

# Data Quality and the Development of National Costing Panels to Better Utilize Administrative Data

**Larry Svenson**

**Senior Manager, Epidemiology and Surveillance  
Surveillance and Assessment Branch  
Alberta Health and Wellness**

# Administrative Data – The Basics

---

**Administrative Data – Data routinely collected for the payment, monitoring, and evaluation of the provision of health (clinical) services.**

- The primary purpose is not research
- Can be a rich source for secondary analyses such as research and surveillance (but you need to know and understand the limits)

# Administrative Data

---

## Limitations

- In general, it is not possible to follow-up with specific individuals
- Information on SES at the individual level is lacking
- Case ascertainment is based on utilization of services
- Utilization patterns can differ for reasons other than incidence, prevalence of severity
- Accuracy of diagnosis can't be confirmed
- Variable accuracy and precision of data
- Sample size

## Strengths

- Data are population-based
- Information can be obtained without the need to contact individuals
- No recall bias
- Information is readily available
- Sample size

# Common Administrative Health Data Myths

---

- **Myth 1** – Administrative data is of a poor quality and therefore of limited value for research, surveillance, policy development, and evaluation of health service outcomes
- **Myth 2** – Physicians only record the diagnosis that gets them paid the most
- **Myth 3** – A family physician practice can be defined by 10-20 diagnostic codes and less than 20 procedure codes.

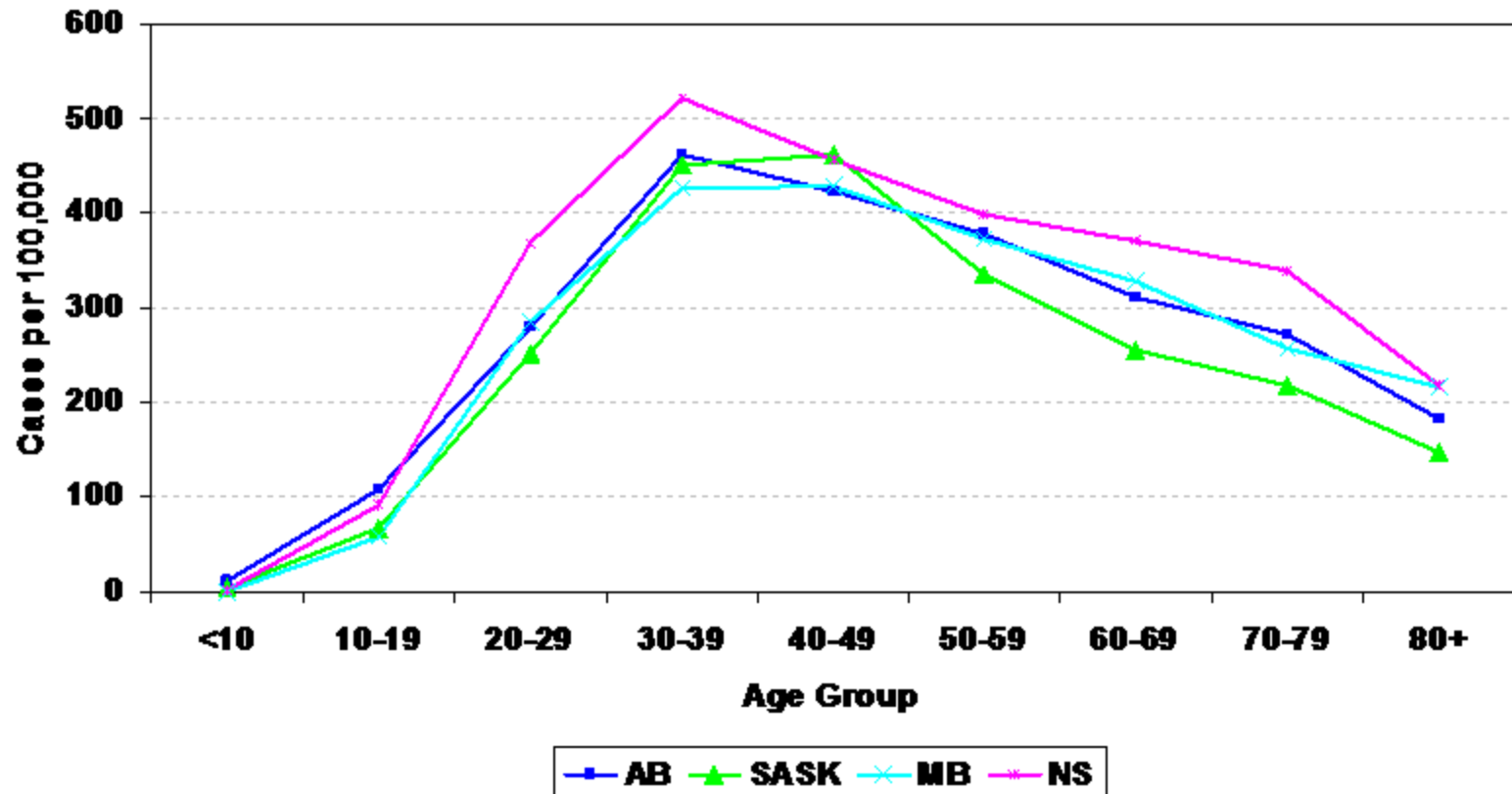
# Addressing Myth 1

## Case Definition Performance – Examples

Condition	Sensitivity / Specificity	PPV	Reference
Ischaemic Heart Disease	Sens. 77% Spec. 98%		Tu et al., Can J Cardiol 2010
Chronic Kidney Disease	Sens. 80% Spec. 90%		Grams et al., Am J Kidney Dis 2010
Osteoporosis	Sens. 93% Spec. 91%		Leslie et al., Osteoporos Int 2010
Lupus		89 to 92%	Chibnik et al., Lupus 2010
Primary Biliary Cirrhosis	Sens. 94%	73 to 89%	Myers et al., Can J Gastroenterol 2010
Hypertension	Sens. 74% Spec. 94%	81%	Quan et al., Hypertension 2009
Lupus	Sens. 42-67% Spec. 99.9%		Bernatsky et al., Rheumatology 2007

# Provincial Consistency – Example

## Age-Specific Prevalence of Crohn's Disease



# Predictors of Chronic Renal Failure

---

	<b>Odds Ratio</b>	<b>(95% Confidence Interval)</b>
Female	1.00 (reference)	
<b>Male</b>	<b>1.68</b>	<b>(1.61, 1.77)</b>
No Subsidy	1.00 (reference)	
<b>Subsidy</b>	<b>1.27</b>	<b>(1.20, 1.34)</b>
<b>First Nations</b>	<b>2.43</b>	<b>(2.16, 2.72)</b>
<b>Social Assistance</b>	<b>4.65</b>	<b>(4.26, 5.09)</b>
<b>Hypertension</b>	<b>7.14</b>	<b>(6.71, 7.60)</b>
<b>Diabetes</b>	<b>3.16</b>	<b>(3.01, 3.32)</b>

---

## **Myth 2 – Physicians only record the diagnosis that gets them paid the most**



# Does Administrative Data Reflect the Chart?

---

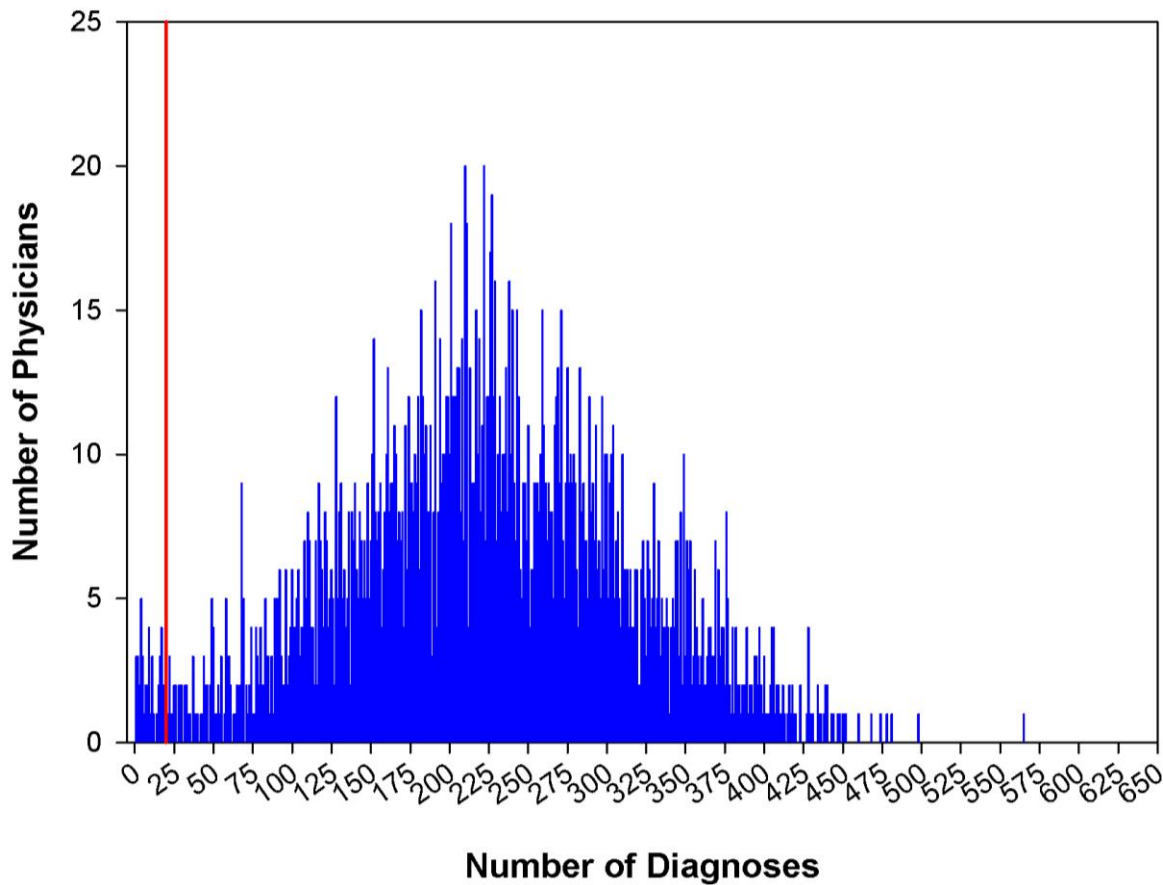
- Kikano et al. (2000) found
  - High concordance between what was on charts and what appeared in the administrative record
- Wilchesky et al. (2004) compared charts with fee-for-service data from Quebec
  - They found that the fee-for-service data were highly specific, but that there was wide variation in sensitivity
- Within Alberta, fee-for-service claims are submitted by electronic billing services which has resulted in better quality of submitted data

**Sources:** Kikano et al., Arch Fam Med 2000 | Wilchesky et al., J Clin Epidemiol 2004

---

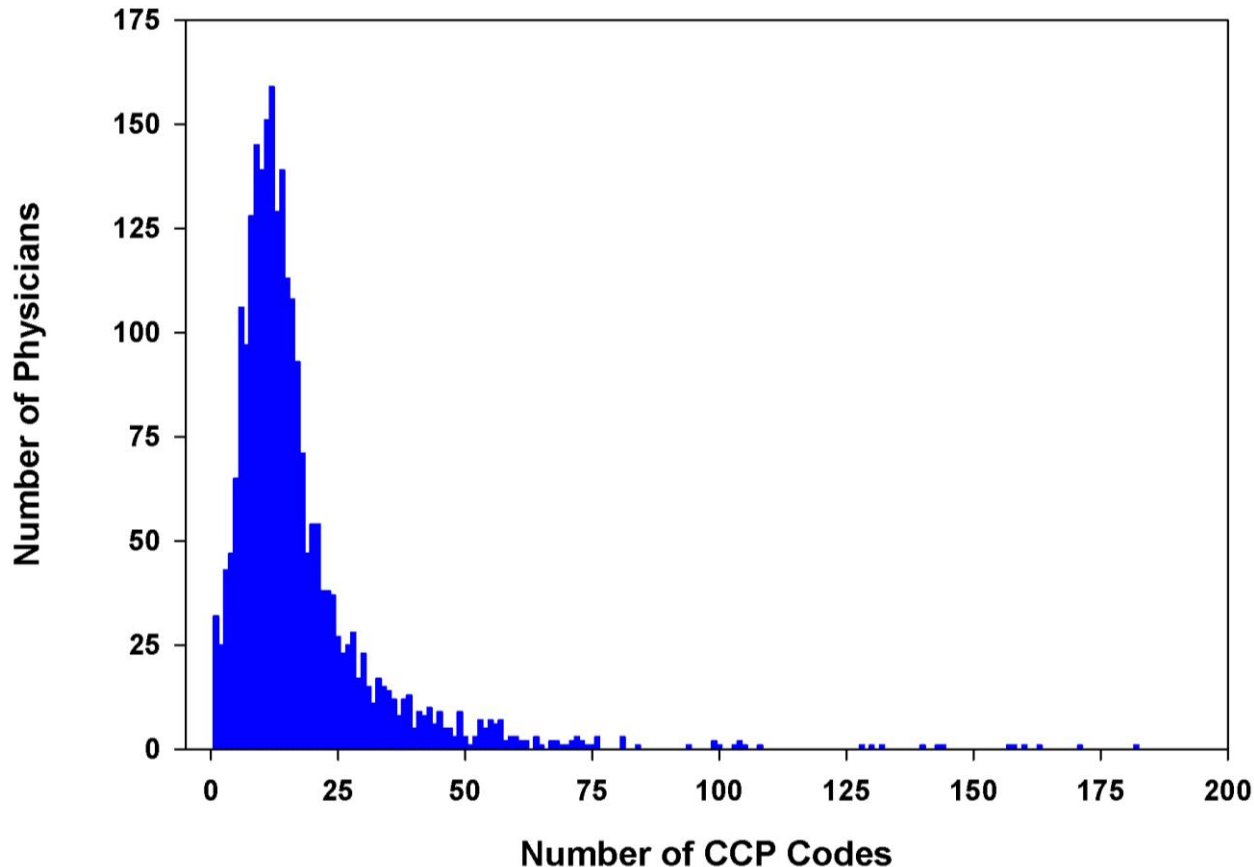
**Myth 3 – A family physician practice can be defined by 10-20 diagnostic codes and less than 20 procedure codes.**

# Distribution of the Number of ICD-9-CM Codes used by General Practitioners



Mean: 240 Diagnoses  
Median: 242 Diagnoses  
Min: 1  
Max: 533

# Distribution of the Number of CCP Codes used by General Practitioners



Mean: 60 Codes  
Median: 43 Diagnoses  
Min: 1  
Max: 391

# Considerations in the Use and Interpretation of Administrative Data

---

## Geography

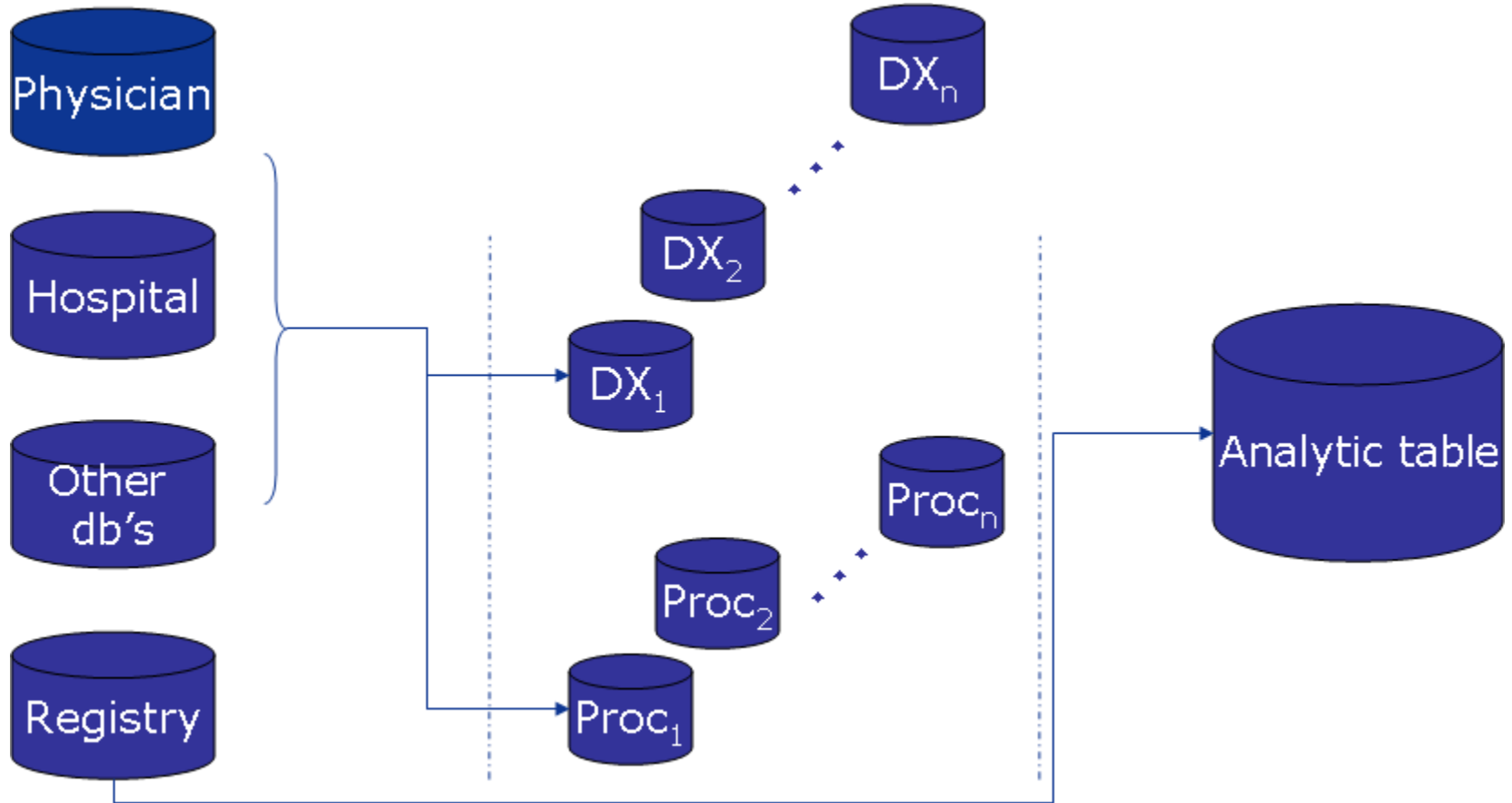
- Case definition algorithms appear to perform independent of geography, however, wide variations exist for incidence and prevalence estimates
- The age/sex distribution across geographic areas were similar
- The specificity of the coding was different with urban facilities recording more specific codes than the rural facilities.

## Socio-Economic Status

- Use of health services varies across socio-economic groups as does health status

**Sources:** Yiannakoulias et al., Cerebrovasac Dis 2003 | Yiannakoulias et al. Chron Dis Can 2009 | Sin et al., Chest 2003

# Generic Analytic Model



# Privacy Legislation and Data Release Policies

---

- Understanding the intent of legislation versus the wording
- Research Ethics Boards often over interpret legislation (err on the side of caution)
- General misunderstanding of the research process (Government) and the obligation to protect privacy (Academia)
- Policies for improving access to data tend to fail due to the implementation of the policy and not the policy itself

# National Standards and Guidelines for the Use of Administrative Data

---

- Need to establish the target audience for standards and guidelines (Who will benefit?)
- Need to demonstrate that it will result in comparable information of a consistent quality (Where do we sit relative to other jurisdictions?)
- Needs to be viewed as a good investment of staff time (Are my staff doing someone else's work?)
- Needs to be of direct value to the federal, provincial, and territorial governments, not just for academic research interests (What is the benefit?)



# Conclusions

---

- Administrative data are valuable, but understand the limits
  - Clinical ambiguity leads to ambiguity in the data
- Linkage across data sources enhances the data and outcomes
- Understand the geographic differences as they relate to access to diagnostic tools and specialists
- There is a need to understand health seeking behaviours and their impact on the completeness of data
- There is a need to increase the use of these data for longitudinal designs
- Purpose drives the analysis and ultimately determines if the value of the data
- Increased use of the data will improve its quality over time