



UNC

THE NORTH CAROLINA
TRANSLATIONAL & CLINICAL
SCIENCES INSTITUTE

Data-Driven Health and Clinical Research :

The Carolina Data Warehouse for Health and Clinical Data Research Networks

Emily Pfaff, MS

IT Director, NC TraCS Institute, UNC School of Medicine

September 22, 2016

What is NC TraCS?

- TraCS is UNC's NIH CTSA (**C**linical & **T**ranslational **S**cience **A**ward), one of 62 CTSA's across the US.
- Much more than research data provisioning; we touch all areas of *translational* science at UNC.
- Translational science: “Bench to bedside”



TraCS Informatics' Mission (in short)

- Serve as a technical and regulatory liaison between the UNC Health Care System and the university for *research* uses of Carolina Data Warehouse for Health data.
- Support approved secondary uses of those data in informatics and clinical research projects.



The Carolina Data Warehouse for Health

What is the CDW-H?

- The Carolina Data Warehouse for Health
- Aggregate of electronic health record data collected in UNCHCS, live as of 2009
- Data on ~4.8+ Million unique patients, 900K+ continuous, expanding with UNCHCS
- Data collection dates back to July 2004



The CDW-H Data Model

- The CDW-H contains data in all of the following domains (and more):
 - Patient demographics
 - Encounter details
 - Diagnoses
 - Procedures
 - Providers
 - Patient vitals
 - Lab tests
 - Medications
 - Orders
 - Notes
 - Charges and Payors
 - Surgery
 - Labor and delivery
 - Medical and social history
 - Patient-reported data
 - Custom data elements



“Computable Phenotypes”

- Can you pull data from the CDW-H that will show me all patients between ages _____ and _____, who have been diagnosed with _____, but haven't had a _____ in the last 6 months, but have had ___ visits in the _____ clinic over the past year? I also need to know if they're taking _____, or have had any ____, ____, or ___ lab values over ___ mg/ml in the past year.



Example uses of CDW-H data

- Support clinical operations and quality improvement efforts
- Collect multiple data points for analysis from patients meeting specific criteria, for operations, QI, or research purposes
- Create patient registries/data marts to support population health, QI, and research
- Embark on data-sharing projects with investigators at peer institutions



About Clinical Data Research Networks (CDRNs)

Clinical Data Research Networks

- A CDRN allows UNC to “share” clinical data with participating institutions.
- Federated data model
 - Data and data warehouses stay distinct, each organization controls use of its own data
- Generally based within large integrated delivery systems
- Supported by PCORI, NIH, Foundations (Duke Endowment)

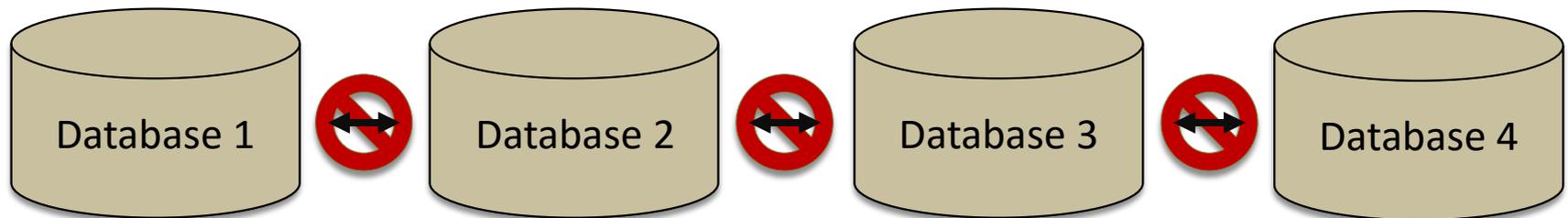


How do we federate data with other systems?

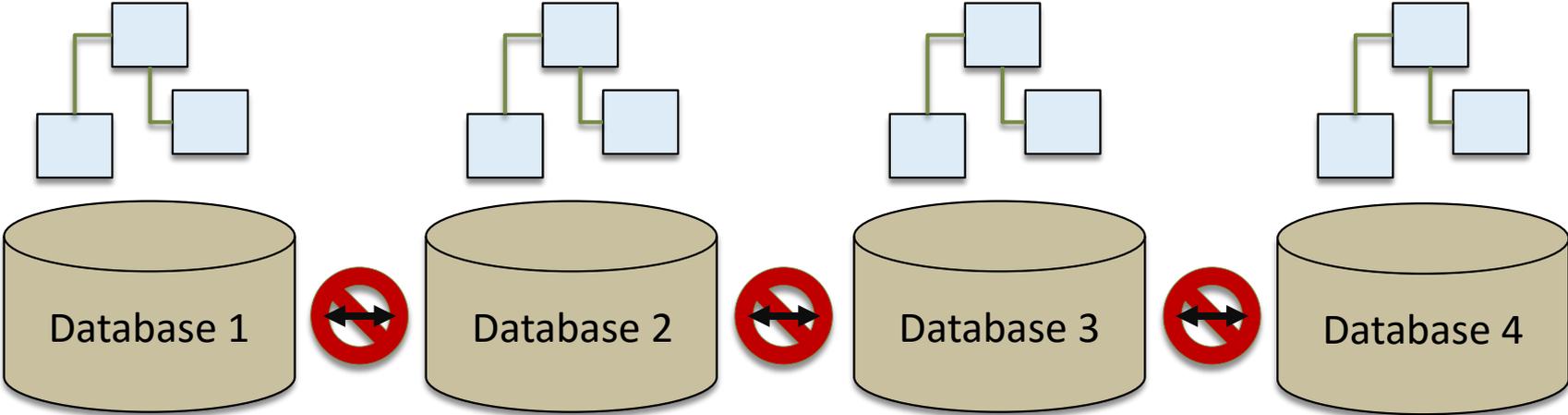
- Through use of a common data model, or CDM.
- Data tables are small in number and generic in structure.
- Each institution is responsible for transforming their data from their electronic health records into the CDM chosen by the CDRN. Then the data can be queried as one, yet kept separate.



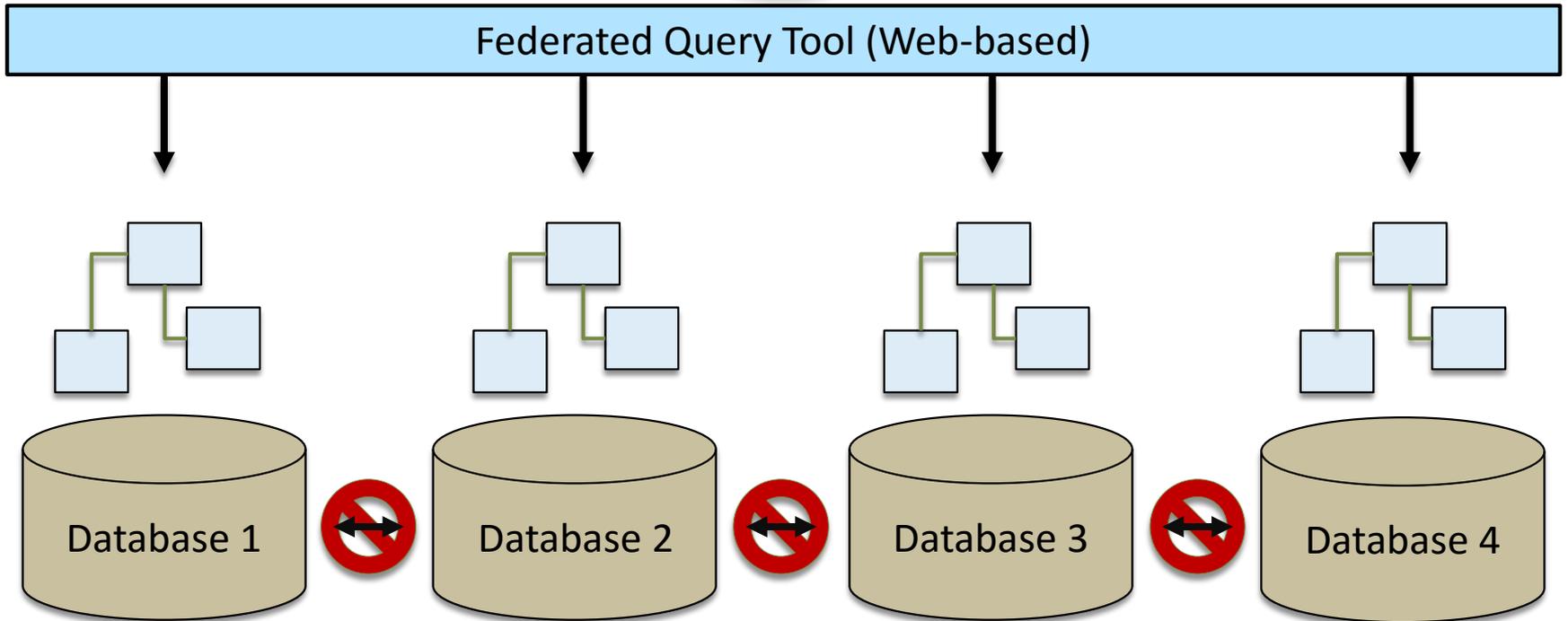
Data Federation



Data Federation



Data Federation



PCORnet Common Data Model

PCORnet Common Data Model v3.0

New to v3.0

DEMOGRAPHIC
PATID
BIRTH_DATE
BIRTH_TIME
SEX
HISPANIC
RACE
BIOBANK_FLAG

Fundamental basis

ENROLLMENT
PATID
ENR_START_DATE
ENR_END_DATE
CHART
ENR_BASIS

DISPENSING
DISPENSINGID
PATID
PRESCRIBINGID (optional)
DISPENSE_DATE
NDC
DISPENSE_SUP
DISPENSE_AMT

DEATH
PATID
DEATH_DATE
DEATH_DATE_IMPUTE
DEATH_SOURCE
DEATH_MATCH_CONFIDENCE

DEATH_CONDITION
PATID
DEATH_CAUSE
DEATH_CAUSE_CODE
DEATH_CAUSE_TYPE
DEATH_CAUSE_SOURCE
DEATH_CAUSE_CONFIDENCE

Data captured from processes associated with healthcare delivery

VITAL
VITALID
PATID
ENCOUNTERID (optional)
MEASURE_DATE
MEASURE_TIME
VITAL_SOURCE
HT
WT
DIASTOLIC
SYSTOLIC
ORIGINAL_BMI
BP_POSITION
SMOKING
TOBACCO
TOBACCO_TYPE

CONDITION
CONDITIONID
PATID
ENCOUNTERID (optional)
REPORT_DATE
RESOLVE_DATE
ONSET_DATE
CONDITION_STATUS
CONDITION
CONDITION_TYPE
CONDITION_SOURCE

PRO_CM
PRO_CM_ID
PATID
ENCOUNTERID (optional)
PRO_ITEM
PRO_LOINC
PRO_DATE
PRO_TIME
PRO_RESPONSE
PRO_METHOD
PRO_MODE
PRO_CAT

Data captured within multiple contexts: healthcare delivery, registry activity, or directly from patients

ENCOUNTER
ENCOUNTERID
PATID
ADMIT_DATE
ADMIT_TIME
DISCHARGE_DATE
DISCHARGE_TIME
PROVIDERID
FACILITY_LOCATION
ENC_TYPE
FACILITYID
DISCHARGE_DISPOSITION
DISCHARGE_STATUS
DRG
DRG_TYPE
ADMITTING_SOURCE

DIAGNOSIS
DIAGNOSISID
PATID
ENCOUNTERID
ENC_TYPE (replicated)
ADMIT_DATE (replicated)
PROVIDERID (replicated)
DX
DX_TYPE
DX_SOURCE
PDX

PROCEDURES
PROCEDURESID
PATID
ENCOUNTERID
ENC_TYPE (replicated)
ADMIT_DATE (replicated)
PROVIDERID (replicated)
PX_DATE
PX
PX_TYPE
PX_SOURCE

Data captured from healthcare delivery, direct encounter basis

LAB_RESULT_CM
LAB_RESULT_CM_ID
PATID
ENCOUNTERID (optional)
LAB_NAME
SPECIMEN_SOURCE
LAB_LOINC
PRIORITY
RESULT_LOC
LAB_PX
LAB_PX_TYPE
LAB_ORDER_DATE
SPECIMEN_DATE
SPECIMEN_TIME
RESULT_DATE
RESULT_TIME
RESULT_QUAL
RESULT_NUM
RESULT_MODIFIER
RESULT_UNIT
NORM_RANGE_LOW
NORM_MODIFIER_LOW
NORM_RANGE_HIGH
NORM_MODIFIER_HIGH
ABN_IND

PRESCRIBING
PRESCRIBINGID
PATID
ENCOUNTERID (optional)
RX_PROVIDERID
RX_ORDER_DATE
RX_ORDER_TIME
RX_START_DATE
RX_END_DATE
RX_QUANTITY
RX_REFILLS
RX_DAYS_SUPPLY
RX_FREQUENCY
RX_BASIS
RXNORM_CUI

PCORNET_TRIAL
PATID
TRIALID
PARTICIPANTID
TRIAL_SITEID
TRIAL_ENROLL_DATE
TRIAL_END_DATE
TRIAL_WITHDRAW_DATE
TRIAL_INVITE_CODE

Associations with PCORnet clinical trials

HARVEST
NETWORKID
NETWORK_NAME
DATAMARTID
DATAMART_NAME
DATAMART_PLATFORM
CDM_VERSION
DATAMART_CLAIMS
DATAMART_EHR
BIRTH_DATE_MGMT
ENR_START_DATE_MGMT
ENR_END_DATE_MGMT
ADMIT_DATE_MGMT
DISCHARGE_DATE_MGMT
PX_DATE_MGMT
RX_ORDER_DATE_MGMT
RX_START_DATE_MGMT
RX_END_DATE_MGMT
DISPENSE_DATE_MGMT
LAB_ORDER_DATE_MGMT
SPECIMEN_DATE_MGMT
RESULT_DATE_MGMT
MEASURE_DATE_MGMT
ONSET_DATE_MGMT
REPORT_DATE_MGMT
RESOLVE_DATE_MGMT
PRO_DATE_MGMT
REFRESH_DEMOGRAPHIC_DATE
REFRESH_ENROLLMENT_DATE
REFRESH_ENCOUNTER_DATE
REFRESH_DIAGNOSIS_DATE
REFRESH_PROCEDURES_DATE
REFRESH_VITAL_DATE
REFRESH_DISPENSING_DATE
REFRESH_LAB_RESULT_CM_DATE
REFRESH_CONDITION_DATE
REFRESH_PRO_CM_DATE
REFRESH_PRESCRIBING_DATE
REFRESH_PCORNET_TRIAL_DATE
REFRESH_DEATH_DATE
REFRESH_DEATH_CAUSE_DATE

Process-related data

Bold font indicates fields that cannot be null due to primary key definitions or record-level constraints.

Federated Data versus HIE

- Form follows function.
- Federated database infrastructure is useful for ad hoc querying across institutions.
 - Ex: How many patients in this network are <65 and have end-stage renal disease?
- Health Information Exchange is useful to support clinical care, in real time.
 - Ex: What care has Ms. Johnson received at Duke that may have an impact on my clinical decisions at her appointment today?



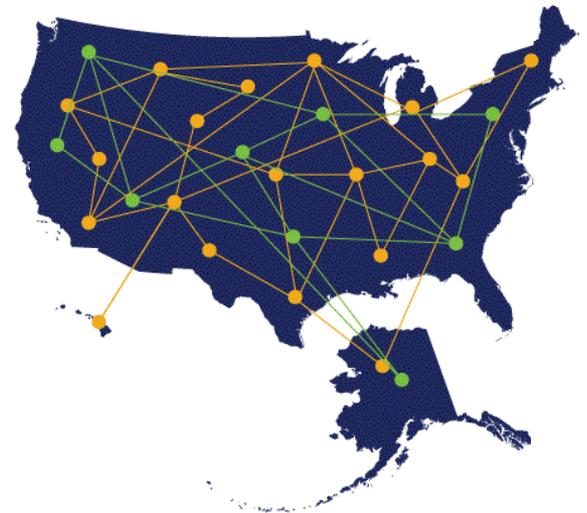
UNC and CDRNs

- UNC is participating in two CDRNs, the **Carolinas Collaborative** (UNC, Duke, Wake Forest, and Health Sciences South Carolina) and the **Mid-South CDRN** (Vanderbilt, UNC, Duke, HSSC).
- We are pooling our (deidentified) clinical data with these collaborators so that a query can be run across multiple institutions and health systems at once.
- Why pool data?
 - to do multi-institutional quality improvement work.
 - to improve recruitment for multi-site trials.
 - to make “big data” actually “big.”



PCORnet

- PCORI's National Patient-Centered Clinical Research Network
 - 13 clinical data research networks
- Designed to make it **faster, easier, less costly** to conduct clinical research...by harnessing the power of large amounts of health data and patient partnerships
- ~145 million lives

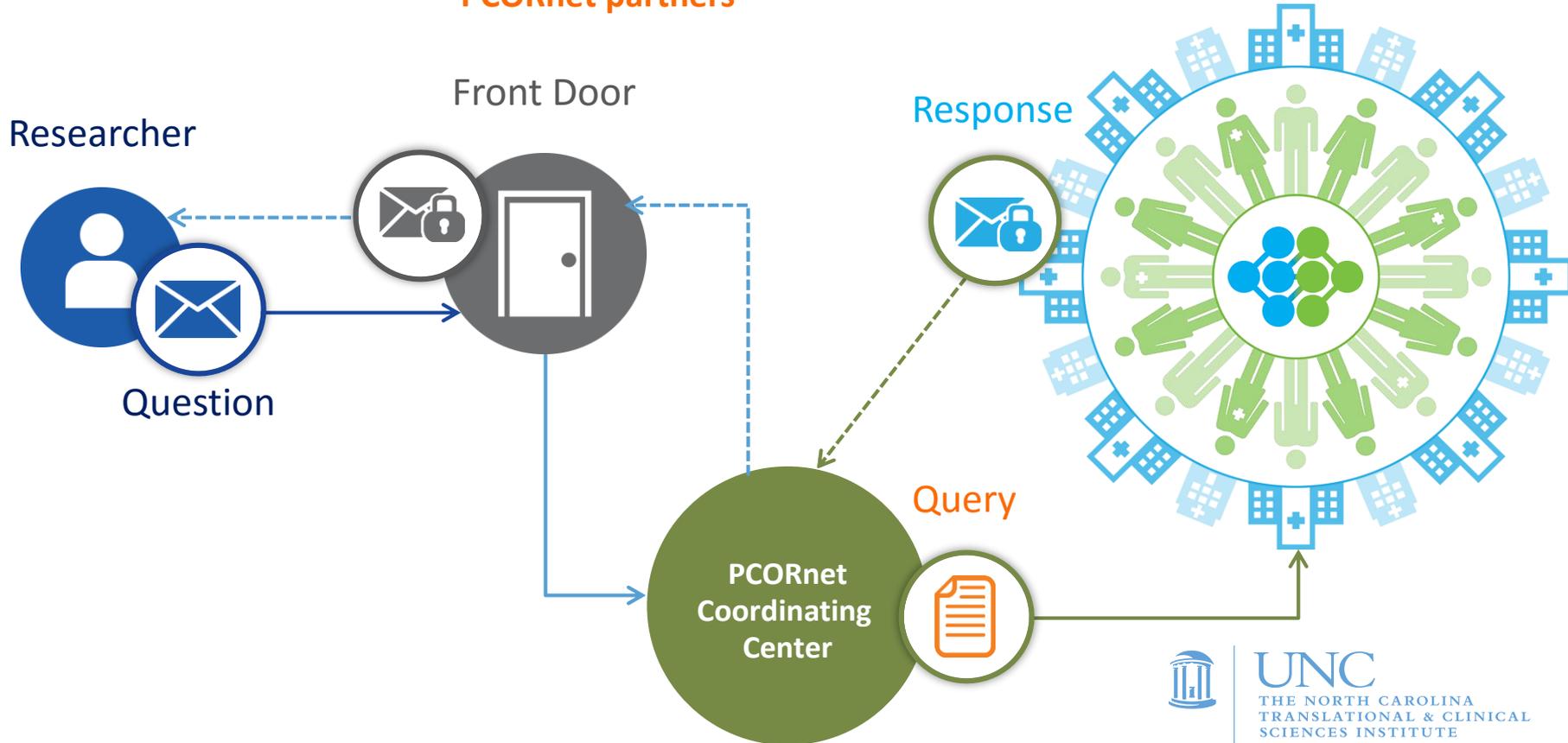


Here's how PCORnet's distributed research network works

The Researcher sends a question to the PCORnet Coordinating Center through the Front Door

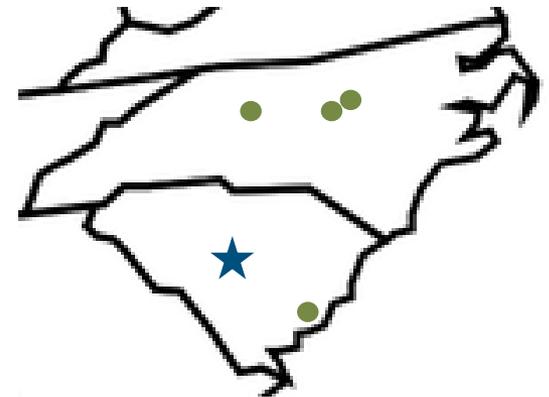
The Coordinating Center converts the question into a query with an underlying executable code, and sends it to PCORnet partners

PCORnet partners review the query and provide a response, which is sent back through the Front Door to the Researcher



Carolinas Collaborative

- A “local” collaborative for the Carolinas, member of the Mid-South PCORnet
- Designed to complement the Mid-South CDRN
- Involved sites:
 - Health Sciences South Carolinas (lead), including Medical University of South Carolina
 - UNC-Chapel Hill
 - Duke University
 - Wake Forest Baptist Medical Center
- Funded by The Duke Endowment



Technical challenges

- Agreement on common data elements
- De-duplication of patients when delivery systems are adjacent to each other (UNC, Duke)
- Development of computable disease phenotypes for multiple clinical conditions
- Expectations for rapid turnaround of requests
- Linkages with vital status records, geocodes, claims
- Secure computing and data environments while maintaining appropriate access



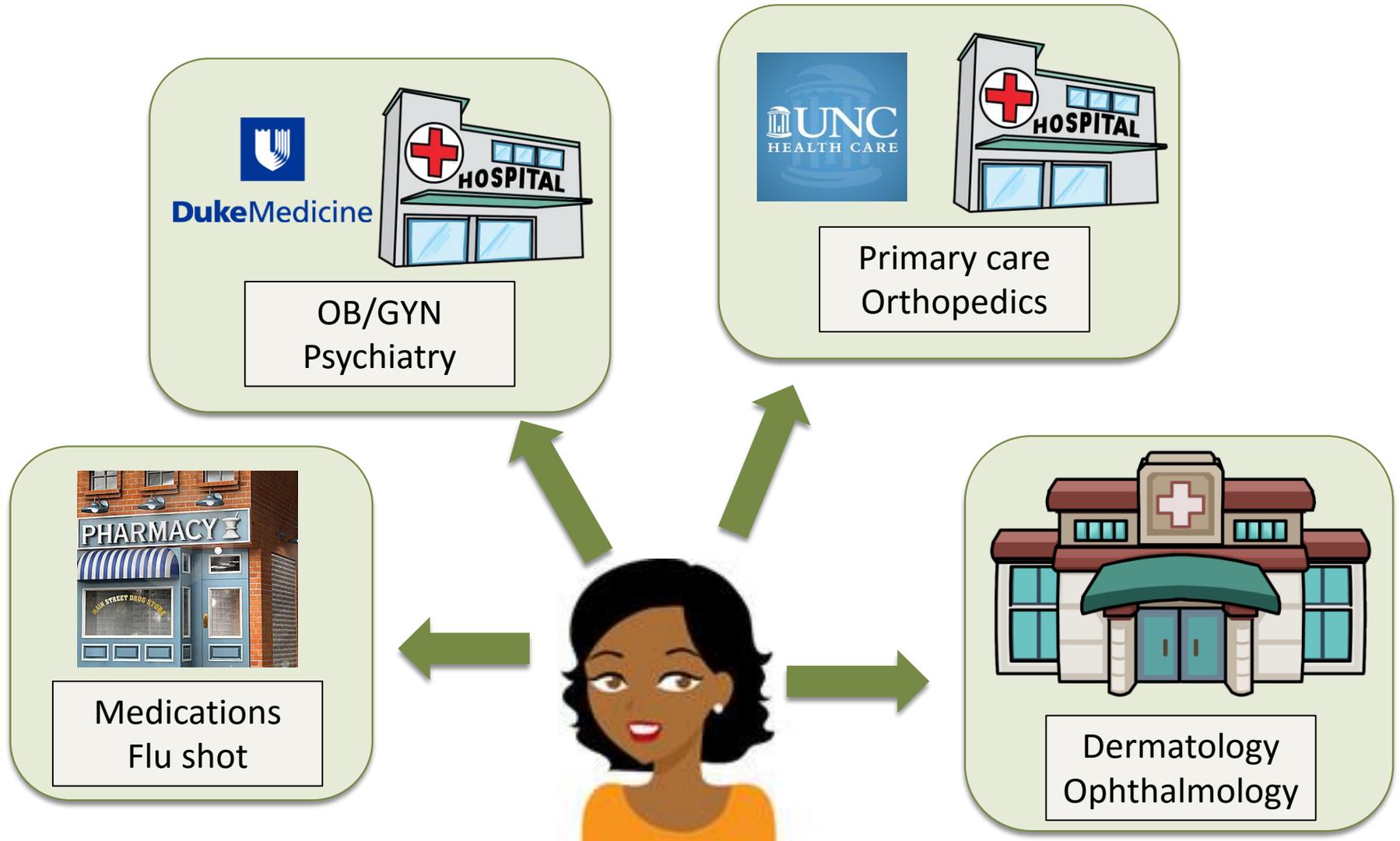
Where does claims data
come in?

Linkage with Claims Data

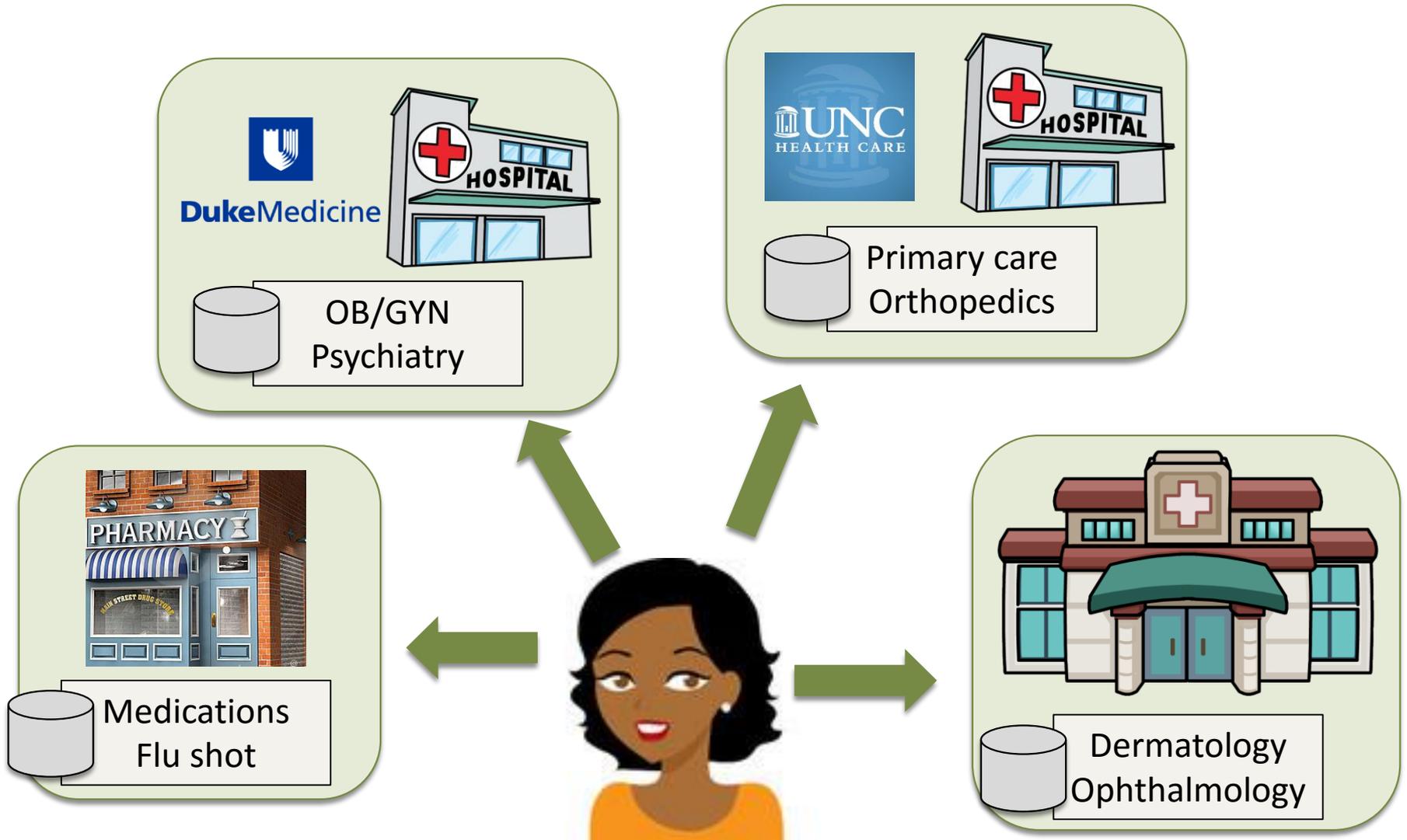
- Integrated delivery systems are ‘open,’ delivering variable proportions of care for a given patient
- Claims data, while less detailed, is more comprehensive
- Linkage of EMR, claims and direct patient reports may provide comprehensive portrait of care



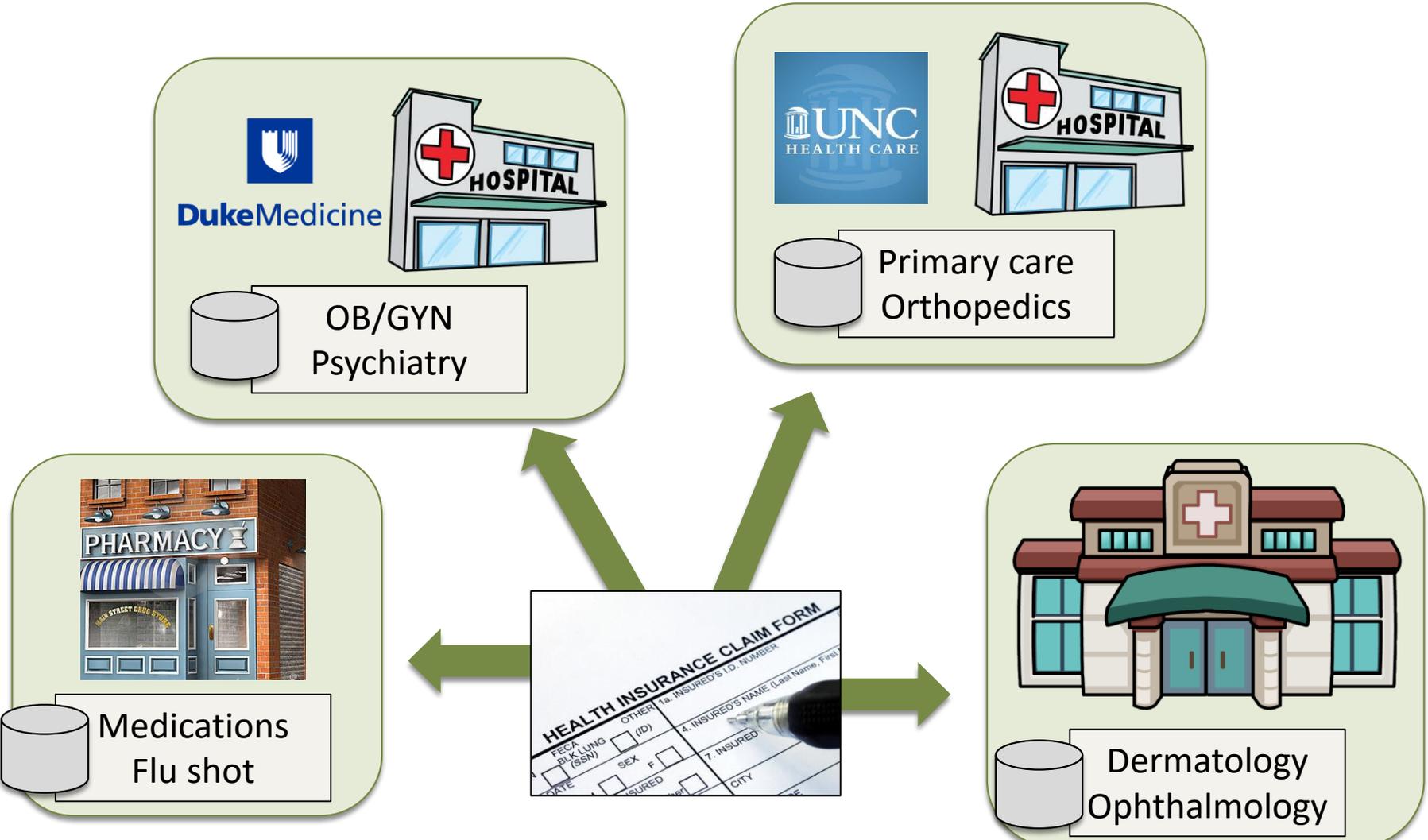
Linkage with Claims Data



Linkage with Claims Data



Linkage with Claims Data



Linkage with Claims Data

- Possible to use linkage methodologies to combine EHR and claims data.
- Provides a ***comprehensive, longitudinal*** view of patient's health care utilization and outcomes.
- More complete data = better patient care, better quality, better research.
- Governance is paramount, but once precedent is set, future projects are easier.



Thank you!

Questions?

Discussion?