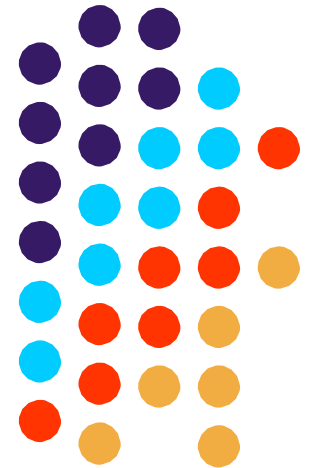


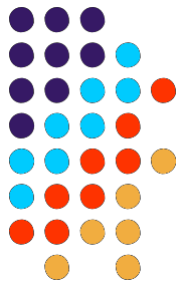
# DATA STANDARDS

[www.OntarioHealthProfiles.ca](http://www.OntarioHealthProfiles.ca)

Ontario Community Health  
Profiles Partnership (OCHPP)

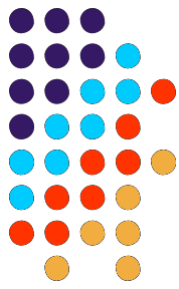


# Purpose of the Data Standards



This guide describes the steps taken to ensure that the information on the [www.OntarioHealthProfiles.ca](http://www.OntarioHealthProfiles.ca) website is accurate, complete and useful and that users are aware of the limitations. Since users are interested in looking for differences between areas, the objective is to reduce the amount of difference that may be due to the quality of the data (variability, small numbers, small sample size, calculation errors, representativeness of the sample, misunderstanding or misinterpretation of the meaning of the indicator, etc.). Epidemiological practice standards and small area analysis guidelines used by other organizations were consulted in developing these data standards.

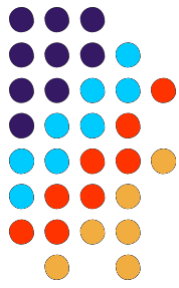
# Using All Available Information to Draw Conclusions



- The data, maps and other materials found on this site may be used in research to identify the range of possible reasons for observed differences in health.

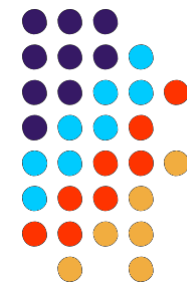
Caution is advised in drawing conclusions based on limited data.

# Data Standards: Outline



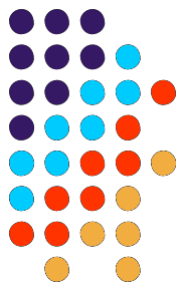
- Selecting Indicators
- Health Indicators
- Spatial Issues
- Demographic Effects
- Random Effects
- Reporting Standards
- Quality Assurance (QA)
- Data Sources & Limitations

# Selecting Indicators: Criteria



|   |   |
|---|---|
| Indicators are Relevant to Reducing health Inequalities | Comparable across time and place;<br>Common or “core” indicator (APHEO, CPHI, DHC, ICES, FCM, Statistics Canada, Other cities, CTPHC);<br>Address complexity, size and diversity of Ontario;<br>Indicator has demonstrated relationship to health inequalities or is being used in a similar context.   |
| Indicators are of Good Quality                          | Consistent definition and data collection methods are used, data integrity is maintained, data is obtained from reliable sources, calculation is transparent and can be reproduced, missing cases are identifiable, standards for small area analysis and reporting are applied, data is available for all areas of the city, limitations on use and interpretation are reported. |
| Supplements existing sources of information             | Indicators supplement other community profile initiatives and are developed in coordination with other information providers.   |
| Useful  | User reviewed, user input, user-identified priorities.  |

# Health Indicators

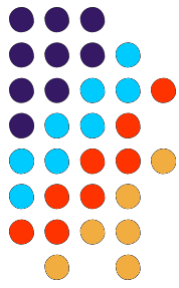


- Age specific or age standardized rates
- Indicator definitions
- Health indicators across the life span will be included: e.g. determinants of health, perceived health, use of prevention and treatment, and -health outcomes (such as, -mortality, disease prevalence, medication use.)

## Prevalence and Incidence Rates

- Rate Ratios: area rate divided by reference area rate to identify policy significance. Rates higher (H), lower (L) or not significantly different (NS) from city rate reported.

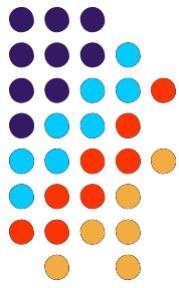
# Data are based on Place of Residence



- All health data are based on residence\* of individuals not where service was provided.
- All health data are geocoded to small area units such as dissemination areas or census tracts which are then aggregated up to other geographic levels.
- The total in the profile tables is the aggregate of all geocoded data.

\*The only data based on place of occurrence is police data and mapping of service sites

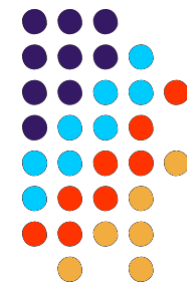
# Area versus Individual Measures



- Neighbourhoods and planning area rates represent an “average” of the individuals living in the area. Individual, family and households incomes can vary widely as many areas are mixed income communities.
- Area rates cannot be assumed to apply to all the individuals living in the area. For example if 40% of a neighbourhood’s residents are low income, and 40% of residents report using a health care service, it cannot be assumed that all those using the service were the low income residents.
- Average area SES characteristics cannot be attributed to the individuals living within this area.



# Accounting for Demographic Effects



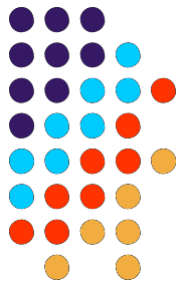
## Demographic Composition

- Variations based on the age and sex make up of an area can explain the observed differences in health events that are known to vary by age and sex.
- Example: a neighbourhood with a high proportion of older adults 75+ will have higher rates of chronic diseases and disabilities that may be explained by these age differences

## Strategy for accounting for age/sex effects

- Age standardized rates by male, female
- Age specific rates where the events or indicators are concentrated (e.g. mammograms among females age 50-69)
- Identify sites located in an area that include a concentration of specific populations (e.g. residences for pregnant teens, long term care facilities. etc.)

# Minimizing Random Effects



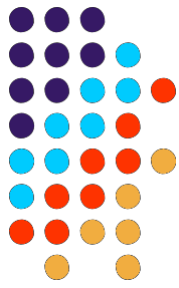
## Random noise:

- Variations based on size of numerator and denominator that can lead to instability in rates because the event is infrequent (rare events) or the number of people in the area that the rate applies to is small.
- Example: A small increase in the number of births among a small population of female teens could double the rate but it reflects too small a number of events to be important for planning. It could be a one-time thing.

## Strategy:

- Reporting standards
- Combine up to 5 years to obtain reportable information
- Combine geographic areas – report only for larger areas
- Coefficient of Variation used in Canadian Community Health Survey (CCHS) survey data
- Confidence intervals

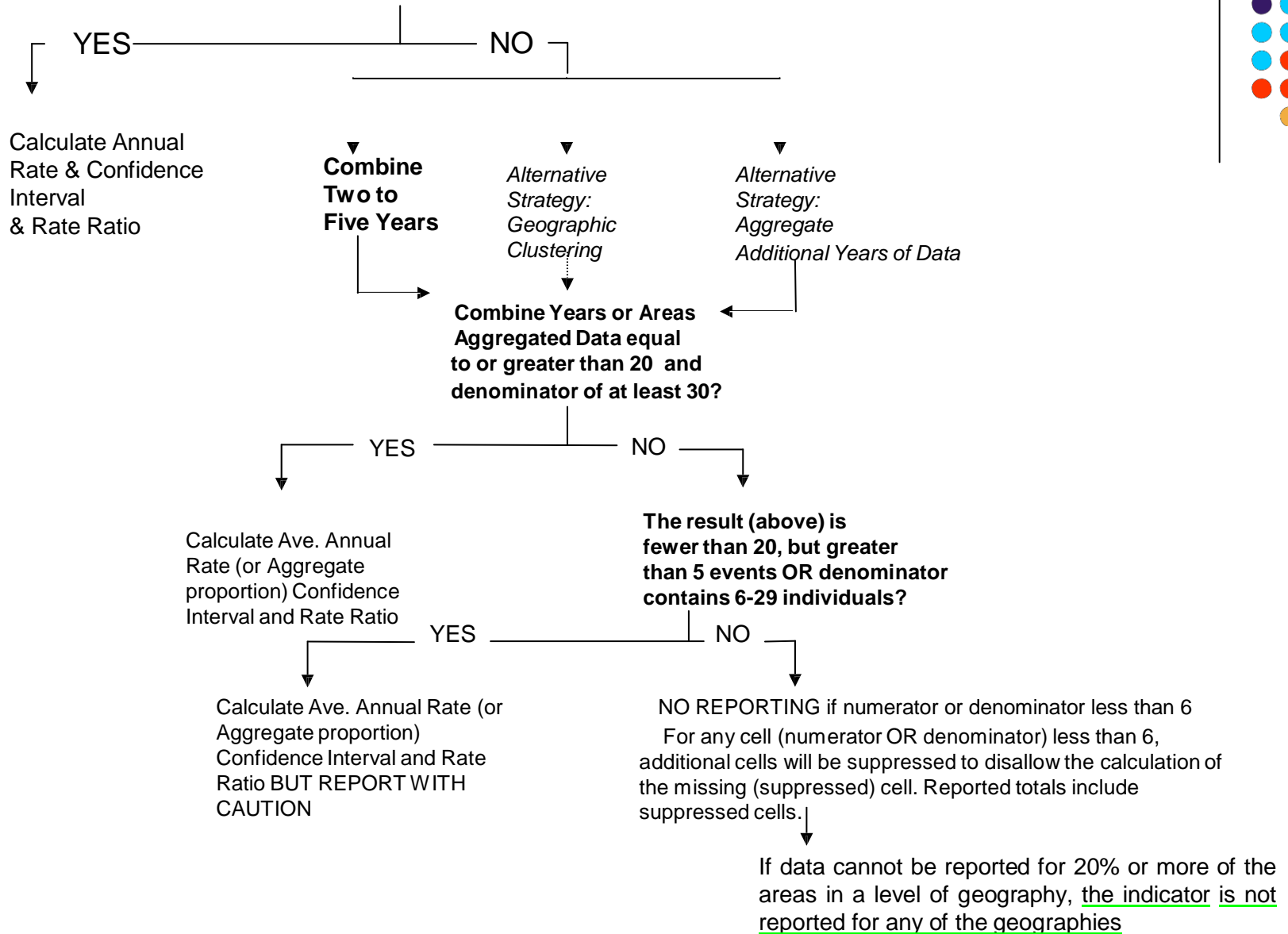
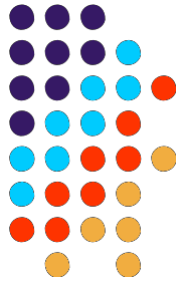
# Ethics & Reporting Standards



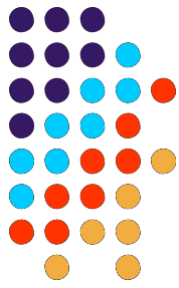
- Full reporting if numerator at least 20 and denominator at least 30
- Reporting with caution if numerator contains 6-19 events OR denominator contains 6-29 individuals.
- No reporting if numerator or denominator less than 6.  
For any cell (numerator OR denominator) less than 6, additional cells will be suppressed to disallow the calculation of the missing (suppressed) cell. Reported totals include suppressed cells.
- Aggregate data for areas or years (2-5 years) for larger sample or population
- No individual level data

# OCHPP SMALL NUMBERS FLOW CHART

Annual # of cases equal to/greater than 20 and denominator of at least 30?

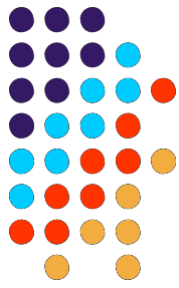


# Mapping Standards



- Map variable at the smallest geographic level for which the majority of extreme, policy relevant rates (e.g. 20% > or < than total rate) are statistically significant (95% confidence intervals)
- Data must be reportable for at least 80% of the units in the geographic level (e.g. if rates for one of the minor areas cannot be reported, the variable will not be mapped at that geographic level).
- For Health Indicators, identify which rates are statistically significant

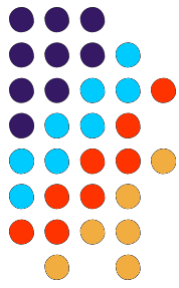
# CCHS Reporting Standards



Use of the Canadian Community Health Survey (CCHS) data requires:

- 1. Checking the unweighted estimates to make sure that the numerator of each cell is not less than 10 for the Ontario Share File, or 30 for the PUMF.
- 2. Checking the efficient of variation (either using CV look-up tables or bootstrapping to create CVs) and follow the release guidelines

# Apply CCHS Sampling Variability Standards



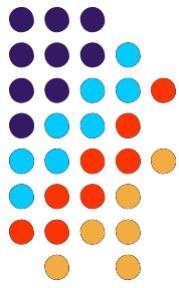
Unqualified (CV= 0.0 – 16.5) Estimates general unrestricted release.

Marginal (CV=16.6 – 33.3) Estimates considered for general unrestricted release but should be accompanied by a warning of high variability associated with estimates. (Footnote on table)

Unacceptable (CV> 33.3) Estimates of unacceptable quality. Conclusions based on these data will be unreliable and most likely invalid and should not be reported.

*The CCHS 1.1 data used on the website was prepared for this purpose by Statistics Canada.*

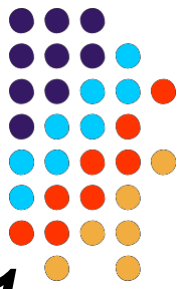
# Data Quality (QA)



## Data Checks:

- Consistent with published data
- Consistent with internal reports/analysis
- Confirmed by independent analysis
- Confirmed by rerunning program
- Do manual computations
- Incorporate formula checks to worksheets
- State data limitations, % missing, representativeness of sample
- Documentation of QA checks





# Data Sources & Limitations

## ***Canada Census 1991, 1996, 2001, 2006, 2011***

(Statistics Canada)

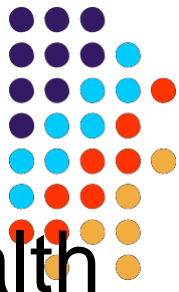
### **Strengths:**

- Best (and only) source of social and demographic info for the entire population (some exceptions)
- Large # of variables: over 1,500

### **Limitations:**

- Census undercount: 5.17% for the Toronto Census Metropolitan Area (CMA); underrepresented groups
- Data suppression, particularly at Dissemination area (DA) level
- Census tracts (CT) only in urban areas limit comparisons
- Only every 5 years

# Data Sources & Limitations (cont'd)



## ***Physician Claims*** (OHIP-Ontario Health Insurance Plan)

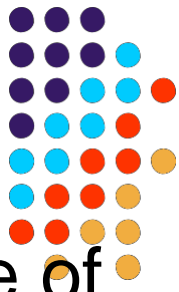
### **Strengths:**

- Can answer: “Who is using services and what kind?”
- Only comprehensive source of population health coverage & provision of publicly-paid health services
- Laboratory and radiology claims include Community Health Centres (CHCs)

### **Limitations:**

- Excludes CHCs for physician visits (e.g. diabetes)
- Health insurance addresses out-of-date
- No individual level socio-economic or cultural info available

# Data Sources & Limitations (cont'd)



## ***Vital Statistics*** - (births and deaths)

Original source: Vital Statistics, Ontario Office of Registrar General (ORG)

Distributed by: Ontario Ministry of Health and Long Term Care (MOHLTC): IntelliHEALTH ONTARIO

## ***Live Birth Database***

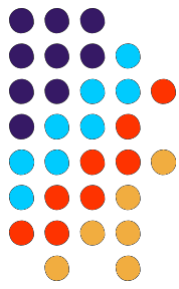
### **Strengths:**

- Includes country of birth
- Links baby to mother for analysis of singleton Low Birthweight (LBW) by age, parity, pregnancy type

### **Limitations:**

- Missing unregistered births
- Births to out-of-province mothers are included in IntelliHEALTH total counts; whereas, births occurring outside of the province to Ontario mothers are not available.
- Missing postal codes
- 2 - 3 year time lag in data availability

# Data Sources & Limitations (cont'd)



## *Vital Statistics* - (births and deaths)

### ***Mortality Data***

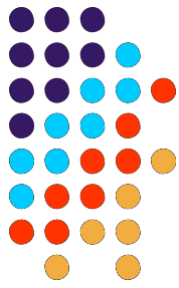
#### **Strengths:**

- Since the registration of deaths is a legal requirement, reporting is virtually complete.
- The death records include demographic as well as medical information about the deceased.

#### **Limitations:**

- Missing postal codes
- 2 - 3 year time lag in data availability
- Variation in data collection procedures over time and/or geography may reduce the validity of time and/or place-specific comparisons
- There are possible errors in the mortality data file related to assignment of municipality of residence/census subdivision of deceased.

# Data Sources & Limitations (cont'd)



***Hospital Inpatient Data*** - Canadian Institute for Health Information (CIHI); Provincial Health Planning Database (PHPDB) [Canada]

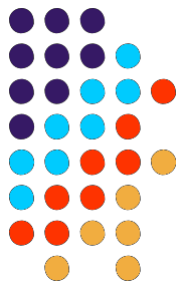
## Strengths:

- Up-to-date postal codes
- Current –1 year time lag

## Limitations:

- Excludes out of hospital births
- Missing postal codes approximately 2%
- No SES or ethnicity info available

# Data Sources & Limitations (cont'd)



## ***Ontario Mental Health Reporting System***

**(OMHRS)** - The Ontario Mental Health Reporting System

(OMHRS) in Ontario officially collects data on patients in adult designated inpatient mental health beds.

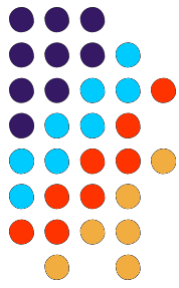
### **Strengths:**

- Captures hospitalizations in General, Provincial Psychiatric, and Specialty Psychiatric facilities.

### **Limitations:**

- Only available since 2006
- No SES or ethnicity info available

# Data Sources & Limitations (cont'd)



## ***National Ambulatory Care Reporting***

***System (NACRS)*** - NACRS is a data collection tool used to capture information on patient visits to hospital and community based ambulatory care: day surgery, outpatient clinics and emergency departments.

### **Strengths:**

- Current postal code

### **Limitations:**

- Only available since 2001
- No SES or ethnicity info available

# Data Issues



- Balancing making the information user-friendly with the providing detail needed for accurate and appropriate use and understanding of the information
- Sustainability, capacity to update data in the future
- Reducing the resources required for data conversion through developing a user-driven interactive site
- Responding to potential health inequalities that are identified on the site
- Comprehensiveness across the range and breadth of health planning needs