## Report on

# Exploring Geographic Scale for Cancer Incidence Mapping 

Demonstration Project on

Geographic Patterns of Cancer Incidence and
Environmental Factors in New Jersey

Program 03074, Environmental and Health Effects Tracking<br>National Center for Environmental Health<br>Centers for Disease Control and Prevention

## Submitted to the CDC January 31, 2007

Prepared by the
New Jersey Department of Health and Senior Services
Consumer and Environmental Health Services
Environmental Public Health Tracking Project
and
Cancer Epidemiology Services
In cooperation with the
New Jersey Department of Environmental Protection

## Table of Contents

Overview of the Demonstration Project ..... 3
Study Team ..... 4
Summary ..... 5
Introduction ..... 6
Approach ..... 7
Practical Minimum Limit on Number of Cases ..... 8
Population Sizes to Generate Minimum Numbers and Comparison to New Jersey Population Units. ..... 8
Conclusion ..... 9
References ..... 10

Figures 1 and 2
Tables 1, 2 and 3
Appendix Tables 1 and 2

## Overview of the Demonstration Project

The New Jersey Department of Health and Senior Services (NJDHSS) was awarded funding from the Centers for Disease Control and Prevention (CDC) to conduct three demonstration projects under the program, "Environmental and Health Effects Tracking," in cooperation with the New Jersey Department of Environmental Protection (NJDEP). The purpose of these demonstration projects is to develop and evaluate methods for linking ongoing, existing health effects and human exposure surveillance systems with existing systems for monitoring environmental hazards and exposures.

One of the three demonstration projects by NJDHSS and NJDEP is to link cancer incidence data with data on environmental hazards and exposures. Environmental factors are known or suspected to play an important role in the etiology of several cancer types. This demonstration project will allow NJDHSS and NJDEP to proactively evaluate the geographic relationships among the incidences of selected cancer types and specific environmental hazards or exposures.

The project was conducted in two phases. Phase 1 involves identification of specific cancer types of interest, and descriptive analysis of incidence data for these cancers, specifically for temporal trends and spatial patterns. The second phase involves the linkage of the cancer incidence and environmental databases to examine specific relationships suggested in the first phase. Separate reports describe the methods, findings and conclusions of these phases of the demonstration project.

During Phase 1, the interagency study team began consideration of issues related to geographic scale in the mapping of cancer incidence data. This report discusses the findings of that effort.

This demonstration project was conducted by the Environmental Public Health Tracking Project (EPHT) in Consumer and Environmental Health Services, NJDHSS, in partnership with Cancer Epidemiology Services (CES) and the New Jersey Department of Environmental Protection (NJDEP).

## Study Team

## New Jersey Department of Health and Senior Services

Consumer and Environmental Health Services
Jerald Fagliano, M.P.H., Ph.D., Principal Investigator
Michael Berry, M.P.H.
Perry Cohn, Ph.D., M.P.H.
Barbara Goun, M.P.H., Ph.D., Project Coordinator
Patricia Haltmeier, L.P.N.
Richard Opiekun, M.S., M.A., Ph.D.

Cancer Epidemiology Services
Kevin Henry, PhD.
Betsy Kohler, M.P.H., C.T.R
Lisa Roche, M.P.H., Ph.D.

## New Jersey Department of Environmental Protection

Steven Anderson, M.S.
Linda Bonanno, Ph.D.
Judith Klotz, M.S., Dr.P.H., consultant

## Summary

This report explores the presentation and mapping of actual disease rates or relative rate information at small geographic scales. Specifically this report considers the question of what should be an appropriate geographic incidence scale for presentation and mapping of cancer incidence relative rate data in New Jersey, given that cancer types occur with widely varying frequency, and given the large variation in population size distributions of population units in the state.

The report sequentially discusses three fundamental questions relevant to the geographical presentation and mapping of cancer incidence:

- What is a practical minimum limit on the expected number of cancer cases in a mapped geographic unit that would generate reliable rate ratios?
- What size population (or person-time) is needed to generate a practical minimum number of cases for mapping, at varying cancer rates?
- How does the distribution of real population units in New Jersey compare to these population sizes?

The report concludes that in New Jersey, most cancer types could be practically mapped for all counties and for a small number of New Jersey municipalities using 5 year incidence intervals. For most municipalities and some census tracts, reliable rates could only be mapped for the most common cancers (prostate, breast, lung and bronchus, and colon and rectum) using 5 year incidence intervals. For the many cancer groupings with annual incidence rates between 10 and 50 newly diagnosed cases per 100,000 population, the report concludes that five year incidence rates or relative rates could only be reliably mapped at the county level, and for perhaps the 25 or 30 largest municipalities in the state.

## Introduction

In 2003, the New Jersey Department of Health and Senior Services (NJDHSS) convened a Cancer Cluster Task Force to examine trends in cancer incidence and mortality, evaluate cancer cluster investigation protocols, and make recommendations for implementation of best practices.

Task Force members, representing government, universities, and public health advocates, expressed interest in the mapping of incidence data at a smaller geographic scale (e.g., municipality or census tract) than typically presented (county or large municipality). One of the reasons for presentation of data at smaller geographic scales is to examine it for local clustering. Mapping of data might allow for the detection of locally elevated rates or other geographic patterns that might point to undiscovered environmental or socioeconomic risk factors. In part, the interest in small scale disease data is also motivated by a community's "right to know" local disease rates.

Also in 2003, the NJDHSS began conducting demonstration projects under the Environmental and Health Effects Tracking (EPHT) project cooperative agreement funded by the U.S. Centers for Disease Control and Prevention (CDC). The Cancer Team established under the New Jersey EPHT project decided to explore the question of appropriate mapping scale for cancer. One of the goals of EPHT is to make public health and environmental data more widely available to the public and researchers, which coincides with the interest expressed by the members of the Cancer Cluster Task Force.

However, there is concern in the public health community about the dissemination and display of information on local disease counts or rates (Rudolph et al., 2006). There are two reasons for this concern: 1) potential violation of ethical and legal protections of the confidentiality of personal health information; and 2) statistical unreliability of the information. Several approaches to decreasing the disclosure risk of confidential data have been proposed and are in use, including administrative controls on access to data, and aggregation of data with rules for cell suppression. Similarly, there are different approaches to increasing statistical reliability of data, including aggregation and geographic smoothing techniques (Rudolph et al, 2006). This report considers the question of statistical reliability only, specifically in relation to the handling of count, rate or relative rate information based on small numbers of cancer cases.

Small area analyses may produce a high proportion of rates or rate ratios that are elevated by chance alone (Neutra et al., 1992; Weinstein and Klotz, 2000). Mapping of unreliable rates may mislead by falsely highlighting areas by chance, or masking an important underlying pattern (Pickle, 2000). While geographic smoothing techniques may be very useful for the visual impact of dampening random variation and revealing geographic patterns, these techniques may not satisfy the "right to know" local disease rates.

This report explores the presentation and mapping of actual disease rate or relative rate information at small geographic scales. Specifically this report considers the
question of what should be an appropriate geographic incidence scale for presentation and mapping of cancer incidence relative rate data in New Jersey, given that cancer types occur with widely varying frequency, and given the large variation in population size distributions of population units in the state.

In practice, public health agencies have used a variety of rules for inclusion of statistics based on low case counts. For example, in its cancer mortality maps and tables, the National Cancer Institute considers data to be too sparse for display if there are fewer than 6 observed deaths, fewer than 12 observed deaths if the rate is not statistically different from the U.S., or fewer than 6 expected deaths (NCI, 2006). The NCI’s SEER program requires at least 16 observed or expected cases (Bleyer et al., 2006). In its county-level cancer incidence maps and tables, the New York State Department of Health marks as unstable any rate that is based on fewer than 20 observed cases (NYSDH, 2006).

## Approach

An initial question to consider is:

- What is a practical minimum limit on the expected number of cancer cases in a mapped geographic unit that would generate reliable rate ratios?

One approach is to identify a minimum limit such that it would be rare for some predesignated multiple (e.g., a 50\% increase, or a doubling) of the expected rate to be exceeded in a geographic unit by chance alone. Based on the Poisson distribution, we examined the cumulative probabilities of observed numbers for a series of expected numbers ranging from 1 to 20 (Figure 1 and Appendix Table 1). For all possible observed numbers for each expected number, we then computed ratios of observed to expected numbers (Figure 2 and Appendix Table 2). This permitted us to determine the expected number that would result in a sufficiently small probability (e.g., 0.05 , or 0.001 ) of an observed number exceeding a rate ratio (RR) of a pre-determined level (e.g., RR $\geq$ 1.5 or $\geq 2.0$ ). These probabilities could be interpreted as the proportion of "false positives" in tables or thematic mapping. This proportion, of course, should be minimized to the degree practical.

Once this question has been answered, a second question to consider is:

- What size population (or person-time) is needed to generate a practical minimum number of cases for mapping, at varying cancer rates?

Finally:

- How does the distribution of real population units in New Jersey compare to these population sizes?


## Practical Minimum Limit on Number of Cases

Table 1 shows the probability of a mapped unit having a RR above 1.5 or 2.0 by chance alone, for a series of expected numbers from 2 to 20 (only even numbers shown). Suppose you wanted to produce map in which there is a less than $5 \%$ probability that mapped unit RR values exceed 1.5 by chance. From Table 1, you can see that a RR of 1.5 would be exceeded in less than $5 \%$ of observations by chance alone (a $5 \%$ false positive rate), when expected numbers are 10 or more. An expected number of 10 would result in less than $0.2 \%$ of observations exceeding a RR of 2.0 by chance.

Suppose that you wanted a map in which there was less than a $1 \%$ chance of mapped unit RR values exceeding 1.5 by chance. In this case you would need an expected number that is larger than 20. Similarly, suppose you wanted to produce a map in which there is a less than $1 \%$ probability that mapped unit RR values exceed 2.0 by chance. From Table 1, you can see that a RR of 2.0 would be exceeded in less than $1 \%$ of observations by chance alone when the expected number is 6 or more. However, an expected number of 6 would result in about $8 \%$ of observations exceeding a RR of 1.5 by chance.

A practical minimum limit of 10 or more expected cases may be chosen to ensure low probabilities of "false positives" and reliability for presentation of small area rates. A larger minimum number, perhaps an expected number of 16 , would result in fewer "false positives." Minimum expected numbers larger than about 16 do not seem to appreciably reduce the false positive proportion. This suggests that selecting a minimum expected number in the range of 10 to 16 cases is a reasonable choice.

## Population Sizes to Generate Minimum Numbers and Comparison to New Jersey Population Units

For surveillance of incidence data for temporal-geographic clustering of cancers, we assumed that a five year interval would be a reasonable time scale for analysis of data. Tables 2 and 3 show the expected number of cases generated over a five-year interval for a range of incidence rates and for a range of population sizes. Table 3 also shows example cancer types within incidence rate ranges for reference. For the many cancer types whose incidence rates exceed about 10 per 100,000 per year, reliable rates based on expected numbers in the 10 to 16 range can be mapped for population units exceeding 50,000 total persons (or 20,000 to 30,000 per sex). The most frequent cancers can be reliably mapped in smaller total populations of about 5,000 to 10,000 . For rarer cancers with incidence rates between 5 and 10 per 100,000 per year, a minimum total population size to generate expected numbers in the range of 10 to 16 cases in five years exceeds 100,000 , or about 50,000 per sex).

Tables 2 and 3 also show the distribution of New Jersey county, municipality and census tract populations using the same ranges of population sizes. From the tables, it is
apparent that cancers with incidence rates below 10 per 100,000 can be reliably mapped at the county level in New Jersey, but that this is feasible for only a small number of municipalities. Of course, analysis of incidence data over longer time frames than fiveyear intervals would be more reliable. Five-year cancer data for the most common cancer type groupings (prostate, breast, lung and bronchus, and colon and rectum), may be reliably mapped for many municipalities and some census tracts in New Jersey. For the many important cancer type groupings with incidence rate between 10 and 50 per 100,000 per year, rates or relative rates may be reliably mapped at the county level, and for perhaps the 25 or 30 largest municipalities in the state.

## Conclusion

In New Jersey, most cancer types can be practically mapped for all counties and for a small number of municipalities. For most municipalities and some census tracts, reliable rates can only be mapped for the most common cancers.

## References

Bleyer A, O’Leary M, Barr R, Ries LAG (eds.) (2006). Cancer Epidemiology in Older Adolescents and Young Adults 15 to 29 years of Age, Including SEER Incidence and Survival: 1975-2000. Chapeter 2, Methods. National Cancer Institute, NIH Pub. No. 065767, Bethesda, MD.

NCI (2006). NCI Cancer Mortality Maps and Graphs: Dictionary. Available at: http://www3.cancer.gov/atlasplus/dictionary.html Accessed December 2006.

Neutra R, Swan S, Mack T (1992). Clusters galore: insights about environmental clusters from probability theory. Sci Total Environ 1992;127:187-200.

NYSDH (2006). About Age Adjusted Rates, 95\% Confidence Intervals and Unstable Rates. Available at: http://www.health.state.ny.us/statistics/cancer/registry/age.htm Accessed December 2006.

Pickle LW (2000). Mapping mortality data in the United States. In: Spatial Epidemiology: Methods and Applications, Elliott P et al., eds., Oxford University Press, New York.

Rudolph BA, Shah GH, Love D (2006). Small numbers, disclosure risk, security, and reliability issues in web-based data query systems. J Public Health Management Practice 2006;12(2):176-183.

Weinstein RB, Klotz JB (2000). Local cancer excesses and follow-up studies. J Registry Management 2000;27(3):98-100.

Figure 1. Cumulative probabilities of observed number of cases in a mapped unit, for expected numbers ranging from 1 to 20.


Figure 2. Range of possible rate ratios for expected numbers ranging from 1 to 20.


Table 1. Probability of a map unit having a rate ratio of greater than 1.5 or 2.0 by chance alone, assuming a Poisson distribution, for expected numbers ranging from 2 to 20 (only even numbers shown).

Rate Ratio > 1.5

| Expected <br> Number | Observed <br> Number for <br> RR = 1.5 | Observed <br> Numbers for <br> RR > 1.5 | Probability <br> of Observing <br> RR > 1.5 |
| :---: | :---: | :---: | :---: |
| 2 | 3 | $\geq 4$ | 0.14 |
| 4 | 6 | $\geq 7$ | 0.11 |
| 6 | 9 | $\geq 10$ | 0.084 |
| 8 | 12 | $\geq 13$ | 0.064 |
| 10 | 15 | $\geq 16$ | 0.049 |
| 12 | 18 | $\geq 19$ | 0.037 |
| 14 | 21 | $\geq 22$ | 0.029 |
| 16 | 24 | $\geq 25$ | 0.022 |
| 18 | 27 | $\geq 28$ | 0.017 |
| 20 | 30 | $\geq 31$ | 0.014 |

Rate Ratio > 2.0

| Expected <br> Number | Observed <br> Number for <br> RR $=\mathbf{2 . 0}$ | Observed <br> Numbers for <br> RR > 2.0 | Probability <br> of Observing <br> RR $>\mathbf{2 . 0}$ |
| :---: | :---: | :---: | :---: |
| 2 | 4 | $\geq 5$ | 0.053 |
| 4 | 8 | $\geq 9$ | 0.021 |
| 6 | 12 | $\geq 13$ | 0.0088 |
| 8 | 16 | $\geq 17$ | 0.0037 |
| 10 | 20 | $\geq 21$ | 0.0016 |
| 12 | 24 | $\geq 25$ | 0.0007 |
| 14 | 28 | $\geq 29$ | 0.0003 |
| 16 | 32 | $\geq 33$ | 0.0001 |
| 18 | 36 | $\geq 37$ | 0.0001 |
| 20 | 40 | $\geq 41$ | $<0.0001$ |

Table 2. Number of expected cases occurring in a 5-year observation period (rounded down to integer) by select population sizes and cancer incidence rates.

| Age-Standardized <br> Annual Rate Per 100,000 Population | Total or Sex-specific Population Estimate |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 5,000 | 10,000 | 20,000 | 30,000 | 40,000 | 50,000 | 60,000 | 70,000 | 80,000 | 90,000 | 100,000 |
| 50 | 12 | 25 | 50 | 75 | 100 | 125 | 150 | 175 | 200 | 225 | 250 |
| 45 | 11 | 22 | 45 | 67 | 90 | 112 | 135 | 157 | 180 | 202 | 225 |
| 40 | 10 | 20 | 40 | 60 | 80 | 100 | 120 | 140 | 160 | 180 | 200 |
| 35 | 8 | 17 | 35 | 52 | 70 | 87 | 105 | 122 | 140 | 157 | 175 |
| 30 | 7 | 15 | 30 | 45 | 60 | 75 | 90 | 105 | 120 | 135 | 150 |
| 25 | 6 | 12 | 25 | 37 | 50 | 62 | 75 | 87 | 100 | 112 | 125 |
| 20 | 5 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |
| 15 | 3 | 7 | 15 | 22 | 30 | 37 | 45 | 52 | 60 | 67 | 75 |
| 10 | 2 | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 |
| 5 | 1 | 2 | 5 | 7 | 10 | 12 | 15 | 17 | 20 | 22 | 25 |
| $\begin{aligned} & \text { NJ Municipalities } \\ & \text { (n=566) with } \\ & \text { Populations } \\ & \text { as Large or Larger: } \end{aligned}$ |  |  |  |  |  |  |  |  |  |  |  |
| Total | 374 | 239 | 120 | 69 | 46 | 29 | 22 | 12 | 9 | 6 | 4 |
| Male | 234 | 118 | 43 | 18 | 9 | 4 | 3 | 3 | 2 | 2 | 2 |
| Female | 240 | 125 | 52 | 23 | 11 | 4 | 4 | 3 | 2 | 2 | 2 |
| ```NJ Census Tracts ( \(\mathrm{n}=1,944\) ) with Populations as Large or Larger:``` |  |  |  |  |  |  |  |  |  |  |  |
| Total | 625 | 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Male | 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Female | 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Table 3. Number of expected cases in males or females occurring in a 5 -year observation period by cancer rate range and total population range.

| Age-Standardized Annual Rate Per 100,000 | Example Cancer Type By <br> New Jersey Rate, 2000 <br> By Sex, All Races Combined |  | Expected Number of Cases Based on Low Ends of Rate and Population Ranges |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | Total Population Size Range (Males plus Females) |  |  |  |  |  |  |  |  |  |
|  |  |  | $\begin{aligned} & \hline \text { Less } \\ & \text { than } \\ & \mathbf{1 0 0 0} \\ & \hline \end{aligned}$ | $\begin{aligned} & 1000 \text { to } \\ & 2500 \end{aligned}$ | $\begin{aligned} & \hline 2500 \text { to } \\ & 5000 \end{aligned}$ | $\begin{aligned} & \hline 5000 \text { to } \\ & 10000 \end{aligned}$ | $\begin{aligned} & \hline 10000 \\ & \text { to } \\ & 25000 \\ & \hline \end{aligned}$ | $\begin{aligned} & 25000 \\ & \text { to } \\ & 50000 \\ & \hline \end{aligned}$ | $\begin{aligned} & \hline 50000 \\ & \text { to } \\ & \mathbf{1 0 0 0 0 0} \\ & \hline \end{aligned}$ | $\begin{aligned} & \hline 100000 \\ & \text { to } \\ & 250000 \\ & \hline \end{aligned}$ | $\begin{aligned} & 250000 \\ & \text { to } \\ & 500000 \\ & \hline \end{aligned}$ | 500000 to 1 million |
| Greater than 100 | Prostate | Breast |  | 2 | 6 | 12 | 25 | 62 | 125 | 250 | 625 | 1,250 |
| 50 to 100 | Lung and bronchus, colon and rectum | Lung and bronchus, colon and rectum |  |  | 3 | 6 | 12 | 31 | 62 | 125 | 312 | 625 |
| 20 to 50 | Urinary bladder, nonHodgkin lymphoma | Corpus uteri |  |  |  | 2 | 5 | 12 | 25 | 50 | 125 | 250 |
| 10 to 20 | Melanoma of the skin, kidney, leukemias, pancreas, stomach | Ovary, non-Hodgkin lymphoma, thyroid, urinary bladder, melanoma of the skin, pancreas |  |  |  |  | 2 | 6 | 12 | 25 | 62 | 125 |
| 5 to 10 | Esophagus, brain and other nervous system, larynx, liver, myelomas, testis, thyroid | Cervix uteri, leukemias, kidney, stomach, brain and other nervous system, myelomas |  |  |  |  |  | 3 | 6 | 12 | 31 | 62 |
| 2 to 5 | Soft tissue (including heart), Hodgkin lymphoma | Hodgkin lymphoma, soft tissue (including heart), esophagus, liver |  |  |  |  |  |  | 2 | 5 | 12 | 25 |
| 1 to 2 | Nasopharynx, bones and joints | Larynx |  |  |  |  |  |  |  | 2 | 6 | 12 |
| \# Census tracts (of 1,944) |  |  | 48 | 266 | 1,004 | 610 | 16 |  |  |  |  |  |
| \# Municipalities (of 566) |  |  | 27 | 72 | 93 | 135 | 150 | 60 | 25 | 3 | 1 |  |
| \# Counties (of 21) |  |  |  |  |  |  |  |  | 1 | 5 | 7 | 8 |

Appendix Table 1. Cumulative Poisson probabilities of observed number or less, given expected numbers of cases between 1 and 20. Cells highlighted in pink are cumulative probabilities for $R \mathrm{R} \leq 1.25$, green are cumulative probabilities for $\mathrm{RR} \leq 1.5$, and cells highlighted in yellow are cumulative probabilities for $R \mathrm{R} \leq 2.0$.

|  | Expected |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| OBS | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 14 | 16 | 18 | 20 |
| 0 | 0.3679 | 0.1353 | 0.0498 | 0.0183 | 0.0067 | 0.0025 | 0.0009 | 0.0003 | 0.0001 | 0.0000 | 0.0000 |  |  |  |  |
| 1 | 0.7358 | 0.4060 | 0.1991 | 0.0916 | 0.0404 | 0.0174 | 0.0073 | 0.0030 | 0.0012 | 0.0005 | 0.0001 | 0.0000 |  |  |  |
| 2 | 0.9197 | 0.6767 | 0.4232 | 0.2381 | 0.1247 | 0.0620 | 0.0296 | 0.0138 | 0.0062 | 0.0028 | 0.0005 | 0.0001 | 0.0000 |  |  |
| 3 | 0.9810 | 0.8571 | 0.6472 | 0.4335 | 0.2650 | 0.1512 | 0.0818 | 0.0424 | 0.0212 | 0.0103 | 0.0023 | 0.0005 | 0.0001 | 0.0000 |  |
| 4 | 0.9963 | 0.9473 | 0.8153 | 0.6289 | 0.4405 | 0.2851 | 0.1730 | 0.0996 | 0.0550 | 0.0293 | 0.0076 | 0.0018 | 0.0004 | 0.0001 | 0.0000 |
| 5 | 0.9994 | 0.9834 | 0.9161 | 0.7852 | 0.6160 | 0.4457 | 0.3007 | 0.1912 | 0.1157 | 0.0671 | 0.0203 | 0.0055 | 0.0014 | 0.0003 | 0.0001 |
| 6 | 0.9999 | 0.9955 | 0.9665 | 0.8894 | 0.7622 | 0.6063 | 0.4497 | 0.3134 | 0.2068 | 0.1301 | 0.0458 | 0.0142 | 0.0040 | 0.0010 | 0.0003 |
| 7 | 1.0000 | 0.9989 | 0.9881 | 0.9489 | 0.8666 | 0.7440 | 0.5987 | 0.4530 | 0.3239 | 0.2202 | 0.0895 | 0.0316 | 0.0100 | 0.0029 | 0.0008 |
| 8 |  | 0.9998 | 0.9962 | 0.9787 | 0.9319 | 0.8472 | 0.7291 | 0.5925 | 0.4557 | 0.3328 | 0.1550 | 0.0621 | 0.0220 | 0.0071 | 0.0021 |
| 9 |  | 1.0000 | 0.9989 | 0.9919 | 0.9682 | 0.9161 | 0.8305 | 0.7166 | 0.5874 | 0.4579 | 0.2424 | 0.1094 | 0.0433 | 0.0154 | 0.0050 |
| 10 |  |  | 0.9997 | 0.9972 | 0.9863 | 0.9574 | 0.9015 | 0.8159 | 0.7060 | 0.5830 | 0.3472 | 0.1757 | 0.0774 | 0.0304 | 0.0108 |
| 11 |  |  | 0.9999 | 0.9991 | 0.9945 | 0.9799 | 0.9467 | 0.8881 | 0.8030 | 0.6968 | 0.4616 | 0.2600 | 0.1270 | 0.0549 | 0.0214 |
| 12 |  |  | 1.0000 | 0.9997 | 0.9980 | 0.9912 | 0.9730 | 0.9362 | 0.8758 | 0.7916 | 0.5760 | 0.3585 | 0.1931 | 0.0917 | 0.0390 |
| 13 |  |  |  | 0.9999 | 0.9993 | 0.9964 | 0.9872 | 0.9658 | 0.9261 | 0.8645 | 0.6815 | 0.4644 | 0.2745 | 0.1426 | 0.0661 |
| 14 |  |  |  | 1.0000 | 0.9998 | 0.9986 | 0.9943 | 0.9827 | 0.9585 | 0.9165 | 0.7720 | 0.5704 | 0.3675 | 0.2081 | 0.1049 |
| 15 |  |  |  |  | 0.9999 | 0.9995 | 0.9976 | 0.9918 | 0.9780 | 0.9513 | 0.8444 | 0.6694 | 0.4667 | 0.2867 | 0.1565 |
| 16 |  |  |  |  | 1.0000 | 0.9998 | 0.9990 | 0.9963 | 0.9889 | 0.9730 | 0.8987 | 0.7559 | 0.5660 | 0.3751 | 0.2211 |
| 17 |  |  |  |  |  | 0.9999 | 0.9996 | 0.9984 | 0.9947 | 0.9857 | 0.9370 | 0.8272 | 0.6593 | 0.4686 | 0.2970 |
| 18 |  |  |  |  |  | 1.0000 | 0.9999 | 0.9993 | 0.9976 | 0.9928 | 0.9626 | 0.8826 | 0.7423 | 0.5622 | 0.3814 |
| 19 |  |  |  |  |  |  | 1.0000 | 0.9997 | 0.9989 | 0.9965 | 0.9787 | 0.9235 | 0.8122 | 0.6509 | 0.4703 |
| 20 |  |  |  |  |  |  |  | 0.9999 | 0.9996 | 0.9984 | 0.9884 | 0.9521 | 0.8682 | 0.7307 | 0.5591 |
| 21 |  |  |  |  |  |  |  | 1.0000 | 0.9998 | 0.9993 | 0.9939 | 0.9712 | 0.9108 | 0.7991 | 0.6437 |
| 22 |  |  |  |  |  |  |  |  | 0.9999 | 0.9997 | 0.9970 | 0.9833 | 0.9418 | 0.8551 | 0.7206 |
| 23 |  |  |  |  |  |  |  |  | 1.0000 | 0.9999 | 0.9985 | 0.9907 | 0.9633 | 0.8989 | 0.7875 |
| 24 |  |  |  |  |  |  |  |  |  | 1.0000 | 0.9993 | 0.9950 | 0.9777 | 0.9317 | 0.8432 |
| 25 |  |  |  |  |  |  |  |  |  |  | 0.9997 | 0.9974 | 0.9869 | 0.9554 | 0.8878 |
| 26 |  |  |  |  |  |  |  |  |  |  | 0.9999 | 0.9987 | 0.9925 | 0.9718 | 0.9221 |
| 27 |  |  |  |  |  |  |  |  |  |  | 0.9999 | 0.9994 | 0.9959 | 0.9827 | 0.9475 |
| 28 |  |  |  |  |  |  |  |  |  |  | 1.0000 | 0.9997 | 0.9978 | 0.9897 | 0.9657 |
| 29 |  |  |  |  |  |  |  |  |  |  |  | 0.9999 | 0.9989 | 0.9941 | 0.9782 |
| 30 |  |  |  |  |  |  |  |  |  |  |  | 0.9999 | 0.9994 | 0.9967 | 0.9865 |
| 31 |  |  |  |  |  |  |  |  |  |  |  | 1.0000 | 0.9997 | 0.9982 | 0.9919 |
| 32 |  |  |  |  |  |  |  |  |  |  |  |  | 0.9999 | 0.9990 | 0.9953 |
| 33 |  |  |  |  |  |  |  |  |  |  |  |  | 0.9999 | 0.9995 | 0.9973 |
| 34 |  |  |  |  |  |  |  |  |  |  |  |  | 1.0000 | 0.9998 | 0.9985 |
| 35 |  |  |  |  |  |  |  |  |  |  |  |  |  | 0.9999 | 0.9992 |
| 36 |  |  |  |  |  |  |  |  |  |  |  |  |  | 0.9999 | 0.9996 |
| 37 |  |  |  |  |  |  |  |  |  |  |  |  |  | 1.0000 | 0.9998 |
| 38 |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 0.9999 |
| 39 |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 0.9999 |
| 40 |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 1.0000 |

Appendix Table 2. Ratios (observed/expected) for expected numbers of cases between 1 and 20.

| Observed | 1 | 2 | 3 | 4 | 5 | Expected |  |  |  | 10 | 12 | 14 | 16 | 18 | 20 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  | 6 | 7 | 8 | 9 |  |  |  |  |  |  |
| 0 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1 | 1.00 | 0.50 | 0.33 | 0.25 | 0.20 | 0.17 | 0.14 | 0.13 | 0.11 | 0.10 | 0.08 | 0.07 |  |  |  |
| 2 | 2.00 | 1.00 | 0.67 | 0.50 | 0.40 | 0.33 | 0.29 | 0.25 | 0.22 | 0.20 | 0.17 | 0.14 | 0.13 |  |  |
| 3 | 3.00 | 1.50 | 1.00 | 0.75 | 0.60 | 0.50 | 0.43 | 0.38 | 0.33 | 0.30 | 0.25 | 0.21 | 0.19 | 0.17 |  |
| 4 | 4.00 | 2.00 | 1.33 | 1.00 | 0.80 | 0.67 | 0.57 | 0.50 | 0.44 | 0.40 | 0.33 | 0.29 | 0.25 | 0.22 | 0.20 |
| 5 | 5.00 | 2.50 | 1.67 | 1.25 | 1.00 | 0.83 | 0.71 | 0.63 | 0.56 | 0.50 | 0.42 | 0.36 | 0.31 | 0.28 | 0.25 |
| 6 | 6.00 | 3.00 | 2.00 | 1.50 | 1.20 | 1.00 | 0.86 | 0.75 | 0.67 | 0.60 | 0.50 | 0.43 | 0.38 | 0.33 | 0.30 |
| 7 | 7.00 | 3.50 | 2.33 | 1.75 | 1.40 | 1.17 | 1.00 | 0.88 | 0.78 | 0.70 | 0.58 | 0.50 | 0.44 | 0.39 | 0.35 |
| 8 |  | 4.00 | 2.67 | 2.00 | 1.60 | 1.33 | 1.14 | 1.00 | 0.89 | 0.80 | 0.67 | 0.57 | 0.50 | 0.44 | 0.40 |
| 9 |  | 4.50 | 3.00 | 2.25 | 1.80 | 1.50 | 1.29 | 1.13 | 1.00 | 0.90 | 0.75 | 0.64 | 0.56 | 0.50 | 0.45 |
| 10 |  |  | 3.33 | 2.50 | 2.00 | 1.67 | 1.43 | 1.25 | 1.11 | 1.00 | 0.83 | 0.71 | 0.63 | 0.56 | 0.50 |
| 11 |  |  | 3.67 | 2.75 | 2.20 | 1.83 | 1.57 | 1.38 | 1.22 | 1.10 | 0.92 | 0.79 | 0.69 | 0.61 | 0.55 |
| 12 |  |  | 4.00 | 3.00 | 2.40 | 2.00 | 1.71 | 1.50 | 1.33 | 1.20 | 1.00 | 0.86 | 0.75 | 0.67 | 0.60 |
| 13 |  |  |  | 3.25 | 2.60 | 2.17 | 1.86 | 1.63 | 1.44 | 1.30 | 1.08 | 0.93 | 0.81 | 0.72 | 0.65 |
| 14 |  |  |  | 3.50 | 2.80 | 2.33 | 2.00 | 1.75 | 1.56 | 1.40 | 1.17 | 1.00 | 0.88 | 0.78 | 0.70 |
| 15 |  |  |  |  | 3.00 | 2.50 | 2.14 | 1.88 | 1.67 | 1.50 | 1.25 | 1.07 | 0.94 | 0.83 | 0.75 |
| 16 |  |  |  |  |  | 2.67 | 2.29 | 2.00 | 1.78 | 1.60 | 1.33 | 1.14 | 1.00 | 0.89 | 0.80 |
| 17 |  |  |  |  |  | 2.83 | 2.43 | 2.13 | 1.89 | 1.70 | 1.42 | 1.21 | 1.06 | 0.94 | 0.85 |
| 18 |  |  |  |  |  |  | 2.57 | 2.25 | 2.00 | 1.80 | 1.50 | 1.29 | 1.13 | 1.00 | 0.90 |
| 19 |  |  |  |  |  |  | 2.71 | 2.38 | 2.11 | 1.90 | 1.58 | 1.36 | 1.19 | 1.06 | 0.95 |
| 20 |  |  |  |  |  |  |  | 2.50 | 2.22 | 2.00 | 1.67 | 1.43 | 1.25 | 1.11 | 1.00 |
| 21 |  |  |  |  |  |  |  |  | 2.33 | 2.10 | 1.75 | 1.50 | 1.31 | 1.17 | 1.05 |
| 22 |  |  |  |  |  |  |  |  | 2.44 | 2.20 | 1.83 | 1.57 | 1.38 | 1.22 | 1.10 |
| 23 |  |  |  |  |  |  |  |  |  | 2.30 | 1.92 | 1.64 | 1.44 | 1.28 | 1.15 |
| 24 |  |  |  |  |  |  |  |  |  |  | 2.00 | 1.71 | 1.50 | 1.33 | 1.20 |
| 25 |  |  |  |  |  |  |  |  |  |  | 2.08 | 1.79 | 1.56 | 1.39 | 1.25 |
| 26 |  |  |  |  |  |  |  |  |  |  | 2.17 | 1.86 | 1.63 | 1.44 | 1.30 |
| 27 |  |  |  |  |  |  |  |  |  |  | 2.25 | 1.93 | 1.69 | 1.50 | 1.35 |
| 28 |  |  |  |  |  |  |  |  |  |  |  | 2.00 | 1.75 | 1.56 | 1.40 |
| 29 |  |  |  |  |  |  |  |  |  |  |  | 2.07 | 1.81 | 1.61 | 1.45 |
| 30 |  |  |  |  |  |  |  |  |  |  |  | 2.14 | 1.88 | 1.67 | 1.50 |
| 31 |  |  |  |  |  |  |  |  |  |  |  |  | 1.94 | 1.72 | 1.55 |
| 32 |  |  |  |  |  |  |  |  |  |  |  |  | 2.00 | 1.78 | 1.60 |
| 33 |  |  |  |  |  |  |  |  |  |  |  |  |  | 1.83 | 1.65 |
| 34 |  |  |  |  |  |  |  |  |  |  |  |  |  | 1.89 | 1.70 |
| 35 |  |  |  |  |  |  |  |  |  |  |  |  |  | 1.94 | 1.75 |
| 36 |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 1.80 |
| 37 |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 1.85 |
| 38 |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 1.90 |
| 39 |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 1.95 |
| 40 |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 2.00 |

