

# Quality of Administrative Data for Chronic Disease Research and Surveillance

Lisa M. Lix, PhD P.Stat.  
School of Public Health  
University of Saskatchewan

November 17, 2010 Health Data Forum Regina

# Outline

- ✧ Background
- ✧ Strategies and Techniques
- ✧ Recommendations

# Background

- ✧ Administrative data are widely used in chronic disease research and surveillance
  - Public Health Agency of Canada Chronic Disease Surveillance System
  - Circumpolar Chronic Disease Observatory: Yukon, Northwest Territories
- ✧ These initiatives cover a broad range of conditions
  - diabetes, hypertension, cardiovascular disease, arthritis, osteoporosis, gastrointestinal conditions

# Background

- ✧ Aims of chronic disease research and surveillance projects:
  - accurately estimate disease prevalence and incidence
  - investigate factors associated with variation in disease rates – socioeconomic, geographic, temporal
  - compare health outcomes, cost, and utilization in chronic disease cohorts and healthy controls

# Background

- ✧ Ascertaining cases of chronic diseases using administrative data is challenged by:
  - Inaccuracy and incompleteness in the recording of diagnoses
  - Incompleteness of the data sources

# Strategies and Techniques to Assess the Quality of Administrative Data

- ✧ ***Validation studies*** that compare administrative data to a reference standard
  - Medical charts
  - Clinical test or registry
  - Self-report
  
- ✧ ***Capture-recapture analyses*** to assess data completeness

# Validation Studies: Estimating Sensitivity and Specificity

## Reference Standard

Admin  
Data

	Has Disease	Does Not Have Disease
Has Disease	A	B
Does Not Have Disease	C	D

$$\text{Sensitivity} = A/(A+C)*100$$

$$\text{Specificity} = D/(B+D)*100$$

# Validating Osteoporosis Case Definitions using Bone Mineral Density Data, Manitoba, 2001/02 – 2002/03

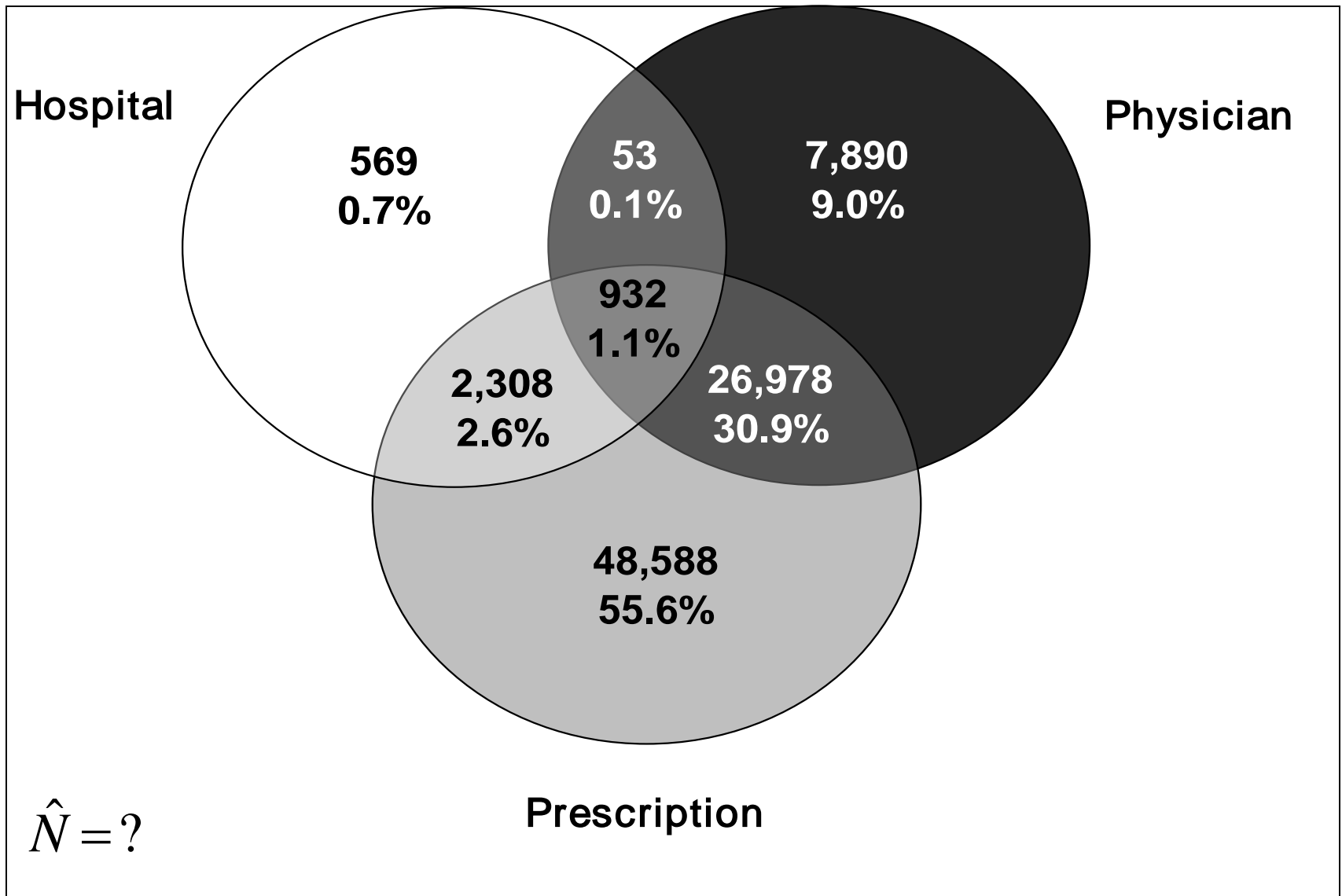
# Yrs data	Algorithm	Sensitivity (%)	Specificity (%)
1	1+H or 1+P	69.4	69.0
	1+H or 2+P	34.1	91.4
2	1+H or 1+P	74.0	64.9
	1+H or 2+P	43.9	86.0

Note: P = physician claims; H = hospital records

# Model-Based Approaches to Address Validity Problems

Type of Model	Description	Specific Examples
Classification	Linear or non-linear rules that predict the probability of group membership (i.e., disease present/absent) based on multiple attributes of the patient or the physician	Logistic regression Classification and regression trees Latent class
Calibration/Missing data	Statistical models that “fill in” missing values (i.e., “missing” diagnosis)	Measurement error/multiple imputation

# Asthma Case Ascertainment, Illustration for Capture-Recapture Method, Manitoba



# Capture-Recapture Analyses

- ✧ Multinomial logistic regression models
  - To identify covariates associated with incompleteness of data sources
- ✧ Log-linear regression models
  - To identify dependencies between data sources

# Relationship between Validity and Completeness

- ✧ Nanan & White (1997) in *Chronic Diseases in Canada*
  - “Even though capture-recapture is a valuable method for enhancing existing surveillance data, there is an ongoing need to strengthen more ‘traditional’ surveillance systems and data collection sources. This involves such activities as improving and validating case definitions...”

# Recommendations

- ✧ **DO** take steps to understand the characteristics of administrative data sources
  - Consult with data providers, clinicians, statisticians
- ✧ **DO** perform validation sub-studies
- ✧ **DO** conduct sensitivity analyses
  - Do the results change when a different approach is used to ascertain chronic disease cases?

# Funding Acknowledgements

- ✧ Canadian Institutes of Health Research
- ✧ Saskatchewan Health Research Foundation
- ✧ Lupina Foundation