general idea of the range of benzene exposures in our study.

- Based on preliminary exposure assessment, a total of 259 patients were either definitely or potentially exposed to benzene, representing an

overall benzene exposure rate of 7% in all patients (cases and controls). The exposure frequency (7%) is slightly lower than that used for power calculation (10%) in the original study design.

- The distribution of benzene-exposed patients by major industry is as follows:

Industry	Percent
Painting	32%
Machine maintenance	20%
Rubber	14%
Home remodeling	13%
Shoe making	12%
Printing	9%

- Based on preliminary exposure assessment, some benzene exposure data of study subjects (cases and controls) are presented in the table below:

Earliest exposure	1936
Occupations/industries with highest exposure	rubber plank manufacture, brush & spray painting
Highest maximum exposure	515.7 ppm *
Highest minimum exposure	12.1 ppm *
Highest typical exposure	161.9 ppm *
Longest exposure	40 years

* preliminary estimates

- In addition to the patients identified to have been exposed to benzene, approximately 2800 patients were identified to have been exposed to other substances in a variety of industries. The largest category is farming with potential exposure to agricultural chemicals. Other frequently encountered chemicals are toluene and xylene.

- Plans have been underway to develop exposure profiles for 8 to 10 key industries using data from the 5 or 6 most industrialized districts of

Shanghai. For example, the exposure assessment team has been working in close collaboration with the Yang Pu District IPHS and has been given access to the exposure records maintained at the local IPHS.

- The exposure assessment team continues to develop exposure data/estimates using the "sector analysis" approach. The analysis for two commonly encountered industries (shoe and rubber) has been completed.

- Exposure data from other sources are being collected, including published literature. According to Professor Liang, who heads the literature review task, preliminary exposure data for the following occupations/industries have been completed: shoemakers, painters, paint manufacturers, printers, dyers, and rubber workers. These literature data will supplement the Shanghai Municipal IPHS database and district IPHS data in the development of job-exposure matrix (JEM).

- Benzene exposure data for the shoe industry derived from the Chinese medical literature have been collected and analyzed. A report summarizing benzene exposure data in the shoe industry in China has been prepared ("Benzene exposure in the shoemaking industry in China, a literature survey, 1978-2004" by Wang L, Zhou Y, Liang Y, Wong O, Armstrong T, Wu Q, Fang J, Ye X, Schnatter R, Irons R, Fu H. To appear in *Regulatory Toxicology and Pharmacology*) and has been accepted for publication in a special issue of *Regulatory Toxicology and Pharmacology* "The Development and Regulation of Occupational Exposure Limits in Asia" (Otto Wong, Guest Editor). The special issue, consisting of papers discussing occupational exposure limits in countries/regions in Asia, is expected to be published around October 2006.

- Preliminary data for another two industries, paint manufacturing and painting, have been assembled and will be finalized by August 2006. Preparation of similar reports for these two industries is now underway.

<u>QA/QC</u>

- Pei Xiaodan, who is responsible for QA/QC, has been providing monthly reports on a regular basis. These monthly reports include accuracy checks of the following items: primary and secondary questionnaires, data entry and initial exposure assessment.

- A recent QA/QC of questionnaires taken between September 2005 and June 2006 indicated that 93.7% of primary questionnaires sampled for review were "qualified" (i.e., no problems were found) and 98.4% of

secondary questionnaires were "qualified." We have discussed with Chen Xiaobao, who is in charge of interviews, ways to further improve the quality of primary questionnaires.

- The overall data entry error from September 2005 to June 2006 was 2.7%, which is acceptable.

- During my recent trips to Shanghai I conducted three *ad hoc* QA checks of samples of both primary and secondary questionnaires taken in the last six months. I identified only 1 or 2 minor mistakes in these questionnaires (with literally thousands of items/entries) per QA check. These results reinforce/validate the routine QA/QC checks by Pei Xiaodan.

Data Retrieval

- Instead of retrieving the necessary data for the case-control study from the overall Shanghai Health Study database ourselves, an arrangement for data retrieval has been made. We will not have direct access to the finalized data, but we will be provided with the necessary data by Gail Jorgensen of ExxonMobil (under subcontract to the University of Colorado).

- Gail Jorgensen will extract all the necessary case-control study data (questionnaires, diagnoses, family history, medical history, lifestyle, employment and exposure histories) for us. Our analyses will be based on these extracted data.

- To validate the extracted data, we will conduct a sample audit. Using study ID numbers, we will be able to compare the extracted data to the original source documents (such as questionnaires and exposure assessment reports). The results of the data audit will be included in our final report.

Test Data Sets

- We received four test data sets from Gail Jorgensen in mid-July. The purpose of these data sets is to allow us to get familiar with the format and contents of the data we will eventually receive and to allow us to develop analytical programs. The first two data sets were samples of AML and NHL patients (150 each). The remaining two data sets consisted of questionnaire data for the selected AML cases, NHL cases and their matched controls (300 each).

- The first two data sets contained diagnostic categories, exposure assessment status (but no exposure information itself), subject ID, matched controls' IDs, and date recorded.

- The two questionnaire data sets were incomplete. Not every case or control in the data sets had a questionnaire. The following table shows the number of questionnaires found for the subjects in the test data sets:

Study subjects	No. of questionnaires found
150 AML cases	143 (95%)
300 AML controls	286 (95%)
150 NHL cases	61 (41%)
300 NHL controls	123 (41%)
Total = 900	613 (68%)

- Even among those with questionnaires, the information available in the data sets was very incomplete. There was some information on smoking and drinking. There was no information on family histories (Part 2 of the Primary Questionnaire), medical or medication histories (Part 3), employment history (Part 4), occupational exposures (Part 5), and non-occupational exposures other than smoking and drinking (Part 6). These data sets in their current form are incomplete and not useful for us to develop analytical programs.

- No information of the Secondary Questionnaire or exposure assessment was available for any subject in the test data sets. We did not know the occupations of the subjects in the data sets and were not able to tell which subjects in the test data sets were exposed to benzene.

- I contacted and discussed the problems of the data sets with Gail Jorgensen, who basically agreed that the data were not ready. She explained that she was not able to reconstruct questionnaires for some study subjects in the study. It would take some time to reconstruct all the questionnaires and to assemble the database properly. Both of us agree that we need to examine the various sources of the database and to carry out a comprehensive audit of the data in Shanghai. We have tentatively planned the following trips:

    - In September, I will go to Shanghai and, with the assistance of the Shanghai staff, I will do the necessary preparation for a comprehensive audit in October.

- In October, Gail Jorgensen and I will go to Shanghai and conduct a comprehensive audit of the data and the database in Shanghai.

- In my opinion, we should put the database at a fairly high priority and make sure that data from different sources (diagnostic, clinical, exposure, questionnaire) will be linked properly and will be ready for analysis in 2007.

Project Management

- Communication with the Fudan team in Shanghai remains a high priority item. In addition to email and telephone calls, I made 3 trips to Shanghai in the first half of 2006.

- At the last annual meeting in Fairfax, Virginia, the following timeline for the case-control study was presented (also attached to our last progress report dated January 23, 2006). The current plan is that case enrollment will be completed by December 2006. Final diagnoses, control selection, interviews and data entry will most likely be completed by the end of the first quarter of 2007. Exposure assessment of newly enrolled patients (cases and controls) will be conducted in the first and second quarters of 2007. Assuming that both diagnostic and exposure data for all patients will be finalized and available by June 2007, we will be performing epidemiologic analyses of the final data in the third quarter of 2007.

- To facilitate data transfer/linkage and to develop plans for statistical analyses, however, we needed relatively large and representative samples of exposure estimates and diagnoses of study subjects by mid-2006 (see table below). These samples would allow us to develop plans for data transfer/linkage and to test-run analysis programs. We received some test data sets from Gail Jorgensen in July 2006. We have reviewed the test data and have discussed the problems of the data with Gail Jorgensen.

Time	Activities
July 2006	Samples of exposure estimates (from the Exposure Assessment Team) and diagnoses of study subjects (from JCML) provided to AHS.*
3Q & 4Q 2006	Case-control study data management and plans for statistical analyses (based on samples provided).*

December 2006	End of case enrollment.*
1Q & 2Q 2007	Selection of controls. Interviews, final diagnosis and exposure assessment of last group of patients.
June 2007	Final diagnosis and exposure assessment of all patients completed.
2Q & 3Q 2007	Final data editing/auditing. Analysis and report preparation.

* See notes below.

Notes Added after the Houston Meeting

- As discussed on pages 7-9, the preliminary test data sets provided to us in July were incomplete and inadequate for us to develop any analytical programs for data management or statistical analysis. The test data sets, however, underscore the inadequacy of the database in its current form.

- There were some suggestions at the August 1-2 meeting in Houston that case accrual may be extended to mid-2007. Any extension will obviously affect the timeline and budget of the case-control study. If the decision of extending case accrual to mid-2007 is made, we need to be informed immediately, so that we can revise our study plan appropriately.