Analytical Studies: Methods and References

Zeno: A Tool for Calculating Confidence Intervals of Rates in Health

by Philippe Finès and Gisèle Carriere

Release date: January 19, 2017

Statistics Canada   Statistique Canada

Canada

Zeno: A Tool for Calculating Confidence Intervals of Rates in Health

by

Philippe Finès and **Gisèle Carriere**
Health Analysis Division
Statistics Canada

Analytical Studies: Methods and References

Papers in this series provide background discussions of the methods used to develop data for economic, health, and social analytical studies at Statistics Canada. They are intended to provide readers with information on the statistical methods, standards and definitions used to develop databases for research purposes. All papers in this series have undergone peer and institutional review to ensure that they conform to Statistics Canada's mandate and adhere to generally accepted standards of good professional practice.

The papers can be downloaded free at www.statcan.gc.ca.

Table of contents

Abstract

Hospitalization rates are among commonly reported statistics related to health-care service use. The variety of methods for calculating confidence intervals for these and other health-related rates suggests a need to classify, compare and evaluate these methods. Zeno is a tool developed to calculate confidence intervals of rates based on several formulas available in the literature. This report describes the contents of the main sheet of the Zeno Tool and indicates which formulas are appropriate, based on users' assumptions and scope of analysis.

Keywords: Confidence intervals, health, hospitalization, rates

1   Introduction

Hospitalization rates are among commonly reported statistics related to health-care service use. A recurrent problem encountered by health analysts is the choice of a method to compute confidence intervals (CIs) for these rates.

When hospital events are independent of one another (only one event is allowed per person; for example, a discharge) and when the rate is neither extremely low nor high, the CI can be computed using an approximation based on the normal distribution. However, circumstances different from these conditions often exist. For example, unlike vital events such as births or deaths, for a given person, repeat events may occur, even if they have a low probability of occurrence. These events are not independent, a situation that has implications for the validity of using approximations based on the normal distribution to calculate CIs. Also, an examination of an individual's hospital visits might consider all visits, or only the first visit for a specific disease. Depending on the scope of analysis, different assumptions are made, and therefore, different methods are needed to calculate CIs.

When only one event per person is examined and the probability of occurrence of the event is low, exact calculations based on specific distributions are recommended. Frequently, a Binomial or Poisson distribution is assumed. Other techniques have been proposed (Glynn et al. 1993; Carriere and Roos 1997; Fay and Feuer 1997; Kegler 2007), which assume specific distributions such as Gamma or Chi-square. A second issue, in addition to rare events, is that analyses of recurrent events (such as hospitalizations) as opposed to non-recurrent events (such as death) violate the Poisson assumption of independence (Carriere and Roos 1997).

The literature on CIs for rates that measure these different circumstances has expanded rapidly and can be daunting for researchers who must decide which formula to use in a specific context. However, few comparative studies of methods of calculating CIs for rates have been conducted. Typically, authors report their work (with their idiosyncratic notations) with little discussion of compatibility and comparability with the work of other researchers.

The objective of a previous version of this article was to catalogue all the methods available to compute CIs for rates. The challenges were numerous and included the reconciliation of notations between authors and the need to ensure that all pertinent literature had been taken into account. The task rapidly became unwieldy. Some papers described exact methods, while others presented approximations; some described how to compare rates (with rate differences or rate ratios); some focused on one event per person, whereas others allowed for multiple occurrences; some described formulas for crude rates, while others focused on standardized rates, etc. A systematic review of published formulas was needed. It became apparent that regrouping a family of formulas related to the general problem of calculating CIs for rates into a single tool would be useful.

Therefore, the original objective shifted from the development of a *catalogue* of formulas to the development of a *tool* that enables researchers to view the effects of applying one or another of several different methods. It addresses the calculation of CIs for rates, but omits formulas for hypothesis testing on rates (such as comparison between groups, or assessment of over-dispersion or zero-inflated distributions [van den Broek 1995]). This tool (worksheet), Zeno, is now available upon request. Since rates and number of events are related, CIs for both metrics are available. Although users have access to several formulas for CIs, the most appropriate formula is not necessarily the one that yields the narrowest CI. On the contrary, the most appropriate formula is the one that satisfies the conditions of assumptions used to describe event distribution; calculations are derived based on these assumptions.

The objective of this article is to describe the Zeno Tool. The next three sections present: the references for the original formulas (Section 2); notations used in the Tool and in this article

(Section 3); and the contents of the main sheet of the Tool (Section 4). The Data and results section (Section 5) contains a pivot table extracted from the Tool. The Conclusion (Section 6) summarizes the description and suggests potential enhancements.

2   Literature review

Although a systematic review of all papers on this domain was not conducted, the references used are up-to-date and pertinent for the purpose of the Tool. The corollary is that, as mentioned before, notation varies in all references. To appreciate the intricacies of a given method and compare it with another, readers must mentally translate the notation used by one author into the terminology used by the other. For example, the concept of weight has surprisingly diverse definitions. The assumptions of the formulas selected for inclusion in the Tool are listed in Table 1. When the proportion of recurrent cases is relatively small, all the measures identified as "One event per person" can be used for the "All events" analyses.

Table 1

Formulas used in Zeno Tool

Formula	Appropriate for one event per person or for all events?	Initially devised for rates or for numbers of events?	References (equations)
Based on Poisson distribution			
Exact formula	One event per person	Numbers of events	(5) in Fay and Feuer (1997)[1] (also [14] [15], [20], [21] in Daly [1992][2])
Normal approximation	One event per person	Numbers of events	(7) in Daly (1992)
Lognormal transformation	One event per person	Rates	(1a) in Kegler (2007)[3]
Based on Binomial distribution			
Exact formula	One event per person	Rates	(4) and (5) in Daly (1992)
Normal approximation	One event per person	Rates	(3) in Daly (1992)
Normal approximation for small proportions	One event per person	Rates	(1.26) and (1.27) in Fleiss (1981)[4]
For analysis of all events			
Compound Poisson distribution	All events	Rates	(1b) in Kegler (2007)
Negative Binomial assumption	All events	Rates	Glynn et al. (1993, p. 780)[5]

1. M.P. Fay and E.J. Feuer, 1997, "Confidence intervals for directly standardized rates: A method based on the gamma distribution."
2. L. Daly, 1992, "Simple SAS macros for the calculation of exact binomial and Poisson confidence limits."
3. S.R. Kegler, 2007, "Applying the compound Poisson process model to the reporting of injury-related mortality rates."
4. J.L. Fleiss, 1981, *Statistical Methods for Rates and Proportions, 2nd Edition.*
5. R.J. Glynn, T.A. Stukel, S.M. Sharp, T.A. Bubolz, J.A. Freeman, and E.S. Ficher, 1993, "Estimating the Variance of Standardized Rates of Recurrent Events, with Application to Hospitalizations Among the Elderly in New England."

3   Notations and formulas

The main concepts and their symbols are listed in Table 2. When all events are considered, rates are denoted by r, and numbers of events are denoted by $\#E$. For age-standardized outcomes, the symbols become ASR and $AS\#E$. Within an age-stratum i, suffix $\_i$ is appended to the symbol. When only one event per person is considered (for example, the first event), the same notations are bracketed by "[" and "]", so that r becomes $[r]$, ASR becomes $[ASR]$, etc. In addition, other symbols are needed: $\#N(\#N\_i)$ is the total (stratum-specific) size of the population under study; and $\#NR(\#NR\_i)$ is the total (stratum-specific) size of the reference population, from which weights $w\_i = \#NR\_i / \#NR$ are computed.

Table 2
Symbols used for concepts in Zeno Tool

Concept	Sizes and weights	Number of events	Rates
Size of population			
By age stratum	#N_i	…	…
Total over age strata	#N	…	…
Size of reference population			
By age stratum	#NR_i	…	…
Total over age strata	#NR	…	…
Weights used for age standardization			
In each stratum	w_i = #NR_i /#NR	…	…
Scope of events examined — method			
All events			
Crude outcomes (by age stratum *i*)	…	#E_i	r_i
Crude outcomes (for all age strata combined)	…	#E	r
Age-standardized outcomes	…	AS#E	ASR
One event per person only			
Crude outcomes (by age stratum *i*)	…	[#E_i]	[r_i]
Crude outcomes (by age strata combined)	…	[#E]	[r]
Age-standardized outcomes	…	[AS#E]	[ASR]

… not applicable

In cases where figures represent both single events for some persons and recurrent, non-independent events for others, a more complicated notation is needed. The notation must be able to describe the contents of cells that contain values that encode the number of persons, and also, the number of events experienced per given person. Using notation from Kegler (2007), $C\_k = h$ is the number of persons for whom the number of events was equal to h. In Kegler (2007), h ranges from 1 to 2; in the Tool presented here, h ranges from 1 to 14. Thus, $\#\{n: n \text{ in stratum } i \text{ and } C\_k(n) = j\}$ reads "the size $[\#]$ of the set $[\{ .. \}]$ made up of all the persons n who $[n:]$ belong to stratum i and for whom the number of events is equal to j $[C\_k(n) = j]$." In other words, the formula expresses the number of persons in stratum i who had j events. For consistency with the other notations in Table 2, $\#E\_j\_i$ denotes $\#\{n: n \text{ in stratum } i \text{ and } C\_k(n) = j\}$; by extension, $\#E\_0\_i$ denotes the number of persons in stratum i with no event, and $\#E\_0$ denotes the number of persons overall with no event.

Results can be defined according to the intended metrics, namely, the rates or the number of events based on the population (specific age strata or totals) and the scope (all events or only one event per person). Rates and number of events are related according to the general formula:

$$Rate = Number\ of\ events\ /\ Size\ of\ population$$

which, using the notations of Table 2, gives, for example:

$$r = \#E / \#N$$

and

$$ASR = AS\#E / \#N.$$

Throughout, population sizes are considered fixed, as this is standard for several authors. Therefore, the rate or number of events can be calculated when the other metric is known. In fact, some of the formulas implemented in the Tool were initially introduced in referenced material to calculate the CIs for rates only (or number of events). The formulas corresponding to the other metric were then developed.

Although several of the referenced formulas focus on a specific calculation, it is possible to expand the formulas to apply to other cases. For example, from references that present the formulas for multiple visits, simplification to one visit per person is straightforward—only minor modifications of the original formulas are needed.

Also, an age-standardized rate is a weighted sum of the age stratum-specific rates.

$$ASR = \sum i\ w\_i * r\_i$$

(where the weights w_i are equal to $\#NR\_i / \#NR$). If in this formula, instead of weights w_i, "pseudo-weights" w'_i (each equal to $\#N\_i / \#N$), are used, the result is:

$$\sum i\ w'\_i * r\_i = \sum i\ \#N\_i / \#N\ *\ \#E\_i / \#N\_i = \sum i\ \#E\_i\ / \#N = \#E / \#N\ = r.$$

This shows that r (crude rate) can be expressed as a weighted sum of stratum-specific rates; in other words, r is the age-standardized rate obtained when using "pseudo-weights." This property has been used to convert the formula for CI of ASR into a formula for CI of r when the latter was not directly provided in the references.

To summarize, the Tool contains a complete series of formulas for the six metrics in the last two columns of Table 2, using the different assumptions in Table 1.

4   Contents of main sheet

The Zeno Tool is essentially a sheet ("main sheet") that contains the data and the results. In fact, there may be as many "main sheets" as desired; the label of the sheet being analyzed must be indicated in sheet "Prep," which reformats the results to produce the appropriate pivot charts and tables.

Cells of a main sheet are denoted by concatenation of column (letter) and row (number). The Excel sheet allows for 21 strata, for which data will be entered by the user on rows $i* = 11$ to 31.

As mentioned earlier, stratum-specific symbols are denoted by concatenation of symbol, " $\_$ " and row $i = 1$ to 21. Rows and strata are linked by this relation: $i* = i + 10$. The essential components of any main sheet are described in Tables 3-1 to 3-4, where the identifier of the cell is described by what follows the "=" sign. For example: cell D11 contains the number of events for stratum 1; r_1 is the rate for stratum 1. The only cells to be modified by users are:

- C1 and D1 to T1 (keyword and title)
- C2: unit of reference (denominator) for rates
- C3: Alpha level of CIs (for example, 0.05)
- area D11 to T31: respectively, for each age stratum $i = 1$ to 21:
  $\#E\_i$, $\#N\_i$, $\#NR\_i$, $\#E\_j\_i$ for $j = 1$ up to 14 (because up to 14 events per person are allowed in the Tool).

Table 3-1

Essential components of the Zeno Tool main Excel sheet, by row and column, with color codes — Part I

Rows	Columns A-C	Column D	Column E	Column F	Columns G-T	Column V	Columns W-Z	Columns AI-AL	Column AM	Column AW
1..10	Labels	Labels	Labels	Labels	Labels	Labels	Labels	Labels	Labels	Labels
$i*=11..31$	Labels	$Di*= \#E_i$	$Ei*= \#N_i$	$Fi*= \#NR_i$	$Gi*= \#E\_1\_i$, up to $Ti*=\#E\_14\_i$ (from Kegler [2007])	$Vi*=r_i$	$Wi*=w_i$	$ALi*=Gi + 2*Hi + 3*Ii + \ldots + 14*Ti$ (from Kegler [2007])	$AMi*= Gi + 4*Hi + 9*Ii + \ldots + 196*Ti$ (from Kegler [2007])	$AWi*= \ln(r_i)$
33	Labels	$D33=\#E$	$E33=\#N$	$F33=\#NR$	$G33=\Sigma$ over i of $\#E\_1\_i$, up to $T33=\Sigma$ over i of $\#E\_14\_i$	Empty	$W33=\Sigma$ over i of w_i $=1$, $Y33=ASR$	$AI33=[\#E]$, $AK33=[ASR]$	Empty	Empty
44	Labels	Empty	Empty	Empty	$G44=G33*1$, up to $T44=T33*14$	$V44=r$	$Y44=AS\#E$	$AJ44=[r]$, $AK44=[AS\#E]$	Empty	$AW44= \ln(r)$
45	Labels	Empty	Empty	Empty	$G45=G33*1^2$, up to $T45=G33*14^2$	Empty	$Z45=s^2$ (from Glynn et al. [1993])	Empty	Empty	Empty
47	Labels	Empty	Empty	Empty	Empty	Empty	$Z47=k$-hat (from Glynn et al. [1993])	Empty	Empty	$AW47= \ln(ASR)$
50	Labels	Empty	Empty	Empty	Empty	Empty	Empty	Empty	Empty	$AW50= \ln([R])$
53	Labels	Empty	Empty	Empty	Empty	Empty	Empty	Empty	Empty	$AW53= \ln([ASR])$

The notes are located at the bottom of Table 3-4.

Table 3-2

Essential components of the Zeno Tool main Excel sheet, by row and column, with color codes — Part 2

Rows	Columns AX-BA	Columns BB-BE	Columns BF-BI	Columns BM-BP	Columns BQ-BT	Columns BU-BX
1..10	Labels	Labels	Labels	Labels	Labels	Labels
i*=11..31	AXi*-BAi*=CI limits, CI width, CV for r_i using Poisson distribution, exact formula	BBi*-BEi*=CI limits, CI width, CV for r_i using Poisson distribution, normal approximation	BFi*-BLi*=estimate of r_i, CI limits of ln(r_i), CI limits, CI width, CV for r_i using Poisson distribution, Lognormal transformation formula	BMi*-BPi*=CI limits, CI width, CV for r_i using Binomial distribution	BQi*-BTi*=CI limits, CI width, CV for r_i using Binomial distribution, normal approximation	BUi*-BXi*=CI limits, CI width, CV for r_i using Binomial distribution, normal approximation
33	Empty	Empty	Empty	Empty	Empty	Empty
44	AX44-BA44=CI limits, CI width, CV for r using Poisson distribution, exact formula	BB44-BE44=CI limits, CI width, CV for r using Poisson distribution, normal approximation	BF44-BL44=estimate of r, CI limits of ln(r), CI limits, CI width, CV for r using Poisson distribution, Lognormal transformation formula	BM44-BP44=CI limits, CI width, CV for r using Binomial distribution	BQ44-BT44=CI limits, CI width, CV for r using Binomial distribution, normal approximation	BU44-BX44=CI limits, CI width, CV for r using Binomial distribution, normal approximation
45	Empty	Empty	Empty	Empty	Empty	Empty
47	AX47-BA47=CI limits, CI width, CV for ASR using Poisson distribution, exact formula	BB47-BE47=CI limits, CI width, CV for ASR using Poisson distribution, normal approximation	BF47-BL47=estimate of ASR, CI limits of ln(ASR), CI limits, CI width, CV for ASR using Poisson distribution, Lognormal transformation formula	BM47-BP47=CI limits, CI width, CV for ASR using Binomial distribution	BQ47-BT47=CI limits, CI width, CV for ASR using Binomial distribution, normal approximation	BU47-BX47=CI limits, CI width, CV for ASR using Binomial distribution, normal approximation
50	AX50-BA50=CI limits, CI width, CV for [r] using Poisson distribution, exact formula	BB50-BE50=CI limits, CI width, CV for [r] using Poisson distribution, normal approximation	BF50-BL50=estimate of [r], CI limits of ln([r]), CI limits, CI width, CV for [r] using Poisson distribution, Lognormal transformation formula	BM50-BP50=CI limits, CI width, CV for [r] using Binomial distribution	BQ50-BT50=CI limits, CI width, CV for [r] using Binomial distribution, normal approximation	BU50-BX50=CI limits, CI width, CV for [r] using Binomial distribution, normal approximation
53	AX53-BA53=CI limits, CI width, CV for [ASR] using Poisson distribution, exact formula	BB53-BE53=CI limits, CI width, CV for [ASR] using Poisson distribution, normal approximation	BF53-BL53=estimate of [ASR], CI limits of ln([ASR]), CI limits, CI width, CV for [ASR] using Poisson distribution, Lognormal transformation formula	BM53-BP53=CI limits, CI width, CV for [ASR] using Binomial distribution	BQ53-BT53=CI limits, CI width, CV for [ASR] using Binomial distribution, normal approximation	BU53-BX53=CI limits, CI width, CV for [ASR] using Binomial distribution, normal approximation

The notes are located at the bottom of Table 3-4.

Table 3-3

Essential components of the Zeno Tool main Excel sheet, by row and column, with color codes — Part 3

Rows	Columns BY-CF	Columns CG-CJ	Columns CM-CO	Columns CP-CR	Columns CS-CU	Columns CV-CX
1..10	Labels	Labels	Labels	Labels	Labels	Labels
i*=11..31	BYi*-CFi*=Calculations related to Compound Poisson distribution, CI limits, CI width, CV for r_i using Compound Poisson distribution	CGi*-CJi*=CI limits, CI width, CV for r_i using Negative Binomial assumption	CMi*-COi*=CI limits, CI width for #E_i using Poisson distribution, exact formula	CPi*-CRi*=CI limits, CI width for #E_i using Poisson distribution, normal approximation	CSi*-CUi*=CI limits, CI width for #E_i using Poisson distribution, Lognormal transformation formula	CVi*-CXi*=CI limits, CI width for #E_i using Binomial distribution
33	Empty	Empty	Empty	Empty	Empty	Empty
44	BY44-CF44=Calculations related to Compound Poisson distribution, CI limits, CI width, CV for r using Compound Poisson distribution	CG44-CJ44=CI limits, CI width, CV for r using Negative Binomial assumption	CM44-CO44=CI limits, CI width for #E using Poisson distribution, exact formula	CP44-CR44=CI limits, CI width for #E using Poisson distribution, normal approximation	CS44-CU44=CI limits, CI width for #E using Poisson distribution, Lognormal transformation formula	CV44-CX44=CI limits, CI width for #E using Binomial distribution
45	Empty	Empty	Empty	Empty	Empty	Empty
47	BY47-CF47=Calculations related to Compound Poisson distribution, CI limits, CI width, CV for ASR using Compound Poisson distribution	CG47-CJ47=CI limits, CI width, CV for ASR using Negative Binomial assumption	CM47-CO47=CI limits, CI width for AS#E using Poisson distribution, exact formula	CP47-CR47=CI limits, CI width for AS#E using Poisson distribution, normal approximation	CS47-CU47=CI limits, CI width for AS#E using Poisson distribution, Lognormal transformation formula	CV47-CX47=CI limits, CI width for AS#E using Binomial distribution
50	BY50-CF50=Calculations related to Compound Poisson distribution, CI limits, CI width, CV for [r] using Compound Poisson distribution	Empty	CM50-CO50=CI limits, CI width for [#E] using Poisson distribution, exact formula	CP50-CR50=CI limits, CI width for [#E] using Poisson distribution, normal approximation	CS50-CU50=CI limits, CI width for [#E] using Poisson distribution, Lognormal transformation formula	CV50-CX50=CI limits, CI width for [#E] using Binomial distribution
53	BY53-CF53=Calculations related to Compound Poisson distribution, CI limits, CI width, CV for [ASR] using Compound Poisson distribution	Empty	CM53-CO53=CI limits, CI width for [AS#E] using Poisson distribution, exact formula	CP53-CR53=CI limits, CI width for [AS#E] using Poisson distribution, normal approximation	CS53-CU53=CI limits, CI width for [AS#E] using Poisson distribution, Lognormal transformation formula	CV53-CX53=CI limits, CI width for [AS#E] using Binomial distribution

The notes are located at the bottom of Table 3-4.

Table 3-4

Essential components of the Zeno Tool main Excel sheet, by row and column, with color codes — Part 4

Rows	Columns CY-DA	Columns DB-DD	Columns DE-DG	Columns DH-DJ	Column DK
1..10	Labels	Labels	Labels	Labels	Labels
i*=11..31	CYi*-DAi*=CI limits, CI width for #E_i using Binomial distribution, normal approximation	DBi*-DDi*=CI limits, CI width for #E_i using Binomial distribution, normal approximation	DEi*-DGi*=CI limits, CI width for #E_i using Compound Poisson distribution	DHi*-DJi*=CI limits, CI width for #E_i using Negative Binomial assumption	DKi*= min(n_i*p_i,n_i*(1-p_i)): criterion used for validity of test based on Binomial distribution
33	Empty	Empty	Empty	Empty	Empty
44	CY44-DA44=CI limits, CI width for #E using Binomial distribution, normal approximation	DB44-DD44=CI limits, CI width for #E using Binomial distribution, normal approximation	DE44-DG44=CI limits, CI width for #E using Compound Poisson distribution	DH44-DJ44=CI limits, CI width for #E using Negative Binomial assumption	DK44= min(np,n(1-p)): criterion used for validity of test based on Binomial distribution
45	Empty	Empty	Empty	Empty	Empty
47	CY47-DA47=CI limits, CI width for AS#E using Binomial distribution, normal approximation	DB47-DD47=CI limits, CI width for AS#E using Binomial distribution, normal approximation	DE47-DG47=CI limits, CI width for AS#E using Compound Poisson distribution	DH47-DJ47=CI limits, CI width for AS#E using Negative Binomial assumption	DK47= criterion used for validity of test based on Binomial distribution
50	CY50-DA50=CI limits, CI width for [#E] using Binomial distribution, normal approximation	DB50-DD50=CI limits, CI width for [#E] using Binomial distribution, normal approximation	DE50-DG50=CI limits, CI width for [#E] using Compound Poisson distribution	Empty	DK50= criterion used for validity of test based on Binomial distribution
53	CY53-DA53=CI limits, CI width for [AS#E] using Binomial distribution, normal approximation	DB53-DD53=CI limits, CI width for [AS#E] using Binomial distribution, normal approximation	DE53-DG53=CI limits, CI width for [AS#E] using Compound Poisson distribution	Empty	DK53= criterion used for validity of test based on Binomial distribution

Notes: The meaning of the color codes used in the Excel sheet is as follows: orange: labels; light purple: data; light green: estimates and intermediate results per stratum; dark green: confidence intervals (CIs) for rates per stratum; pink: CI for number of events per stratum; dark purple: verification. The formulas mentioned in this table are explained in Table 1. The references for the citations indicated in this table are in Table 1. The meaning of the symbols is given in Table 2. CV: coefficient of variation.

5 Data and results

The data used to test the Tool were originally gathered from administrative cancer databases in Manitoba. These data were chosen to illustrate the value of the Tool and to demonstrate a situation in which interest may focus on all the events (total rate of hospitalization for cancer including recurrences) or on only one event per person (rate of initial hospitalization for cancer among residents of the province).

Table 4 gives a sample of the results. Users can choose the significance threshold (for example, 95%, 90%)[1] of the CIs and specify the metrics (for example, r, ASR, $[r]$, $[ASR]$), the statistics (estimate, lower [L] and upper [U] limits of CIs), the methods (formulas in Table 1), as well as whether these are required globally or for specific strata. Users must be aware of the appropriateness of the method based on the scope, as shown in Table 1.

Table 4

Pivot table extracted from Zeno Tool

Row labels	Estimate	Exact CI (Poisson[1])	CI (Poisson Lognormal[2])	CI (Compound Poisson[3])
r				
Estimate	1860.2	…	…	…
L	…	1714.3	1851.9	1845.0
U	…	2083.5	1868.5	1875.5
ASR				
Estimate	1885.5	…	…	…
L	…	1737.6	1877.1	1869.4
U	…	2111.7	1893.8	1901.6
[r]				
Estimate	923.7	…	…	…
L	…	851.3	917.9	917.9
U	…	1034.6	929.6	929.6
[ASR]				
Estimate	964.9	…	…	…
L	…	889.3	959.0	958.0
U	…	1080.7	970.9	971.9

… not applicable

1. Poisson distribution.

2. Poisson distribution, Lognormal transformation formula.

3. Compound Poisson distribution.

Notes: The formulas are explained in Table 1. The meaning of the symbols is given in Table 2. CI: confidence interval; L: lower bound of 95% CI; U: upper bound of 95% CI.

Source: Statistics Canada, authors' calculations. Data originally from administrative cancer databases in Manitoba, but modified for illustrative purposes.

---

1. Significance threshold defined as 1 minus alpha level.

6 Conclusion

The Zeno Tool regroups and extends formulas from several published sources. It presents, on a single sheet, the calculations proposed by different authors and allows users to compare the impact of using various methods on the resulting confidence intervals (CIs). The Tool also expands the referenced formulas. For example, in one reference, the CI for ASR might be present, but not the CI for r; in another reference, the CI for r might be present, but not the CI for $[r]$; or formulas for r may exist, but not those for $\#E$. Thus, the Zeno Tool completes the set of formulas available for use in a broader range of circumstances that may exist in the data.

The Tool could be expanded to include features such as comparisons between groups, or tests on over-dispersion or zero-inflated distributions. Implementation of modelling of rates could also be considered. However, caution should be exercised before introducing additional features. While these features would be useful, ease of use could be compromised. The Tool might become cumbersome because additional analyses, such as hypothesis tests, would require more columns or rows, which might not be needed in all situations. As presented, this generalized tool is applicable to a broad array of circumstances, yet flexible and adaptable to suit other specific needs.

References

Carriere, K.C., and L.L. Roos. 1997. "A Method of Comparison for Standardized Rates of Low-Incidence Events." *Medical Care* 35 (1): 57–69.

Daly, L. 1992. "Simple SAS macros for the calculation of exact binomial and Poisson confidence limits." *Computers in Biology and Medicine* 22 (5): 351–361.

Fay, M.P., and E.J. Feuer. 1997. "Confidence intervals for directly standardized rates: A method based on the Gamma distribution." *Statistics in Medicine* 16 (7): 791–801.

Fleiss, J.L. 1981. *Statistical Methods for Rates and Proportions, 2nd Edition.* New York: John Wiley and Sons Ltd.

Glynn, R.J., T.A. Stukel, S.M. Sharp, T.A. Bubolz, J.A. Freeman, and E.S. Fisher. 1993. "Estimating the Variance of Standardized Rates of Recurrent Events, with Application to Hospitalizations Among the Elderly in New England." *American Journal of Epidemiology* 137 (7): 776–786.

Kegler, S.R. 2007. "Applying the compound Poisson process model to the reporting of injury-related mortality rates." *Epidemiologic Perspectives & Innovations* 4 (1). DOI: 10.1186/1742-5573-4-1.

van den Broek, J. 1995. "A score test for zero inflation in a Poisson distribution." *Biometrics* 51 (2): 738–743.