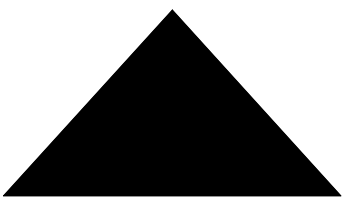
Central East
Health Information Partnership

Data Quality in RDIS: Issues Related to Combining Data Sets

October, 2000



CEHIP

Data Quality in RDIS: Issues Related to Combining Data Sets

Katharine Dawson, MHSc
University of Toronto

*Funded by CEHIP Student Award Program**

October, 2000

**This report, produced by Kathy Dawson as an MHSc Community Health and Epidemiology practicum student in the winter of 2000, was a co-operative effort between the Communicable Disease Surveillance Unit at Toronto Public Health (TPH) and the Central East Health Information Partnership.*

17250 Yonge Street, Box 234, Newmarket, Ontario L3Y 4X1 Tel: (905) 764-6346 Fax: (905) 895-0848 URL: http://www.cehip.org

Table of Contents

1. Overview and Purpose

This report evaluates the quality of data contained within the Reportable Disease Information System (RDIS), with particular attention to issues arising as a consequence of attempts to combine data from six separate information systems into one data set. The need to integrate data from six RDIS systems arises as a result of the amalgamation of the former six municipalities (East York, Etobicoke, North York, Scarborough, Old Toronto, and York) into the new mega-city of Toronto. Data quality issues discussed within this report were encountered during the analysis of the trend of disease over time for four specific reportable diseases.

Overall, RDIS is a well-designed system that provides essential basic information on cases of communicable diseases in Ontario, including person, place and time. This data as well as other important information for communicable disease surveillance is contained in the mandatory fields of the RDIS system. These mandatory fields are the most complete and accurate of the RDIS data fields, and thus are most useful to combine for the 6 former health units of Toronto. This combined data may be analyzed to determine past and present trends of disease for the entirety of the region encompassed by the new mega-city of Toronto.

Nonetheless, the quality of the data contained in the RDIS system could be improved, in particular that contained within the non-mandatory data fields. Improved data may allow for enhanced planning for resource allocation, as well as the evaluation of public health programs and initiatives already in place, that were originally designed to decrease communicable disease rates for Toronto and Ontario. Until such a time that the quality of data in fields not designated as mandatory improves considerably, the RDIS system is limited in its' use to that purpose for which it was originally intended: basic surveillance of cases of communicable disease.

2. Background

2.1 Reporting of Notifiable Diseases to the Health Department

Case reports for reportable diseases may be received by TPH by telephone, facsimile, mail, courier, or in person, on a twenty-four hour, seven day a week basis. Cases may be reported by laboratories, physicians, hospitals or other institutions, health departments, schools, and childcare centres (RDIS Guidelines and Procedures Manual, 1992). Laboratories have their own reporting forms, however physicians and hospitals often use a form supplied by the health unit. As soon as possible following the reporting of a case, one of several data entry clerks enters the initial information into the RDIS system on the computer, and assigns an incident number to the case. Information entered into the system by the clerks includes, as available:

1. the name, telephone number, address and postal code (to determine responsible health unit), and birth date of the case (age is then calculated by RDIS);
2. the reporting physician and his/her phone number and address;
3. the disease;
4. episode date and type;
5. reported date;
6. diagnosis date and type (including any lab tests and results performed, date and type of specimen collected);

If the address/ date of birth/ telephone number/ other pertinent information regarding the case is not available on the report submitted, data entry personnel are responsible for following up and obtaining this information for entry into RDIS, usually by contacting the physician attending the case. The clerks verify the address and postal code information manually at this stage of data entry. As well, the clerk checks to ensure the case is not already entered in RDIS for that particular disease (within a certain time limit), to avoid duplicate records. Duplication is confirmed if the two reports describe cases with the same name, date of birth, and reporting physician. The clerks then forward a copy of a worksheet (Appendix A)

containing the entered data, as well as any lab reports or information received from hospitals or physicians, to the Public Health Investigator at the appropriate health office for follow-up.

The Public Health Investigators are responsible for case management and follow-up, as well as the updating and completion of information on the worksheet. The Investigators must seek out the remaining information required by RDIS, such as method of diagnosis and detection, and may also obtain information for additional relevant but optional data fields, such as the risk setting of the case and the complications of disease experienced by the case. A description of some of these fields is available in Appendix B. The information necessary to complete the RDIS worksheet is obtained through consultation with the case and/ or their physician. The Investigators also determine whether a reported case meets the case definition, the occurrence of clusters and outbreaks, and when a case may be closed. Once the Investigator completes the worksheet, he or she forwards the work sheet to a clerk for entry into RDIS. RDIS is updated when and if new information is available for a case.

All disease episodes in which the Classification field designates the individual is a case of the disease, and in which the Episode Status designates that the individual meets the case definition, are transmitted to the Ministry of Health weekly by the responsible health unit. Specific fields are compulsory for transmission to the Ministry of Health by local public health offices. As well, other fields exist that may be transmitted if the data has been collected and entered into RDIS, and additional fields are present for optional use at the local health office level but which are not transmitted to the Ministry. Future updates or changes to cases within RDIS pertaining to compulsory or transmissible fields for a disease episode are re-transmitted to the Ministry.

The degree to which the transmissible and optional data fields in RDIS are utilized has been left to the discretion of individual health units and offices. All disease episodes with the classification of ‘case’ and the episode status of ‘meets case definition’ are transmitted to the Ministry. Cases must be transmitted to the Ministry of Health within 30 days of being confirmed as a case.

2.2 Relevant Legislation

Reportable diseases are those diseases which have been specified as being reportable to the local medical officer of health by Ontario regulation 599/91 under the Health Protection and Promotion Act, 1983 (RDIS Guidelines and Procedures Manual, 1992). There are currently 57 reportable diseases.

2.3 The Reportable Disease Information System

Information recorded in RDIS pertaining to reportable diseases in Ontario is transmitted electronically from local health departments to the Ministry of Health and Long Term Care. From the provincial level, information on the Federally notifiable diseases is transmitted to the Laboratory Centre for Disease Control, Health Canada, and eventually information on internationally notifiable diseases are communicated to the World Health Organization.

Various component databases exist within RDIS, however this report will focus on data contained within the Registry database pertaining to each of the communicable diseases to be examined. The Registry database contains information specifically on the individual and their disease episode (RDIS Guidelines and Procedures Manual, 1992).

2.4 Data Quality Checks

Each health department is responsible for the integrity of its own RDIS data (RDIS Guidelines and Procedures Manual,1992). This includes data capture, entry, maintenance, verification, retrieval, access and analysis. Much of the monitoring of quality assurance is system-generated; for instance, the

incidence of double-counting is minimized since the system will warn the operator that a file already exists for a client with similar attributes (eg name and birthdate).

It is the responsibility of clerks to verify address and postal code information manually during data entry, and to ascertain that the case is not already entered in RDIS for the same disease episode to avoid duplicate records. Duplication is confirmed if the two reports describe cases with the same name, date of birth, and reporting physician.

As well, the Ministry of Health evaluates the presence of duplicates within RDIS by performing a matched analysis on the Personal Identifier (PID) (RDIS Guidelines and Procedures Manual, 1992). If a duplicate record is found, the health departments are phoned and asked to investigate whether these are true duplicates or not.

3. Methods

Data sets from the six separate RDIS systems of the former health units of Toronto, namely East York, Etobicoke, North York, Scarborough, Old Toronto, and York, were examined for 4 notifiable diseases (influenza, meningococcal disease, invasive Group A streptococcal infections (GAS), and pertussis). Data quality issues identified while combining these data sets are discussed in this report, including an overview of which data fields can be combined across the 6 RDIS systems, and which data fields are not reliable to describe trends of disease.

Major issues to be aware of when using data combined from separate RDIS systems are discussed, including specific cautions regarding the future use of the data for the 6 former health units by Toronto Public Health. Recommendations are included to improve the consistency in usage and reliability of data collected in the data fields of the RDIS system. These recommendations are applicable for any community in Ontario, as well as for the Ontario Ministry of Health and Long Term Care and others attempting to combine data from separate RDIS systems.

Population data for the 6 former health units of Toronto as well as the new mega-city was obtained from the Central East Health Information Partnership and Statistics Canada. The estimates are Metropolitan Toronto population estimates by Census Subdivision for the years 1990 to 1999. Tables displaying the distribution of the population at risk of becoming a disease case for Toronto within such categories as age and region of residence for each year are available in Appendix C. The population data was utilized to calculate rates as well as compare the relative percentage of cases within each risk category to determine which subgroups contained disproportionate numbers of cases.

4. Results

4.1 Differences in Interpretation and Definitions of RDIS Data Fields

It is apparent that the various former health units of Toronto did not always interpret the data fields in the same manner, or did not interpret the data fields correctly. Variations in the definition of data fields used by the different health units affects the ability to combine the separate RDIS databases into one system. Although the data contained in any one RDIS system belonging to a single health unit remains consistent within that health unit, differences in definitions affects the reliability of the data across the different RDIS systems. Errors in the interpretation of a particular data field affect the accuracy and reproducibility of the data contained within the RDIS system.

The use of different definitions for data fields across health units is particularly apparent for the date fields, especially the 'Episode Date' and the 'Date of Diagnosis'. The 'Episode Date' is intended to describe the earliest date the disease was *suspected* in the case, and the 'Diagnosis Date' would describe the date the disease was *confirmed*. Depending on the particular disease and health unit in question, the 'Episode Date' may represent for a substantial proportion of cases either the onset date of symptoms, the clinic visit or specimen collection date, the date the data was entered into RDIS, the report form completion date, or one of several other reference dates. Similarly, the 'Date of Diagnosis' may refer predominantly to the date the specimen was collected, the date the specimen was received by the lab, the date the report was received by the health unit, the date of clinical diagnosis, or the date of diagnosis reported by the client. Although the presence of the 'Type of Date' field allows the definition of each date entry to be clarified somewhat, a question of bias affecting the data exists when particular definitions are used disproportionately by one health unit over another. Without accurate and consistent date fields, the trend of disease over time cannot be clearly identified.

An additional error apparent for the date fields recorded in RDIS exists due to definitional discrepancies as well (Table 1). In particular, the intended order of the dates recorded in the 'Episode Date', 'Date of Diagnosis', and 'Report Date' data fields is frequently not maintained. For example, many disease cases contain diagnosis dates occurring prior to episode dates, or report dates, etc. The incorrect ordering of the date fields can bias the trend over time of data, particularly for the 'Episode Date' data field as this field is relied upon to determine at what point in time the disease episode occurred for analysis and reporting purposes.

Table 1. Number and Percentage of Disease Cases with Dates that Appear in the Incorrect Order in RDIS.

	Sporadic Cases of Influenza		Cases of Meningococcal Disease		Cases of Group A Streptococcal Infection		Cases of Pertussis	
	No.	%	No.	%	No.	%	No.	**%**
Episode Date after Diagnosis Date	76	6.4	8	3.7	7	3.3	50	**3.9**
Episode Date after Report Date	13	1.1	2	0.9	3	1.4	14	**1.1**
Diagnosis Date after Report Date	90	7.6	21	9.7	17	8.0	70	**5.5**
Birth Date after Report Date	1	0.1	0	0.0	0	0.0	0	**0.0**
Birth Date after Diagnosis Date	5	0.4	0	0.0	0	0.0	0	**0.0**

As well, it may be important to some CD Managers and Medical Officers of Health to be able to recognize the delay between symptom onset and the diagnosis and reporting of the disease episode, as this time period permits the potential communication of the disease to contacts of the case. The lack of accurate, consistently recorded dates prohibits the analysis of the reporting delay for cases of communicable diseases recorded in RDIS.

Additional difficulties in the interpretation of data fields arose from an examination of the 'Complications' field for meningococcal disease. Many cases were reported as developing meningococcemia as a complication of their meningococcal disease, however for only a small number of these cases was meningococcemia not the original diagnosis. This error occurred disproportionately in Scarborough, but also occurred for disease cases in Old Toronto and North York. The misinterpretation of the complications data field leads to an overestimate of the burden of disease on both the community and the health care system within particular areas of Toronto, and for the entire region in general. The allocation of resources in response to such an overestimate may inadvertently prevent adequate resources or funding for other communicable diseases or public health sectors.

The interpretation of the correct postal code to enter into RDIS is not always achieved by all health units. The consistent recording of the postal code of the residence of the case is required to ensure the correct health unit is monitoring the case, ensuring that cases and case management services are not duplicated across separate RDIS systems and health units. This prevents the overestimate of the burden of disease within a particular region and for Toronto as a whole, and creates an inefficient use of public health resources. As well, correctly identified postal codes allow geographic trends and clustering

of disease within Toronto to be clearly identified and responded to through public health services and prevention programs. However at times, cases within particular health offices have postal code designations that lie within another health unit's jurisdiction. This may occur when a postal code is incorrectly entered, when a case is followed up by the incorrect health unit, when a case moves from one health unit jurisdiction to another, or when a case is entered by a health unit other than the one presiding over the area where the case resides but does not mark the case for transmission to the Ministry of Health. The most likely explanations are either a typographic error when the postal code was recorded, or a 'No' designation in the 'Transmit to Ministry?' data field.

4.2 Usage of Mandatory RDIS Data Fields Within and Between RDIS Systems

The mandatory data fields are those variables designated by the Ministry of Health and Long Term Care as essential for the surveillance of each particular notifiable disease. It is important that these data fields are complete and accurate, as this data provides the principle information on which to base disease prevention programs and allocate funds and other resources to communicable disease control within and across health units.

Variations in the proportions of missing data among mandatory field items across health units may bias results. For example, data may be more likely to be complete for certain cases such as those with more severe disease episodes, and less likely to be complete for those cases which are more difficult to follow, such as homeless persons or persons who are uncooperative with public health staff. The overall trend of disease depends on the ability of public health staff to gather information on *all* cases of disease and not just the most serious or convenient cases, otherwise the resulting inferences for the population at risk and subsequent programs designed to reach this population may be inaccurate.

The use of mandatory data fields is fairly complete and quite consistent across the former health units for most items, for the four diseases examined. For no case was information missing from data fields such as 'Incident Number', 'Disease', 'Responsible Health Unit', and 'Episode Date'. In almost all cases, the 'Type of Episode Date', 'Associated Organism/ Agent', 'Gender', 'Date of Birth/ Age', and 'Date of Diagnosis' were complete. Notably, the 'Lab Report Date' data field was missing information for 100.0% of cases, for all diseases.

Postal code information was fairly complete, although North York and Toronto had a few cases with unknown postal codes for several of the diseases examined. Unfortunately a tool to determine whether the street address entered into RDIS agreed with the recorded postal code was not available, thus the accuracy of data entry for valid postal codes could not be evaluated. Nonetheless, postal code information was fairly complete. Similarly, the 'Type of Diagnosis Date', 'Method of Detection', and 'Method of Diagnosis' data fields generally had quite low proportions of missing information.

The mandatory data fields with considerable proportions of missing data include the 'Outbreak Associated?', the 'Physician Report Date', the 'Deceased?', and the 'Lab Report Date' data fields. The proportion of missing data for the 'Outbreak Associated?' data fields varied considerably, from 12.2 to 31.0% for the four different notifiable diseases examined. A large variability was observed between health units as well, with Etobicoke consistently recording the greatest proportion of missing data for the 'Outbreak Associated?' data field regardless of the communicable disease examined. For example, for sporadic cases of influenza the proportion of missing data ranged from 1.4% in Scarborough to 94.7% in Etobicoke. For meningococcal disease, the proportion of missing data for the 'Outbreak Associated?' field ranged from 0.0% in York to 72.7% in Etobicoke, while for GAS the range was 0.0% in East York to 68.4% in Etobicoke. For cases of pertussis, information was missing for the 'Outbreak Associated?' data field for a low of 0.7% of cases in Scarborough, and a high of 63.0% of cases in Etobicoke.

Similarly, the 'Physician Report Date' data field had a large proportion of missing data, from 37.5% to 47.9% depending on the particular communicable disease examined. Influenza and Group A streptococcal infections had the greatest proportion of missing data for this field, both with 47.9% of cases missing data. Consistently for each of the four diseases examined, East York had the lowest proportion

of missing information for the 'Physician Report Date' field, and York had the greatest proportion. For influenza, the percentage of missing data ranged from 14.9% in East York to 99.0% in York. For GAS, 0.0% of cases were missing information for this data field in East York while 100.0% of cases were missing in York. For cases of pertussis, the amount of missing data ranged from 8.7% in East York to 96.9% in York. Data for this field was most complete for meningococcal disease, although a significant amount of data was missing even for this disease at 37.5%, ranging from 7.7% in East York to 93.8% in York.

The 'Deceased?' field also recorded a substantial percentage of cases missing information. A difference in the completeness of data within this field appeared to be influenced by the severity of the disease in question. Missing data was quite high for pertussis and influenza at 95.0% and 93.8%, respectively. However, information was far more complete for the more serious and less common diseases, meningococcal disease and Group A streptococcal infections, at 31.0% and 3.8% respectively. For pertussis, missing information ranged from 86.2% in York to 99.5% in Old Toronto, while for influenza missing data varied from 76.6% in East York to 98.2% in Toronto. Missing information ranged from 18.2% in Etobicoke to 46.2% in East York for cases of meningococcal disease. For GAS, 0.0% of cases were missing data for the 'Deceased?' field in East York, Etobicoke, and Scarborough, while 37.5% of cases in York were missing information.

With regards to the mandatory fields overall, data is most complete for the former Scarborough health unit, and least complete for the North York health unit. The proportion of missing data for each data field depends on both the specific disease and health unit recording the information. For influenza, data was most complete for the East York health unit, and least complete for the North York and York health units. For meningococcal disease, data was most complete for the Etobicoke, Scarborough, and York health units, and least complete for East York and North York. In regards to invasive Group A streptococcal infection, East York and Scarborough maintained the most complete data while information from North York and York was most lacking. For pertussis, data was most complete for the Scarborough and Old Toronto health units and, again, least complete for North York.

Over time, the completeness of most of the mandatory data fields has either remained unchanged, or has been improving for each of the diseases examined. It is important that the quality of data within RDIS be as complete as possible, and if anything improve over time, so that long term trends can be examined and new trends in disease identified as they occur due to an accurate basis of comparison against disease patterns from previous years. Of great concern are those data fields that are quite incomplete and not improving, as well as any mandatory data field that is being filled out less often than in the past. For example, the 'Physician Report Date' field is still not well completed overall, and for meningococcal disease is being filled out even less consistently than in previous years. Similarly for GAS, both the 'Deceased?' and the 'Method of Diagnosis' data fields are not being completed as well in recent years as they have in the past. The completeness of the 'Deceased?' data field has remained relatively unchanged for sporadic cases of influenza, which is also of concern due to the high proportion of missing values for this data field each season. Similarly the percentage of missing data for 'Outbreak Associated?' data field has not decreased over time for cases of GAS, despite the large amount of missing information each year for this variable.

Many data fields are left blank within RDIS, leaving open to interpretation whether the information for this field was unavailable, not applicable, or simply a 'No' response. For example, the 'Deceased?' data field rarely contains the 'No' or 'Unknown' response, however appears to accurately designate 'Yes' when a death has occurred. Nonetheless, it cannot be confidently assumed that a case has not resulted in death for all fields with a blank entry. Thus, the large proportion of cases missing information for the 'Deceased?' field render the field almost impossible to use in data trend analysis. As well, the inconsistency in mortality designation affects the ability to combine data fields from different health offices because it is difficult to be sure that all health units were equally vigilant in attempting to obtain mortality information and therefore that the data from each health unit is equally reliable.

4.3 Usage of Non-Mandatory RDIS Data Fields

The use of non-mandatory data fields varies considerably for certain data items across the former health units. This is in part due to the preference of Investigators and CD Managers, who may have expressed a desire to collect information on specific indicators that are of interest locally but are not required by the Ministry of Health and Long Term Care. Other variables may have been collected only when the information was readily available, and not actively sought after. Thus, it would be difficult to combine data from separate RDIS systems for non-mandatory fields, as for most the proportion of missing and/or unknown or unspecified data varies considerably across health units and thus has a high potential for bias. For the purposes of this discussion, the term 'missing' data/ information will include information which is unknown or unspecified.

Each of the non-mandatory data fields was examined for each of the four communicable diseases as well. An evaluation of each of the non-mandatory data fields reveals that the quantity of missing information alone prohibits the combining of data from different health units. This is both because the overall proportion of missing information is quite high for these data fields, and because the amount of missing information varies considerably across the individual health units. Overall, East York and Toronto appear to have the most complete data for non-mandatory fields, and York the most incomplete (Table 2). Caution should be practiced when combining and analyzing data with such a large proportion of missing information, as the potential for bias is considerable.

A minimum of 35% of data is missing for at least one former health unit for all of the non-mandatory data fields (Table 2). This is a significantly high proportion of missing data, and thus the combining of data for these variables from each of the former health units for all years would be cautioned against. In addition, since the 'Hospitalized?', 'Lab Test', 'Vaccination', and 'Treatment' data fields have related variables, these related variables should not be combined for the separate health units either. These related variables include 'Hospital Admission Date', 'Hospital Discharge Date', 'Hospital Name', 'Lab Test Result', 'Lab Specimen', 'Lab Test Date', 'Vaccination Status', 'Vaccination Date', 'Treatment Start Date', and 'Treatment End Date'.

However, for certain data fields the completeness of the data has been improving over time and for recent years, is sufficiently complete to warrant consideration of combining data from separate health units. For example, the 'Sites' variable is missing 0.0% of the time for meningococcal disease from 1997 to 1999 for all 6 former health units, thus data from these years could be combined for analysis. For the 'Outbreak Code' data field, although no outbreaks occurred in 1999, there were no missing outbreak codes for those outbreaks of GAS that occurred in 1997 and 1998 and no missing codes for pertussis outbreaks occurring between 1996 and 1998. However, although the 'Outbreak Code' variable has become sufficiently complete in some instances to warrant combining the data for the former 6 health units for certain years, the mandatory 'Outbreak Associated?' field on which the 'Outbreak Code' data field is based has too large a proportion of missing information to allow this to happen.

Table 2. Range of Missing Data for the Non-Mandatory Data Fields by Disease.

	Sporadic Cases of Influenza		Cases of Meningococcal Disease		Cases of Group A Streptococcal Infection		Cases of Pertussis	
	LOW	**HIGH**	**LOW**	**HIGH**	**LOW**	**HIGH**	**LOW**	**HIGH**
Symptoms	12% (EY)	92% (Yk)	13% (NY)	75% (Yk)	0% (OT)	100% (Yk)	15% (OT)	91% (Yk)
Risk Setting	24% (Sc)	78% (Et)	30% (Sc)	85% (EY)	13% (OT)	75% (Yk)	12% (OT)	68% (Yk)
Hospitalization	28% (EY)	98% (Yk)	3% (NY)	88% (Yk)	2% (NY)	100% (Yk)	38% (EY)	95% (Et)
Complications	33% (OT)	100% (Yk)	49% (Sc)	94% (Et)	0% (EY, Et)	38% (Yk)	73% (OT)	100% (Et)
Outbreak Code	N/A	N/A	N/A	N/A	25% (OT)	N/A	N/A	N/A
Vaccination	16% (OT)	90% (NY)	74% (NY)	100% (Et, EY, Sc)	N/A	N/A	12% (Yk)	83% (NY)
Sites	N/A	N/A	12% (Et)	54% (EY)	0% (EY, Sc, OT)	50% (Yk)	N/A	N/A
Case Disposition	N/A	N/A	4% (OT)	100% (Sc)	0% (Et)	97% (Sc)	3% (EY)	63% (Sc)
Meningitis Subtype	N/A	N/A	15% (Sc)	69% (EY)	N/A	N/A	N/A	N/A
Treatment	N/A	N/A	N/A	N/A	N/A	N/A	49% (EY)	97% (Yk)
Lab Test	N/A	N/A	63% (OT)	92% (EY)	0% (EY, Sc)	50% (Yk)	N/A	N/A
Source of Infection	N/A	N/A	N/A	N/A	32% (NY)	100% (Yk)	N/A	N/A
Risk Factors	N/A	N/A	N/A	N/A	32% (NY)	100% (Yk)	N/A	N/A

*EY=East York; Et=Etobicoke; NY=North York; Sc=Scarborough; OT=Old Toronto; Yk=York
**reported percentages may be biased by small numbers of cases for the following health units: East York, Etobicoke, and York

4.4 Use of the 'Other' Data Fields and Limited Choice of Responses

There is often no 'free' field to record information for a particular data item that does not have the appropriate response available as an option. This is true of several data fields, including the 'Symptoms', 'Method of Diagnosis', 'Risk Setting', 'Method of Detection', 'Complications', 'Disease Site', 'Risk Factors', and other variables. Information contained in these 'Other' categories may be useful to health departments by suggesting new trends of disease or the emergence of different forms or degrees of severity of a particular disease. At times, the proportion of cases reportedly belonging to the 'Other' category for specific data fields is substantial. More useful for disease surveillance would be either the availability of more response options that would reflect the categories indicated by cases recording 'Other' as their response, or a 'free' data field that would allow data entry personnel to manually enter any responses that are not already available in the RDIS system.

4.5 Multiple Data Fields Available to Record the Same Information

The availability of multiple data fields to record the same information results on occasion in data not being recorded where it is mandatory. The 'Deceased?' field is a mandatory field in RDIS to record the death of a case. However, this information may also be recorded in the 'Case Disposition' data field and in the 'Complications' data field. The problem arises when mortality information is entered in only one or two of these fields, and in particular when this information is entered somewhere other than the mandatory 'Deceased?' field. The use of any one of these fields likely will not provide a complete total of the number of deaths that have occurred as a result of a particular disease. Nonetheless, the mandatory 'Deceased?' field is most accurate, and only rarely is mortality information missing from this field.

4.6 Errors in Data Entry

At times, incorrect information is mistakenly entered into the RDIS system. This was obvious for some of the date fields, where for example a case may have had a birth date recorded later than the diagnosis or report date. For example, 6 cases of influenza had such an inaccuracy.

As well, occasionally information that should have been available has not been entered into the RDIS system. For example, certain cases may have had a hospital admission date recorded but are missing a 'Yes' response in the 'Hospitalized?' data field. Similarly, several cases did not have a lab test recorded as occurring, but information elsewhere such as a lab specimen, date or test result suggested that a lab test did occur. These errors occurred for cases of meningococcal disease and invasive Group A streptococcal infection cases.

An additional consideration with the RDIS system is that most errors in data entry will be difficult to detect for the numerous data fields with pull-down menus from which to select the intended option for each information category. Pull-down windows likely decrease the quantity of data entry errors, however they also likely decrease the ability of errors to be detected in the absence of a direct comparison to the original paper reports for each specific case.

5. Summary of Data Quality Concerns: Consequences for Data Usage

Overall, data from the six RDIS systems of the former health units of Toronto Public Health is quite similar, with the main differences observed in the completeness of the data and the definitions favoured for certain data fields.

For certain health units, many of the data fields and in particular the non-mandatory data fields have a high proportion of missing data, thus making the data that is available unreliable. When there is a wide range in the proportion of missing data for a particular data field across several health units, those with higher percentages missing are likely to be biased and not representative of the communicable disease cases for the community. For example, the data may be more likely to be available for the serious cases of disease, or for specific health care professionals who are more vigilant at seeking out and providing public health Investigators with related data for disease cases, etc. If some former health units have a high proportion of missing data for certain fields, it may be advisable that data for these variables not be combined with data from health units with much lower percentages of data missing. It is quite likely that these data sets do not represent similar populations of disease cases.

5.1 Sources of Differences Among the Six RDIS Data Sets

There are several possible causes for the differences among the six RDIS data sets for each communicable disease examined. Firstly, there is an absence of continuous training by designated personnel, from either within each health unit or from the Ministry of Health and Long Term Care, of Public Health Investigators and data entry personnel. In the absence of designated training personnel, health unit staff have themselves trained new employees on the RDIS system. They have therefore trained new staff in an unstructured manner, leaving open the possibility that individual preferences and biases may be passed along from existing staff members to new ones, broadening the differences in data across various health units.

As well, the individual preferences of the specific Communicable Disease Manager(s) and Medical Officer(s) of Health for each health unit likely played a large role in which non-mandatory data fields were filled out on a regular basis. The large abundance of available data fields for each disease prohibits most health units from actively seeking all of the case information all of the time for non-mandatory fields, and to save time and resources often many fields are filled out only if the information is readily available. Thus, there would be a low reproducibility of the non-mandatory portion of the data sets across health units, and with such a high proportion of missing data many of these fields would be rendered useless even within the same health unit.

An additional reason for the large discrepancy is that the definitions for filling out the data fields for cases of communicable disease contained in the RDIS Guidelines and Procedures Manual are too

general and not disease-specific to a great enough degree. Thus, individual health units are again free to interpret these instructions as they may, and these interpretations are not necessarily similar across the various health units.

5.2 Which Data Fields Can Be Combined?

For data fields to be combined, the definitions and interpretations used by each of the health units must be identical for each of the data fields intended to combine. As well, the data fields must be sufficiently complete so that it is highly likely that information contained in each of the data fields is representative of the population from which the data was collected.

Overall, the majority of the mandatory data fields can be combined and analyzed for the 6 former health units of Toronto Public Health. However, there are several mandatory fields that should preferably not be combined, but nevertheless require caution before combining should grouped analysis be desired. These specific data fields include the 'Outbreak Associated?', 'Deceased?', and 'Physician Report Date' data fields. In particular, the 'Outbreak Associated?' and 'Deceased?' data fields provide essential information to the monitoring of the epidemiology and severity of communicable diseases over time, and the main factor prohibiting their usefulness exists in the large proportion of missing data for these fields, in particular for certain health units.

The majority of the non-mandatory data fields cannot be combined for analysis due to a large proportion of missing data, with few exceptions. The only noted exception is contained in the last paragraph of Section 4.3, whereby the 'Sites' variable may be combined for meningococcal disease for the years 1997 to 1999.

Although definitional inconsistencies for the date fields remain a concern, these date fields may still be combined together as the 'Type of Date' data field allows health departments to document the exact source of the date information, and exclude certain date types if they choose. As well, for most communicable diseases the dates are rarely more than a few days to a few weeks apart, leaving a small window of error regardless of the type of date entered. The only time the exact date is essential to note is for the onset of disease in an outbreak situation, as this date would allow susceptible contacts of cases to be identified.

6. Recommendations to Improve Data Quality in the Future

1. Standardization of Definitions for the RDIS Data Fields.

   The definitions employed for each data field must be standardized within and across health departments, and the categories within each field must be evaluated to ensure they appropriately represent information relevant to the data field. This is particularly important for the date fields. To assist in this process, a manual containing detailed protocols for each notifiable disease or category of diseases should be developed, explicitly defining the case definition criteria for the disease, and what information is mandatory to obtain regarding each disease case.

   As well, the data fields and each of their related options should be clearly defined as they relate to the particular disease, so that there is no confusion as to which is the appropriate category to place a response. This manual should be available to each data entry clerk and Investigator, to clarify any questions that may arise during data entry. Any remaining uncertainties should be addressed by the Communicable Disease Manager(s).

2. Training of Public Health Staff

The training of public health staff involved with the use of RDIS must be standardized. Training would help to ensure staff are interpreting data fields in the same manner, and appropriately. Specific personnel should be designated within each health department to train new staff in a consistent manner, so that all personnel receives the same training regardless of the health department in which they are employed, or the period in time in which they were trained. Training staff in an identical manner will help ensure that the interpretation and entry of data is as consistent as possible across the province of Ontario. Training staff from the various health departments should meet regularly to ensure practices and any changes or improvements in training methods are implemented provincially in a standardized manner.

3. Improving Upon the Completeness of Mandatory Data Fields.

The considerable amount of missing data within several of the mandatory data fields in RDIS should be addressed. One method to improve the completeness of the data fields would be to ensure that for each case, absolutely all mandatory data fields must have a response entered and if not, an 'Unknown', 'Unspecified', or 'Not Applicable' designation must be entered. Further the Ministry of Health should designate a certain maximum percentage missing allowed for each data field, estimated perhaps from the average completeness of each data field for the top 20% of health units with the most complete data. These benchmarks may be utilized by health departments to monitor the performance of staff and to evaluate the efficacy of the RDIS system and surveillance procedures. Percentages missing could be examined and compared to benchmarks monthly to ensure prompt action is taken if problems are detected.

4. Improving Upon the Consistency in Use and Completeness of Non-Mandatory Data Fields.

The use of non-mandatory data fields must be standardized. Communicable Disease Managers and Medical Officers of Health must determine which of these fields would be most useful to monitor, as well as how many fields for which it is reasonable to collect information given limited time and resources available to public health staff. As the high proportion of missing data for many non-mandatory fields renders them useless at times, it would be far more prudent too choose a limited number of non-mandatory fields to include in communicable disease surveillance efforts. Staff must then ensure that the data for these designated fields is collected completely and accurately and is therefore useful to the health department.

5. Improving the Flexibility of the RDIS System.

The RDIS system is quite a rigid system in which to record data about disease cases, as many data fields are too general and not disease-specific. This system is not sufficiently flexible to accommodate changes in disease presentation or severity. The system must become more flexible to new and emerging characteristics of communicable diseases, and this may be achieved in two steps.

First, the RDIS system should be altered to provide 'free' space in which to record information for which a pre-existing category does not exist, particularly for those data fields with an 'Other' category. This would allow for the monitoring of changes in the trend of disease, such as new disease sites or increasingly common but previously unrecognized symptoms of a disease.

Second, CD Managers should designate an individual to periodically examine the entries into the 'Other' fields, and suggest possible additional data categories to be added to the data field in RDIS. These categories would then incorporate into the existing system the new trends of disease, to ensure that data entry remains as standardized and up-to-date as possible in the future and to alert other health departments of the emergence or increase in a specific characteristic of disease. As well, adding categories to the existing RDIS system allows for quicker detection and response to changing trends as the 'Other' categories would no longer have to be sorted through.

6. Accurately Documenting Mortality Information.

Improved accuracy in the recording of cases that result in death may result from designating only one data field in which to record this information. As the 'Deceased?' field is the only mandatory data field which records death information, it should also be the *only* data field in which to record it. Thus, the 'Case Disposition' and 'Complications?' data fields should no longer record death information concerning disease cases. The availability of multiple fields for this purpose increases the opportunity for this information to be recorded somewhere other than the mandatory 'Deceased?' field.

7. Data Quality Checks to Detect and Prevent Errors in RDIS Datasets.

To prevent errors in the data contained within RDIS, it is essential that health units perform regular data quality checks, perhaps weekly or monthly, to ensure that the information that has been entered is sensible. For example, data contained in RDIS may be imported into a statistical analysis system such as SPSS. The correct and sensible entry of dates can be verified by subtracting each of the date fields from each other to ensure that the 'Birth Date' occurs before the 'Episode Date', which occurs before the 'Diagnosis Date', which occurs before the 'Physician Report Date', etc.

As well, errors can be detected by ensuring that, for example, whenever information relevant to a laboratory test is present, the 'Lab Test' data field records 'Yes' for a test having been performed. These errors may be detected by performing a cross-tabulation with 'Lab Test' versus each of the related lab fields. Similarly, the 'Hospitalized?' data field, 'Outbreak Associated?' field, 'Vaccination' field, and the 'Treatment' data field may be cross-referenced with their related data fields.

Additional errors may be detected by performing random comparisons between data contained in the RDIS system to the original paper reports on which case information was recorded. This would allow CD Managers to determine how accurately data is being entered into RDIS.

7. Conclusions Regarding Combining Separate RDIS Systems

The Reportable Disease Information System successfully meets the purposes for which it was originally intended for: providing basic information on cases of communicable diseases in Ontario to allow recognition of trends in person, place and time. The information required for such disease surveillance is contained within the mandatory data fields in RDIS, which are the most complete and accurate of the data fields and which for the most part can be combined into one data set from the former 6 health units.

However, before the role of RDIS could be expanded to enhance resource allocation, program planning and evaluation purposes, the completeness of the specific non-mandatory data fields that

provide information useful for these purposes must be improved. At present, these fields are quite incomplete and definitions aren't standardized sufficiently to allow combination of non-mandatory fields, nor to allow these fields to be useful for purposes beyond basic surveillance.

References

Benenson, AS (Ed.). (1995). *Control of Communicable Diseases Manual*, 16th ed. American Public Health Association, Washington, DC. 577p.

Ontario Ministry of Health. *Reportable Disease Information System Guidelines and Procedures Manual.* April 1992.

Statistics Canada. Population Statistics. www.statcan.ca. April 18, 2000.

Appendix A

Influenza Worksheet

CITY OF TORONTO DEPARTMENT OF PUBLIC HEALTH
REPORT OF INFLUENZA

Health Area: - Date Reprted: Incident #:

Patient: DOB: Age:
Sex:

Address: P.C.

Tel: (H)
(B) Physician:
Tel:
other:

Episode Date: (20) ___/___/___ (y/m/d)
Date Type: (25)
___(1) Onset date of symptoms
___(2) Clinic visit or specimen collection date
___(3) Treatment date
___(4) Specimen received date
___(5) Final lab report date
___(6) Report form completion date
___(97) Estimate
___(98) Other: _____

Associated Agent/Organism: (170)
___(397) Influenza virus
___(398) Influenza virus type A
___(399) Influenza virus type B
___(400) Influenza virus type C

Classification: (177)
___(1) Case

Case Disposition: (840)
___(50) Untraceable
___(53) Lost to Follow-up
___(65) Discharged
___(98) Other: _____
___(99) Unknown

Date of Diagnosis: (412) ___/___/___
Date Type: (414)
___(1) Date Specimen Collected
___(2) Date Specimen Received by Lab
___(3) Date of Clinical Diagnosis
___(4) Date of Diagnosis Reported by Client
___(5) Date Report Received by HU
___(99) Unknown
___(98) Other: _____

Method of Diagnosis: (410)
___(22) Virus Isolation
___(26) 4-fold increase/decrease in anitbody titres
___(32) Antigen detection
___(80) Epi link to lab confirmed cases(s)
___(98) Other

Method of Detection: (450)
___(1) Symptoms
___(2) Contact Tracing
___(3) Routine Screening
___(4) Post-Mortem
___(98) Other: _____
___(99) Unknown

Symptoms: (425) _____

CITY OF TORONTO DEPARTMENT OF PUBLIC HEALTH

Patient's Name: Incident #:

Risk Setting: (100450)

___(3) Day Care
___(4) Residential Facility
___(5) Correctional Facility
___(6) Facility For the Developmentally Disabled
___(7) Shelter/Rooming House
___(9) Hospital
___(10) Travel/ Lived in Endemic Area
___(11) Medical Office
___(12) Workplace
___(15) Home
___(16) School
___(97) Unspecified
___(98) Other
___(99) Unknown

Specify Setting:(100455)_____

Risk of Transmission: (200450)

___(10) Institution ___(20) Travel ___(98) Other
___(15) School/Daycare ___(25) Occupational ___(99) Unknown

Comments: _____

Outbreak Associated? (860) ___Yes ___No ___Unspecified
Outbreak Code Number: (862)

Prior Vaccination:

___(70) Influenza Vaccine Date: _____
___(97) No Prior Vaccination
___(99) Unknown

Prior Prophylaxis:

___(10) Amantadine Date: _____
___(97) No Prior Prophylaxis
___(99) Unknown

Complications: (420)

___(1) Pneumonia ___(20) Reye Syndrome ___(98) Other
___(15) Death Date: _____ ___(60) None ___(99) Unknown

Hospitalized?: (400) ___No ___Yes Admission: ___/___/___ (yy/mm/dd)
Hospital: _____ Discharge: ___/___/___

Episode Status: ___Meets case definition ___Does not meet case definition

Comments: _____

_____

_____

Note to file: ___Yes ___No

Signature: _____ Date: _____

(Personal information contained on this form is collected under the authority of the Health Protection and Promotion Act 1983, and is used to follow Reportable case investigations and for statistical purposes. Questions about this collection should be directed to the Department of Public Health F.O.I. Coordinator at 392-7407. [ ] _____)

CITY OF TORONTO DEPARTMENT OF PUBLIC HEALTH
REPORT OF MENINGOCOCCAL MENINGISTIS / MENINGOCOCCEMIA

Health Area: - Date Reported: Incident #:

Patient: DOB: Age:
Sex:

Address: P.C.

Tel: (H) Physician:
(B) Tel:
other:

Episode Date: (20) ___/___/___ (y/m/d)
Date Type: (25)
___(1) Onset date of symptoms
___(2) Clinic visit or specimen collection date
___(3) Treatment date
___(4) Specimen received date
___(5) Final lab report date
___(6) Report form completion date
___(97) Estimate
___(98) Other: _____

Subtype: (175)
___(76) Not groupable/typeable
___(82) Group A
___(84) Group C
___(89) Group W-135
___(91) Group Y
___(999) Unspecified

Classification: (177) ___(1) Case

Case Disposition: (840)
___(50) Untraceable
___(53) Lost to Follow-up
___(65) Discharged
___(98) Other: _____
___(99) Unknown

Date of Diagnosis: (412) ___/___/___
Date Type: (414)
___(1) Date Specimen Collected
___(2) Date Specimen Received by Lab
___(3) Date of Clinical Diagnosis
___(4) Date of Diagnosis Reported by Client
___(5) Date Report Received by HU
___(99) Unknown
___(98) Other: _____

Method of Diagnosis: (410)
___(400) N. Meningitides culture/detection - Meningitis
___(402) Clinical case with purpural fulminans - Meningitis
___(404) N. Meningitides culture/detection - Meningococcemia
___(406) Clinical case with purpura fulminant - Meningococcemia

Method of Detection: (450)
___(1) Symptoms
___(2) Contact tracing
___(4) Post Mortem
___(98) Other: _____
___(99) Unknown

Symptoms: (425) _____

Sites: (428)
___(12) Disseminated
___(30) Bone and joint
___(33) Central nervous system
___(98) Other

CITY OF TORONTO DEPARTMENT OF PUBLIC HEALTH

Patient's Name: Incident #:

Laboratory Tests

Type of Test	Specimen	Result	Date	Comments

Risk setting: (100450)

___(3) Day Care
___(4) Residential Facility
___(5) Correctional Facility
___(6) Facility For the Developmentally Disabled
___(7) Shelter/Rooming House
___(9) Hospital
___(10) Travel/ Lived in Endemic Area
___(12) Workplace
___(15) Home
___(16) School
___(19) Sexual rendezvous outside home
___(97) Unspecified
___(98) Other
___(99) Unknown

Specify Setting (e.g. school name):(100455)_____

Risk of Transmission: (200450)

___(10) Institution ___(20) Travel ___(98) Other
___(15) School/Daycare ___(25) Occupational ___(99) Unknown

Comments: _____

Outbreak Associated? (860) ___Yes ___No ___Unspecified
Outbreak Code Number: (862)

Prior Vaccination:

___(74) Meningicoccal Vaccine Date: _____
___(97) No Prior Vaccination (99) Unknown

Prior Prophylaxis:

___(2) Rifampin From Date: _____ To Date: _____
___(97) No Prior Prophylaxis (98) Other (99) Unknown

Complications: (420)

___(15) Death Date: _____ Cause: _____
___(19) Residual neurological deficit ___(60) None
___(24) Meningicoccemia ___(98) Other ___(99) Unknown

Hospitalized?: (400) ___No ___Yes Admission: ___/___/___ (yy/mm/dd)
Hospital: _____ Discharge: ___/___/___

Episode Status: ___Meets case definition ___Does not meet case definition

Note to file: ___Yes ___No

Signature: _____ Date: _____

(Personal information contained on this form is collected under the authority of the Health Protection and Promotion Act 1983, and is used to follow Reportable case investigations and for statistical purposes. Questions about this collection should be directed to the Department of Public Health F.O.I. Coordinator at 392-7407. [ ] _____)

CITY OF TORONTO DEPARTMENT OF PUBLIC HEALTH
REPORT OF GROUP A STREPTOCOCCUS, INVASIVE

Health Area: - Date Reported: Incident #:

Patient: DOB: Age:
Sex:
Address: P.C.
Tel: (H) Physician:
(B) Tel:
other:

Episode Date: (20) ___/___/___ (y/m/d)
Date Type: (25)
___(1) Onset date of symptoms
___(2) Clinic visit or specimen collection date
___(3) Treatment date
___(4) Specimen received date
___(5) Final lab report date
___(6) Report form completion date
___(97) Estimate
___(98) Other: _____

Classification: (177)
___(1) Case ___(2) Carrier/Infection ___(98) Other ___(99) Undetermined

Case Disposition: (840)
___(50) Untraceable ___(65) Discharged ___(99) Unknown
___(53) Lost to Follow-up ___(98) Other: _____

Date of Diagnosis: (412) ___/___/___
Date Type: (414)
___(1) Date Specimen Collected
___(2) Date Specimen Received by Lab
___(3) Date of Clinical Diagnosis
___(4) Date of Diagnosis Reported by Client
___(5) Date Report Received by HU
___(99) Unknown
___(98) Other: _____

Method of Diagnosis: (410)
___(583) Isolation of GAS from a normally sterile site with or without evidence of clinical severity
___(584) Isolation of GAS from a nonsterile site with evidence of clinical severity
___(585) Serological confirmation of GAS infection with clinical evidence of severity
___(99) Undetermined

Method of Detection: (450)
___(1) Symptoms
___(2) Contact Tracing
___(3) Routine Screening
___ (4) Post Mortem
___(98) Other: _____
___(99) Unknown

Symptoms: (425)
List: _____

_____

Laboratory Tests

Type of Test	Specimen	Result	Date

Comments

CITY OF TORONTO DEPARTMENT OF PUBLIC HEALTH

Patient's Name: Incident #:

Sites: (428)

___(33) Central Nervous System
___(41) Bacteremia (with no focus of infection)
___(42) Upper Respiratory Tract
___(43) Lower Respiratory Tract
___(44) Peritonitis (abdomen)
___(45) Arthritis (joints)
___(46) Osteomyelitis (bone)
___(47) Soft tissue (skin, fascia, muscle)
___(48) Surgical wound infection
___(49) Postpartum fever, endometritis
___(98) Other
___(99) Unknown

Source of Infection: (460)
___(4) Person-to-Person ___(8) Wound ___(98) Other ___(99) Unknown
Specify Source (If known): (465)_____

Risk Setting: (100450)
___(3) Day Care
___(4) Residential Facility
___(5) Correctional Facility
___(6) Facility For the Developmentally Disabled
___(7) Shelter/Rooming House
___(9) Hospital
___(11) Medical Office
___(15) Home
___(97) Unspecified
___(98) Other
___(99) Unknown
Specify Setting: (100455) _____

Risk Factor: (470)
___(4) HIV infection / AIDS
___(7) Illicit drug user (inj. or non-inj.)
___(12) Injection drug user
___(16) Underlying medical conditions
___(20) close contact of known GAS infection
___(21) Peripartum
___(22) Trauma within past month
___(23) Varicella within past month
___(24) Chronic dermatological condition
___(46) Dialysis
___(54) Invasive surgical/dental procedure
___(98) Other ___(99) Unknown

Outbreak Associated? (860) ___Yes ___No ___Unspecified
Outbreak Code Number: (862)

Complications: (420)
___(15) Death Date: _____
___(40) Hepatitis or liver function abnormalities
___(44) Adult Respiratory Distress Syndrome(ARDS)
___(45) Soft tissue necrosis including necrotizing faciitis or myositis or gangrene
___(49) Disseminated intravascular coagulation(DIC)
___(41) Hypotenstion
___(42) Renal failure
___(43) Coagulapathy/bleeding disorder
___(46) Amputation
___(50) Erythematous
___(98) Other

Hospitalized?: (400) ___No ___Yes Admission: ___/___/___ (y/m/d)
Hospital: _____ Discharge: ___/___/___

Episode Status: ___Meets case definition ___Does not meet case definition
Note to file: ___Yes ___No

Signature: _____ Date: _____

(Personal information contained on this form is collected under the authority of the Health Protection and Promotion Act 1983, and is used to follow Reportable case investigations and for statistical purposes. Questions about this collection should be directed to the Department of Public Health F.O.I. Coordinator at 392-7407. [ ] _____)

Pertussis Worksheet

CITY OF TORONTO DEPARTMENT OF PUBLIC HEALTH
REPORT OF PERTUSSIS

Health Area: - Date Reported: Incident #:

Patient: DOB: Age:
Sex:

Address: P.C.

Tel: (H) Physician:
(B) Tel:
other:

Episode Date: (20) ___/___/___ (y/m/d)
Date Type: (25)
___(1) Onset date of symptoms
___(2) Clinic visit or specimen collection date
___(3) Treatment date
___(4) Specimen received date
___(5) Final lab report date
___(6) Report form completion date
___(97) Estimate
___(98) Other: _____

Classification: (177)
___(1) Case

Case Disposition: (840)
___(50) Untraceable ___(65) Discharged ___(99) Unknown
___(53) Lost to Follow-up ___(99) Other _____

Date of Diagnosis: (412) ___/___/___
Date Type: (414)
___(1) Date Specimen Collected
___(2) Date Specimen Received by Lab
___(3) Date of Clinical Diagnosis
___(4) Date of Diagnosis Reported by Client
___(5) Date Report Received by HU
___(99) Unknown
___(98) Other: _____

Method of Diagnosis: (410)
___(20) Positive culture
___(28) Positive fluorescent antibody test
___(420) Epi link to lab confirmed case and characteristic cough lasting 2+ weeks
___(422) Clinical case including characteristic cough
___(98) Other

Method of Detection: (450)
___(1) Symptoms
___(2) Contact Tracing
___(98) Other: _____
___(99) Unknown

Symptoms: (460)
List: _____

_____

CITY OF TORONTO DEPARTMENT OF PUBLIC HEALTH

Patient's Name: Incident #:

Risk setting: (100450)

___(3) Day Care
___(4) Residential Facility
___(5) Correctional Facility
___(6) Facility For the Developmentally Disabled
___(7) Shelter/Rooming House
___(9) Hospital
___(10) Travel/ Lived in Endemic Area
___(11) Medical Office
___(12) Workplace
___(15) Home
___(16) School
___(17) Local Camping
___(23) Vacation Property
___(97) Unspecified
___(98) Other
___(99) Unknown

Specify Setting (e.g. school name):(100455)_____

Risk of Transmission:

___(200) Institution ___(215) Occupational ___(99) Unknown
___(205) School/Daycare ___(98) Other

Comments: _____

Outbreak Associated? ___Yes ___No ___Unsp. Outbreak #:

Prior Vaccination/Prophylaxis:

Vaccine	Number	Date	Status	Comments
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____

Vaccine Choices:

(8) DPT-Polio
(9) DPT
(30) Pertussis vaccine
(97) No Prior Vaccination
(99) Unknown
(130) PENTA (DPT-Polio/Act-HIB)

Status Choices:

(1) Complete as recommended, adequate dose
(2) Complete as recommended, inadequate dose
(3) Not completed, medical contraindication
(4) Not completed, prior immunity by disease
(5) Not completed, personal exemption
(6) Not completed, no other reason
(7) Not completed, inadequate compliance
(8) Not completed, side effects
(98) Other (99) Unknown

Treatment Type/Prescribed Date Started

___(7) Erythromycin _____ ___(97) No drugs prescribed
___(41) Trimethoprim-Sulfamethoxazole _____ ___(99) Uknown
___(98) Other drug _____

Complications: ___(1) Pneumonia ___(3) Encephalitis ___(15) Death Dt: _____
___(18) Seizures ___(60) None ___(98) Other ___(99) Unknown

Hospitalized?: (400) ___No ___Yes Admission: ___/___/___ (yy/mm/dd)
Hospital: _____ Discharge: ___/___/___

Episode Status: ___Meets case definition ___Does not meet case definition
Note to file: ___Yes ___No

Signature: _____ Date: _____

(Personal information contained on this form is collected under the authority of the Health Protection and Promotion Act 1983, and is used to follow Reportable case investigations and for statistical purposes. Questions about this collection should be directed to the Department of Public Health F.O.I. Coordinator at 392-7407. [ ] _____)

Appendix B

Glossary of Terms

The following glossary of terms is divided into three sections. The first section discusses the definitions related to the surveillance of disease in Ontario. These definitions were taken from the RDIS Procedure Manual (1992). The second section, again deriving from the RDIS Procedure Manual, details the definitions of each of the data fields used in the surveillance of the four communicable diseases described in this report. The third section of the report describes the four communicable diseases whose trends are summarized in this report. The definitions for this third section of the glossary were taken from the Control of Communicable Diseases Manual (Benenson, 1995).

1. Disease Surveillance Terms

Case Definition – Criteria established by the Ministry of Health and Long Term Care for surveillance purposes. Each reportable disease has a case definition which outlines the criteria necessary to confirm that episode of disease.

Client – Refers to the individual reported to have a notifiable disease, and for whom the disease record is created.

Clinical Diagnosis – Verification of signs and symptoms by an appropriately qualified health professional.

Closed Case – A case in which further investigation is no longer necessary or possible. Mandatory items must be entered before a case is closed, although an override is available for unobtainable information in non-category list mandatory fields.

Confirmed Case – An episode which meets the case definition established by the Ministry of Health and Long Term Care. This generally consists of laboratory confirmation of an organism or epidemiologic linkage to one or more confirmed case(s).

Database – A computerized collection of data organized for rapid search and retrieval. In RDIS, the Registry database stores the surveillance and case management information.

Epidemiologically Confirmed Case – An episode lacking confirmation, but where an epidemiological association exists to one or more confirmed case(s).

Episode – An episode is the general term that is used in RDIS to denote a possible or confirmed occurrence of a reportable disease. One client (person) can experience multiple disease episodes (of different diseases or if possible, the same disease more than once).

Investigator – Health department staff person who has primary responsibility for the direct investigation of an episode.

Mandatory Item – An item specified by the Ministry of Health as one which must be entered as part of the basic surveillance dataset. All mandatory items must be entered before an episode can be closed.

Notifiable Diseases – Though not specifically referred to in legislation, this term is used in the RDIS manual to include those diseases designated as reportable, communicable, or virulent under the Health Protection and Promotion Act, 1983.

RDIS – Reportable Disease Information System. An information system designated to support the surveillance, reporting, and control of notifiable diseases in Ontario. As a decentralized system, components exist within the Ministry of Health and within the 42 Ontario health departments.

Surveillance – Scrutiny of and watchfulness over the factors contributing to a health and/or safety hazard or disease or the spread of infection.

2. Definitions of the Data Fields

Associated Organism/ Agent – The organism or agent that is the cause of the disease episode.

Case Disposition – This item lists several options with regard to the eventual outcome of the episode. More than one option may be chosen if appropriate.

Classification – Identifies if the disease episode refers to a case, carrier, or other, from the options available. All diseases have case as an option. This item is intended to capture information regarding the type of episode reported.

Complications – Medical complications attributed to the disease episode.

Date Admitted to Hospital – For those cases that were hospitalized, the date the case was admitted to hospital.

Date Discharged from Hospital – For those cases that were hospitalized, the date the case was released from hospital.

Date of Birth (Age) – The client's date of birth, or age in years during this particular disease episode. If the date of birth is entered, the RDIS system automatically calculates the age. If only the age is entered, the RDIS system automatically approximates the date of birth by comparing the age of the client to the episode date.

Date of Diagnosis (and Type) – The earliest date the disease episode was first diagnosed. This date may be one of the following: date specimen collected; date specimen received by lab; date of clinical diagnosis; date of diagnosis reported by client; date report received by health unit; other; unknown.

Date Vaccine Administered – The date the individual received the vaccine specified in the 'Vaccination' data field.

Deceased? – Is the client deceased?

Disease – The name of the disease.

Disease Sites – The area(s) of the body that are affected by the disease episode.

Drug (Prophylactic) – Any drugs that have been taken in order to prevent this particular infection.

Episode Date (and Type) – Intended to represent the earliest date of onset of the disease episode, most often represented by the date of the onset of symptoms. If the date of the onset of symptoms is not known, the next available date should be selected from the following list: clinic visit or specimen collection date; treatment date; specimen received date; final lab report date; today's date (date entered); or other.

Episode Status – captures information on whether the episode meets the case definition, is pending confirmation or is not confirmed. Only the cases that meet the case definition are of interest to the Ministry of Health for surveillance purposes.

Gender – The sex of the client, male, female, or unknown.

Hospitalized? – Was the client hospitalized for this disease episode?

Hospital Name – The name of the hospital the case was admitted to.

Incident Number – The number of the disease episode case to which the report is linked.

Lab Report Date – The date the specimen was submitted to the lab. If this date is unavailable, the date that the specimen arrived at the lab or the date of the lab report is used.

Lab Specimen Type or Site – The specimen type or area of the body from which the lab test was obtained.

Lab Test – The type of lab test.

Lab Test Date – The date the lab test specimen was submitted to the lab. If this date is not available, the date the specimen arrived at the lab or the lab report date is used.

Lab Test Result – The result of the lab test.

Method of Detection – The method by which the disease episode was discovered.

Method of Diagnosis – For an episode to be classified as a case, this item/field must match the case definition criterion that was used to confirm the case.

Outbreak Associated? – Was this episode associated with an outbreak?

Outbreak Code Number – The outbreak number assigned by the health department for this outbreak. It is recommended that health departments adopt the following system for numbering outbreaks: “aaaa” is the Health Unit Master Number, “bb” is the file number given to the outbreak by the Health Unit and “yy” is the year (last two digits).

Physician Report Date – The date the physician reported the disease episode to public health authorities.

Postal Code – The postal code of the residence of the client. If unknown, this field is to be left blank as no other postal codes should be substituted.

Prophylaxis – Any drugs that have been taken in order to prevent this particular infection.

Responsible Health Unit – The health department in which the client resides, and that is reporting the disease episode.

Risk Factors – Factors that may have had an impact on the client increasing their risk of acquiring the disease episode.

Risk Setting – The place/environment in where the client may have acquired the infection.

Sites – The area(s) of the body that are affected by the disease episode.

Source of Infection – Indicates how the client most likely acquires the infection.

Subtype – The A more detailed typing for the organism/ agent.

Symptoms – The symptoms/ signs experienced by the client during the disease episode.

Transmit to Ministry? – Do you want to transmit the disease episode to the Ministry of Health system?

Treatment End Date – The date the treatment was stopped or completed.

Treatment Start Date – The date treatment began for the case.

Treatment Status – The level or state of the treatment administered to the client for the disease episode.

Treatment Type/ Prescribed – The type of treatment of drugs prescribed to the client for the disease episode.

Vaccine – The name of the vaccine given for the prevention of this specific disease episode.

Vaccination Date Administered – The date in which the vaccine was given.

Vaccination Status – The level/state of the vaccine induced protection of the client for the particular disease in question.

3. Definitions of the Communicable Diseases

Influenza

An acute viral disease of the respiratory tract characterized by fever, headache, myalgia, prostration, eoryza, sore throat, and cough. Three types of influenza virus are recognized: A, B, and C. Type A includes three subtypes (H1N1, H2N2, H3N2) that have been associated with widespread epidemics and pandemics. Type B has been associated with regional/ widespread epidemics, and Type C has been associated with sporadic cases and minor localized outbreaks. The virus type is determined by the antigenic properties of the two relatively stable structural proteins, the nucleoprotein and the matrix protein.

Airborne spread predominates among crowded populations in enclosed spaces, however transmission of the influenza virus may also occur by direct contact under certain conditions. Influenza illness may occur in pandemics, epidemics, localized outbreaks, and as sporadic cases. Epidemics occur in Canada almost every year, and are caused mainly by Type A viruses, often by Type B viruses, or by both viruses at the same time. In temperate zones such as Canada, epidemics tend to occur in the winter.

Meningococcal Disease

An acute bacterial disease, characterized by sudden onset with fever, intense headache, nausea and often vomiting, stiff neck and, frequently, a petechial rash with pink macules or, very rarely, vesicles. Delerium and coma often appear; occasional fulminating cases exhibit sudden prostration, ecchymoses and shock at onset. Formerly, case-fatality rates exceeded 50%, but with early diagnosis, modern therapy and supportive measures, the case-fatality rate is between 5 and 15%.

Not everyone who acquires infection will progress to invasive disease. A small minority will progress to invasive disease, however, characterized by one or more clinical syndromes, include bacteremia, sepsis, meningitis, or pneumonia.

Pertussis (Whooping Cough)

An acute bacterial disease involving the respiratory tract. The initial catarrhal stage has an insidious onset with an irritating cough that gradually becomes paroxysmal, usually within 1 to 2 weeks, and lasts for 1 to 2 months or longer. Paroxysms are characterized by repeated violent coughs; each series of paroxysms has many coughs without intervening inhalation and can be followed by a characteristic crowing or high-pitched inspiratory whoop. Paroxysms frequently end with the expulsion of clear, tenacious mucus, often followed by vomiting. Transmission of the diseases is primarily be direct contact with discharges from respiratory mucous membranes of infected persons by the airborne route, probably by droplets. The incubation period is usually 6 to 20 days.

Pertussis is an endemic disease common to children, especially young children, with outbreaks occurring periodically. Morbidity is slightly higher in females than males. A marked decline in the incidence and mortality rates of pertussis has occurred during the last 4 decades. The number of case fatalities is low for immunized populations; pneumonia is the most common cause of death. However, fatal encephalopathy, probably hypoxic, and inanition from repeated vomiting occasionally occur. Approximately 80% of deaths are among children less than 1 year of age, and 70% are under 6 months. The case-fatality rate is less than 1% in infants less than 6 months old.

Streptococcal Diseases Caused by Group A (Beta Hemolytic) Streptococci

Group A streptococci cause a variety of diseases, most frequently streptococcal sore throat and streptococcal skin infection (impetigo or pyoderma). Other diseases include scarlet fever, puerperal fever, septicemia, erysipelas, cellulitis, mastoiditis, otitis media, pneumonia, peritonsillitis, wound infections, and rarely, necrotizing fasciitis, rheumatic fever, and a toxic shock-like syndrome. In outbreaks, one form of clinical disease predominates.

Appendix C

Population Data

Population data used in this report for the calculation of rates were obtained from the Central East Health Information Partnership and were produced by Statistics Canada. The estimates are for the Metropolitan Toronto population by Census Subdivision for the years 1990 to 1999.

Sections of this report summarizing the four communicable diseases examined involve categories of cases, for example by age group or responsible health unit. Such categorization frequently resulted in small numbers from which stable, reliable rates could not be calculated. Thus numbers and percentages were employed to describe the distribution of cases instead. To aid in the understanding of the distribution of the population within these categories, the following tables may be consulted:

Table 1. Percentage of Population of Toronto Residing in Each Former Health Unit Jurisdiction by Year, 1990 to 1999.

	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
East York	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.6
Etobicoke	13.6	13.6	13.6	13.7	13.7	13.7	13.8	13.8	13.8	13.8
North York	24.9	24.7	24.7	24.7	24.7	24.7	24.7	24.7	24.7	24.5
Scarborough	22.9	23.0	23.1	23.2	23.3	23.3	23.4	23.5	23.6	23.6
Old Toronto	28.0	28.0	27.9	27.8	27.7	27.6	27.5	27.4	27.3	27.3
York	6.2	6.2	6.2	6.2	6.1	6.1	6.1	6.1	6.1	6.2

It is important to remember that differences observed in the number of cases reported under each former health unit are not only reflective of population differences across these regions. Other factors that may influence the number of cases reported in a region include differences in reporting practices of physicians in the area and differences in individual tendency to seek medical treatment.

Table 2 may be consulted when examining the number of cases within each age category for specific communicable diseases. Note that in the summaries, some of these age groups are combined due to very low numbers of cases.

Table 2. Percentage of Population of Toronto Within Each Age Group by Year, 1990 to 1999.

Age Group (years)	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
0 – 9	11.3	11.3	11.6	11.9	11.9	12.0	12.2	12.4	12.6	12.7
10 – 19	11.1	10.9	10.9	10.8	10.9	10.9	11.0	10.9	10.8	10.7
20 – 29	20.5	19.7	19.1	18.4	17.7	17.0	16.2	15.9	15.1	13.8
30 – 39	17.7	17.8	17.7	17.9	18.2	18.6	18.9	18.8	19.0	19.2
40 – 49	12.5	12.9	13.2	13.5	13.7	14.0	14.3	14.4	14.5	15.0
50 – 59	9.9	9.9	9.8	9.8	9.8	9.8	9.8	10.1	10.4	10.7
60 – 69	9.1	9.3	9.2	9.1	9.0	8.8	8.6	8.4	8.3	8.4
70 – 79	5.2	5.5	5.7	5.8	5.9	6.0	6.0	6.1	6.3	6.4
80 +	2.6	2.7	2.8	2.9	2.9	2.9	2.9	3.0	3.0	3.1