

# **Characteristics of Chronic Lymphocytic Leukemia Patients Part I: Demographics**

**April 13, 2007**

**Kenneth O. Cogger, PhD  
President, Patient Databases, Inc.  
www.patientdatabases.org**

## **Introduction and Background Material**

In 1998, Professor Sheldon Messenger (University of California, Berkeley) conducted the first systematic survey of CLL patients. His survey was directed to the then approximately 750 subscribers to the CLL email list (<http://listserv.acor.org/archives/ctl.html>). Responses to this survey numbered 340, representing about 45% of the list membership in 1998. This survey is available at <http://www.acor.org/leukemia/survtoc.html>.

In 2001, I developed a web-based pilot project to collect CLL patient information on a continual basis. The feasibility of this project led to the creation of Patient Databases, Inc., a 501(C) (3) charitable organization based in Washington, DC which now administers the data collection website at <http://patientdatabases.org>. Some preliminary data on patient demographics was posted on this site in 2003. The present report greatly expands on this preliminary data.

Currently, there are 2,596 subscribers to the CLL email list. That list, as well as other on-line sites, has encouraged CLL patients to participate in the database project. At the present time, our database has 1,074 participants, most but not all of whom are CLL list subscribers. This is a significant large number of patients, and our results are based on a larger number of CLL patients than has ever been subject to analysis in any scientific published study. Our participants have entered demographic information such as age, gender, race, time of diagnosis, time of first treatment, etc. as well as blood test results and treatment information.

The current report is the first of two studies and concentrates on the demographics of patient participants. Later, a second report will deal with blood test results and their distribution, prognostic indicators, and other information gleaned from the database.

It is important to note that these 1,074 participants do not constitute a 'random sample' amenable to the usual statistical calculations of confidence intervals and tests of significance. Our participants are self-selected, and there are likely biases in all of our reported statistics. It is impossible to state the direction or magnitude of such potential biases. For example, all participants are evidently internet knowledgeable. It might be argued that younger people have greater knowledge of internet usage, thereby skewing our participant group towards younger ages. However, it could also be argued that age differences in our electronic era are rapidly diminishing. It might also be argued that there are gender or race differentials in internet access and knowledge, further skewing results in unknown directions. Alternatively, it could also be argued that gender differentials are becoming increasingly small.

The National Cancer Institute's SEER data on cancer, and CLL in particular, also is subject to potential criticism. The SEER data is not based on complete US coverage, but only selected areas of the country. Some of the SEER data is based on patients who have died, skewing the results to older patients. Our database consists largely of living patients or, at least, patients who have not been identified as dead. Of course, since a dead patient cannot enter a time of death, we also cannot with certainty determine the life status of any participant. We have, from the CLL list postings, noted and recorded the death of many participants in an attempt to keep the database as accurate as possible.

These caveats are important to keep in mind when interpreting our results. The statistics presented are best viewed as descriptions of our participants, not as estimates of characteristics of the CLL population at large. The results are still very important, first because of the size of our study, second because of the continuing gathering of information on living patients, and third because of the depth of information collected.

## Demographic Characteristics of CLL Patient Participants

Demographic characteristics are important for two reasons. First, they may be of interest to the medical community. For example, is CLL largely a disease of old white males? If so, the interest of pharmaceutical companies in developing treatments for such a narrow group of patients may be economically limited. Second, CLL patients often question just how their disease fits into the overall characteristics of CLL patients. For example, if a female is diagnosed at age 55, is that really unusual in terms of either age or gender or both? It may be reassuring for a CLL patient to know that their case is not an unusual one and, further, to know how other patients similar to themselves are faring. Our demographic characteristics provide clues to answering these and other questions.

Table 1. Age at Diagnosis (All Patients)

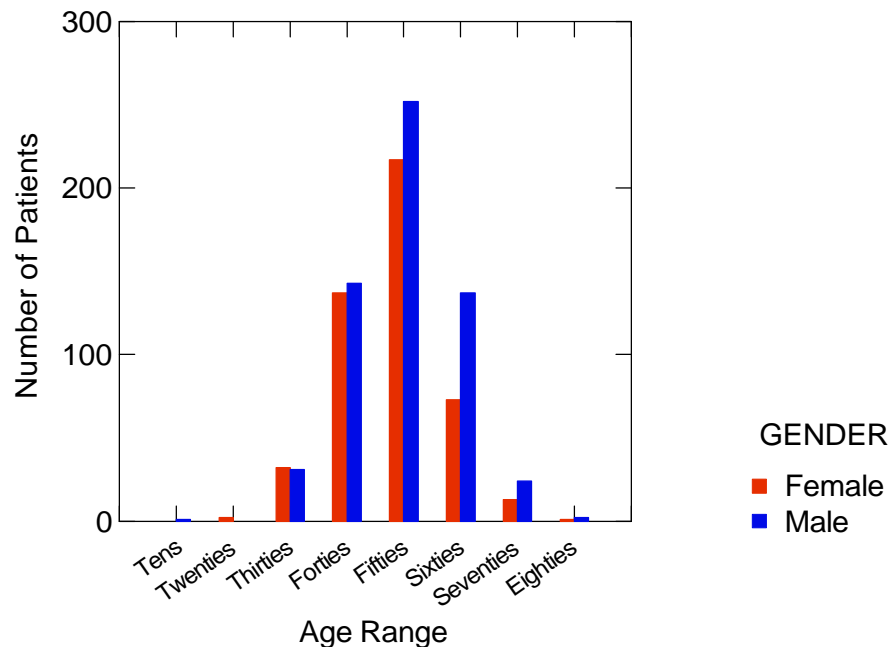


Table 1 indicates that the diagnosis of CLL peaks for both males and females in their fifties, not elderly by modern standards. Patients also included one male teenager and two females in their twenties. I've also had personal contact with both men and women in their twenties who are not part of our database. Participants in the database are clearly not “old” either on average or in terms of potential diagnosis at an early age.

In terms of gender, females actually outnumber males for diagnosis in their thirties or younger and are nearly equivalent for diagnosis in their forties. Only for those in their fifties or older do males exceed females in their diagnosis of CLL. For participants in our database, the stereotype of CLL as being a disease of “old men” can be disputed.

Table 2. Gender (All Patients)

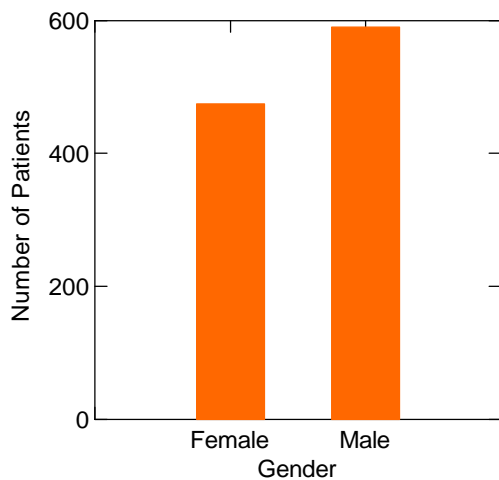


Table 2 combines all diagnostic ages to give a summary picture by gender alone. It suggests that males outnumber females, but not by a large margin. And, from Table 1, in some age categories females outnumber males.

Table 3. Age at Diagnosis (All Patients)

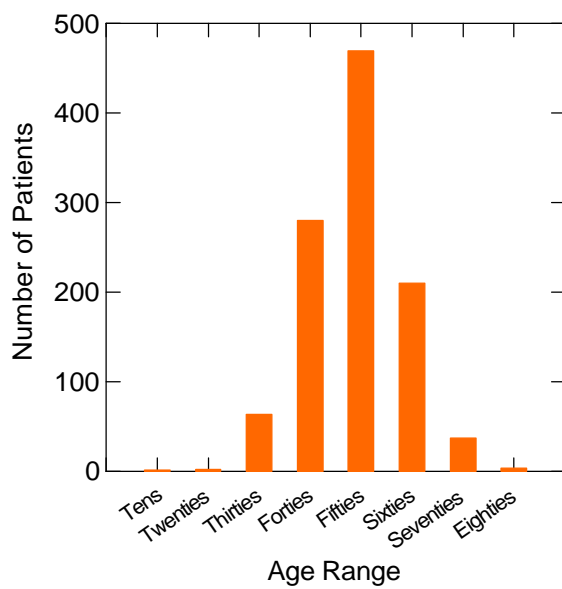


Table 3 combines genders to present an overall picture of diagnostic age. As in Table 1, it is clear that the diagnosis of CLL peaks for people in their fifties and then declines. These three tables describe characteristics for the entire database, which is predominately US patients, but also includes entries from many other countries.

Table 4. Age at Diagnosis (US Patients)

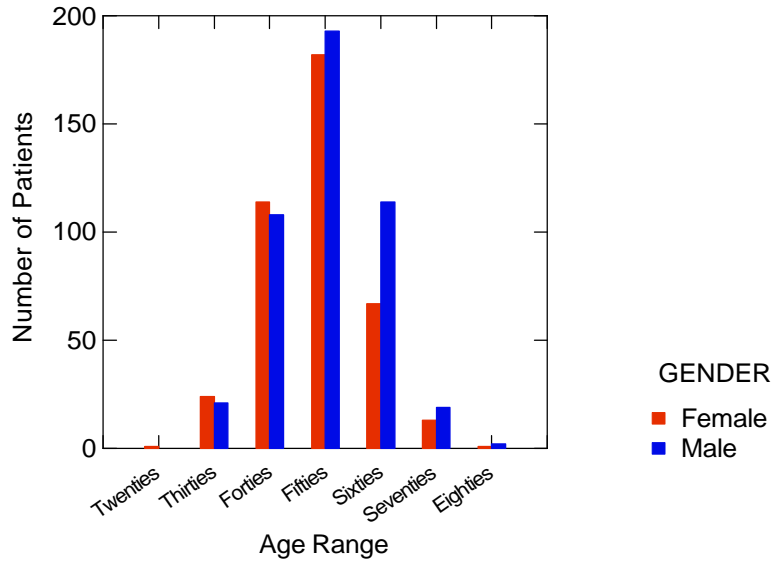


Table 5. Gender (US Patients)

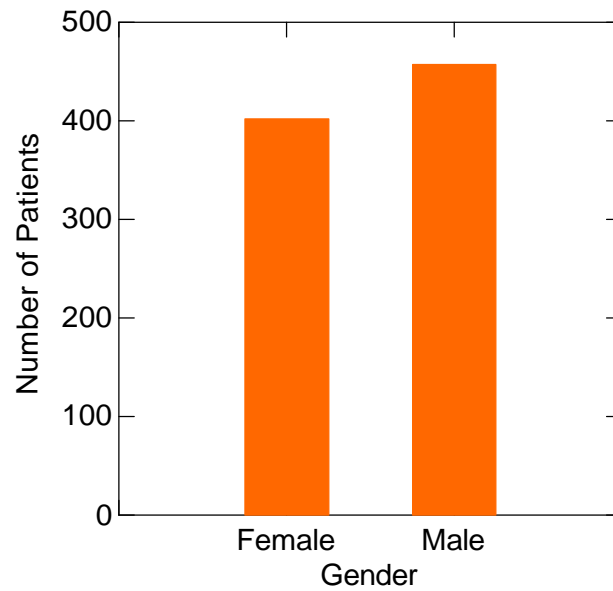
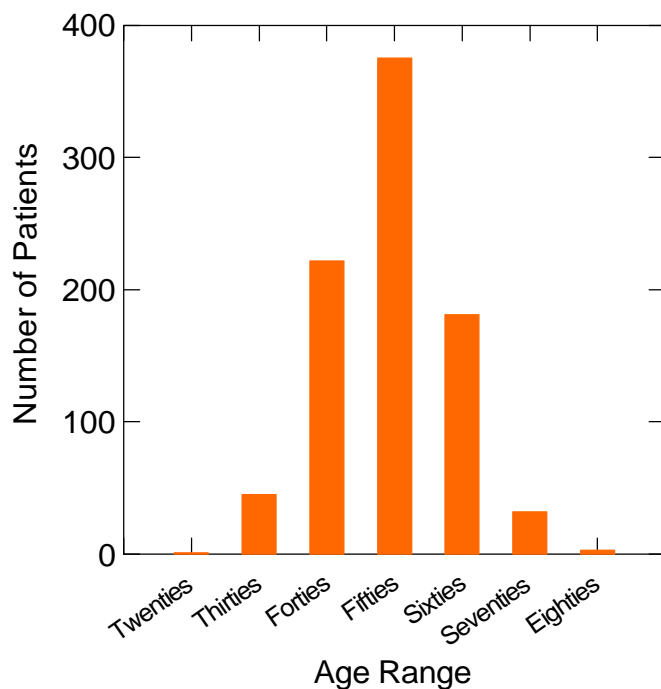


Table 6. Age at Diagnosis (US Patients)



Tables 4-6 simply reproduce the first three tables, but are restricted to US patients only. Obviously, similar conclusions can be drawn as for the other tables. No striking differences appear between the entire participant database and US participants.

Table 7. Distribution of Patients by Country

Number of Patients	Country
1	Czech Republic
1	Iceland
1	India
1	Ireland
1	Greece
1	Uruguay
1	Lithuania
1	Macedonia
1	Mexico
1	Hungary
1	Denmark
1	Japan
1	Philippines
1	Poland
1	Romania
1	Slovenia

Number of Patients	Country
1	China
1	Spain
1	Switzerland
1	Thailand
1	Chile
1	Mozambique
2	Finland
2	Italy
2	South Africa
3	Egypt
3	Belgium
5	Germany
5	Netherlands
8	Israel
10	New Zealand
28	Australia
32	Great Britain
84	Canada
868	United States

Table 7 indicates that the bulk of our database is from the British Commonwealth (New Zealand, Australia, Great Britain, and Canada) and the United States. Participants come from many other countries also. This table is consistent with observations in the medical literature on the frequency of CLL diagnosis in terms of ethnicity as well as the fact that the Patient Databases website is in English only. It is perhaps remarkable that more Scandanavian countries do not appear, nor does Russia, since English is a common second language and their ethnicity would perhaps suggest there should be more CLL participants from those countries.

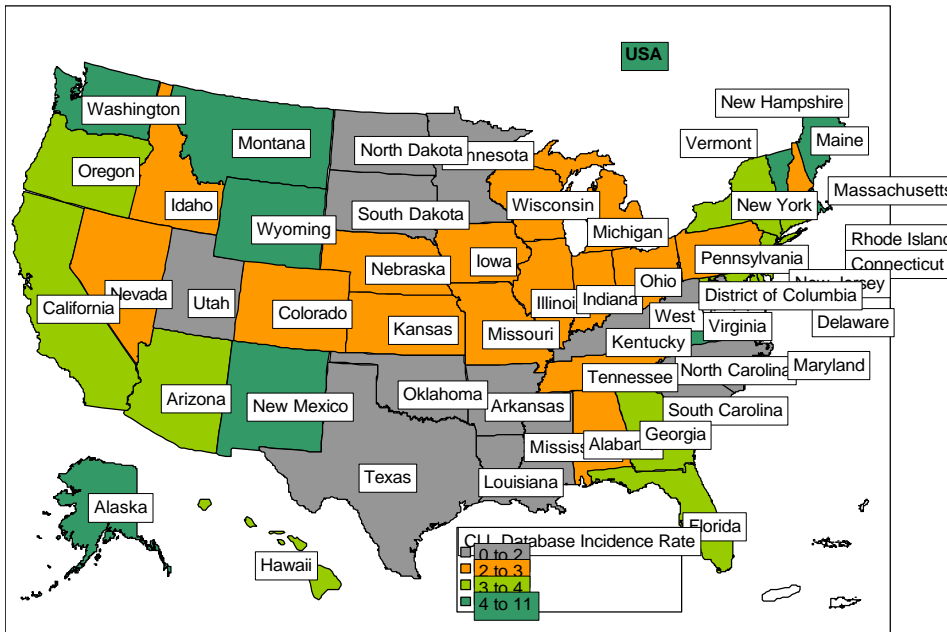
Table 8. Distribution of US Patients by State

Number of Patients	State
1	North Dakota
1	South Dakota
2	Arkansas
2	West Virginia
3	Delaware
3	Idaho
3	Louisiana
3	New Hampshire
3	Rhode Island
3	Utah
4	Mississippi
4	Montana
4	Nebraska
4	Vermont
4	Wyoming
5	District of Columbia
5	Hawaii

6	Alaska
6	Nevada
6	Oklahoma
7	Iowa
7	Maine
7	Minnesota
8	Kansas
8	South Carolina
9	Kentucky
10	New Mexico
11	Colorado
12	Connecticut
12	Wisconsin
13	Alabama
14	Indiana
14	Oregon
15	North Carolina
16	Tennessee
17	Missouri
18	Maryland
19	Arizona
28	Illinois
30	Georgia
31	Michigan
32	Washington
33	Massachusetts
33	New Jersey
35	Ohio
35	Texas
37	Virginia
38	Pennsylvania
55	Florida
62	New York
130	California

Table 8 indicates that all USA states are represented, with more populous states having more patients. This is not surprising. To get a clearer picture, the map below color codes each state according to the number of patients in the database per million state population. Grey states have the lowest rates, orange a bit higher, light green a bit higher, and dark green has the highest participation rates.

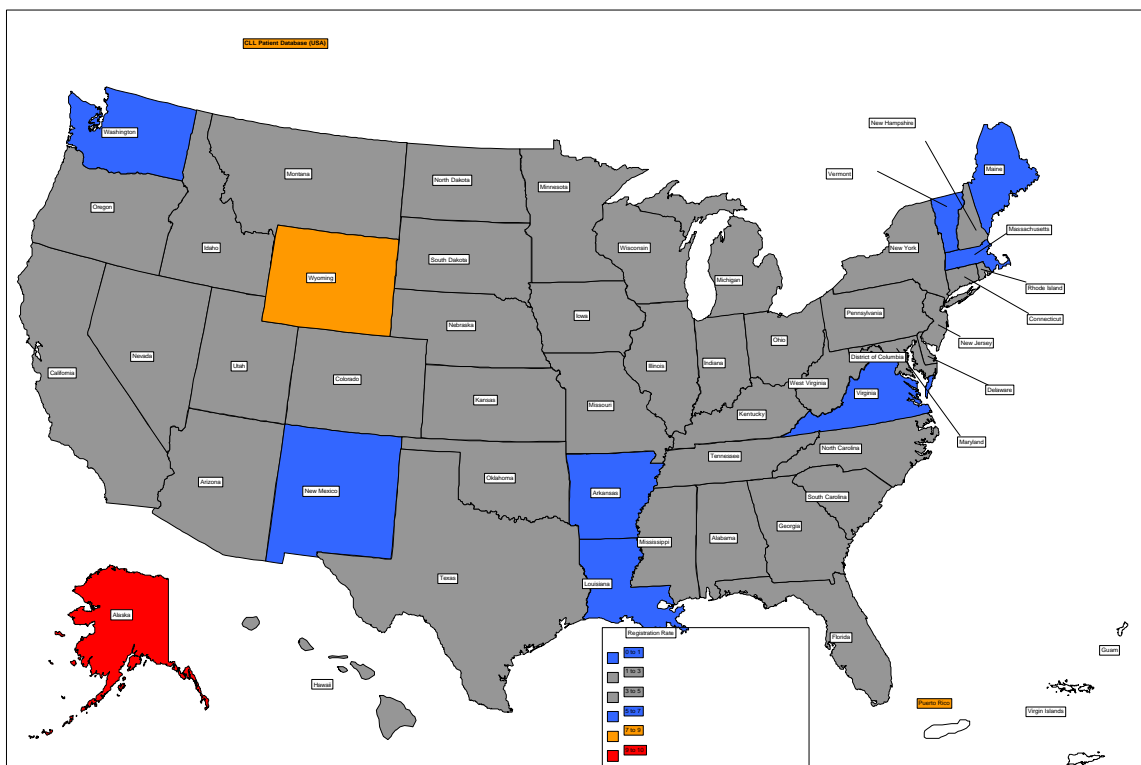
## Unadjusted Patient Counts by State, per million population



This map simply reveals that patient participation rates per million vary from state to state. That is to be expected. But, are the differences between rates (and colors) within normal statistical bounds of variation? Is our database representative, or are there some states with unusually high or low CLL participation?

To answer these questions, we group states statistically with only four colors. States having participation rates close to the average state rate are coded grey; “close” is defined as within one standard deviation, a common statistical measure of variability. Those states differing from the average state rate by more than one but less than two standard deviations are coded blue, those between two and three standard deviations are coded orange, and those more than three standard deviations away from the national average are coded red.

## Statistically Standardized Rates By State



The “Gaussian”, “Normal”, or “Bell-Shaped” probability distribution almost universally applied in situations like this suggests that about 68% of the states should be grey, 27% of the states should be blue, 4% of the states should be orange, and less than 1% red. The above map is almost exactly how we would expect it to be colored.

Using the statistical percentages indicated above with 50 states, the table below shows what our database actually has in terms of number of states of each color, and what the expected number of states would be assuming normal statistical variation.

	Gray	Blue	Orange	Red	# of States
Database	40	8	1	1	50
Normally Expected	34	14	2	0	50

The database is quite representative in terms of participants registering on the basis of state populations assuming there are no differences between states in the incidence of CLL patients. Grays and Blues total to exactly what would be expected statistically and only the one red state, Alaska, might be viewed as unusual. However, there are only six patients from Alaska, and if that number had been five instead of six this state would be colored orange rather than red. The patient database is remarkably representative across states, lending credence to many of our conclusions regarding demographics.

Table 9. Distribution of Patients by Race

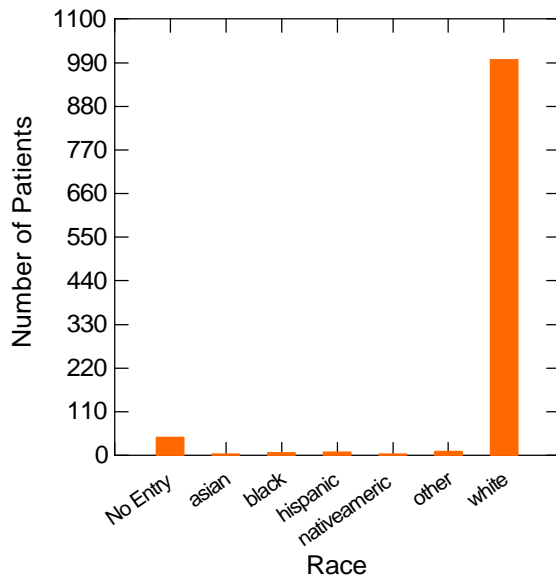


Table 9 confirms medical studies that CLL is predominately a disease of White people. Very few Asians, Hispanics, or Blacks are found in our database.

Table 10. Distribution of Education Level

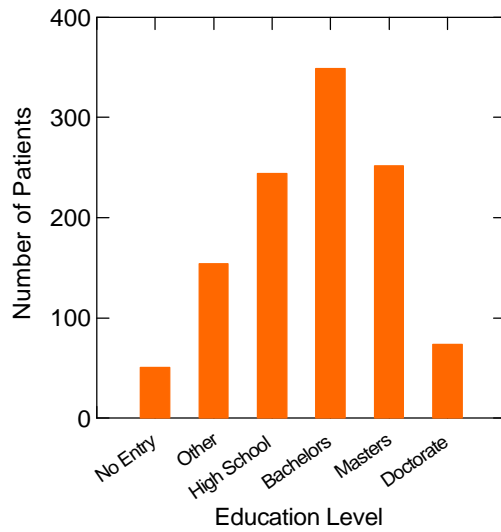


Table 10 indicates that the majority of our participants with a recorded education level have at least a Bachelor's degree and that many have a Master's degree. Our participants are relatively highly educated, perhaps explaining their ability and knowledge to participate on the internet, but perhaps also not related to their disease status.

Table 11. Time to Treatment by Age at Diagnosis

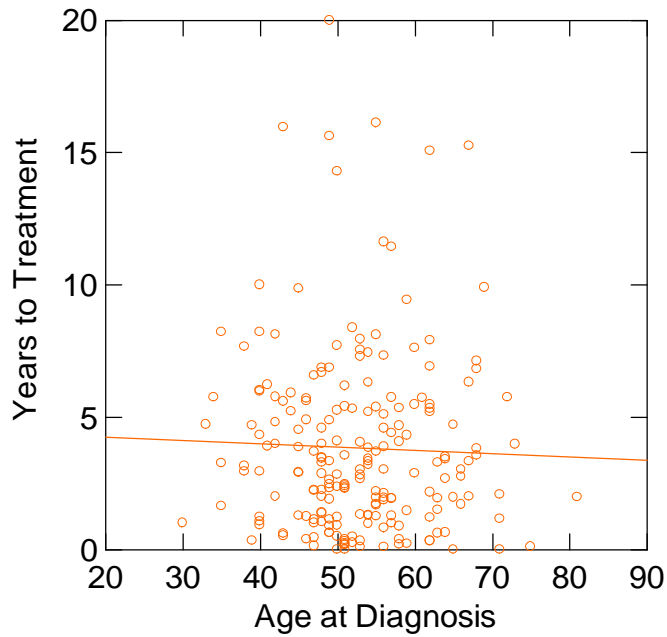


Table 11 is based on the 206 patients in our database who have received treatment of some kind. The vast majority of our participants have evidently not received treatment of any kind. Clearly there is a great deal of variability in time to treatment for all diagnostic ages. Superimposed on this scatter plot is a straight line estimated with linear regression which reveals a weak but inverse relation between age at diagnosis and the time to treatment. Older patients tend to have a slightly shorter time to treatment than younger patients. This trend is quite small, however, providing little evidence that young or old patients diagnosed with CLL can expect that the time until treatment of any kind is related to their age at diagnosis. Time to treatment is on average a bit shorter for older patients, but the evidence is far from conclusive.

Table 12. Time to Treatment by Gender

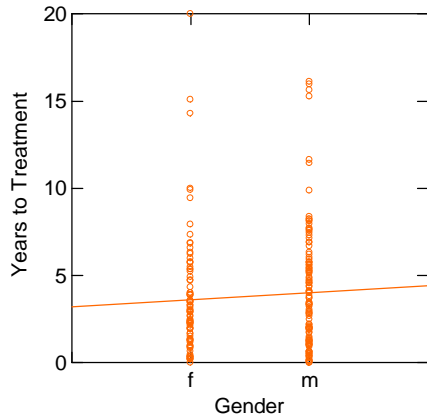


Table 12 shows a scatter plot of times to treatment by gender. Again, note the high variability in times to treatment for both men and women. A linear regression line is superimposed on this data, revealing a slight tendency for females to have earlier treatment than males. This relationship is quite weak, however, and standard tests indicate the association is not statistically significant.

## **Concluding Remarks**

This report on the basic demographics of the patient database has several surprising results and some reassuring results for patients. First, our patient participants are not “old”. By modern standards, CLL patients are moderately young, with patients in their teens, twenties, and thirties being represented. Second, CLL is not a “male” disease. It appears to be nearly gender-neutral. Third, the time until treatment also appears to be independent of both gender and age at diagnosis. This should be reassuring to patients who are young in age or female, both categories being sometimes dismissed as insignificant. Finally, the broad spectrum of CLL patients in our database (gender and age) should be an economic incentive to those researching and developing treatment protocols that the living CLL patient population is relatively young with many years ahead of them with opportunities for successful treatment options, and diverse enough to warrant such investment. The maps of participation rates support the conclusion that our database representation is relatively uniform across the United States.

## **Acknowledgments**

This report and the work behind it would not have been possible without the efforts of many unpaid volunteers and, especially, our participants. Shelly Messenger, now deceased, had the original idea to survey CLL patients in 1998. His idea led directly to the current ongoing website project. Irene Murphy led the successful effort to establish a permanent organization to manage operations. Diane MacKinnon has been tireless in her assistance to me. The entire Board of Directors has been patient and supportive during the long process of producing this report. The distinguished Advisory Committee has also been invaluable. I alone am responsible for any errors and mistakes in this report. None of the individuals named here are responsible for my errors in judgment or execution of the report. Finally, I must acknowledge “Granny Barb” Lackritz, deceased, who initiated so many projects beneficial to CLL patients and is remembered and revered, especially by those of us who personally met her.

### **Patient Databases, Inc. Board of Directors**

Ken Cogger, PhD, President, Conifer, CO  
Karen Goodell, Treasurer, Lawrence, KS  
Diane T. MacKinnon, Chevy Chase, MD  
Carol Lucas, Beaufort, SC  
Marianne McGhan, Soap Lake, WA  
David Thomas, Calgary, Alberta, Canada  
Marcia Weller Weinhold, Ph.D., Hammond, IN

### **Patient Databases, Inc. Advisory Committee**

Gerald Marti, M.D. Ph.D., CLL Program, National Cancer Institute, Bethesda, MD;  
Susan J. Leclair, Ph.D., Department of Medical Laboratory Science, University of Massachusetts Dartmouth  
Dr. Terry Hamblin, Professor of Haematology at the University of Southampton, England,