TEXAS TECH UNIVERSITY
HEALTH SCIENCES CENTER™

Secondary Use of Data Resources for Research and Education

Chip Shaw, EdD, MPH
Executive Director Clinical Research Data Warehouse

What is Secondary Use

Data gathered and recorded previously for purposes other than the current project.

- Usually historical and already assembled
- Require no access to respondents or subjects
- Cost and time savings

Secondary Data Sources

Public data	Data sets available to anyone. May require permission from the owner to access the data
Registry data	A collection of information about individuals, usually focused around a specific diagnosis or condition.
Administrative data	Data generated at each patient encounter for billing and claims purposes.
EMR data	Data sets extracted from electronic medical record systems.

Public Data

Publically available research data.

Public Health Methodology

PUBLIC HEALTH Reports

Measuring the Prevalence of Diagnosed Chronic Obstructive Pulmonary Disease in the United States Using Data From the 2012-2014 National Health Interview Survey

Public Health Reports
2017, Vol. 132(2) 149-156

DOI: 10.1177/0033354916688197
journals.sagepub.com/home/phr
SAGE

Brian W. Ward, PhD[1], Colleen N. Nugent, PhD[1], Stephen J. Blumberg, PhD[1], and Anjel Vahratian,

Trends in Obesity and Severe Obesity Prevalence in US Youth and Adults by Sex and Age, 2007-2008 to 2015-2016

Craig M. Hales, MD[1]; Cheryl D. Fryar, MSPH[1]; Margaret D. Carroll, MSPH[1]; et al

» Author Affiliations | Article Information

JAMA. 2018;319(16):1723-1725. doi:10.1001/jama.2018.3060

Registry Data

Clinical data used to improve the quality and safety of the care, compare the effectiveness of different treatments, and to monitor the safety of implanted devices.

Arthritis Care & Research
Vol. 69, No. 11, November 2017, pp 1659–1667
DOI 10.1002/acr.23197

ORIGINAL ARTICLE

Lifetime Risk of Primary Total Hip Replacement Surgery for Osteoarthritis From 2003 to 2013: A Multinational Analysis Using National Registry Data

ILANA N. ACKERMAN,[1] MEGAN A. BOHENSKY,[2] RICHARD DE STEIGER,[3] CAROLINE A. ANTTI ESKELINEN,[5] ANNE MARIE FENSTAD,[6] OVE FURNES,[7] STEPHEN E. GRAVES,[8] JAASON HAAPAKOSKI,[9] KEIJO MÄKELÄ,[10] FRANK MEHNERT,[11] SZILARD NEMES,[12] SØREN OVERGAARD,[13] ALMA B. PEDERSEN,[11] AND GÖRAN GARELLICK[14]

Efficacy and Safety of Ponatinib in CML and Ph+ ALL Patients in Real-World Clinical Practice: Data from a Belgian Registry

Timothy Devos, Koen Theunissen, Fleur Samantha Benghiat, Alain Gadisseur, Stef Meers, Dominik Selleslag, Gaëtan Vanstraelen, Pierre Zachée, Marc André, Philippe Lewalle, Mia Janssen, Rik Schots, Koen Van Eygen, Alain Kentos, Marie Lejeune, Agnes Triffet, Inge Vrelust, Carolina Kuipers, and Violaine Havelange

Blood 2018 132:1744; doi: https://doi.org/10.1182/blood-2018-99-114070

Administrative Data

Claims data used to study health care delivery, benefits, harms, and costs.

ELSEVIER

The American Journal of Medicine

Volume 130, Issue 7, July 2017, Pages 809-818

Clinical research study

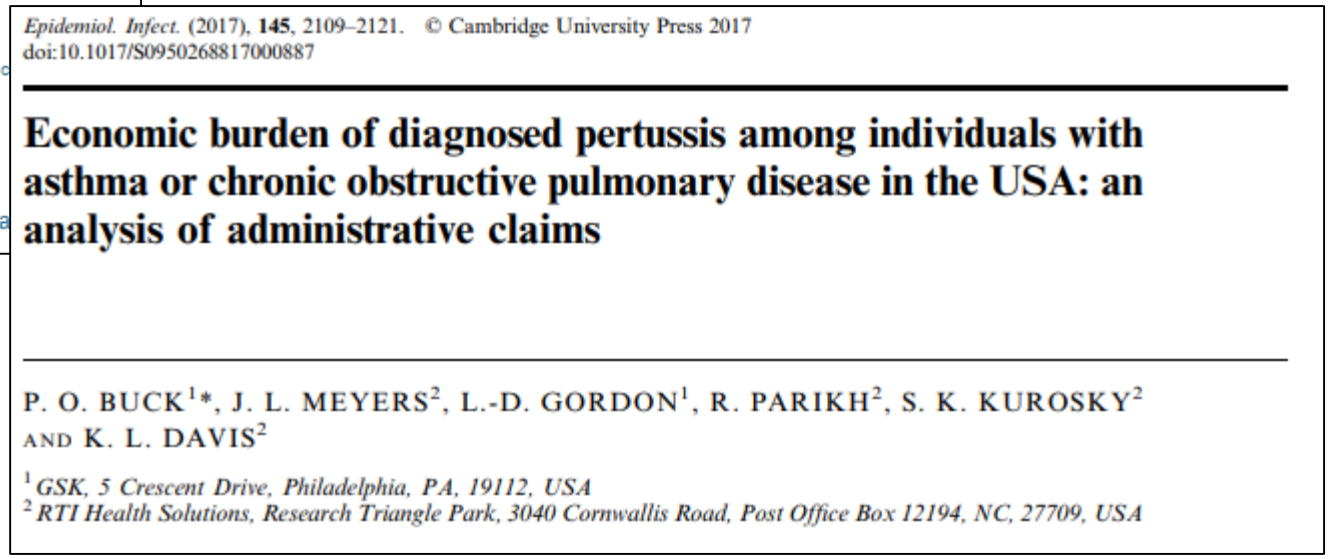
Positive Airway Pressure Therapies and Hospitalization in Chronic Obstructive Pulmonary Disease

Monica M. Vasquez MPH [a], Leslie A. McClure PhD [b], Duane L. Sherrill PhD [a], Sanjay R. Patel MD, MS [c], Krishnan MD, PhD [d], Stefano Guerra MD, PhD [a, e, f], Sairam Parthasarathy MD [e, g]

Show more

https://doi.org/10.1016/j.amjmed.2016.11.045

Get rights a



Epidemiol. Infect. (2017), **145**, 2109–2121. © Cambridge University Press 2017
doi:10.1017/S0950268817000887

Economic burden of diagnosed pertussis among individuals with asthma or chronic obstructive pulmonary disease in the USA: an analysis of administrative claims

P. O. BUCK[1]*, J. L. MEYERS[2], L.-D. GORDON[1], R. PARIKH[2], S. K. KUROSKY[2] AND K. L. DAVIS[2]

[1] *GSK, 5 Crescent Drive, Philadelphia, PA, 19112, USA*
[2] *RTI Health Solutions, Research Triangle Park, 3040 Cornwallis Road, Post Office Box 12194, NC, 27709, USA*

Secondary Use of Clinical Data

- Billing and Cost Analysis
- Morbidity and mortality reporting
- Quality
  - HEDIS Reporting
  - Continuous quality improvement
- Patient safety reporting
  - Adverse event reporting
- Clinical Trials
  - Cohort identification
  - Post-marketing information on drugs and devices
- Clinical Research
- Health population statistics
- Public Health
  - Bio-surveillance
  - Disease reporting
  - Disease registries
- Education
  - Develop data sets for simulation
  - Student research projects
  - Data sets for biostatistics, epidemiology, etc.

EMR Data - Cerner Health Facts

A de-identified patient database from participating healthcare institutions with time-stamped and sequenced information on pharmacy, laboratory, admission and billing data from all patient care locations. As of 2018, the Health Facts database contained

- Over 65 million patients
- Patient information from 750 healthcare facilities across the United States
- Over 500 million encounters
- 4.7 billion laboratory results
- Detailed pharmacy, laboratory, billing and registration data as far back as 2000
- 684 million orders for nearly 4,500 drugs by name and brand.

Cerner Health Facts Data

ELSEVIER

Annals of Epidemiology

Volume 26, Issue 2, February 2016, Pages 151-154.e4

Brief communication

Temporal trends of esophageal disorders by age in the Cerner Health Facts database

Jessica L. Petrick PhD, MPH [a], Tuyet Nguyen MD [b], Michael B. Cook PhD [a]

Show more

https://doi.org/10.1016/j.annepidem.2015.11.004

ELSEVIER

The American Journal of Cardiology

Volume 119, Issue 11, 1 June 2017, Pages 1809-1814

Heart Failure

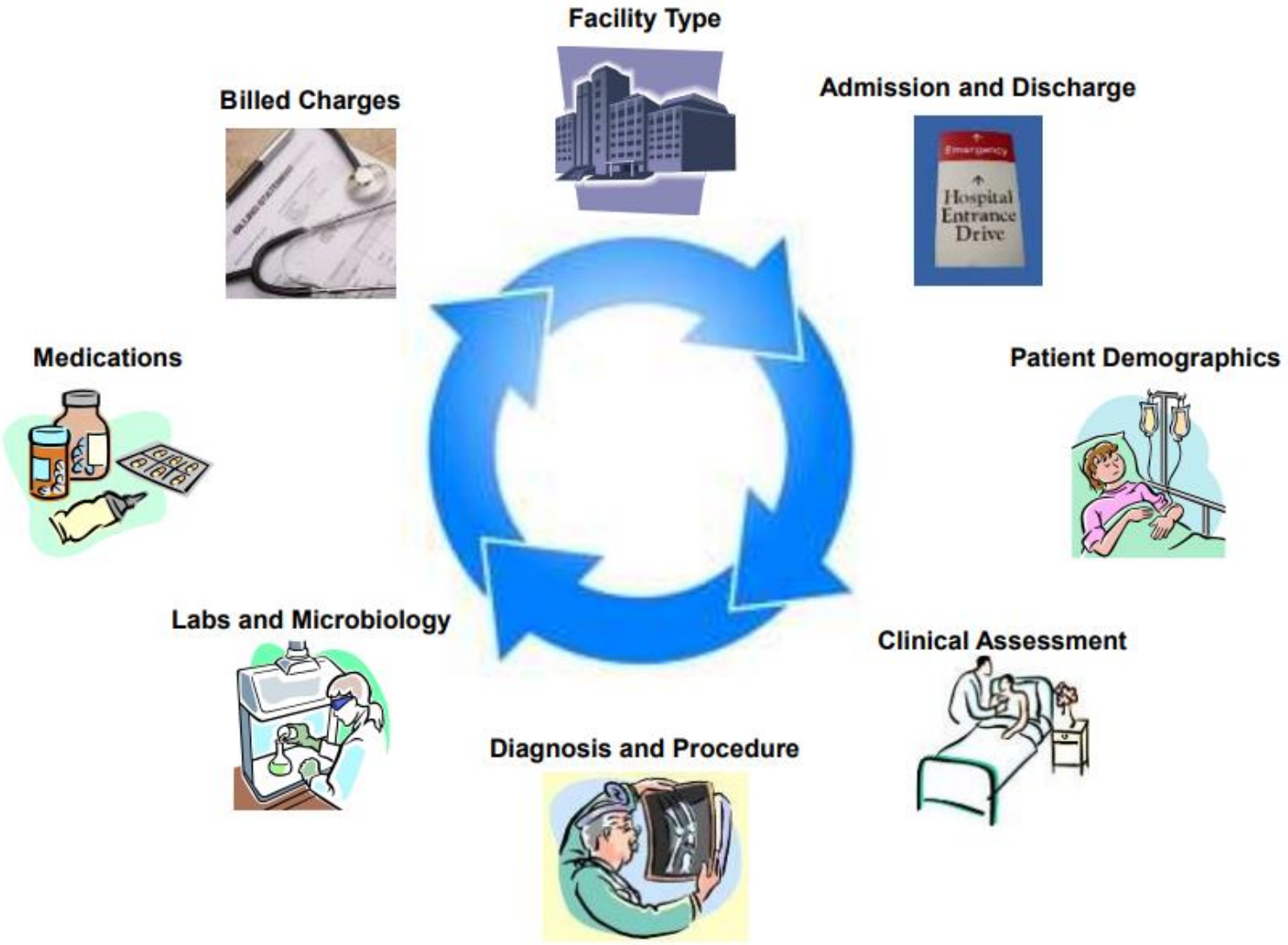
Effect of Transient and Sustained Acute Kidney Injury on Readmissions in Acute Decompensated Heart Failure

Benjamin J. Freda DO [a], Alexander B. Knee MS [b], Gregory L. Braden MD [c], Paul F. Visintainer PhD [b], Charuhas V. Thakar MD [d, e]

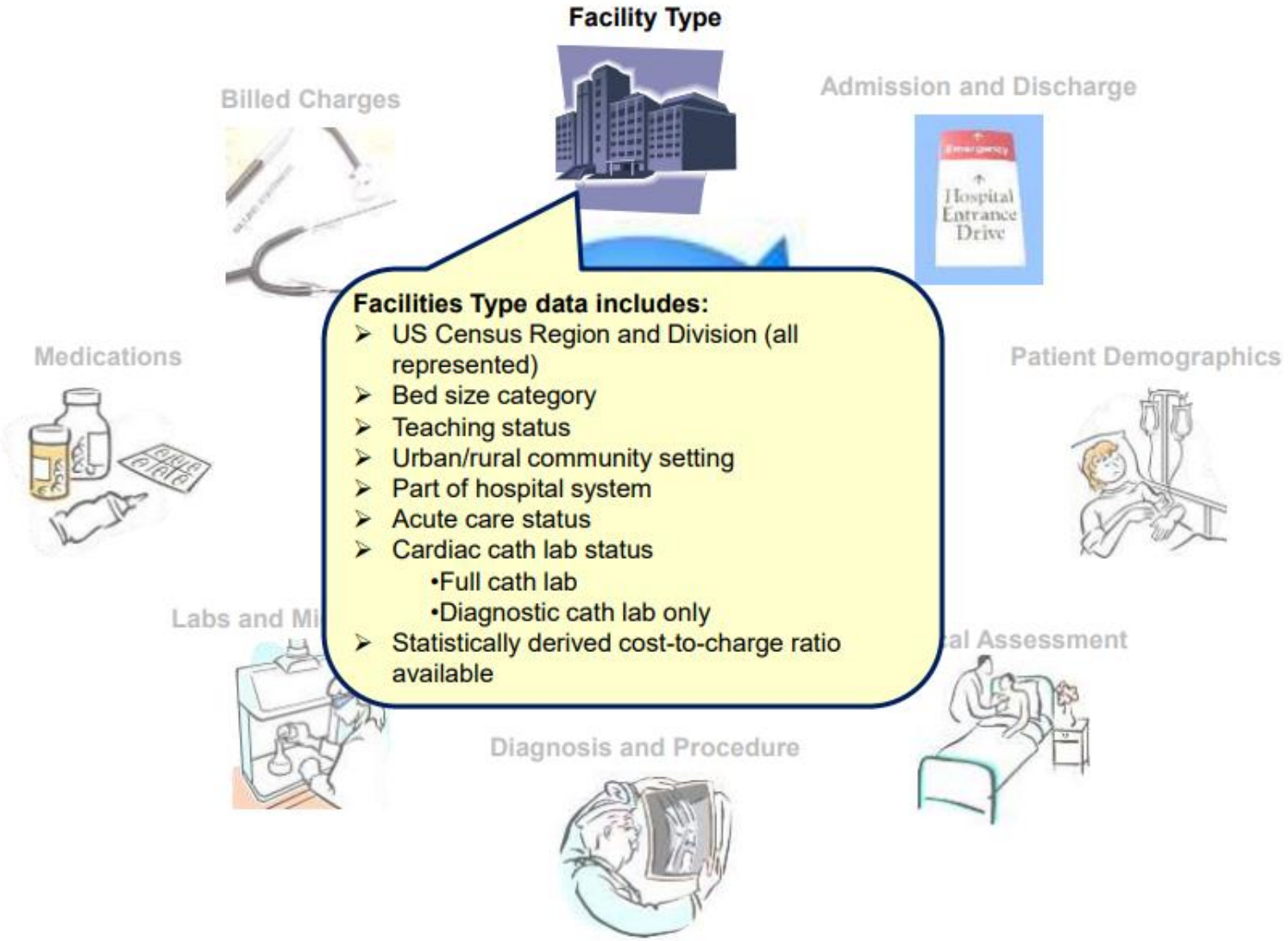
Show more

https://doi.org/10.1016/j.amjcard.2017.02.044

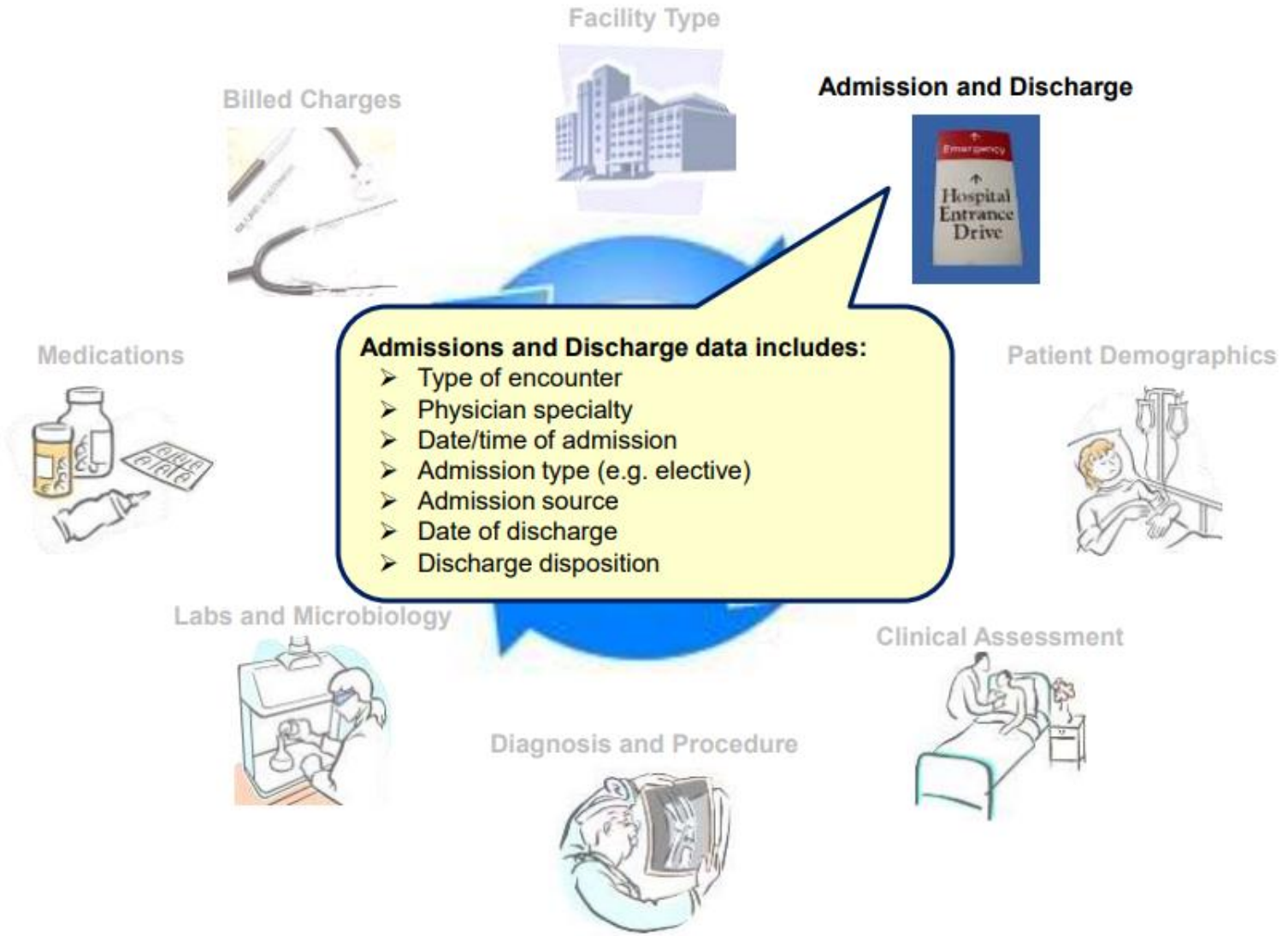
Cerner Health Facts Data Includes:



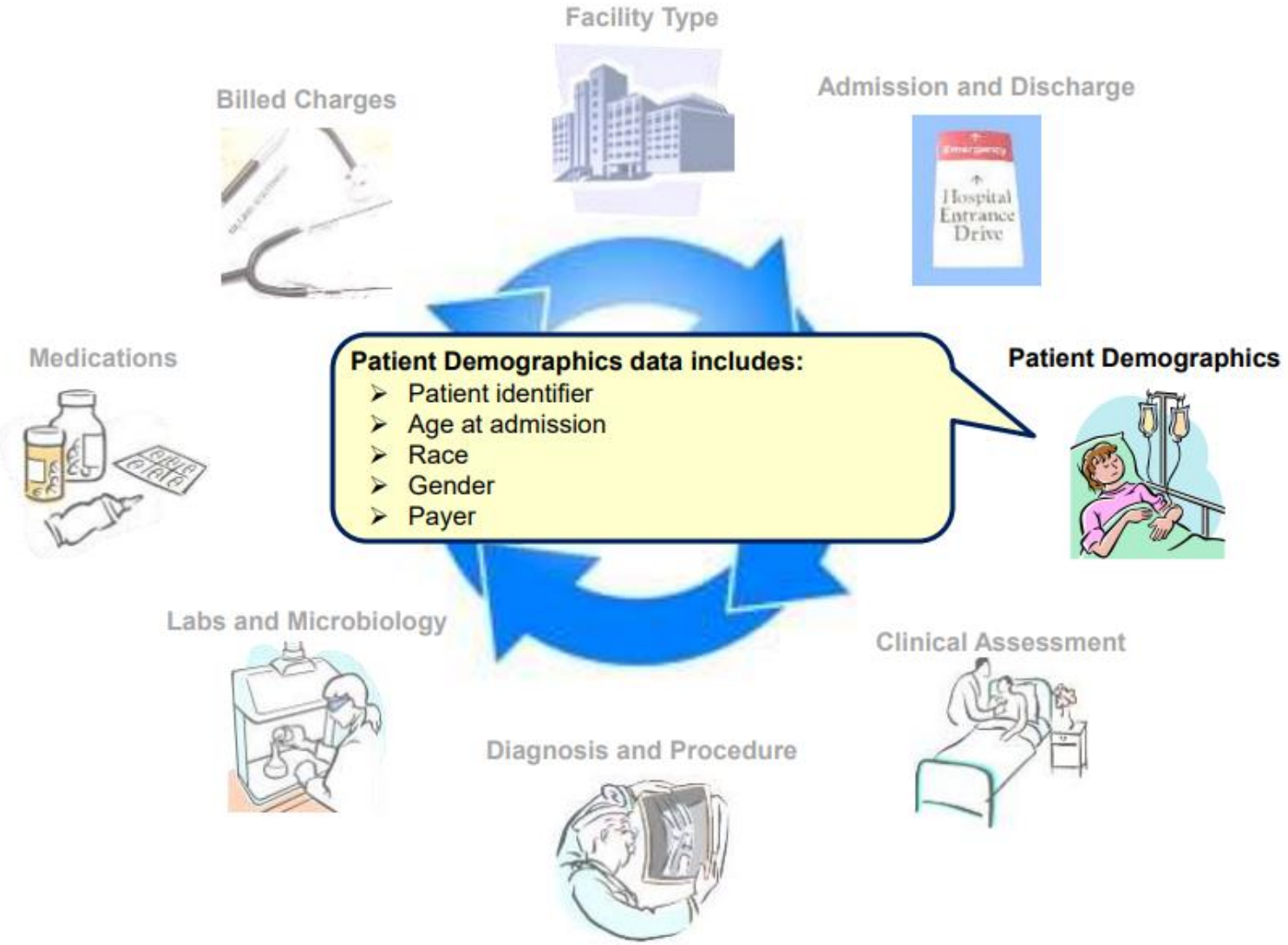
Facility Type Data



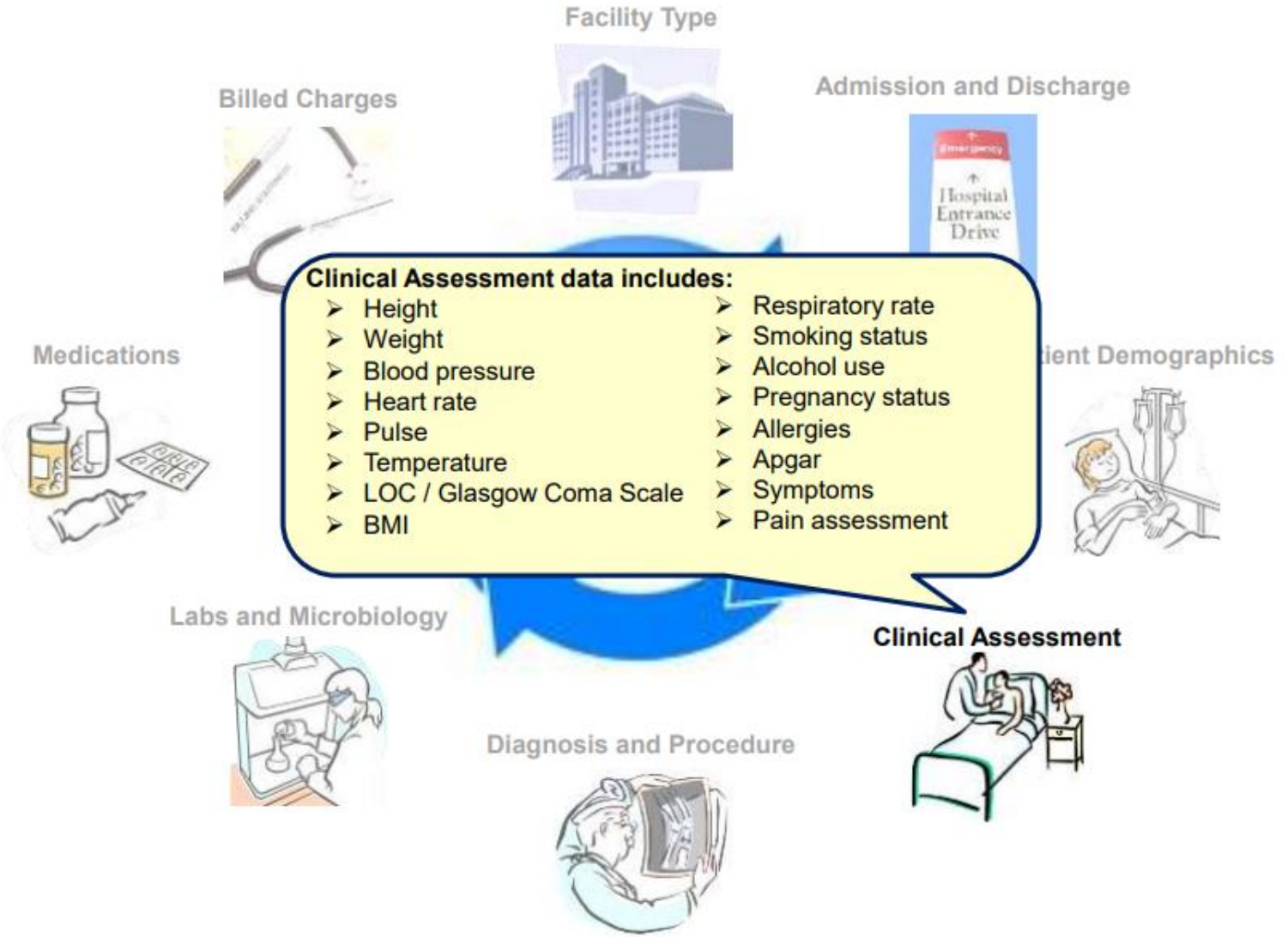
Admissions and Discharge Data



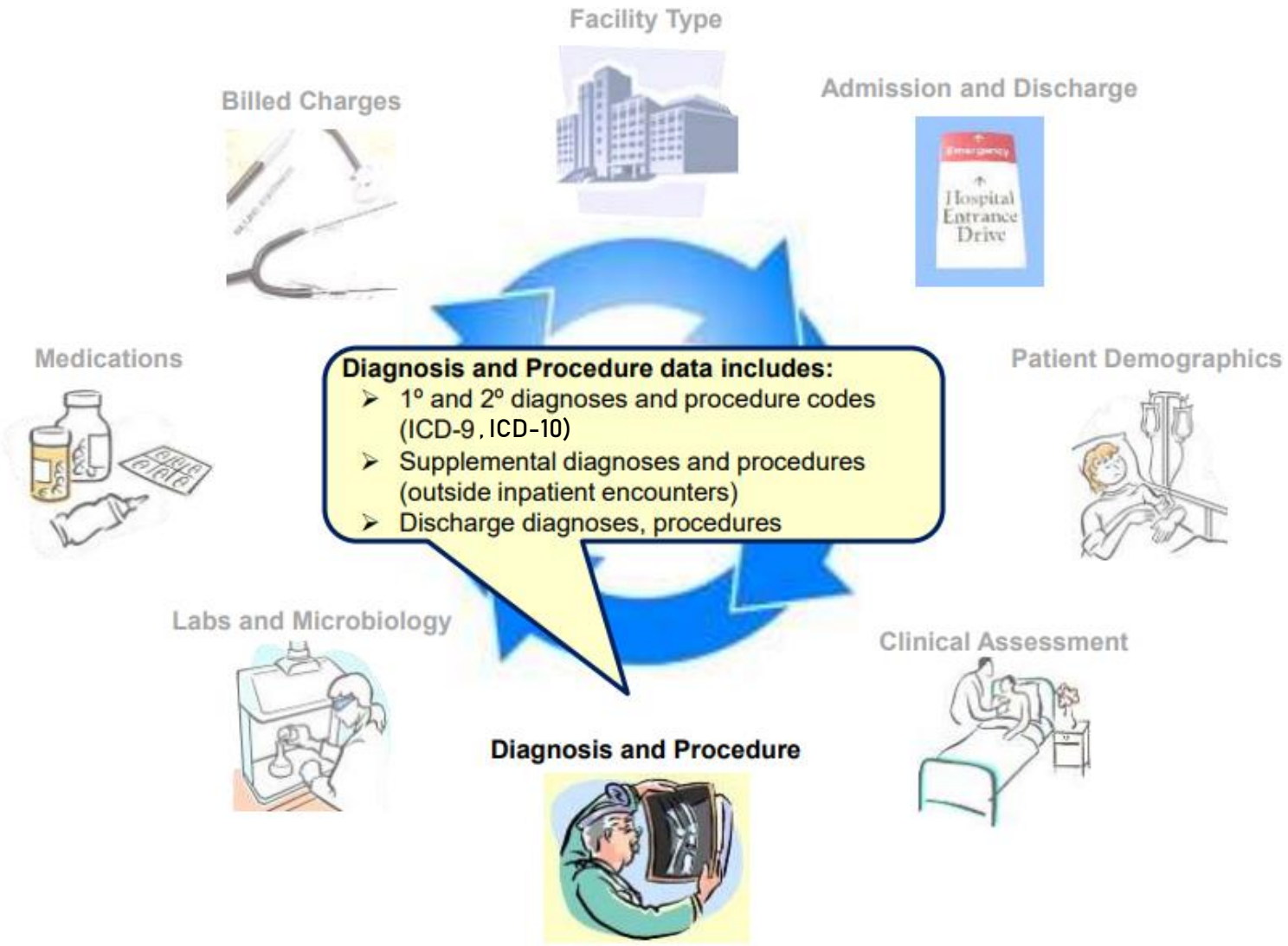
Patient Demographics Data



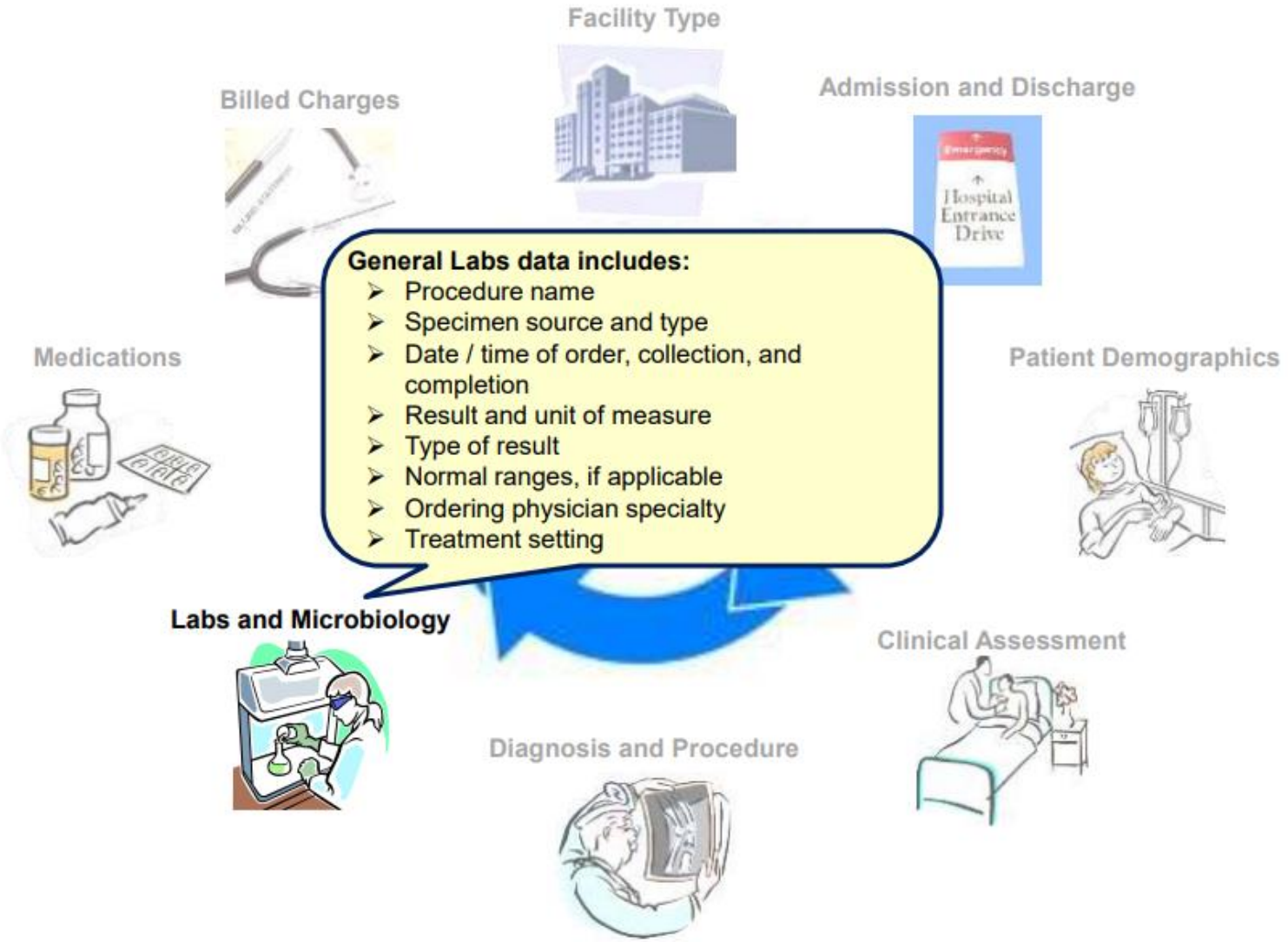
Clinical Assessment Data



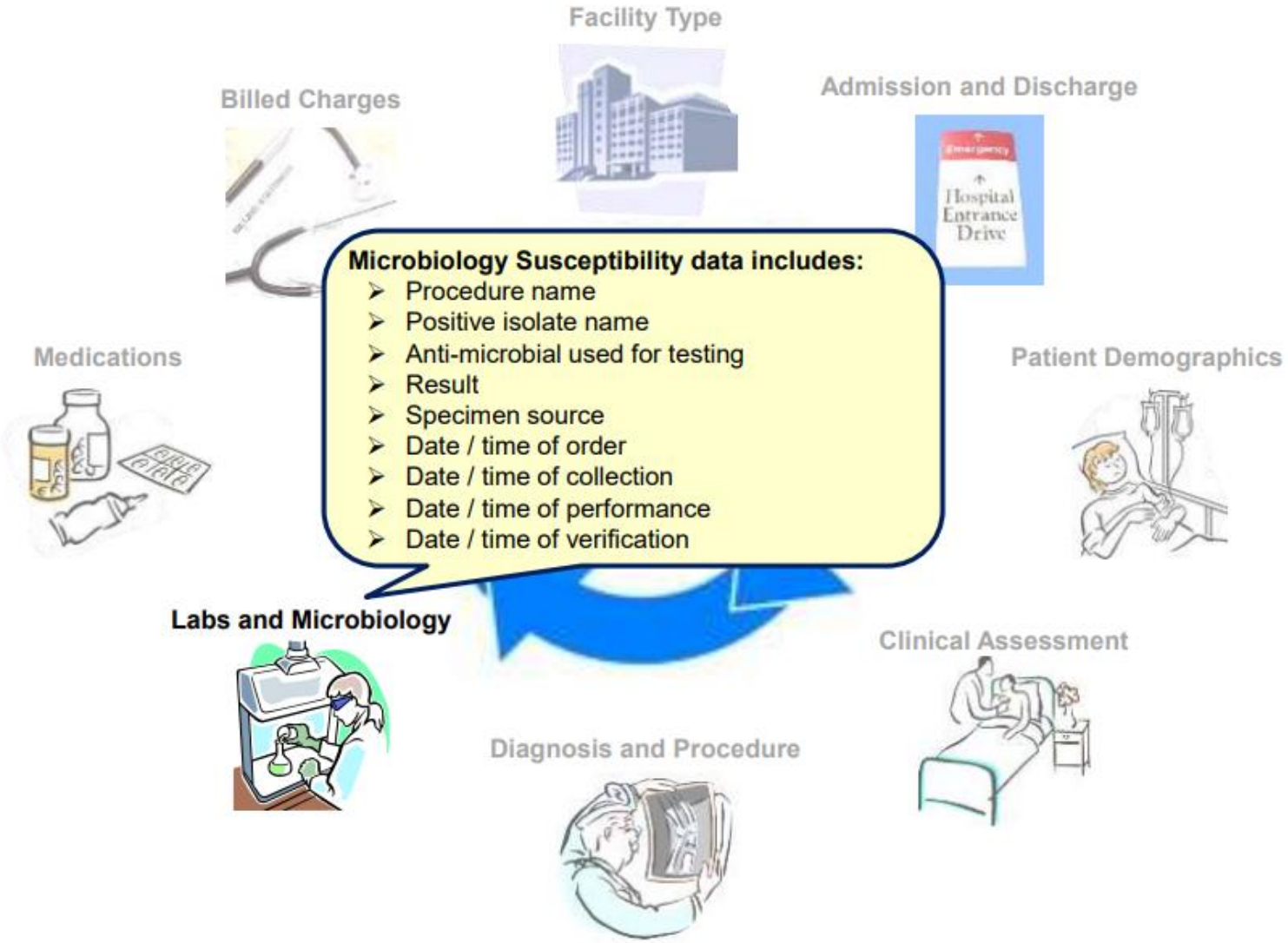
Diagnosis and Procedure Data



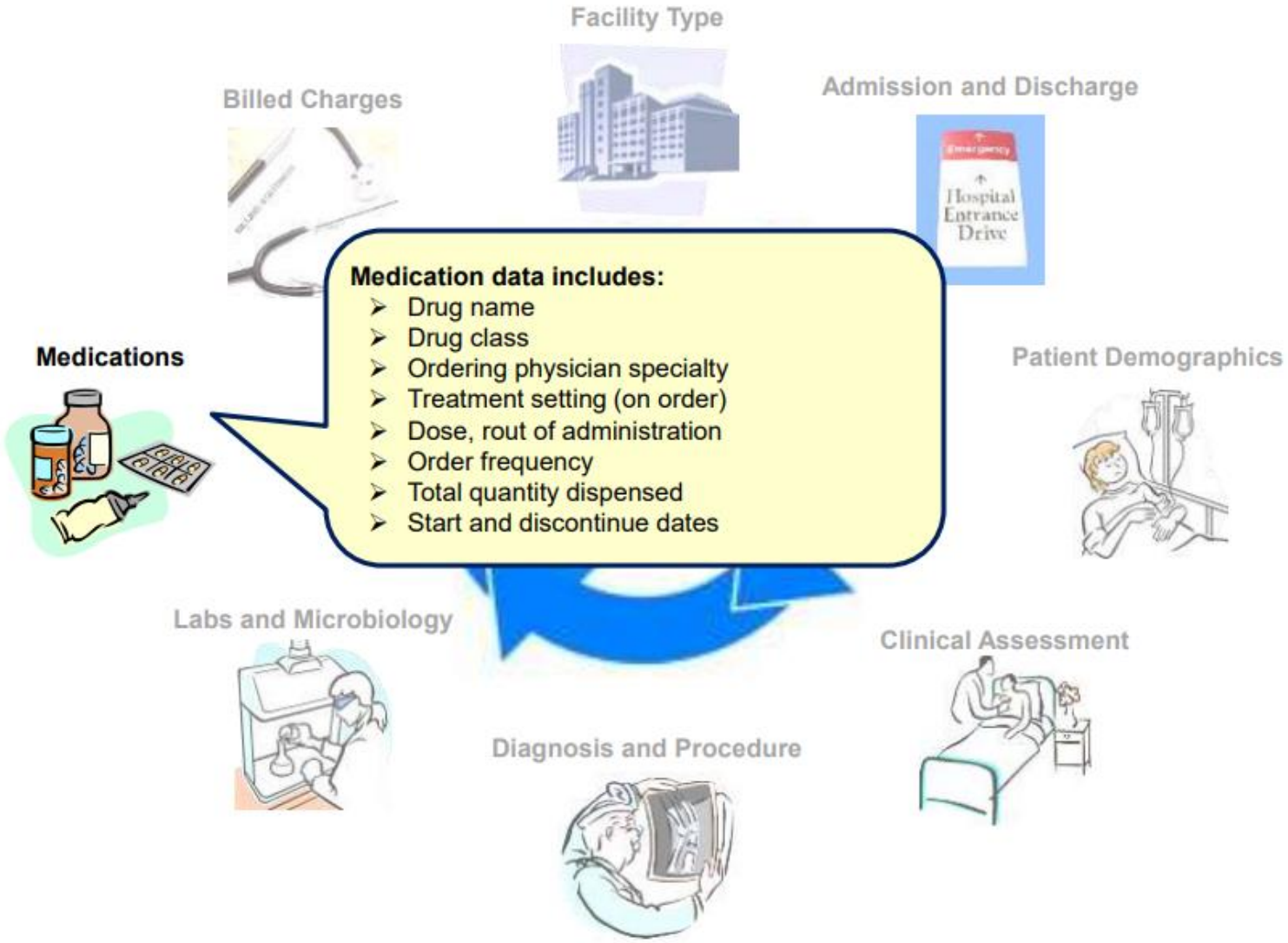
Laboratory Data



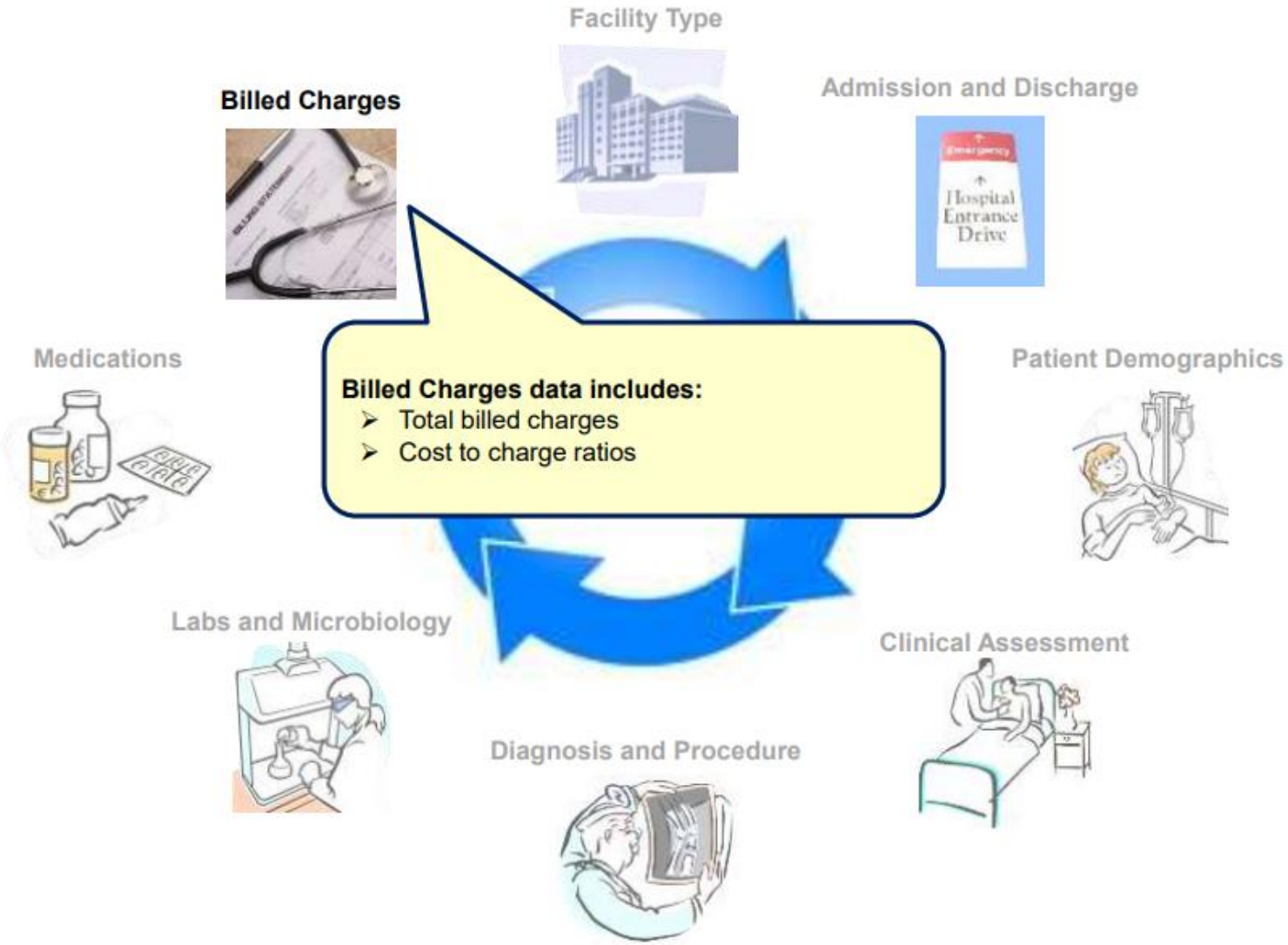
Microbiology Susceptibility Data



Medication Data



Billing Data



Cerner Health Facts Metrics

Metrics Name	
Total Patients	69M
Inpatient Encounters	22M
Emergency Encounters	66M
Outpatient Encounters	431M
Lab results	4.7B
Microbiology	192M
Medication orders	684M
Total diagnoses	971M
Total procedures	119M
Clinical events	5.3B

Metrics Name	
Total Systems	85
Total Facilities	750
Total Inpatient Facilities	388
Geography - Midwest	177
Geography - Northeast	235
Geography - South	263
Geography - West	174

Metrics Name	
Bed Size - ≤ 5	80
Bed Size - 6 to 99	136
Bed Size - 100 to 199	70
Bed Size - 200 to 299	48
Bed Size - 300 to 499	31
Bed Size - 500+	23
Type - Non-Teaching	279
Type - Teaching	108

Number of Encounters for Key Conditions				
Hypertension	17.3M		Diabetes	10.1M
Cardiovascular disease	7.1M		Asthma	3.3M
Atrial fibrillation	5.1M		COPD	2.7M

Cerner Health Facts De-Identification

Identifier	Action
Account numbers	Not extracted for *Health Facts*.
Any other unique identifying number, characteristic, or code	System-assigned numbers are provided to records. These numbers are mapped to *Health Facts* and are not identifiable.
Biometric identifiers, including finger and voice prints	Not extracted for *Health Facts*.
Certificate or license numbers	Not extracted for *Health Facts*.
Date (except year) directly related to an individual (for example, date of birth (DOB), discharge date, date of death) and all ages over 89 and all elements of dates (including year) indicative of age, except that such ages and elements may be aggregated into a single category of age 90 or older.	Dates are shifted by a consistent value across a single patient record.
Device identifiers and serial numbers	Not extracted for *Health Facts*.
Email addresses	Not extracted for *Health Facts*.
Fax numbers	Not extracted for *Health Facts*.
Full face photographic images and any comparable images	Not extracted for *Health Facts*.
Geographic subdivisions smaller than the census division	Not extracted for *Health Facts*.
Health plan beneficiary numbers	Not extracted for *Health Facts*.
IP address numbers	Not extracted for *Health Facts*.
Medical record numbers (MRNs)	Not extracted for *Health Facts*.
Names	Not extracted for *Health Facts*.
Social Security numbers (SSNs)	Not extracted for *Health Facts*.
Telephone numbers	Not extracted for *Health Facts*.
Vehicle identifiers and serial numbers, including license plate numbers	Not extracted for *Health Facts*.
Web URLs	Not extracted for *Health Facts*.

Cerner Health Facts Data

LIMITATIONS

- Bias
  - “Sick people have more data”
    - Selection Bias
    - Confounding
- Data Quality
  - Misclassification
  - Missing data

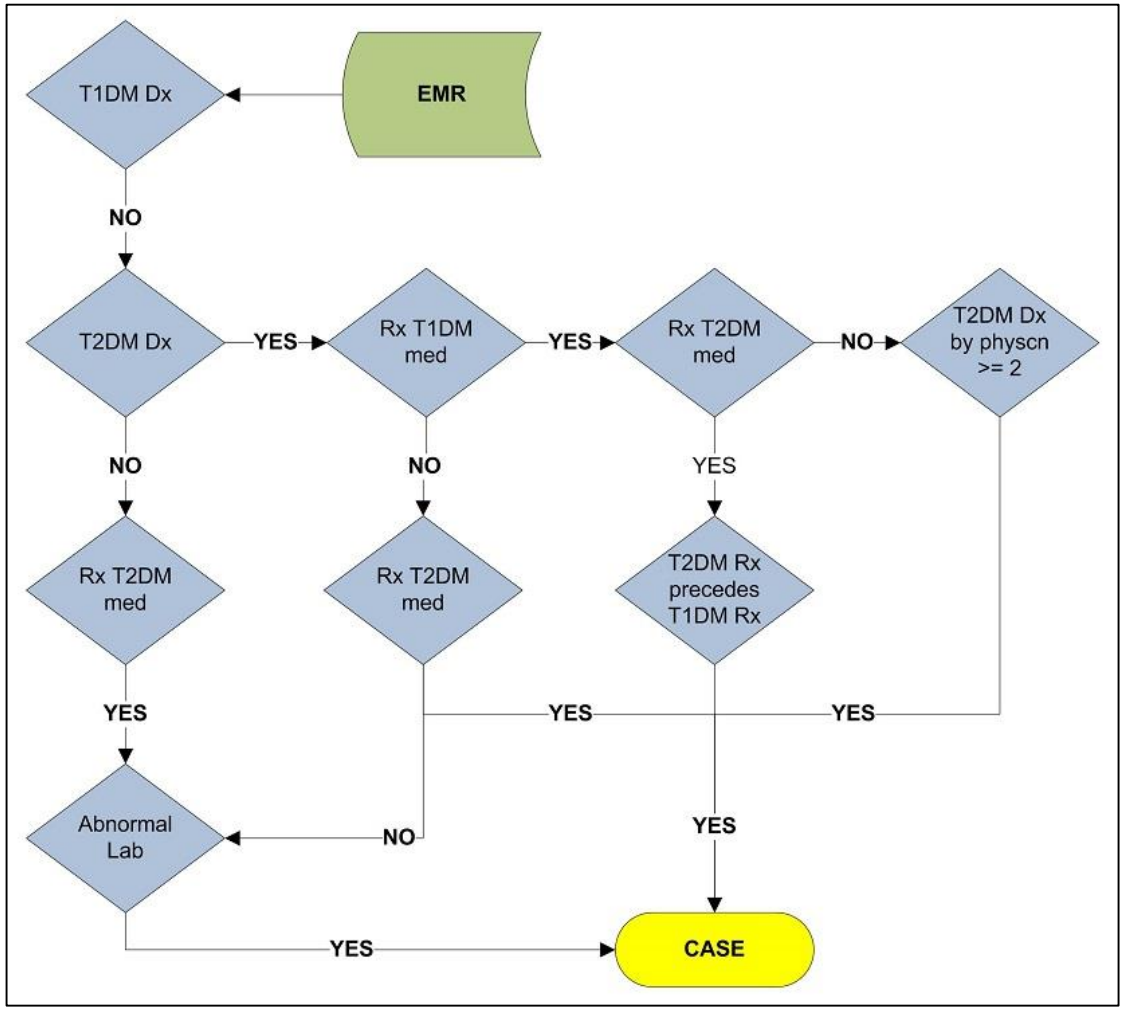
ADDRESSING LIMITATIONS

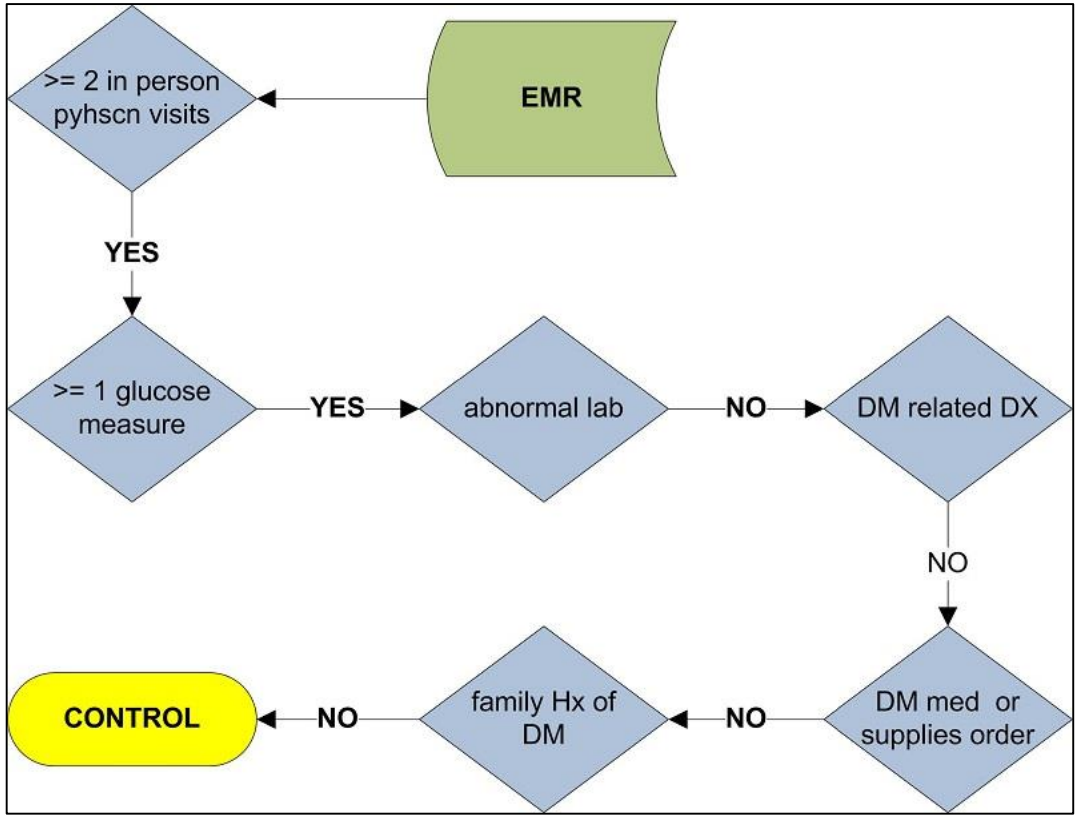
- Validation against a Gold Standard
- Inclusion & exclusion criteria
- Case and Control Selection

Inclusion Exclusion Criteria-
Data request for a hypertension control study

- Which of the ICD-9-CM, ICD-10-CM codes for hypertension should be included?
- Should outpatient, inpatient, and emergency department encounters be included?
- Should automated blood pressure monitoring data be included?
- Should orders, medication reconciliation, and fulfillment data be used?
- How far back in time should data be evaluated?
- Should rolling year versus calendar year be used?
- Should deceased patients be included?
- Should perioperative data be included?
- Should hypertension in the gestational period be included?

Data Set Production-
Type 2 Diabetes Case and Control selection

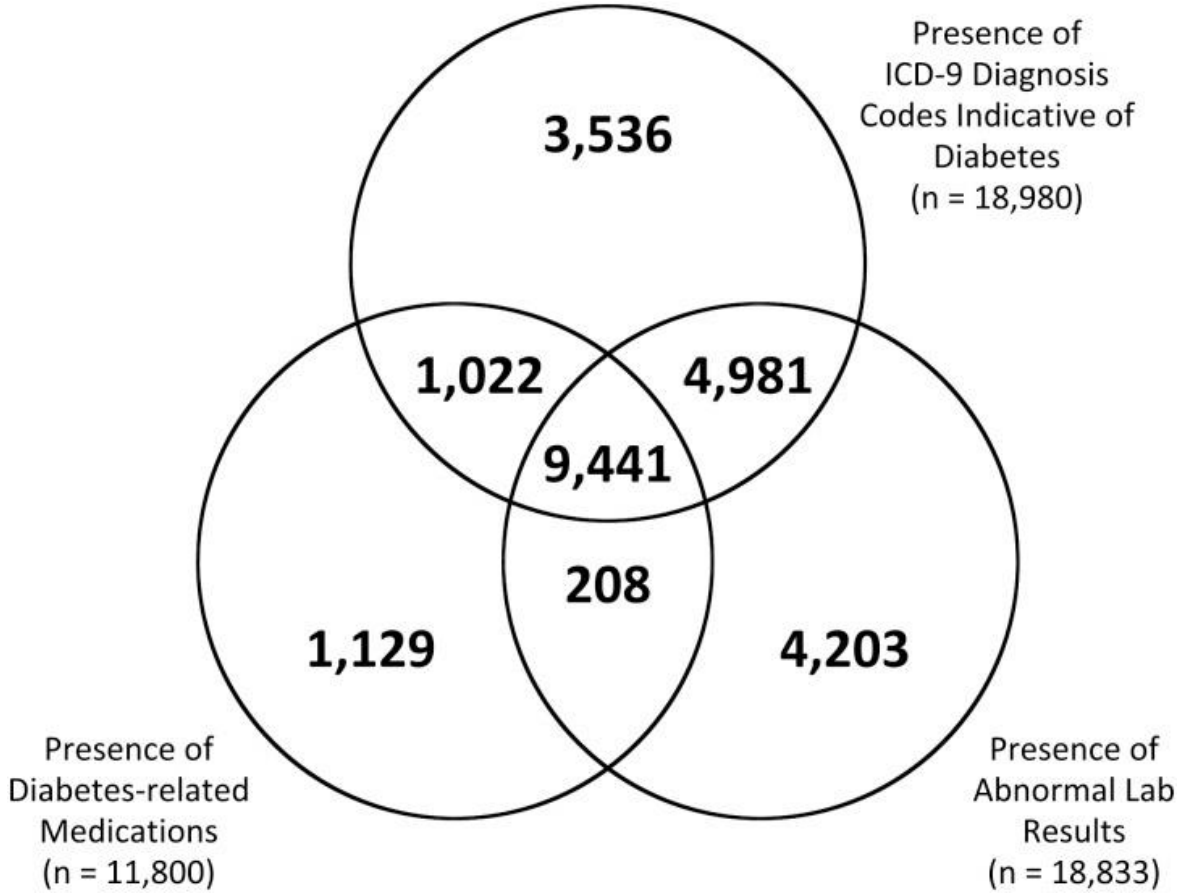




Northwestern University Type 2 diabetes mellitus (T2DM) algorithms for extracting both T2DM cases and T2DM controls from the electronic medical record (EMR).

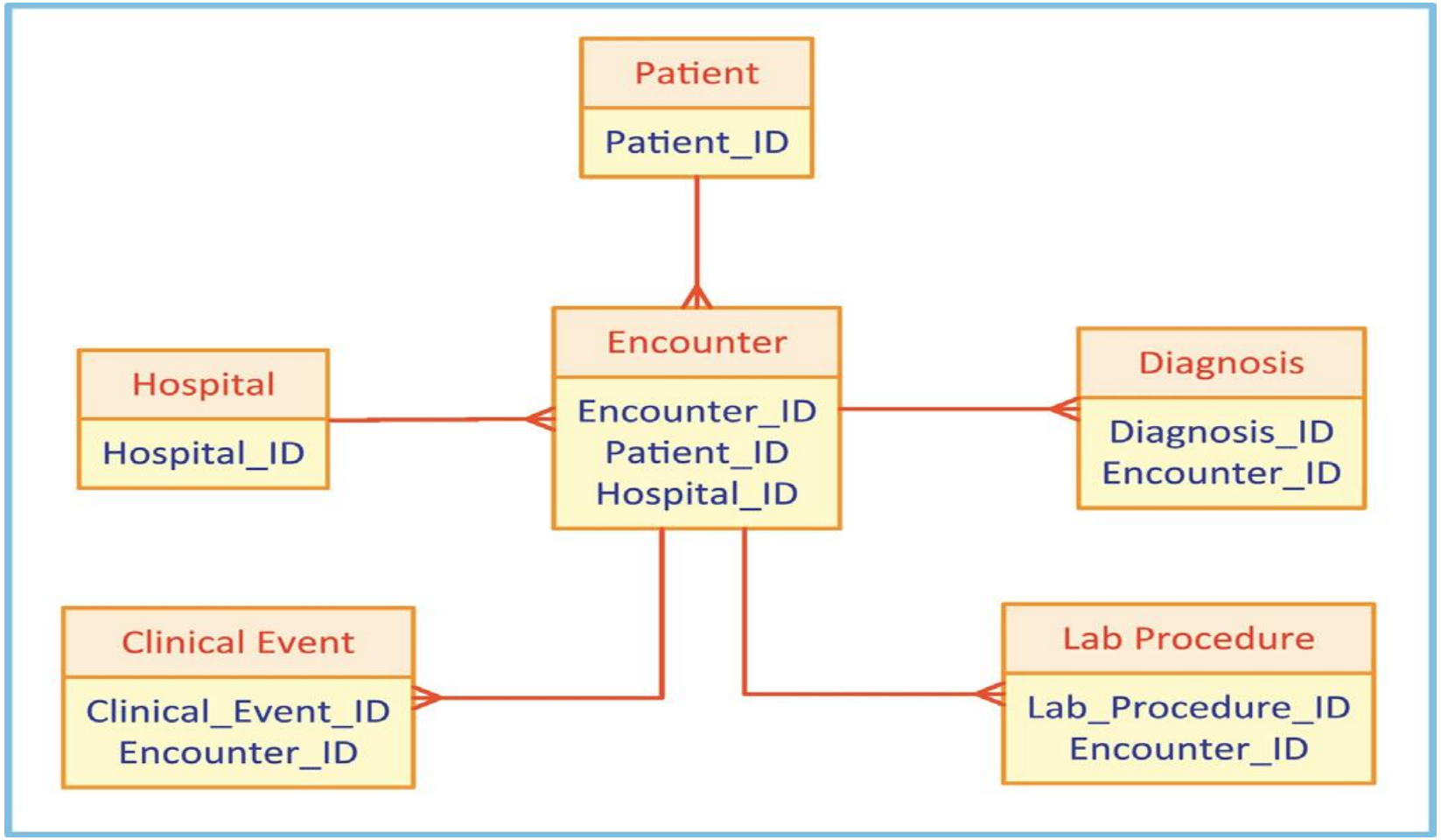
dbGaP Study Accession: phs000888.v1.p1

Using Diabetes Inclusion Exclusion Phenotype



Richesson RL, Rusincovitch SA, Wixted D, et al. A comparison of phenotype definitions for diabetes mellitus. J Am Med Inform Assoc. 2013;20(e2):e319-26.

Conceptual Data Model of Cerner Health Facts

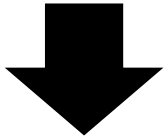


S. Piri, et al., A data analytics approach to building a clinical decision support system for diabetic retinopathy: Developing and deploying a mode..., Decision Support Systems (2017), http://dx.doi.org/10.1016/j.dss.2017.05.012

Data Set Processing

Aggregation

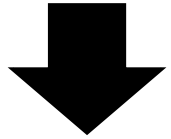
Hosp	Encounter	Labs / Meds	
ID	ID	Description	Value
1	1	Lab Name 1 (Variant 1)	value
2	2	Lab Name 1 (Variant 2)	value
3	3	Lab Name 2	value



Hosp	Encounter	Labs / Meds	
ID	ID	Description	Value
1	1	Lab Name 1	value
2	2	Lab Name 1	value
3	3	Lab Name 2	value

Table Transposition

Encounter	Labs / Meds	
ID	Description	Value
1	Lab Name 1	Some Value
1	Lab Name 2	Some Value
1	Lab Name 3	Some Value



Encounter ID	Lab Name 1	Lab Name 2	Lab Name 3
1	Some Value	Some Value	Some Value

Identifying Appropriate Data

- Define the question you want to study
- Specify the population you want to study
- Specify what variables you want to include in your analysis
  - Exclusion and inclusion criteria
- What kind of data is most appropriate for your research
- Contact the Clinical Research Data Warehouse at CRDW@ttuhsc.edu and we can help to obtain the proper data and datasets

Clinical Research Data Warehouse Website