NDSS

National Diabetes Surveillance System Methods Documentation, 2008

April 2009

Module for Interpreting NDSS Data, version 208 (v208)

Table of Contents

1 EXECUTIVE SUMMARY **3**

A. PURPOSE 3
B. NDSS PRODUCTS 3
C. LIST OF UPDATES - NDSS SOFTWARE AND DOCUMENTATION V207 TO V208 3

2 KEY TERMS: **5**

3 DATA INTERPRETATION MODULE **7**

A. PURPOSE 7
B. NDSS PRODUCTS 7
C. DATA COLLECTION AND REPORTING 7
C1. Participating Provinces and Territories *8*

4 DATA SOURCES **8**

A. HEALTH INSURANCE REGISTRY FILE 8
B. PHYSICIAN CLAIMS FILE 9
C. HOSPITAL FILE 9

5 DIABETES CASE CRITERIA **10**

A. THE CASE DATE 11
B. TYPE 1 AND 2 DIABETES 11
C. CASES OF GESTATIONAL DIABETES ARE EXCLUDED 11

6 DEFINITIONS FOR CALCULATIONS **12**

A. PREVALENCE (PERIOD) 12
B. FORECASTED PREVALENCE 12
C. INCIDENCE RATE 14
D. ALL- CAUSE DEATH RATES 14
D1. Life Expectancy *15*
E. USE OF HEALTH SERVICES 15
F. COMORBID HOSPITALIZATION RATES 16
G. DIRECT AGE-STANDARDIZATION AND CONFIDENCE INTERVALS 19

7 FORMULAE USED FOR CALCULATIONS **20**

A. NOTATION 20
B. PREVALENCE ESTIMATES 21
B1. Age-Specific: *21*
B2. Age-Standardized: *22*
C. FORECASTED PREVALENCE ESTIMATES 23
C1. Age-Specific Counts: *23*
D1. AGE-SPECIFIC: 25
D2. Age-Standardized Rates: *25*
E. ALL-CAUSE DEATH RATES 26
E1. Individuals diagnosed with diabetes: Age-Specific Rates *26*
E2. Individuals with diabetes: Age-Standardized Rates *27*
E3. Individuals without diabetes: Age-Specific Rates *28*
E4. Individuals without diabetes: Age-Standardized Rates *28*
E5. Life Expectancies: *29*
F. HEALTH SERVICES USE AND COMORBID HOSPITALIZATIONS 30
F1. Age-Specific and Age-Standardized Comorbid Hospitalization Rate Ratios: *30*

8 RECORD OF DATA USE DECISIONS FOR NDSS PRODUCTS **32**

A. Products for Policymakers and the General Public 32

A1. Report from the National Diabetes System, Diabetes in Canada, 2008 32

B. Products for Health Researchers and Policymakers 32

B1. Data Files, PDF and CSV Format....... 32

9 DATA INTERPRETATION....... 33

A. False Negatives and False Positives 33

B. Outside the Scope of NDSS....... 33

APPENDIX A: ICD-9, CM, AND CCP CODES....... 34

APPENDIX B: RUN-IN PERIOD....... 35

APPENDIX C: VALIDATION STUDIES 37

REFERENCES 40

1 Executive Summary

A. Purpose

This module can be used when interpreting the data and analysis techniques for National Diabetes Surveillance System (NDSS) products reporting data between the 1995-1996 and 2005-2006 fiscal years, using v208 of the NDSS software and documentation. Basic knowledge of epidemiologic measures is assumed.

B. NDSS Products

Table 1. List of NDSS Products Using v208-available at www.ndss.gc.ca

Products	Formats	Data Years	Audience and Purpose For Use By:
Report from the National Diabetes Surveillance System, Diabetes in Canada, 2008	PDF, HTML	2001-2002 to 2005-2006, Forecast to 2010-2011	Provincial/territorial and federal policymakers and those interested in health information when updating, developing, and evaluating health programs
National Diabetes Surveillance System Methods Documentation, 2008	Word, PDF, HTML	1995-1996 to 2005-2006	Provincial/territorial and Federal policymakers and those interested in health information when interpreting the NDSS data and analysis techniques that were derived using v208 of the NDSS software and documentation
NDSS Data Files	CSV	1998-1999 to 2005-2006	Provincial/territorial and federal policymakers and those interested in health information when reporting provincial/territorial or Canadian data.
Report from National Diabetes Surveillance System, Data Quality Survey, 2008	Word, PDF, HTML	1995-1996 to 2005-2006	Provincial/Territorial and federal policymakers and those interested in health information when evaluating the quality of the NDSS data.

C. List of Updates - NDSS Software and Documentation v207 to v208

C1. The data collection period for this version included fiscal years 1995-1996 to 2005-2006 inclusive.
C2. The age of individuals was determined at the mid-year point, October 1. This is a change from v207, the age of the individual was calculated at fiscal year end (March 31).
C3. Forecasted prevalence was added to NDSS calculations and analysis.
C4. The average number of days that individuals stayed in hospital was added to the NDSS calculations and analysis.

C5. Although data collection began with 1995-1996, data are only used for national analyses beginning with the 1998-1999 fiscal year. It was necessary to designate a run-in period (i.e. a period of time where data were collected, but not reported) to increase the likelihood that cases diagnosed before the start of the collection period (April 1, 1995) were not coded as incident cases. Without a run-in period, prevalence rates would be artificially low and incidence rates artificially high. It had been determined that an appropriate run-in period for the NDSS would be three years. Therefore, the NDSS data available for analysis begins with the 1998-1999 fiscal year. Refer to Appendix B.

2 Key Terms:

1. Age-Specific Rate: The rate calculated for the five-year age group.

2. Average Hospital Days Stayed: The average number of days that an individual stayed in hospital, in the most recent fiscal year.

3. Canadian Classification of Health Interventions/Procedures (CCI/CCP): The Canadian Classification of Health Interventions (CCI) is the new national standard for classifying health care procedures. CCI is the companion classification system to ICD-10-CA. CCI replaces the Canadian Classification of Diagnostic, Therapeutic and Surgical Procedures (CCP) and the intervention portion of ICD-9-CM in Canada.

4. Comorbid Hospitalization: A hospitalization for a health problem other than diabetes, for an individual who had already met the NDSS case criteria for diabetes.

5. Confidence Intervals: The 95% confidence interval is an estimated range of values that are likely to include the true value of a parameter (i.e., rate) 19 times out of 20.

6. Data Collection Period: April 1, 1995 to March 31, 2005

7. Data Collection Start Date: The national data collection began on the first day of the 1995-1996 fiscal year, April 1, 1995, because major updates to all provincial and territorial databases were completed by this date.

8. Data Suppression: Counts and rates not presented to protect an individual's confidentiality.

9. Death Rate: The rate of deaths from all causes among the insured population, in the most recent fiscal year.

10. Direct Age-Standardization: A technique called age-standardization enables fairer comparisons between populations and over time by removing differences in the observed rates that resulted from the underlying differences in age distribution from each province or territory. A directly age-standardized rate represents what the observed rate would have been if the population had the same population as the 1991 Canadian standard population.

11. Discharge Abstract Database: The Canadian Institute for Health Information (CIHI) receives institution separation data (discharges, deaths, sign-outs, transfers) directly from participating Canadian hospitals. These data are recorded in the Discharge Abstract Database (DAD).

12. False Negatives: Individuals who have not met the NDSS case criteria, but have diabetes. The potential proportion of false negatives was indicated by the NDSS validation studies.

13. False Positives: Individuals who have met the NDSS case criteria, but do not have diabetes. The potential proportion of false positives was indicated by the NDSS validation studies.

14. Fee-for-Service: Many doctors and other health care providers were paid a fee for each service that they provided to their patient, such as an office visit, test, procedure, or other health care service. This billing system is called fee-for-service.

15. Fiscal Year: The fiscal year used by the federal, provincial, and territorial governments in Canada is the 12 month period beginning April 1 and ending March 31
For example, the fiscal year 2005-2006 would include data for the 12 month period April 1, 2005 to March 31, 2006 inclusive.

16. Gestational Diabetes: Gestational diabetes is a form of diabetes that develops in women during pregnancy and resolves after delivery. Gestational diabetes occurs in about 4% of all pregnancies; such women have an increased risk of developing type 2 diabetes later in life.

17. Incident Case: An individual in the insured population who has met the case criteria for the first time in the selected year

18. Insured Population: The total number of individuals who had a valid health insurance number within a selected province or territory at any point during the selected year. Individuals who had less than a full year of coverage, due to immigration, emigration, birth, or death during that year are included in the population. Also, some individuals had a valid health insurance number in more than one province or territory during a year because of immigration or emigration. These individuals were included in each of the provincial or territorial insured population.

19. International Classifications of Disease (ICD): The ICD is the international standard diagnostic classification for diseases and health conditions. It is used to classify diseases and other health problems recorded on many types of health and vital records including death certificates and health records.

20. Prevalent Case: An individual, among the insured population, who has met the case criteria prior to or during the selected fiscal year

21. Run-in Period: The period of time where data were collected, but not reported, to increase the likelihood that cases diagnosed before the start of the collection period (April 1, 1995) were not coded as incident cases (Refer to Appendix B.)

22. Shadow Billing: Shadow billing is an administrative process whereby physicians submit service provision information using provincial and territorial fee codes; however, payment is not directly linked to the services reported.

3 Data Interpretation Module

A. Purpose

This module can be used when interpreting the data and analysis techniques for National Diabetes Surveillance System (NDSS) products reporting data between the 1995-1996 and 2005-2006 fiscal years, using v208 of the NDSS software and documentation. Basic knowledge of epidemiologic measures is assumed.

B. NDSS Products

Table 1. List of NDSS Products Using v208-available at www.ndss.gc.ca

Products	Formats	Data Years	Audience and Purpose For Use By:
Report from the National Diabetes Surveillance System, Diabetes in Canada, 2008	PDF, HTML	2001-2002 to 2005-2006, Forecast to 2010-2011	Provincial/territorial and federal policymakers and those interested in health information when updating, developing, and evaluating health programs
National Diabetes Surveillance System Methods Documentation 2008	Word, PDF, HTML	1995-1996 to 2005-2006	Provincial/territorial and Federal policymakers and those interested in health information when interpreting the NDSS data and analysis techniques that were derived using v208 of the NDSS software and documentation
NDSS Data Files	CSV	1998-1999 to 2005-2006	Provincial/Territorial and federal policymakers and those interested in health information when reporting provincial/territorial or Canadian data.
Report from National Diabetes Surveillance System, Data Quality Survey, 2008	Word, PDF, HTML	1995-1996 to 2005-2006	Provincial/Territorial and federal policymakers and those interested in health information when evaluating the quality of the NDSS data.

C. Data Collection and Reporting

Data were collected and reported by fiscal year. The national data collection began with the 1995-1996 fiscal year. This starting point was selected by consensus among the participating provinces and territories because many of them had significant system changes in 1994-1995. Some jurisdictions, however, have data available prior to that period and may use it for their own reporting and validation purposes.

In January 2008, the NDSS requested that all provinces and territories submit data for fiscal years 1995-1996 to 2005-2006, using version 208 of the NDSS software and documentation.

Fiscal year 2005-2006 was the most recent data available at the time. Therefore April 1, 1995 to March 31, 2006 was defined as the data collection period. Although data collection began with 1995-1996, data were only used for national analysis beginning with the 1998-1999 fiscal year, to allow for a run-in period. Refer to appendix B.

C1. Participating Provinces and Territories

All provinces and territories, except Nunavut, submitted aggregated data, in time to be included in the most recent national reports and data products. As Nunavut data become available, they will be posted on the NDSS web site. Other jurisdictions were unable to provide selected data and these exceptions are noted in the applicable sections.

4 Data Sources

NDSS uses three administrative data sources that exist in all provinces and territories: the health insurance registry file, the physician claims file, and the hospital file. Publicly funded health insurance, administered by the provinces and territories, covers almost the entire Canadian population. Exceptions are people covered by Federal jurisdiction such as those in the military, Royal Canadian Mounted Police (RCMP) and federal correctional facilities. In all provinces and territories, individuals are assigned a unique personal health insurance number that must be provided upon receipt of health services. Payment of health services claims, under this system, is documented in person-specific administrative databases in each jurisdiction.

An encryption of the personal health number or other secure methodology was used to link records in these databases while at the same time protecting personal information. The data are linked by the provincial and territorial governments, or its designated agent, and maintained according to jurisdictional custodial obligations.

A. Health Insurance Registry File

The health insurance registry file in each province and territory contains records for each person entitled to coverage under that jurisdiction's health insurance scheme. The NDSS used the health insurance registry file to determine demographic information (age, sex, date of death) and an individual's eligibility for insurance within the province or territory. The total number of people that were eligible (insured population) for health insurance in that province/territory, within the selected fiscal year, was used as the denominator for many NDSS calculations.

The timeliness of the information updates on immigration, emigration, and deaths is not uniform across all provinces and territories; however, several of the provinces and territories have made recent efforts to improve the quality of these data.

Age was calculated using the age at mid-fiscal year (October 1). This is an update from v207, for which age was calculated as of (March 31).

Exceptions:

Quebec: Instead of the insured population, the denominator for rate calculations, represented the census-based population estimate from their vital statistics office, Le Directeur de l'état civil.

Newfoundland and Labrador: Instead of the insured population, the denominator for rate calculations, represented the census-based population estimate from their vital statistics office, Government of Newfoundland and Labrador Department of Government Services and Lands Government Service Centre.

B. Physician Claims File

The physician claims file was used to determine the number of diabetes physician claims and their potential case date for an insured individual, according to part of the NDSS case criteria, using the ICD-9[1] code of 250 or diagnosis text. The physician claims file was also used for the measurement of the health services analyses.

Most physician activity in Canada is captured by physician billing files, either through fee-for-service or shadow-billing claims. In both cases, physicians provide information, such as the health insurance number of the person receiving the service, the date the service was rendered, and the diagnoses and service provided. Physicians remunerated through a fee-for-service payment structure are required to submit a record of services rendered in order to receive payment. Physicians who are remunerated through alternative payment structures (salary, etc.) are encouraged to submit shadow-billing claims for services rendered; however they may not be mandated to do so. Physicians who service remote areas were often salaried.

Because some information about the insured population is not collected through the fee-for-service system and as a result, not available to the NDSS for the ascertainment of diabetes diagnoses, the number of individuals with diabetes may be underreported by the NDSS.

For most provinces and territories, each physician claim contains a single diagnosis, coded using the International Classification of Disease, 9th Edition (ICD-9).[1] The maximum number of diagnoses that a physician may claim, among all the provinces and territories that responded to the 2008 NDSS Data Quality Survey[2] (excluding Nunavut), ranged from 1 to 5.

C. Hospital File

The hospital file was used to determine the number of hospitalizations for an insured individual, according to part of the NDSS case criteria, using an ICD-9[1] or ICD-9-CM[3] code of 250 (diabetes mellitus) or equivalent ICD-10-CA[4] codes: E10 to E14, selected from all available diagnostic codes in the hospital file. The hospital file was also used to identify comorbid hospitalizations. The NDSS used the procedure codes for the number of amputations. The average number of days that individuals stayed in hospital was also calculated using information from the hospital file.

Provinces and Territories used acute care level hospital information from the Discharge Abstract Database (DAD) to submit hospitalization information to the NDSS because all provincial and territorial institutions (except for Quebec) report hospital data to the DAD. Quebec provided the hospitalization information to the NDSS from the MED-ECHO file, provided by The Ministère de la Santé et des Services Sociaux (MSSS). The Canadian Institute for Health Information (CIHI) receives institution separation data directly from participating Canadian hospitals which are recorded in the DAD.[5]

All available diagnosis fields were used for the ascertainment of diabetes cases. Because not all hospital data are reported to the DAD, such as data from emergency room services, the number of individuals with diagnosed diabetes may be slightly underreported.

At the beginning of the data collection period, all provinces and territories used the International Classification of Disease (ICD)-9[1] or ICD-9-Clinical Modification (CM)[3] classification standard to record hospital diagnoses. However, beginning April 1, 2001, some jurisdictions implemented the 10th Edition, for Canada, (ICD-10-CA)[4] coding, see Table 2. As of fiscal year 2006-2007 all provinces will be using this standard.

Table 2. Fiscal Data Year of Provincial and Territorial Implementation of ICD-10-CA

Province or Territory	Fiscal Data Year of Implementation
Alberta (AB)	2002-2003
British Columbia (BC)	2001-2002
Manitoba (MB)	2004-2005
New Brunswick (NB)	2003-2004
Newfoundland and Labrador (NL)	2001-2002
Northwest Territories (NT)	2002-2003
Nova Scotia (NS)	2001-2002
Nunavut (NU)	2002-2003
Ontario (ON)	2002-2003
Prince Edward Island (PE)	2001-2002
Quebec (QC)	2006-2007
Saskatchewan (SK)	2001-2002 and 2002-2003
Yukon Territory (YT)	2001-2002

5 Diabetes Case Criteria

The NDSS diabetes case criteria for diagnosed diabetes,[I] was based on research by Blanchard et al.[6] To meet the case criteria an insured individual, aged **1 year** and older, must have

[I] From this point forward in the report, “diabetes” refers to a diabetes case, as defined by the NDSS case criteria.

EITHER:

One or more hospitalization with an ICD-9[1] or ICD-9-CM[3] code of 250 (diabetes mellitus) or equivalent ICD-10-CA[4] codes: E10 to E14, selected from all available diagnostic codes in the hospital file,

OR

Two or more physician claims with the relevant ICD-9[1] code of 250 or diagnosis text ***within two years***, selected from the ***first*** diagnosis code available on the claim.

Once a person has met these criteria, the person is defined as a case, in all subsequent years when he/she had a valid health insurance number within the same province or territory. However, the NDSS does not track individuals as they move between provinces and territories, therefore, if a person moved to a different province or territory the criteria must be re-met for the person to be defined as a case in the new jurisdiction.

A. The Case Date

The diabetes case date, for an individual, was defined as the date of the hospital admission, or the most recent (i.e., last date rule) of the two physician claims, that contributed to the individual meeting the NDSS case criteria, for the first time, in the selected province or territory.

The NDSS definition for the case date and run-in period selection has changed since the previous report, "Responding to the Challenge of Diabetes in Canada" released in 2003.[7] For the NDSS reports released since 2003, the NDSS used last date method (using the most recent date of the medical claim) instead of first date method (using the first date of the medical claim). Therefore, the run-in period has increased from 2 (1997-1998) to 3 (1998-1999) years.

B. Type 1 and 2 Diabetes

The NDSS did not distinguish between diabetes types (type 1, type 2) due to limitations of the physician claim data and the hospital file. In ICD-9-CM[3], the broad category of diabetes was coded as the 3-digit code of 250. The latest revision of the International Classification of Disease, 10th Edition, Canada, (ICD-10-CA),[4] codes for type 1 and 2 diabetes, using separate alpha-numeric codes (E10 and E11, respectively), are available. It is anticipated that as additional ICD-10-CA[4] coded hospital data are accumulated and validated, it may be possible to analyze and report rates associated with hospitalization stratified by diabetes type, for example, comparing the rate of amputations among those with type 1 diabetes versus those with type 2 diabetes.

C. Cases of Gestational Diabetes are Excluded

Gestational diabetes is a form of diabetes that develops in women during pregnancy and resolves after delivery. Gestational diabetes occurs in about 4% of all pregnancies and there is evidence that this condition increases the risk of these women of developing type 2 diabetes later in life.

Currently, the main focus of NDSS is tracking only type 1 and 2 diabetes and therefore, excluded women with gestational diabetes, which is a temporary condition.

The ICD systems allow for coding gestational diabetes separately (ICD-9[1] 648 ICD-10-CA[4] P70) from other diabetes codes (ICD-9[1] 250 or ICD-10-CA[4] E10-14) however evidence suggests that more stringent criteria are necessary. Therefore, the NDSS case criteria excluded women diagnosed with diabetes 120 days preceding or 90 days after any pregnancy-related visit. The diagnostic codes, indicating pregnancy, are listed in Table 3.

Table 3. Obstetric ICD Codes	
ICD-9	650-669
ICD-10 and ICD-10-CA	O265, O290-O30, O318, O320-O369, O40-O411, O418-O439, O60, O60-O669, O680-O849, O890-O899, O904, O908, O95-O97, Z354-Z356

6 Definitions for Calculations

A. Prevalence (Period)

Prevalence (prior to and during the selected fiscal year) =

$$\frac{\text{Total Number of Prevalent Cases}}{\text{Insured Population}} \times 100$$

Exceptions:

Quebec: Instead of the insured population, the denominator, for rate calculations, represented the census-based population estimate from their vital statistics registry, Registre des événements démographiques.

Newfoundland and Labrador: Instead of the insured population, the denominator for rate calculations, represented the census-based population estimate from their vital statistics office, Government of Newfoundland and Labrador Department of Government Services and Lands Government Service Centre.

B. Forecasted Prevalence[II]

Observed prevalent case estimates were extrapolated for five years, 2006-2007 to 2010-2011. These estimates were constructed based on conservative assumptions. Constant age and sex specific diabetes incidence and death rates were assumed to prevail for the duration of the 5-year projection period. These rates were applied to the moderate growth population projections

[II] Projection method was originally developed by the Population Health Surveillance and Epidemiology, BC Ministry of Healthy Living and Sport and then adapted for the NDSS data by PHAC.

(Statistics Canada scenario 3[III]) by province and territory and then summed to produce the expected number of prevalent diabetes cases in the Canadian population (aged 1 year to 85+). The forecasted prevalence counts were age-specific, using five-year age groups, and rounded to the nearest 100.

The following is an overview of the forecasting process, see section 7.B. Forecasted Prevalence Estimates for more details:

1) The Statistics Canada population projection, for the next fiscal year, was applied to NDSS population for the 2005-2006 fiscal year. (Constant: $w_{ij} = \frac{N_{ij}}{scN_{ij}}$)
2) The NDSS death count was removed from the prevalent count for the given fiscal year.
3) The NDSS prevalent cases, for each age group, were factored by the proportion of the oldest single year of age, for each previous fiscal year five-year age group, from the 2005-2006 Statistics Canada population. For the first age group (0-1 years) the Statistics Canada population for children without diabetes was used for the previous fiscal year five-year age group. For the last age group (85+ years) no end weights were used. (Constant: $\gamma_{ij} = \frac{\text{SYAP}_{\max(k_i)\,j}}{\sum_{i \in k_i}^{\max(k_i)} \text{SYAP}_{ij}}$)
4) The robust incidence rates were also used to produce the NDSS incident cases for the next fiscal year.
5) The NDSS robust incidence rates (for fiscal years 2003-2004 to 2005-2006) were applied to the results of steps 1-4, to produce the NDSS prevalent cases for the next fiscal year. (Constant: $iR_i = \frac{\sum_{j \in \phi} iC_{ij}}{\sum_{j \in \phi} \left(N_{ij} - pC_{ij} + iC_{ij}\right)}$)
6) The NDSS robust death rates (for fiscal years 2003-2004 to 2005-2006) and were applied to the results of steps 1-5 to produce the NDSS death cases for the next fiscal year. (Constant: $mR_i = \frac{\sum_{j \in \phi} mC_{ij}}{\sum_{j \in \phi} pC_{ij}}$)

For the first projected year (2006-2007), the NDSS prevalent count for 2005-2006 was used. For all other years, up to 2010-2011, the previous year of projected prevalence, based on the Statistics Canada population estimate, was used in the calculation. The integers of each number were used, by the Integer function in the Statistical Analysis Software (SAS).

[III] Population Projections for Canada, Provinces, and Territories, Statistics Canada, available from URL: http://www.statcan.ca/bsolc/english/bsolc?catno=91-520-XIE (September 2008)

C. Incidence Rate

The denominator represented an insured individual at risk for diagnosed diabetes in the selected fiscal year. The formula for the incidence rate is included below.

Incidence Rate (for the selected fiscal year) =

$$\frac{\text{Total Number of Incident Cases}}{\text{Total Number of Insured Population} - [\text{Prevalent Cases} - \text{Incident Cases}]} \times 1000$$

Exceptions:

Quebec: Instead of the insured population, the denominator, for rate calculations, represented the census-based population estimate from their vital statistics registry, Registre des événements démographiques.

Newfoundland and Labrador: Instead of the insured population, the denominator, for rate calculations, represented the census-based population estimate from their vital statistics office, Government of Newfoundland and Labrador Department of Government Services and Lands Government Service Centre.

D. All- Cause Death Rates

Death rates: Ratios of death rates (the ratio of the rates of deaths from all causes to individuals with or without diabetes) were calculated.

Death rate (with diabetes) =

$$\frac{\text{Total number of deaths from all causes, among prevalent cases, in the fiscal year}}{\text{Total number of prevalent cases in the fiscal year}}$$

Death rate (without diabetes) =

$$\frac{\text{Total number of deaths from all causes, among individuals without diabetes, in the fiscal year}}{\text{Total number of individuals without diabetes in the fiscal year}}$$

Rate ratio of death rates =

$$\frac{\text{Death rate among prevalent cases in the fiscal year}}{\text{Death rate among individuals without diabetes in the fiscal year}}$$

Exceptions:

Quebec: In Quebec, the number of deaths among individuals without diabetes was calculated by subtracting the number of deaths among individuals with diabetes from the count of all deaths from their vital statistics registry, Registre des événements démographiques. The total number of individuals without diabetes represented the difference between census-based population estimates and the number of persons with diabetes as determined by the NDSS.

D1. Life Expectancy

Life expectancy was defined as the mean number of years remaining until death for individuals surviving to the beginning of age category.[8] It can be used to compare the general levels of mortality between populations over time. Life expectancy was calculated using a life table, which is a widely-used statistical table for demographic, social, and health studies. There are two types of life tables: cohort and current or period.

The period life table was used for the NDSS analysis. In a period life table, the average cross-sectional mortality experience of a population, measured over a short period of time, (usually one to three years) is applied to the life span of a hypothetical population. The NDSS used three fiscal years of death rates: 2003-2004 to 2005-2006. It was assumed that this hypothetical population is stationary and that the number of births is equal to the number of deaths. Using this technique, life expectancy and conditional probabilities of dying may be computed for any age. The NDSS period life tables were constructed using the Chiang method.[9] The Chiang method required two data components: 1) age-specific death rates, for 19 five-year age groups (0-1, 1-4, 5-9….85+) among individuals with and without diabetes and 2) the number of person-years lived within each age group, which was derived from a report by the Institute for Evaluative Clinical Services.[10, 11]

Because the deaths rates were aggregated by five-year age groups, the methods described by Hsieh et al. were used to close the abridged life table.[12] Data were grouped or summarized to facilitate data analyses and interpretation. This technique was also used to protect an individual's confidentiality. Diabetes status was not available for people younger than one year, therefore, the 2002 to 2004 sex-specific death rates for the Canadian population[13] were used to model the mortality experience of individuals with and without diabetes. The 0-1 year sex-specific death rates, used to construct the life table, represented infants without diabetes. Infant death rates for diabetes were unavailable. Because the 0-1 age group experienced a high rate of mortality, we assumed that the number of infants with and without diabetes would be about the same.

E. Use of Health Services

The use of health services by individuals with and without diabetes was measured by calculating the number of physician and specialist visits, and days of hospital stays. These are aggregated and age-standardized rates and rate ratios, computed using similar methods as described for the rates and rate ratios of death rates. The two measures are visits with general practitioners and specialists and the average number of days an individual stayed in hospital.

The categories of general practitioners and specialists did not include health care professionals, such as, chiropractors, naturopaths, dentists, or optometrists.

Each of the two health services are calculated separately. The age-standardized health service utilization rates and rate ratios were calculated separately for persons with and without diabetes as follows:

Rate of health service use (with diabetes) =

$$\frac{\text{Total number of services used, by prevalent cases, in the fiscal year}}{\text{Total number of prevalent cases in the fiscal year}}$$

Rate of health service use (without diabetes) =

$$\frac{\text{Total number of services used, by individuals without diabetes in the fiscal year}}{\text{Insured Population without diabetes in the fiscal year}}$$

Rate Ratio of health service use =

$$\frac{\text{Rate of health service use by prevalent cases}}{\text{Rate of health service use among individuals without diabetes}}$$

Exceptions:

Quebec: The numbers of primary care and specialist visits among individuals without diabetes were not provided for Quebec.

Nunavut: Aggregate numbers of hospitalizations for Nunavut were unavailable.

F. Comorbid Hospitalization Rates

The comorbid hospitalization rate calculation was defined as the proportion of hospitalized individuals with diabetes (with valid health insurance at any time during the fiscal year), who also had at least one select comorbid disease and/or lower limb amputation.

The diseases selected for the analyses are those often associated with diabetes (Table 3). Hospitalizations were collected for the following diseases: cardiovascular disease, hypertensive disease, ischaemic heart disease, acute myocardial infarction, heart failure, cerebrovascular disease, chronic kidney disease, and lower limb amputations.

The type of comorbid diseases were ascertained using the ICD-10-CA[4] and the Canadian Classification of Health Interventions (CCI) classification schemes. The Canadian Classification of Health Interventions (CCI) is the new national standard for classifying health care procedures. CCI is the companion classification system to ICD-10-CA[4] and CCI replaces the Canadian Classification of Diagnostic, Therapeutic and Surgical Procedures (CCP) and the intervention portion of ICD-9-CM[3] in Canada.

The national analyses includes only those provinces and territories using the ICD-10-CA[4] and CCI classification schemes to avoid bias of the results due to differences between the ICD-9[1] and

ICD-10-CA[4] schemes. The specific diagnosis codes and intervention codes are listed in Table 3 below.

All available diagnosis fields and/or procedure/interventions fields were used for the ascertainment of comorbid hospitalizations.

The following rates and rate ratios were calculated for each selected comorbid disease, listed in Table 3, for the most recent fiscal year.

The age-standardized comorbid hospitalization rates and rate ratios were calculated separately for persons with and without diabetes as follows:

Rate of comorbid hospitalizations (with diabetes) =

$$\frac{\text{Comorbid hospitalizations for prevalent cases}}{\text{Total number of prevalent cases in the fiscal year}}$$

Rate of comorbid hospitalizations (without diabetes) =

$$\frac{\text{Comorbid hospitalizations for individuals without diabetes}}{\text{Insured Population without diabetes in the fiscal year}}$$

Rate Ratio of selected comorbid hospitalizations=

$$\frac{\text{Comorbid hospitalization rate for prevalent cases}}{\text{Comorbid hospitalization rate for individuals without diabetes}}$$

Exceptions:

Quebec: Quebec was excluded as the national data includes only provinces and territories that have adopted the ICD-10-CA[4] and CCI classification schemes. Quebec started using this scheme with their systems starting in fiscal year 2006-2007.

Yukon: Amputation data were unavailable.

Table 3. Specifications for Ascertainment of Comorbid Hospitalizations ICD-10-CA and Canadian Classification of Health Interventions (CCI) Coding Schemes		
Comorbid Disease	**ICD-10-CA**	An individual with more than one comorbid hospitalization was counted once in each category. When more than one hospitalization was recorded in one of the highlighted cardiovascular disease categories: ischaemic heart disease, hypertensive disease, heart failure, or cerebrovascular disease, this case was counted only once under the broader "cardiovascular disease" category. Similarly, acute myocardial infarction was counted only once under the broader "ischaemic heart disease" category.
Cardiovascular Disease	I00-I78	
Ischaemic Heart Disease	I20-I25	
Hypertensive Disease	I10-I15	
Acute Myocardial Infarction	I21-I22	
Heart Failure	I50	
Stroke	I60-I69	
Chronic Kidney Disease	N18-N19	
Canadian Classification of Health Interventions (CCI)		
Lower Limb Amputations	1SQ93, 1VA93, 1VC93, 1VG93, IVQ93, 1WA93, 1WE93, 1WJ93, 1WL93, 1WM93	
Excluding cases where one of the following diseases is also coded (ICD-10-CA)	C40, C41, C461, C47, C49, C80, C962, D160, E830, G901, H473, L590, M431, M432, M898, N079, N818, P293, P960, Q, R294, S, T	The count of lower limb amputations excluded those caused by trauma or cancer.

G. Direct Age-Standardization and Confidence Intervals

NDSS software calculated direct age-standardized rates and confidence intervals using an inverse gamma distribution (inverse chi-square distribution) when the rate is greater than zero, based on the work of Anderson and Rosenberg (1998) and Fay/Feuer (1997).[14, 15] The NDSS used the 1991 Canadian population for age-standardization, see Table 4.

Table 4. 1991 Canadian Standard Population Weights

Age Group	1991 Canadian Standard Population Weights/100,000 Population
0-1	1428.7
1-4	5517.7
5-9	6945.4
10-14	6803.4
15-19	6849.5
20-24	7501.6
25-30	8994.4
31-34	9240.0
35-40	8338.8
41-44	7606.3
45-49	5953.6
50-54	4764.9
55-59	4404.1
60-64	4232.6
65-69	3857.0
70-74	2965.9
75-79	2212.7
80-84	1359.5
85+	1023.7

7 Formulae Used for Calculations

A. Notation

1.	iC_{ij}	Incident cases: The number of individuals in age group *i* with a case date in time period *j*
2.	iR_{ij}	Age-specific incidence rates: The incidence rate for individuals in age group *i* for time period *j*
3.	iR_j	Age-standardized incidence rates for time period *j*
4.	mC_{ij0}	Number of deaths (without diabetes): The number of individuals without diabetes in age group *i* with a date of death in time period *j*
5.	mC_{ij1}	Number of deaths (with diabetes): The number of individuals with diabetes in age group *i* with a date of death in time period *j*
6.	mR_{ij0}	Age-specific death rates (without diabetes): The death rate for individuals without diabetes in age group *i* for time period *j*
7.	mR_{ij1}	Age-specific death rates (with diabetes): The death rate for individuals with diabetes in age group *i* for time period *j*
8.	mR_{j0}	Age-standardized death rates for individuals without diabetes for time period *j*
9.	mR_{j1}	Age-standardized death rates for individuals with diabetes for time period *j*
10.	mRR_{ij}	Age-specific rate ratio of death rates: The ratio of age-specific death rates among individuals with and without diabetes for individuals in age group *i* for time period *j*
11.	mRR_j	Age-standardized rate ratio of death rates: The ratio of age-adjusted death rates for individuals with and without diabetes for time period *j*
12.	N_{ij}	Insured persons: The number of individuals in age group *i* with valid health insurance at any time during time period *j*
13.	pC_{ij}	Prevalent cases: The number of people with diabetes at the beginning of time period *j* plus the number of people newly diagnosed with diabetes at any time during time period *j*
14.	pR_{ij}	Age-specific prevalence: The prevalence for individuals in age group *i* at the end of time period *j*
15.	pR_j	Age-standardized prevalence rates at the end of time period *j*
16.	$\sum_i$	Indicates a summation over an index. In this case *i* is the index.
17.	p_{si}	Size of the standard population in age group *i*
18.	w_{si}	Standard population weight for the age group *i*

B. Prevalence Estimates

B1. Age-Specific: An age-specific rate is the rate measured for a particular age group.

Age-specific period prevalence is calculated (deaths in period *j* are included in both the numerator and the denominator).

pC_{ij} Numerator: The number of people diagnosed with diabetes at the beginning of time period *j* plus the number of people newly diagnosed with diabetes at any time during time period *j*

N_{ij} Denominator: The number of individuals in age group *i* with valid health insurance at any time during time period *j*

The computation formulae for rates and their confidence bounds were:

$$pR_{ij} = \frac{pC_{ij}}{N_{ij}}$$

$$\mathrm{var}\left[pR_{ij}\right] = \frac{pC_{ij}}{N_{ij}^2}$$

$$Lower_bound = \frac{\mathrm{var}\left[pR_{ij}\right]}{pR_{ij}} \times invgamma\left(\frac{\alpha}{2}, \frac{pR_{ij}^2}{\mathrm{var}\left[pR_{ij}\right]}\right)$$

$$Upper_bound = \frac{\mathrm{var}\left[pR_{ij}\right]}{pR_{ij}} \times invgamma\left(1 - \frac{\alpha}{2}, \frac{pR_{ij}^2}{\mathrm{var}\left[pR_{ij}\right]} + 1\right)$$

B2. Age-Standardized:

For time period *j* age-standardized rates are computed as weighted averages of age-specific rates. For example, for the 5-9 age group, the age-specific weight for age group *i* is the proportion of the 5-9 age group in the 1991 Canadian standard population.

Age-standardized rate is calculated by multiplying each age-specific rate by the standard weight and summing the weighted age-specific rates.

Age-standardized rates and the associated "1-α" percent confidence intervals were computed using the following formulae:

Age-standardized rate is calculated by multiplying each age-specific rate by the standard weight and summing the weighted age-specific rates.

Age-standardized rates and the associated "1-α" percent confidence intervals were computed using the following formulae:

Let $\sum_i$ denote the summation over the range of age groups of interest. Let p_{si} denote size of the standard population (Canada 1991), and define $w_{si} = p_{si} / \sum_i p_{si}$, which is the proportion of the standard population implied by the symbol $\sum_i$. Then pR_j represents the age-standardized prevalence rate and is defined as:

$$pR_j = \sum_i w_{si} \, pR_{ij} = \sum_i w_{si} \frac{pC_{ij}}{N_{ij}}$$

$$\mathrm{var}\left[pR_j\right] = \sum_i w_{si}^2 \, \mathrm{var}[pR_{ij}] = \sum_i w_{si}^2 \frac{pC_{ij}}{N_{ij}^2}$$

$$Lower_bound = \frac{\mathrm{var}\left[pR_j\right]}{pR_j} \times invgamma\left(\frac{\alpha}{2}, \frac{pR_j^2}{\mathrm{var}\left[pR_j\right]}\right)$$

$$Upper_bound = \frac{\mathrm{var}\left[pR_j\right]}{pR_j} \times invgamma\left(1 - \frac{\alpha}{2}, \frac{\left(pR_j\right)^2}{\mathrm{var}\left[pR_j\right]}\right)$$

C. Forecasted Prevalence Estimates

C1. Age-Specific Counts:

For time period *j* age-specific prevalent counts were computed for each age group i as the forecasted number of individuals with diagnosed diabetes at the end of time period *j*. The forecasts were rounded to the nearest 100.

Constants:

$$w_{ij} = \frac{N_{ij}}{scN_{ij}}$$

$$iR_i = \frac{\sum_{j \in \phi} iC_{ij}}{\sum_{j \in \phi} \left(N_{ij} - pC_{ij} + iC_{ij}\right)} \qquad \phi \in \{2004-2005, 2005-2006, 2006-2007\}$$

$$mR_i = \frac{\sum_{j \in \phi} mC_{ij}}{\sum_{j \in \phi} pC_{ij}}$$

Where w_{ij} -weight of NDSS population to Projected Stats Canada population for the age group *i* and the year *j*.

scN_{ij} -projected Stats Canada population for the age group *i* and the year *j*.

Let *j* = time period starts with the 2006-2007 and ends with the 2010-2011 fiscal years
Let *i* = age groups 1-19
Let $scN_{i,j+1}$ = Projected population for the age group *i* and the year *(j+1)* from Statistics Canada.
Let $N_{i,j+1}$ = NDSS projected population for a *(j+1)* year (Weighted with Statistics Canada projected population)
Let pC_{ij} = NDSS prevalent cases for the age group *i* and the year *j*.
Let mC_{ij} = NDSS mortality cases for the age group *i* and the year *j*.
Let pC^{*}_{ij} = NDSS prevalent cases without mortality cases
Let $pC^{\wedge}_{ij}$ = NDSS prevalent cases without incident cases
Let iR_i = NDSS incidence rates
Let iC_{ij} = NDSS incident cases
Let mR_i = NDSS mortality rates

$$N_{i,j+1} = scN_{i,j+1}\left(\frac{N_{ij}}{scN_{ij}}\right)$$

$$pC^{*}_{ij} = pC_{ij} - mC_{ij}$$

Constant

Let $SYAP_{ij}$ = single year of age population, from 0 to 85+, from Statistics Canada population for the projected year 2006-2007 to 2010-2011

Let γ_{ij} = oldest single year of age proportion of age group i for the year j (for the 5-year age group this value is approximately 0.20).

For all years j,

$$\gamma_{1j} = 1$$

$$\gamma_{ij} = \frac{\mathrm{SYAP}_{\max(k_i)\,j}}{\sum_{i \in k_i}^{\max(k_i)} \mathrm{SYAP}_{ij}}$$

where

$$k_2 = \{1,2,3,4\}$$

$$k_i = \{5(i-2), 5(i-2)+1, 5(i-2)+2, 5(i-2)+3, 5(i-2)+4\}$$

$$\gamma_{19j} = 0$$

$$pC^{\wedge}_{i,j+1} = \gamma_{i-1}\, pC^{*}_{i-1,j} - \gamma_i\, pC^{*}_{ij} + pC^{*}_{ij}$$

$$iC_{i,j+1} = iR_i\left(N_{i,j+1} - pC^{\wedge}_{i,j+1}\right)$$

$$pC_{i,j+1} = pC^{\wedge}_{i,j+1} + iR_i\left(N_{i,j+1} - pC^{\wedge}_{i,j+1}\right)$$

$$mC_{i,j+1} = pC_{i,j+1}\, mR_i$$

Let n = 19 age groups

Let $\sum_{i=1}^{n}$ denote the summation over the range of age groups of interest

Let $pC_{.j}$, $mC_{.j}$ and $iC_{.j}$ = Sum over the index *i. e.i.* $pC_{.j} = \sum_{i=1}^{n} pC_{ij}$

Then it can be shown:

$$pC_{.j+1} = pC_{.j} - mC_{.j} + iC_{.j+1} = \sum_{i=1}^{n}\left(pC_{ij} - \gamma_{i-1}pC_{i-1,j} + \gamma_i pC_{ij} - mC_{ij} + iC_{i,j+1}\right)$$

D. Incidence Rates

D1. Age-Specific:

iC_{ij}Numerator: The number of individuals in age group *i* with a case date in time period *j*

$(N_{ij}\text{-}\ pC_{ij}+\ iC_{ij})$Denominator: The number of individuals with a valid heath insurance number at any time in period *j* less the number of people diagnosed with diabetes at the end of time period *j* plus the number of people newly diagnosed with diabetes at any time during time period *j*

Age-specific incidence rates per 1,000, and the associated "1-α" percent confidence intervals were computed using the following formulae:

Let c denote 1,000. Then:

$$iR_{ij} = \frac{iC_{ij}}{N_{ij} - pC_{ij} + iC_{ij}}c$$

$$\mathrm{var}\left[iR_{ij}\right] = \left(\frac{c}{\left(N_{ij} - pC_{ij} + iC_{ij}\right)}\right)^2 iC_{ij}$$

$$Lower_bound = \frac{\mathrm{var}\left[iR_{ij}\right]}{iR_{ij}} \times invgamma\left(\frac{\alpha}{2}, \frac{iR_{ij}^2}{\mathrm{var}\left[iR_{ij}\right]}\right)$$

$$Upper_bound = \frac{\mathrm{var}\left[iR_{ij}\right]}{iR_{ij}} \times invgamma\left(1 - \frac{\alpha}{2}, \frac{iR_{ij}^2}{\mathrm{var}\left[iR_{ij}\right]} + 1\right)$$

D2. Age-Standardized Rates:

For time period *j* age-standardized rates are computed as weighted averages of age-specific rates. For example, for the 5-9 age group, the age-specific weight for age group *i* is the proportion of the 5-9 age group in the 1991 Canadian standard population.

Age-standardized incidence rates per 1,000 and the associated "1-α" percent confidence intervals were computed using the following formulae:

Let $\sum_i$ denote the summation over the range of age groups of interest. Let p_{si} denote size of the standard population (Canada 1991), and define $w_{si} = p_{si} / \sum_i p_{si}$, which is the proportion of the standard population implied by the symbol $\sum_i$. Let c denote 1,000. Then iR_j represents the age-standardized incidence rate and is defined as:

$$iR_j = \sum_i w_{si} iR_{ij} = \sum_i \frac{cw_{si} iC_{ij}}{N_{ij} - pC_{ij} + iC_{ij}}$$

$$\mathrm{var}[iR_j] = \sum_i w_{st}^2 \ \mathrm{var}[iR_{ij}] = \sum_i \left(\frac{cw_{si}}{N_{ij} - pC_{ij} + iC_{ij}} \right)^2 iC_{ij}$$

$$Lower_bound = \frac{\mathrm{var}[iR_j]}{iR_j} \times invgamma\left(\frac{\alpha}{2}, \frac{iR_j^2}{\mathrm{var}[iR_j]} \right)$$

$$Upper_bound = \frac{\mathrm{var}[iR_j]}{iR_j} \times invgamma\left(1 - \frac{\alpha}{2}, \frac{(iR_j)^2}{\mathrm{var}[iR_j]} \right)$$

E. All-Cause Death Rates

E1. Individuals diagnosed with diabetes: Age-Specific Rates

mC_{ij1}Numerator: The number of individuals with diabetes in age group *i* with a date of death in time period *j*

pC_{ij} Denominator: The number of people with diabetes at the beginning of time period *j* plus the number of people newly diagnosed with diabetes at any time during time period *j*

Age-specific death rates and the associated “1-α” percent confidence intervals were computed using the following formulae:

Let c denote 1,000. Then:

$$mR_{ij1} = \frac{mC_{ij1}}{pC_{ij}} c$$

$$\mathrm{var}[mR_{ij1}] = \left(\frac{c}{pC_{ij}} \right)^2 mC_{ij1}$$

$$Lower_bound = \frac{\mathrm{var}[mR_{ij1}]}{mR_{ij1}} \times invgamma\left(\frac{\alpha}{2}, \frac{mR_{ij1}^2}{\mathrm{var}[mR_{ij1}]}\right)$$

$$Upper_bound = \frac{\mathrm{var}[mR_{ij1}]}{mR_{ij1}} \times invgamma\left(1-\frac{\alpha}{2}, \frac{mR_{ij1}^2}{\mathrm{var}[mR_{ij1}]}+1\right)$$

E2. Individuals with diabetes: Age-Standardized Rates

For time period *j* age-standardized rates are computed as weighted averages of age-specific rates. For example, for the 5-9 age group, the age-specific weight for age group *i* is the proportion of the 5-9 age group in the 1991 Canadian standard population.

Age-standardized death rates per 1,000 and the associated "1-α" percent confidence intervals were computed using the following formulae:

Let $\sum_i$ denote the summation over the range of age groups of interest. Let p_{si} denote size of the standard population (Canada 1991), and define $w_{si} = p_{si} / \sum_i p_{si}$, which is the proportion of the standard population implied by the symbol $\sum_i$. Let c denote 1,000. Then mR_{j1} represents the age-standardized mortality rate and is defined as:

$$mR_{j1} = \sum_i w_{si} mR_{ij1} = \sum_i \frac{cw_{si} mC_{ij1}}{pC_{ij}}$$

$$\mathrm{var}[mR_{j1}] = \sum_i w_{st\,i}^2 \ \mathrm{var}[mR_{ij1}] = \sum_i \left(\frac{cw_{si}}{pC_{ij}}\right)^2 mC_{ij1}$$

$$Lower_bound = \frac{\mathrm{var}[mR_{j1}]}{mR_{j1}} \times invgamma\left(\frac{\alpha}{2}, \frac{mR_{j1}^2}{\mathrm{var}[mR_{j1}]}\right)$$

$$Upper_bound = \frac{\mathrm{var}[mR_{j1}]}{mR_{j1}} \times invgamma\left(1-\frac{\alpha}{2}, \frac{(mR_{j1})^2}{\mathrm{var}[mR_{j1}]}\right)$$

E3. Individuals without diabetes: Age-Specific Rates

mC_{ij0}Numerator: The death rate for individuals without diabetes in age group *i* for time period *j*

$(N_{ij}\text{-}\ pC_{ij})$Denominator: The number of individuals in age group *i* with valid health insurance at any time during time period *j* less the number of people diagnosed with diabetes at the end of time period *j.*

Age-specific death rates per 1,000 and the associated “1-a” percent confidence interval were computed using the following formulae:

Let c denote 1,000. Then:

$$mR_{ij0} = \frac{cmC_{ij0}}{N_{ij} - pC_{ij}}$$

$$\text{var}\left[mR_{ij0}\right] = \left(\frac{c}{N_{ij} - pC_{ij}}\right)^2 mC_{ij0}$$

$$Lower_bound = \frac{\text{var}\left[mR_{ij0}\right]}{mR_{ij0}} \times invgamma\left(\frac{\alpha}{2}, \frac{mR_{ij0}^2}{\text{var}\left[mR_{ij0}\right]}\right)$$

$$Upper_bound = \frac{\text{var}\left[mR_{ij0}\right]}{mR_{ij0}} \times invgamma\left(1 - \frac{\alpha}{2}, \frac{mR_{ij0}^2}{\text{var}\left[mR_{ij0}\right]} + 1\right)$$

E4. Individuals without diabetes: Age-Standardized Rates

For time period *j* age-standardized rates are computed as weighted averages of age-specific rates. For example, for the 5-9 age group, the age-specific weight for age group *i* is the proportion of the 5-9 age group in the 1991 Canadian standard population.

Age-standardized death rates per 1,000 and the associated “1-α” percent confidence interval were computed using the following formulae:

Let $\sum_i$ denote the summation over the range of age groups of interest. Let p_{si} denote size of the standard population (Canada 1991), and define $w_{si} = p_{si} / \sum_i p_{si}$, which is the proportion of the standard population

implied by the symbol $\sum_i$. Let c denote 1,000. Then mR_{j0} represents the age-standardized mortality rate and is defined as:

$$mR_{j0} = \sum_i w_{si} mR_{ij0} = \sum_i \frac{cw_{si} mC_{ij0}}{N_{ij} - pC_{ij}}$$

$$\mathrm{var}\left[mR_{j0}\right] = \sum_i w_{st\,i}^2 \ \mathrm{var}\left[mR_{ij0}\right] = \sum_i \left(\frac{cw_{si}}{N_{ij} - pC_{ij}} \right)^2 mC_{ij0}$$

$$Lower_bound = \frac{\mathrm{var}\left[mR_{j0}\right]}{mR_{j0}} \times invgamma\left(\frac{\alpha}{2}, \frac{mR_{j0}^2}{\mathrm{var}\left[mR_{j0}\right]} \right)$$

$$Upper_bound = \frac{\mathrm{var}\left[mR_{j0}\right]}{mR_{j0}} \times invgamma\left(1 - \frac{\alpha}{2}, \frac{\left(mR_{j0}\right)^2}{\mathrm{var}\left[mR_{j0}\right]} \right)$$

E5. Life Expectancies:

Life expectancy e_x is the average number of years of life remaining to be lived among individuals who survive to the beginning of the age category x. The life expectancies were computed from an abridged life table constructed using Chaing's method.[9]

T_x - the total number of person years remaining to be lived by all individuals who survive to the beginning of age category x.

l_x - the number of persons who survived to the beginning of age category x.

$e_x = \frac{T_x}{l_x}$ - Life expectancy; the mean number of years of life remaining until death among individuals surviving to the beginning of age group x.

Where x = 1, 2, …., 19 and represent age groups 0-1, 1-4, 5-9, …80-84, 85+

F. Health Services Use and Comorbid Hospitalizations

Rates and rate ratios for health services use (general and specialist practitioner visits) and comorbid hospitalizations were computed the same way as those for the death rates, with the appropriate ***event count*** substituted for the number of deaths in the numerator. The denominators, total number of individuals with a valid health insurance number, with or without diabetes, were similar to those used in the mortality calculations. The rates of hospital days stayed were calculated the same way as those for the death rates. For comorbid hospitalization calculations, the event count represented the number of hospitalizations, among individuals with and without diabetes, for which select comorbid diseases and lower limb amputation procedures were present.

F1. Age-Specific and Age-Standardized Comorbid Hospitalization Rate Ratios:

In order to calculate the confidence intervals for all of the rate ratio calculations (deaths, physician and specialist visits, comorbid hospitalizations) the variances were calculated using the delta method.[16] The delta method, in its essence, expands a function of a random variable about its mean, usually with a one-step Taylor approximation, and then applies the variance. The confidence intervals and their age-specific and age-standardized rates were computed using the following formulae:

For ***age-specific*** rate ratios for age group i and time period j, let mR_{ij1} represent rates for people with diabetes, mR_{ij0} rates for people without diabetes and mRR_{ij} the rate ratio. Then:

$$mRR_{ij} = \frac{mR_{ij1}}{mR_{ij0}}$$

$$\mathrm{var}\left[\log\left(mRR_{ij}\right)\right] = \frac{\mathrm{var}\left[mR_{ij1}\right]}{mR_{ij1}^2} + \frac{\mathrm{var}\left[mR_{ij0}\right]}{mR_{ij0}^2}$$

$$Low_bound = \exp\left(\log\left(mRR_{ij}\right) - probit\left(1-\frac{\alpha}{2}\right)\mathrm{var}\left[\log\left(mRR_{ij}\right)\right]^{\frac{1}{2}}\right)$$

$$Upper_bound = \exp\left(\log\left(mRR_{ij}\right) + probit\left(1-\frac{\alpha}{2}\right)\mathrm{var}\left[\log\left(mRR_{ij}\right)\right]^{\frac{1}{2}}\right)$$

Age-standardized rate ratios are computed in an identical fashion; however, the notation is somewhat different. Notably, the index for age is not required.

$$mRR_j = \frac{mR_{j1}}{mR_{j0}}$$

$$\operatorname{var}\left[\log\left(mRR_j\right)\right] = \frac{\operatorname{var}\left[mR_{j1}\right]}{mR_{j1}^2} + \frac{\operatorname{var}\left[mR_{j0}\right]}{mR_{j0}^2}$$

$$Low_bound = \exp\left(\log\left(mRR_j\right) - probit\left(1-\frac{\alpha}{2}\right)\operatorname{var}\left[\log\left(mRR_j\right)\right]^{\frac{1}{2}}\right)$$

$$Upper_bound = \exp\left(\log\left(mRR_j\right) + probit\left(1-\frac{\alpha}{2}\right)\operatorname{var}\left[\log\left(mRR_j\right)\right]^{\frac{1}{2}}\right)$$

8 Record of Data Use Decisions for NDSS Products

A. Products for Policymakers and the General Public

The NDSS reports were designed for use by federal and provincial policymakers for use in planning and developing health programs. They were also designed for those with a general interest in heath and health outcomes.

A1. Report from the National Diabetes System, Diabetes in Canada, 2008

A1a. Aggregation of Incidence and Prevalence Rates

The numbers of infants (age group 0-1) with diabetes were incomprehensible. Age-specific prevalence and incidence rates were presented for age groups (from 1-4 years to 85+ years). The younger age groups (1-4, 5-9, 10-14, and 15-19 years) were aggregated to form the 1 to 19 year age group, in order to improve the robustness of the rates. Rates based on small numbers can fluctuate widely across time and must be interpreted with caution. If an event happens twice in one year and four times the next year, the rate of the event doubles; even though the event is rare.

A1b. Rate Ratio of Death Rates

The numbers of deaths among children with diabetes were too small to interpret and report. Therefore they are excluded from NDSS products for policymakers and the general public.

B. Products for Health Researchers and Policymakers

As part of the NDSS mandate to build capacity among the provincial and territorial governments, the provincial, and territorial data files (CSV) were designed for use by health researchers. These files and methods can be used to analyse and compare NDSS data, at the national, provincial, and territorial levels.

B1. Data Files, PDF and CSV Format

The data files were produced and available in comma separated values (CSV) and PDF file format and the tables contain data among adults and children in five-year age groups (1-4 years to 85+ years). Counts and rates were suppressed when greater than 0 and fewer than 6 cases.

B1a. Data Suppression

The counts and rates were suppressed when the statistic was 6 or fewer cases. However, when the statistic was 0, it was presented.

9 Data Interpretation

A. False Negatives and False Positives

Using administrative data for surveillance, as in the NDSS, often requires a compromise when trying to determine the number of cases of a disease. It is necessary to balance the possibility of collecting false positives and missing false negatives. Validation studies have indicated that the case criteria, used by the NDSS, is reliable at balancing both false-negatives and false-positives in order to depict a relatively accurate picture of the trends of diagnosed diabetes in Canada. Refer to Appendix C.

In any given year false positives and false negatives may balance out. However, as more years of data are collected, false negatives will likely decline as there is more opportunity for false negatives to be captured by the NDSS case definition. The same is not true for false positives, which will continue to accumulate as more years of data are collected.

B. Outside the Scope of NDSS

There are important limitations of the NDSS that should be considered. For example, there are individuals with undiagnosed diabetes in Canada, and these are outside the scope of the NDSS. Reliance solely on administrative health care data for capturing cases has the disadvantage of describing a pattern of disease without offering insight into, or explanation for, differences within or between populations. For example, factors such as the social and economic environment, the physical environment, and individual characteristics (e.g., body weight) and behaviours (e.g., tobacco use), or differences in screening practices, may explain the pattern of disease in a population. The NDSS does not collect these types of data at this time, however some research is ongoing, in some provinces and territories, to link the NDSS with the Canadian Community Health Survey (CCHS) which does include this type of data.

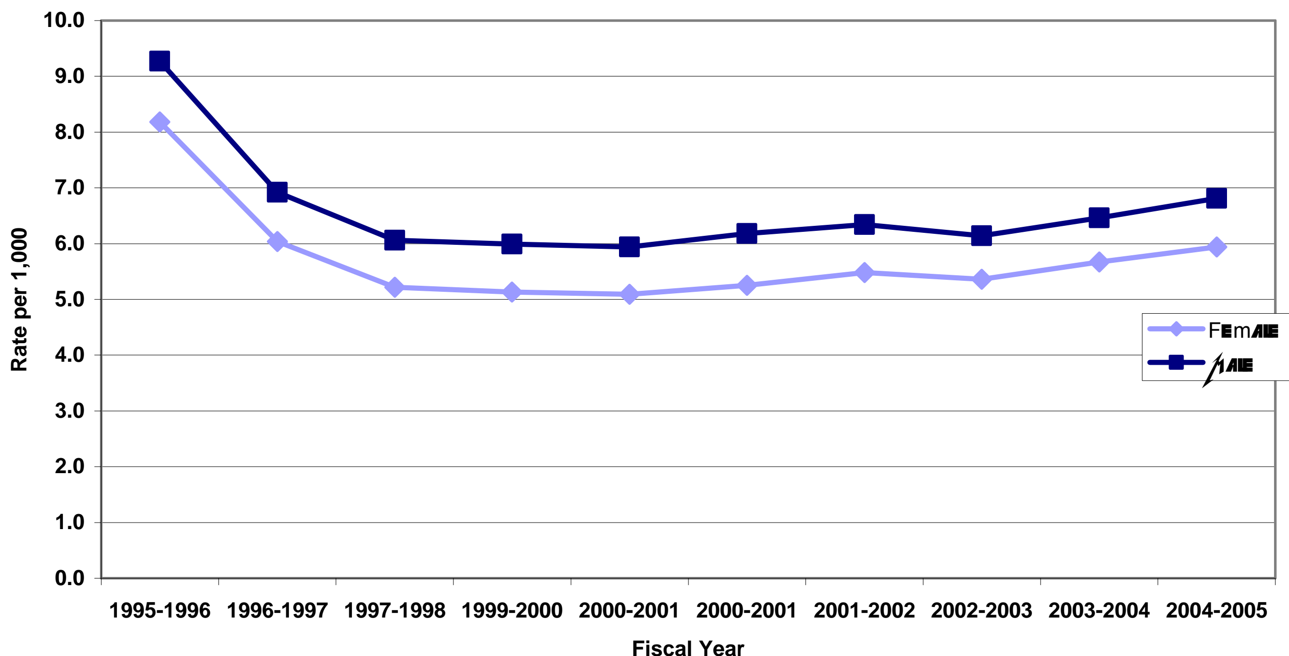
Appendix A: ICD-9, CM, and CCP Codes

Table 3 (a). For Reference Only **ICD-9, ICD-9-CM and CCP**	
Comorbid Disease	**ICD-9**
Cardiovascular Disease	390-448
Ischaemic Heart Disease	410-414
Hypertensive Disease	401-405
Acute Myocardial Infarction	410
Heart Failure	428
Stroke	430-438
Chronic Kidney Disease	585-586
ICD-9-CM **OR** **Canadian Classification of Diagnostic, Therapeutic and Surgical Procedures (CCP)**	
Chronic Kidney Disease	ICD-9-CM: 3995, 5498, 556 CCP: 5195, 6698, 6759
Lower Limb Amputations	ICD-9-CM: 8411-8419 CCP: 9611-9615
Excluding cases where one of the following diseases is also coded (ICD-9)	170, 171, 213, 740-759, 800-900, 901-904, 940-950

Appendix B: Run-in Period

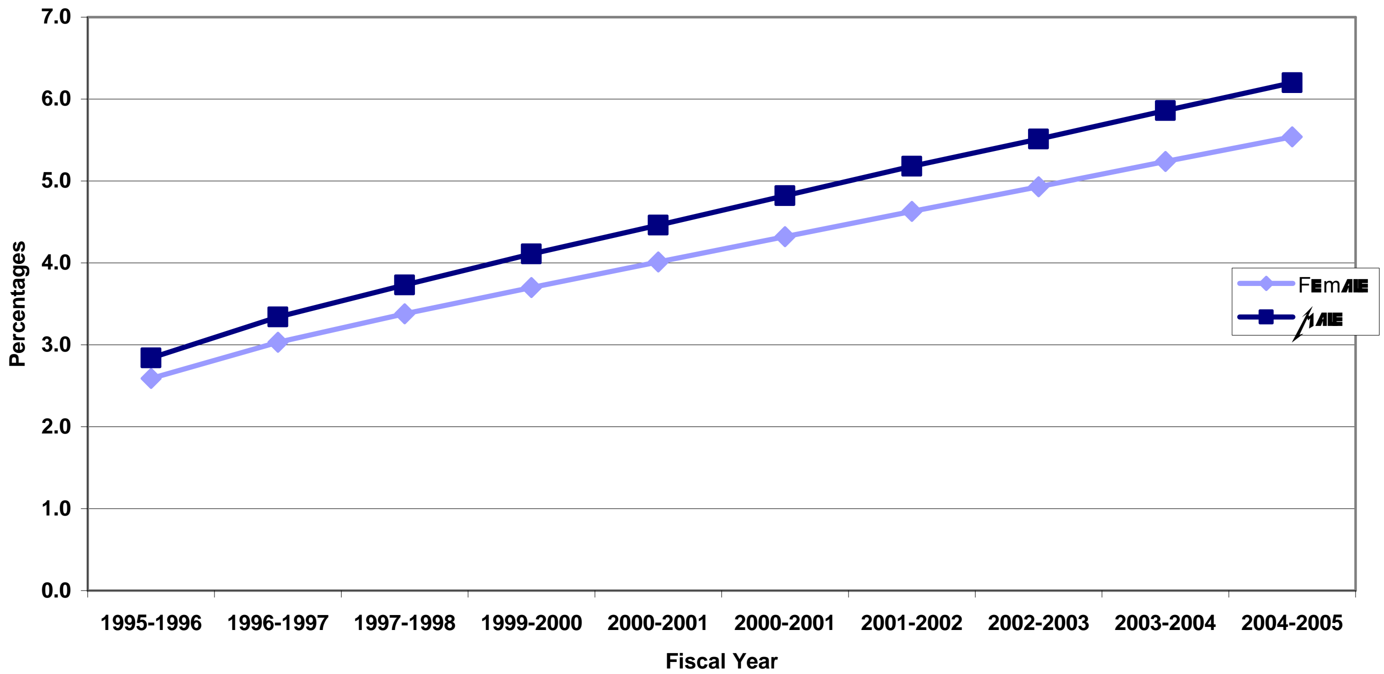
Figure 1. demonstrates that the incidence rates for the first 2 years of the capture period (1995-1996 to 1996-1997) were artificially high. Additionally, because the NDSS case criteria required physician claims and hospital visits within 2 years, the run-in period must be at least 2 years. Due to the implementation of the last date method (using the most recent date for the case date) and adequate measures to reduce artificial affects on the incidence and prevalence rates (Figures 1 and 2.) a run-in period of 3 years was selected by the NDSS Scientific Committee.

Figure 1. Incidence of Diagnosed Diabetes among People Aged 1 Year and Older by Sex and Year, Canada*, 1995-1996 to 2005-2006



Source: Public Health Agency of Canada, using NDSS data files contributed by provinces and territories, as of September, 2008.
*Data for Nunavut were unavailable

Figure 2. Prevalence Percentages of Diagnosed Diabetes among People Aged 1 Year and Older by Sex and Year, Canada*, 1995-1996 to 2005-2006



Source: Public Health Agency of Canada, using NDSS data files contributed by provinces and territories, as of September, 2008
*Data for Nunavut were unavailable.

Appendix C: Validation Studies

The first and essential attribute of the NDSS is the case definition which relies on administrative data for the determination of diabetes in the population. The advantage of using these data is the universal availability of the required data components in the provinces and territories.

The mining of administrative data for the identification of diabetes cases in the population began with work by Young et al. (1991) using the Manitoba health services database to estimate prevalence and incidence in the province. In this study, the operative case definition was a single medical claim or hospital abstract in which a diagnosis of diabetes was recorded. Several years later, in order to estimate diabetes incidence and period prevalence in Manitoba, Blanchard et al. (1996) undertook the development of a provincial diabetes database derived from administrative data. A case rule was determined by review of utilization patterns among diabetics registered in a diabetes education program, in combination with life table analysis. The case definition, subsequently named the “Manitoba Rule”, required that to be labelled as diabetic an individual must have had 2 physician claims within a 2-year period or 1 hospitalization with a diagnosis of diabetes recorded. Upon validating the Manitoba diabetes database against the diabetes education program, agreement of > 95% was found. The high rate of concordance between the two data sources was not unexpected given that the case rule for the diabetes database was derived from observing the education program.

In a project funded by Health Canada and Alberta Health and Wellness, it was demonstrated that a diabetes surveillance system in one province utilizing the Manitoba Rule was successfully reproducible in two other provinces (James et al., 2004). For the first NDSS report in 2003, the Manitoba Rule, with adjustment to the age parameter, was used to ascertain diabetes cases for the 20-years and older population. Since then, refinements to the algorithm have been adopted, the most significant being the lowering of the age threshold from 20-years and older to 1-year and older; the exclusion of gestational diabetes cases; and the designation of the last claim date rather than the first claim date as the *de facto* case date.

For the NDSS, the identification of a diabetes case requires scanning three administrative databases: the physician claims database, the hospitalization database and the health insurance registry. The first two serve to net potential diabetes cases while the registry confirms residency and eligibility for health care coverage. Each of the databases has inherent obstacles which may confound the capture of a diabetic case. For example, the physician claim form in most jurisdictions will allow for the recording of only one diagnosis obliging a physician to choose one of potentially several conditions to record on the claim. If repeated visits to a physician do not result in diabetes appearing as the precedent condition on the claim, a case could go undetected for an extended period of time. An individual with such a claim history would be considered a false negative, that is, the NDSS algorithm was unable to detect a true case. False negatives also arise when a case is sub-clinical; access to health care services is limited meaning fewer opportunities for the diagnosis of interest to be remarked and recorded; or an incorrect diagnostic code is entered or not entered at all. In a study that examined the validity of diagnostic

codes in Quebec medical services claims it was found that among the study physicians, 30.1% of their claims had the diagnosis field left blank (29.9%) or had an invalid code recorded (0.2%) (Wilchesky et al., 2004).

Analogously, a false positive is an individual identified as diabetic according to the NDSS case rule but the condition is in fact not present. For example, an individual is suspected as diabetic and a diagnosis of diabetes is recorded on one or more claims but upon further testing the presence of the condition is rejected. The NDSS methodology is unable to discern such records and over time these minimal evidence cases accrue in the database. We address this issue in a separate section later in the report.

The notion of false positive and false negative presumes that a gold standard or perfect or near-perfect registry of all existing diabetes cases exists against which the NDSS can be measured. In fact, a gold standard does not exist in the majority of provinces and territories, and consequently various proxies have been used in NDSS validation studies including self-reported survey data, diabetes education registries, and medical records.

Diabetes Case Ascertainment and Validation in the Adult Population

Canadian studies designed to validate cases of diabetes ascertained using the Manitoba Rule against a gold standard were compiled and reviewed. Initially, we were interested in determining whether any of these studies were comparable due to very similar or identical design. None were found to be directly comparable although as shown in Table 1, two studies used self-report survey data as the gold standard. All of the studies with the exception of one (VanTil, 2001) used a modified version of the Manitoba Rule for case ascertainment. For example Hux et al. (2002) excluded gestational diabetes while Lix et al. (2006) used a different age threshold. In each study, however, the 2 claim in 2-years or 1-hospitalization criteria, the *2 in 2* rule, was applied. The sensitivity of the *2 in 2* rule varied from 79.5% to 91% while the positive predictive value (PPV), an indication of how effective the case rule is in minimizing false positives ranged from 62.2% to 87.9%. In all studies where specificity was reported, the *2 in 2* rule was shown to be excellent at ruling out non-diabetics.

Table 5: Validation of the *2 in 2* Case Rule, by Provincial Source of Data

	Author	Gold standard	Study cohort	GDM excluded	Admin data years	Sens %	Spec %	PPV %	kappa	Youden's Index
NS	LeBlanc, 1998	Self-report (NSHS 1995)	> 0 years	..	1992-1994	62.7	99.4	..	..	0.62*
PEI	Van Til, 2001	Diabetes registry	≥ 25 years	..	1990 – 1994	82.4	98.8	76.4	0.78	0.81*
PEI	Van Til, 2001	Diabetes registry	≥ 25 years	..	1990 – 1999	89.2	96.6	62.2	0.71	0.86*
ON	Hux 2002	Drug claims	> 65 years	Yes	1991 – 1999	91	..	..	..	
ON	Hux 2002	Physician charts	..	Yes	1991- 1999	86.0	97.0	80.0	0.80*	0.83*
ON	Hux 2002	NPHS	..	Yes	1991 – 1999	85.0	..	64.0	..	
MB	Lix 2006	CCHS 2001	≥19 years	..	†	79.5	99.3	87.9	0.82	0.79

.. not stated
* Calculated from information provided in the article
† Administrative data were searched for the 2-year period prior to the date of the interview.
GDM – gestational diabetes
PPV – positive predictive value
NPHS – National Population Health Survey (year not stated)
NSHS – Nova Scotia Health Survey
CCHS – Canadian Community Health Survey

Youden's Index = (sensitivity + specificity) -1

Landis and Koch (1977) guidelines for the evaluation of kappa are as follows:
• Marginal reproducibility: $\kappa < 0.39$
• Good reproducibility: $\kappa = 0.40$ to 0.75
• Excellent reproducibility: $\kappa > 0.75$

In the Lix study, the interpretation of the kappa value is based on Altman (2001)
• Poor agreement: $\kappa < 0.20$
• Fair agreement: $\kappa = 0.20$ to 0.39
• Moderate agreement: $\kappa = 0.40$ to 0.59
• Good agreement: $\kappa = 0.60$ to 0.79
• Very good agreement: $\kappa = 0.80$ to 1.00

In developing the Manitoba Rule, the selection of a period of 2-years in which to confirm a diabetes case was based on two characteristics found in the population of individuals with diabetes under study (Blanchard et al., 1996). First, the median time between health care contacts, either physician or hospital, among diabetic individuals enrolled in an education program was 30 days. In addition, the probability of having a subsequent claim or contact for diabetes within 2-years of the previous claim for diabetes was 0.96.

References

[1] International Statistical Classification of Diseases, Ninth Revision, World Health Organization, Geneva, 1978.
[2] Public Health Agency, National Diabetes Surveillance System Report, Data Quality Survey, 2008.
[3] International Statistical Classification of Diseases, Ninth Revision, Clinical Modification, World Health Organization, Geneva, 1978.
[4] International Statistical Classification of Disease, Tenth Revision, Canada, Canadian Institute for Health Information, Version 2006.
[5] Canadian Institute for Health Information. Data Quality Documentation: Discharge Abstract Database 2005-2006. Available at URL: http://secure.cihi.ca/cihiweb/dispPage.jsp?cw_page=services_dad_e#quality.
[6] Blanchard JF, Ludwig S, Wajda A, Dean H, Anderson K, Kendall O, Depew N. Incidence and prevalence of diabetes in Manitoba, 1986-1991. *Diabetes Care* 1996; 19(8):807-811
[7] National Diabetes Surveillance System, Responding to the Challenge of Diabetes in Canada, Available at URL: http://www.phac-aspc.gc.ca/ccdpc-cpcmc/ndss-snsd/english/pubs_reports/pdf/WEB_NDSS_English_Report-nocover.pdf, January 2008.
[8] Keyfitz N. A life table that agrees with the data. Part I. *Journal of the American Statistical Association* 1966; 61 (314): 305-12
[9] Chiang CL. The Life Table and its Applications. Malabar, Florida: Robert E. Krieger Publishing Company; 1984.
[10] Chronic Diseases in Canada. Health adjusted life expectancy at the local level in Ontario. Public Health Agency of Canada. Vol. 21, 2-2000 URL: http://www.phac-aspc.gc.ca/publicat/cdic-mcc/21-2/e_e.html.
[11] Institute for Clinical Evaluative Services. Adding years to life and life to years: Life and health expectancy in Ontario, 2001 URL: http://www.ices.on.ca/WebBuild/site/ices-internet-upload/file_collection/Atlas%20%2D%20Adding%20years%20to%20life%20and%20life%20to%20years%2Epdf, September 2008.
[12] Hsieh J. A general theory of life table construction and a precise abridged life table method. *Biomedical Journal* 1991 (2): 143-162.
[13] Life Tables, Canada, Provinces and Territories, 2000 to 2002 URL: http://www.statcan.ca/english/freepub/84-537-XIE/84-537-XIE2006001.htm, August 2007.
[14] Anderson, RN and Rosenberg, HM. Age-standardization of Death Rates: Implementation of the Year 2000 Standard. *National Vital Statistics Reports* Vol. 47, No. 3. Hyattsville, MD: National Center for Health Statistics. October 1998.
[15] Fay MP and Feuer EJ. Confidence intervals for directly standardized rates: a method based on the gamma distribution. *Stat Med* 1997 Apr 15;16 (7):791-801.
[16] Dodge, Yadola. The Concise Encyclopedia of Statistics. Springer: Berlin, Germany, 2008.