Illinois Department of
PUBLIC HEALTH

Pat Quinn, Governor
Damon T. Arnold, M.D., M.P.H., Director

525-535 West Jefferson Street • Springfield, Illinois 62761-0001 • www.idph.state.il.us

September 8, 2009

Patrick McNulty
Public Health Administrator
McHenry County Department of Health
2200 N. Seminary Avenue – Route 47 N.
Woodstock, IL 60098

Dear Mr. McNulty,

We received your letter of August 19, 2009 regarding cancer statistics in McHenry County. Most of the data related to your inquiry have been previously sent to and discussed with your Department by Melinda Lehnherr, our assistant chief of the Division of Epidemiologic Studies, Illinois Department of Public Health. Below is our response to the specific questions you asked in your letter.

1. For the most recent period (2002-2006) does McHenry County have a higher Cancer rates (all sites) than Illinois?

   A: No – the difference is not statistically significant, i.e., the difference, as judged by probability and statistical evidence, is likely to have occurred by chance.

2. For the most recent period (2002-2006) does McHenry County have a higher Cancer rate (Brain) than Illinois?

   A: No – the difference is not statistically significant.

3. For the most recent period (2002-2006) does McHenry County have a higher rate (other central nervous system cancers) than Illinois?

   A: No – the difference is not statistically significant. Brain and CNS are combined in the county-level public data, but the difference is still not statistically significant.

4. Over the last 15 years has there been a significant increase in the rate of brain and other central nervous system cancers in McHenry County?

A: No significant trend, either upward or downward, was observed in McHenry County.

5. Over the last 15 years has there been a significant increase in the rate of brain and other central nervous system cancers in zip codes 60050 and 60051 or in the McCullom Lake census tract?

A: Analyzing data for areas smaller than a county is difficult because of the small numbers of cases as well as of populations. Rates based on small numbers are less stable. For example, when the number of cases is less than or equal to 16, the relative standard error for the rates exceeds 25 percent (the relative standard error, or RSE, measures how unreliable or uncertain a statistical estimate is relative to its size, and at this high level of uncertainty, interpretation of the rate is often meaningless). The lack of population counts from the Census for small geographical areas also contributes enormously to the difficulty, as population counts are needed to accurately calculate and compare rates. [i.e., in order to determine a rate, we must divide the number of cases in an area by the number of people in the same area. Without an accurate count of the number of people in the area, the rate calculated will also be inaccurate.] The Census Bureau provides accurate counts of population for many demographic categories and for many geographic areas, but it does so only for census years (once every decade). In between census years, the Census Bureau only provides population estimates by age, sex, and race at the state and county level, but not for smaller geographic areas. Further, rates cannot be calculated for ZIP code areas or census tracts if their boundaries change over time, which seems to be the case for at least one of the two ZIP codes you mentioned. There were no population data for 60051 for 1990 or 2000.

6. What are the difficulties in the analysis and making conclusions from zip code or census track cancer data?

A: See answer #5 above.

7. Why can't you analyze data on more specific diagnosis of brain or other central nervous system cancers?

A: The number of cases is very small and an analysis of them would produce extremely unstable rates even at the county level. A statistical analysis based on small numbers is on shaky ground. For example, if a very rare cancer occurs in one person out of

200,000 in one year and in two persons out of 200,000 in the next year, the rate will appear to have doubled, but the significance of this change is questionable. Additionally, with such a high specificity of information to display, cases could potentially be re-identified. A fundamental expectation of residents of Illinois is that the personal medical information collected by the Illinois State Cancer Registry (ISCR) from their medical provider will be kept confidential and that information from the Registry will be publicly reported only in aggregate from, where no individual identities can be determined. When the number of cases is very small, the release of specific information creates the risk that the identity of individuals could be inadvertently revealed by comparisons with other publicly available data sources.

8. Why do you analyze cancer data over multiple years rather than single years?

   A: As noted in answer #7 above, when there are only a few cases in a population, rates are unstable over time and the risk of re-identification of patient information is high. By combining data over multiple years, these problems can be reduced.

9. What is the most current year of cancer data? Why is there a delay in the release of data?

   A: 2006 – The national standard for release of cancer data is 24 months after the end of a calendar year. First, data must be collected from multiple sources (e.g., hospitals, doctors, laboratories), both within and outside the State. Out-of-State data is required because cancer registry information is based upon the patient's place of residence, and many Illinois residents obtain health care in other States. Before data can be released, all sources of the cancer data must be combined on one record and de-duplicated (since a diagnosis for the same individual may have been reported by more than one source), out-of-state data exchange must be completed, and cases must pass a set of auditing processes for accuracy of the data, all of which take time to complete. The public data set for 1986-2006 for Illinois was released in April 2009, the same time when the national data were released.

10. Are there source(s) of more current (beyond 2006) and reliable cancer data for McHenry County?

A: No – ISCR is the only source for population-based cancer incidence data in Illinois. Cancer registry data are much more accurate and reliable than other potential sources of data such as surveys or self-reporting.

11. How reliable are the IDPH Cancer Registry Data? Historically?

A: Completeness and quality of ISCR data have been high for many years. The North American Association of Central Cancer Registries (NAACCR), in its effort to benchmark and foster best practices among population-based state registries, has developed a certification process that reviews registry data for completeness, quality, and timeliness of reporting (http://www.naaccr.org/Certification/index.html.) ISCR data met the criteria for gold certification, the highest standard, consistently from 1996 through 2006, and as of November 2008, the statewide completeness of case reporting from all reporting sources, assessed using the NAACCR Standard, is considered to be 100 percent for the time period 1994 through 2006. Data completeness, however, was considerably lower prior to 1994 - for example, it was 91% for 1992 and 92% for 1993.

12. Why is it difficult to do cancer cluster studies?

A: There are many limitations to designing, conducting and interpreting cancer cluster studies. First is the statistical consequence of small numbers (i.e., small numbers produce unstable rates and large relative standard errors, as noted in answers #5, #7, and #8 above). Second is the inability to rule out or even reduce chance occurrences (i.e., the 'uncertainty' can never be reduced to zero as long as random variation exists with cancer statistics). Third is a lack of statistical power to tell a true difference (i.e., studies based on small numbers and limited information often do not have enough 'ability' to tell a difference when there is one). Fourth, the lack of information on history of residence (i.e., while the residence at date of diagnosis is known, the length of time a person has lived at that residence is unknown). Fifth is the lack of information on the genetic background, lifestyle, occupational history, exposure and other cancer risk factors of the diagnosed individuals. In addition, there is often a lack of accurate denominators (or population numbers) needed to calculate rates, and the presence of an often long latency period (from exposure to onset of disease) for most adult cancers. The long latency period confounds the problem of lack of residence

history. One less obvious but equally if not more important limitation is the problem created by comparing many cancer/gender/racial subgroups simultaneously. When performing multiple comparisons on the same data set, there is an inflated probability of finding falsely 'significant' differences. For example, statistical analyses that rely on a "95 percent confidence interval," when performed 100 times, can be expected to produce five false positives, even if no correlation actually exists. There is also a risk of preselection bias in choosing a geographical area to include cases (also called the bull's eye problem). In other words, if one looks a map showing the incidence of cases and draws a geographical area for study tightly around those known cases, the likelihood of finding a statistical connection is artificially created. Finally, with exceptions to only a few cancers, the medical community doesn't quite understand the cause of cancer. The lack of knowledge in cancer etiology makes it impossible to conduct a cancer cluster study by following the familiar investigation procedures of an outbreak investigation for an infectious disease or other outbreak, such as influenza or food poisoning.

13. Can you determine causality for cancer registry statistics?

A: No. An analysis of cancer registry statistics, by design, is to compare cancer incidence across geographic areas or population groups. While a lower number may sometimes calm fears and provide a certain kind of assurance to a community, a high number does not in itself establish a particular cause, either for individuals or for the area. Often, an analysis that is based on small numbers cannot even distinguish a real change from changes that are created by chance. In a general sense, finding causes of any disease is usually a long, slow process. Any scientific study has limitations and therefore no single study is likely to prove that a particular exposure definitely causes a particular disease. Neither can a single study nor even a large number of epidemiologic studies enable a person to know why he or she developed cancer.

14. What is the limitation of your analysis of the cancer registry data?

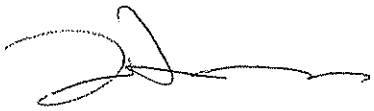
A: See answer #12 and answer #13 above.

15. Is census tract cancer data available to a Health Department?

A: No. When cancer information is reported under law by health providers to ISCR, it contains only a patient's address -- the patient's census tract is not reported. To get the census tract information, the registry has to use a geocoding process to assign census tracts to individual cases according to reported addresses. However, due to missing or incorrect address information from the reported data, only about 90 percent of the ISCR's cases can be so assigned, leaving 10 percent of the cases without any census tract designation. Incomplete data are not suitable for epidemiologic analyses, which is why we have to rely on reported addresses to conduct analyses. Cancer data containing reported addresses are confidential but they can be released for legitimate scientific research purposes under a confidentiality agreement if an application for such access is approved by an Institutional Review Board (IRB) and the Illinois Department of Public Health's Data Research and Release Committee (DRRC).

We hope our response helps you understand the complexity of cancer cluster studies. Should you have any further questions, please do not hesitate to contact Dr. Tiefu Shen, Chief of Division of Epidemiologic Studies, Office of Policy, Planning and Statistics at 217-785-1873, email tiefu.shen@illinois.gov.

Sincerely,



Tiefu Shen, M.D., M.P.H., Ph.D.
Chief
Division of Epidemiologic Studies
Illinois Department of Public Health