Challenges and efforts for big data in health care in the USA

Marianne Huebner

Michigan State University, USA

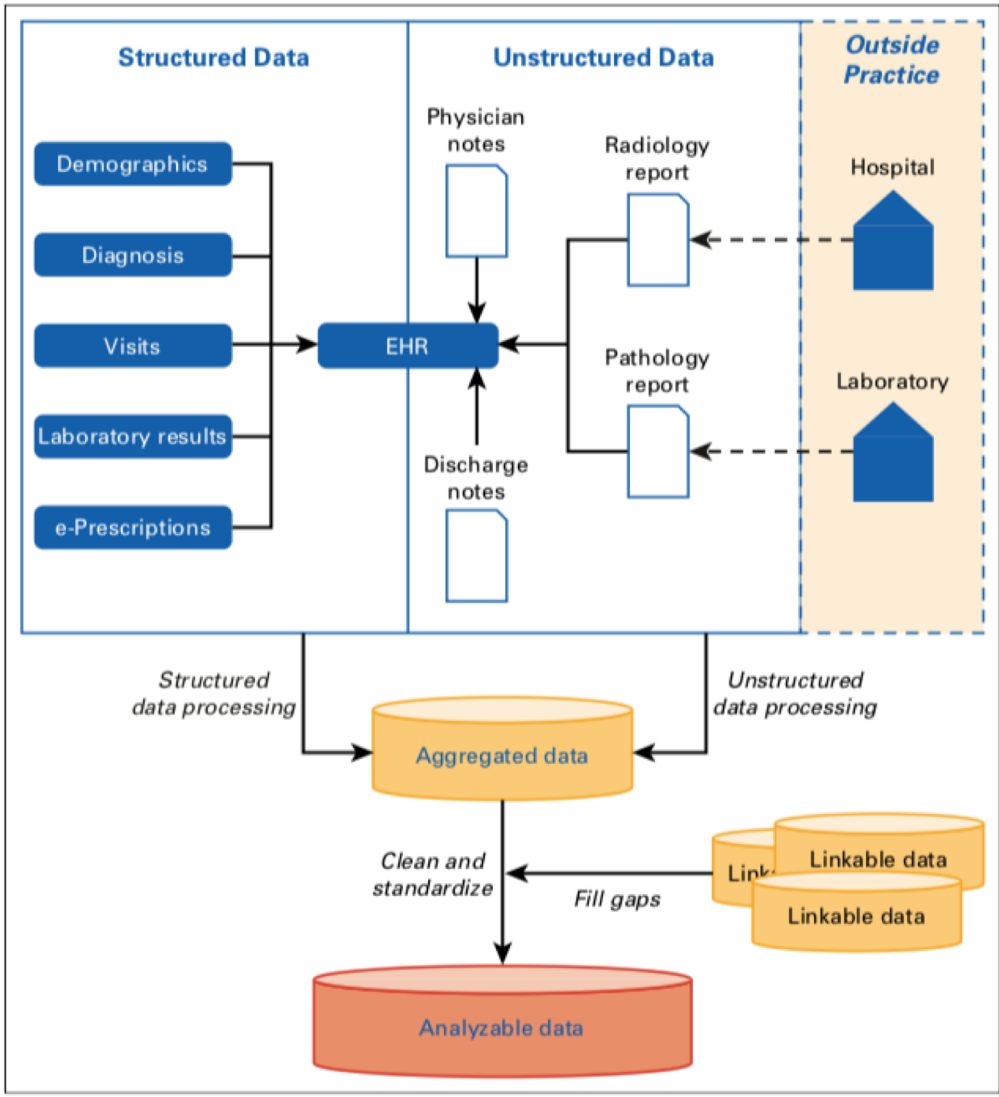
University Medical Center, Hamburg

MICHIGAN STATE
UNIVERSITY

UKE
HAMBURG

From EHR to analyzable data

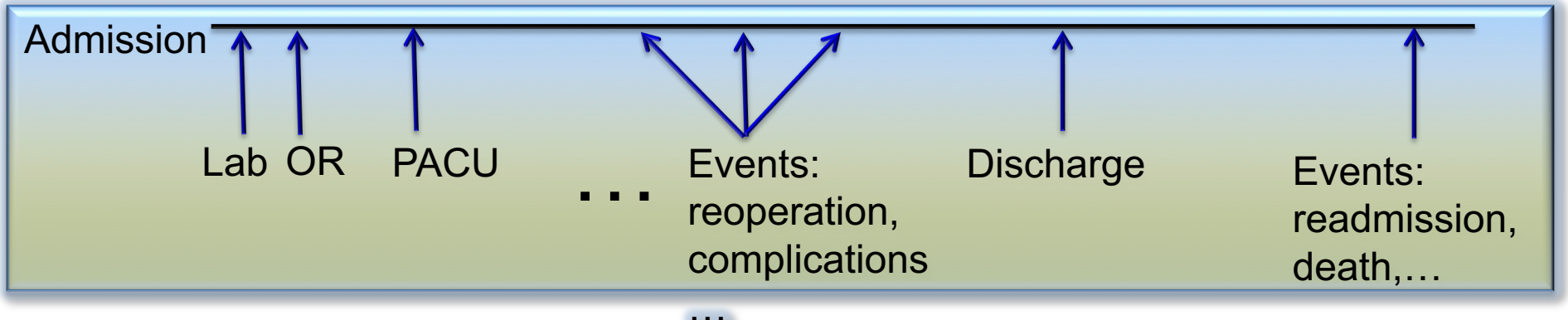
- Facilitate data collection at the point of care
- Retrieval of unstructured data: manual or automatic
- Quality measures: benchmark to gold standards



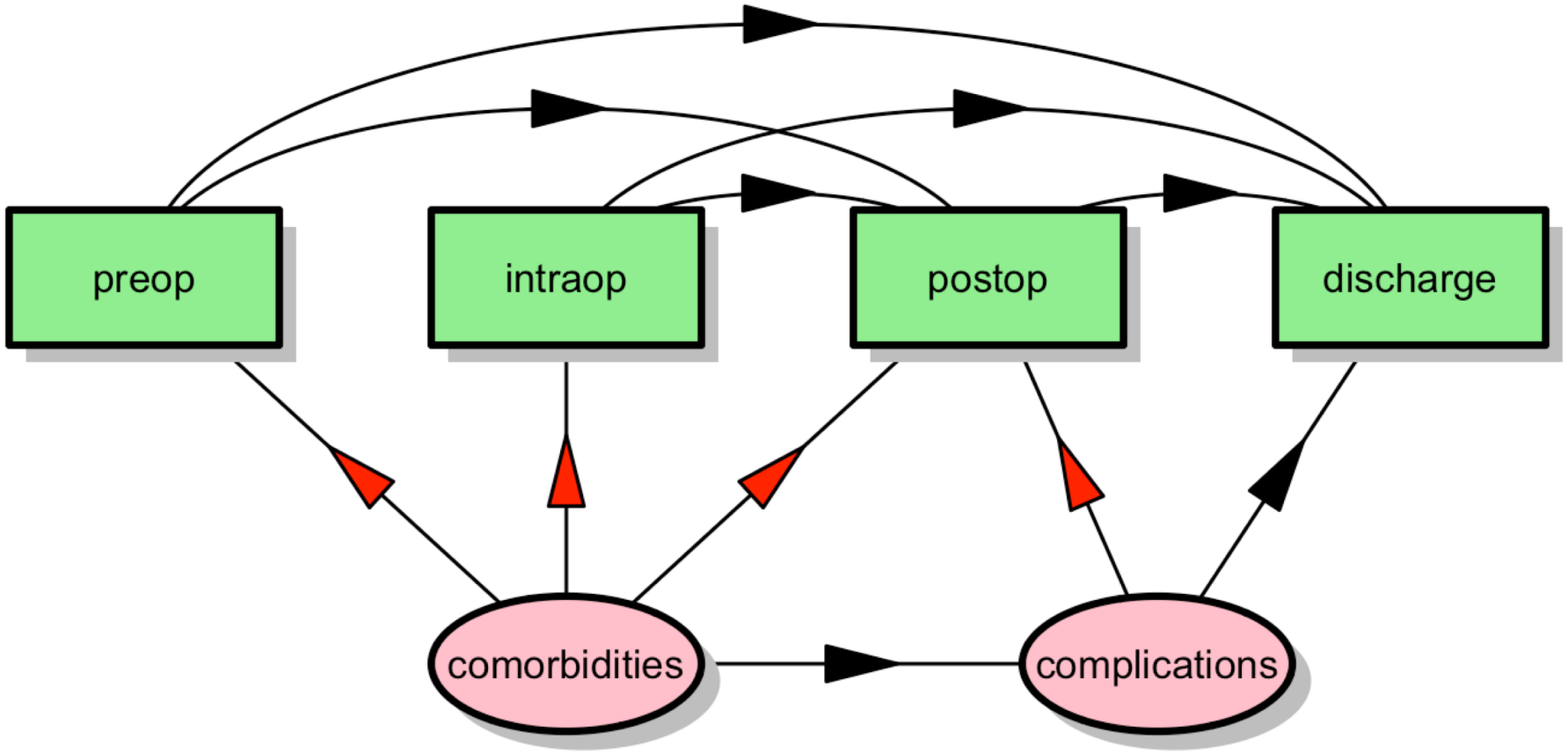
Abernathy et al, JOP 2017; 13: 530-534

Electronic health records: Hospitals

Sources	Size [MB]	Data types	
Cohort	1.21	TS, C, N	
Registration	1.07	C,N	TS= time stamp, C=categorical, N=numeric, T=text
Allergies	1.8	TS, N, T	
Labs	312	TS, N, T	
Pharmacy	174	TS, T, N	
Clinical Notes	3.7	TS, T, N	
Nurses Flow sheet	830	TS, T, C,N	
Orders	405	TS, T	
Insurance payments	…	…	



Enhanced recovery pathway

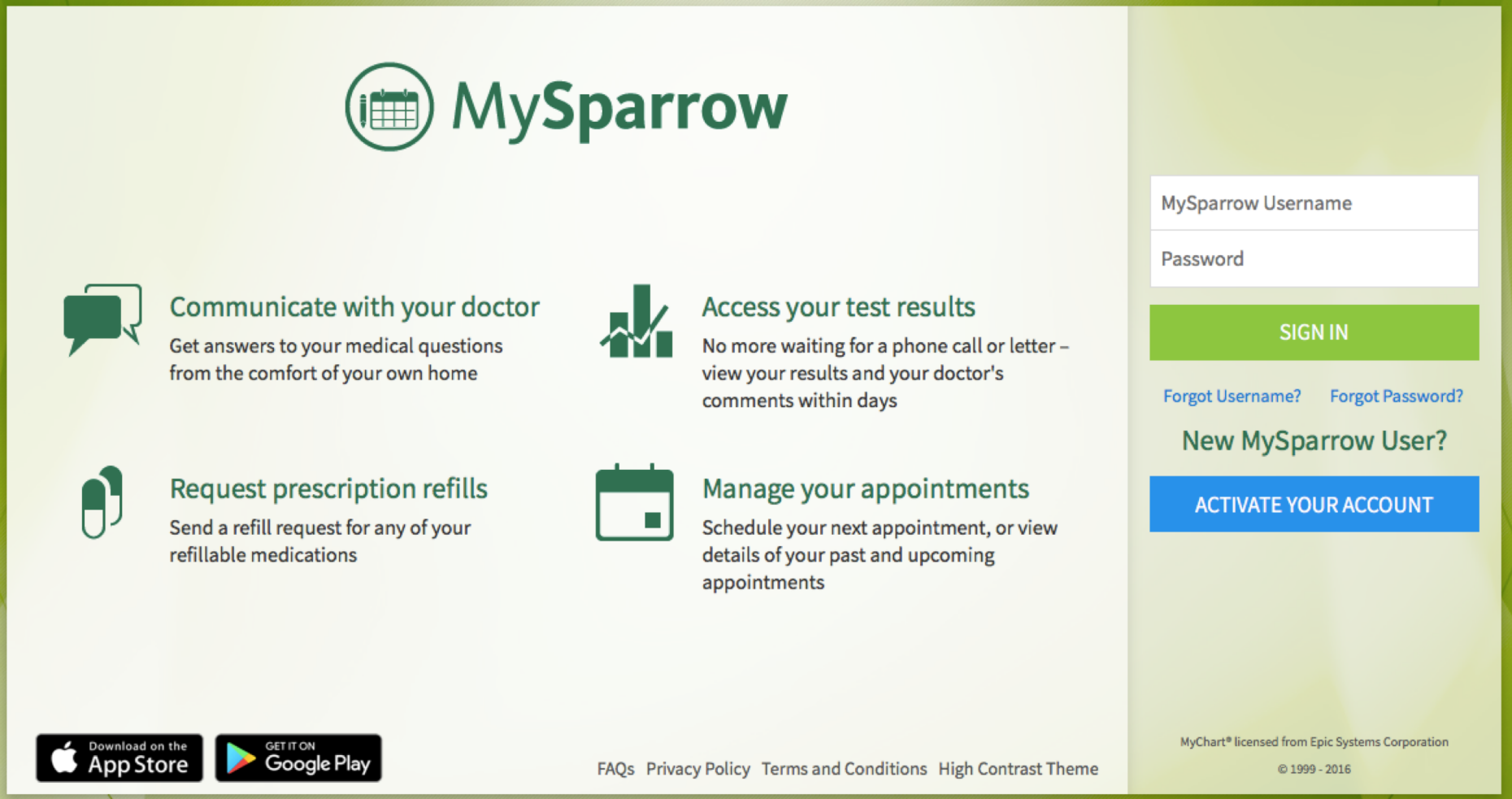


Data to clinicians

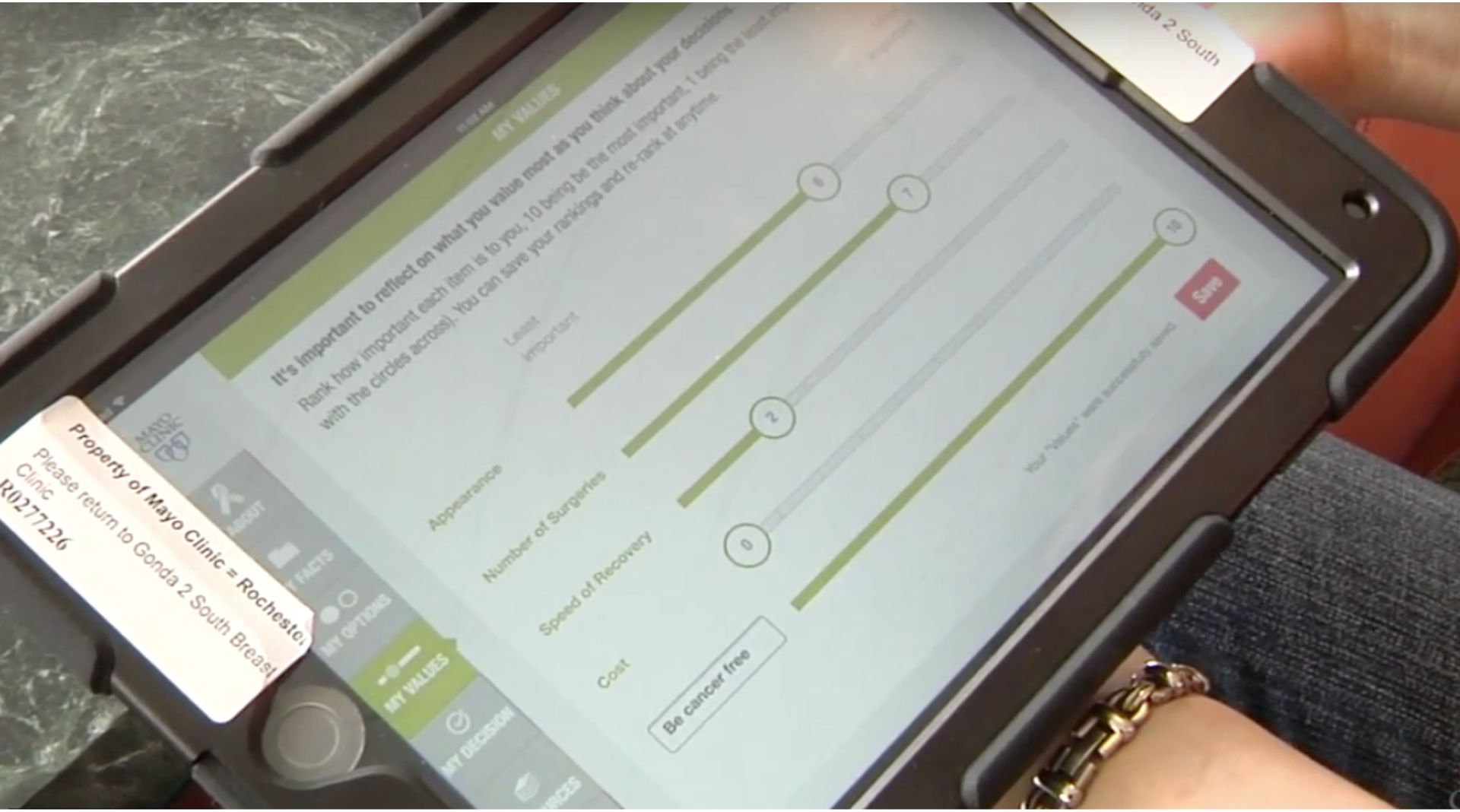
- Phone App
- Identify potential problems (pain, diet management, complications)
- Clinician effort
  - Current medical records: 30 mi (95% navigation, 619 mouse clicks)
  - New Tool: 4 min (25 mouse clicks)

Compliance with clinical pathway elements according to benchmarks

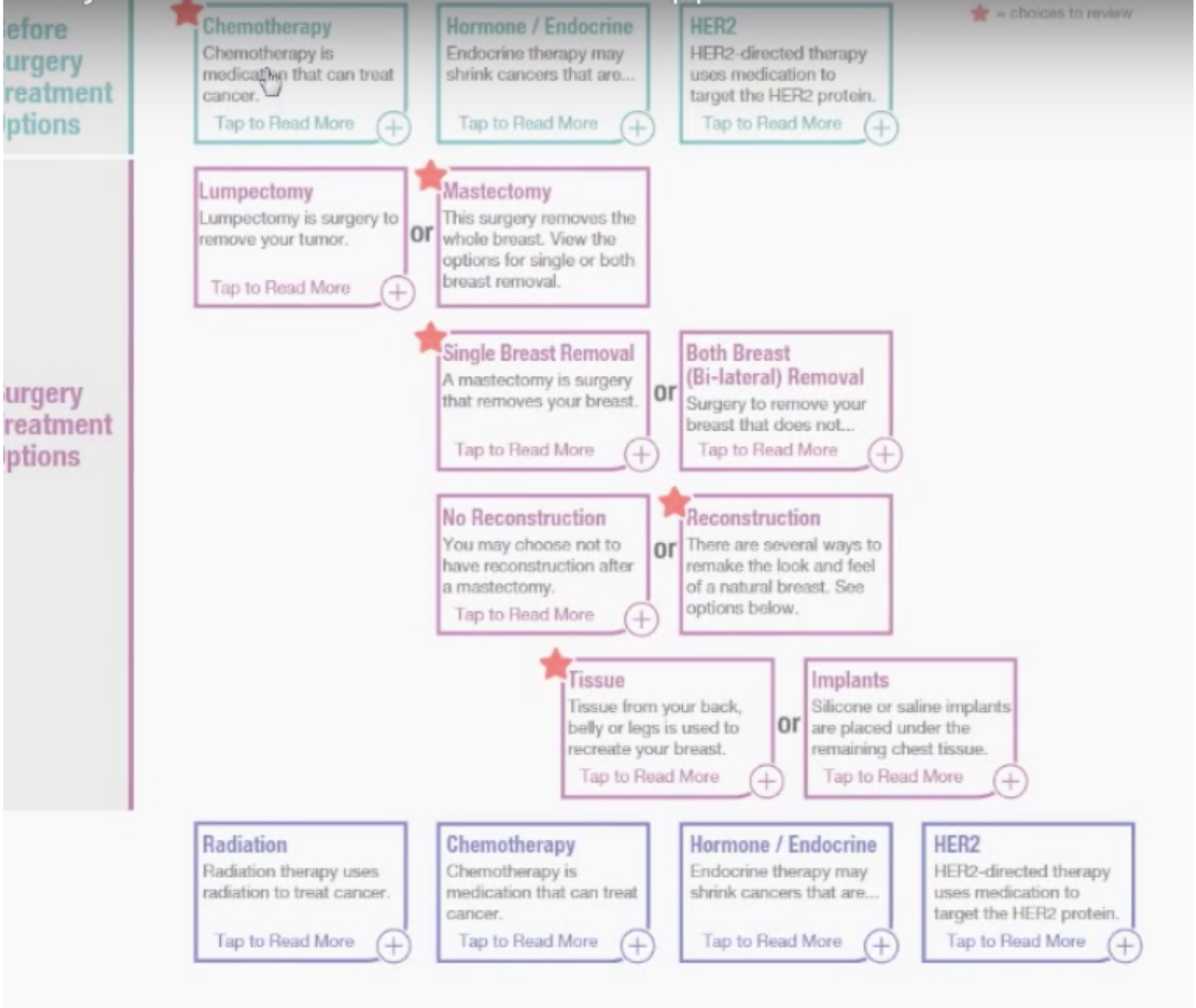
Don’t ignore this feedback loop: Data to patients



Decision making for patients



https://youtu.be/EK5y3EfHKRY


efore
urgery
reatment
ptions
= choices to review
Chemotherapy
Chemotherapy is medication that can treat cancer.
Tap to Read More
Hormone / Endocrine
Endocrine therapy may shrink cancers that are...
Tap to Read More
HER2
HER2-directed therapy uses medication to target the HER2 protein.
Tap to Read More
urgery
reatment
ptions
Lumpectomy
Lumpectomy is surgery to remove your tumor.
Tap to Read More
or
Mastectomy
This surgery removes the whole breast. View the options for single or both breast removal.
Single Breast Removal
A mastectomy is surgery that removes your breast.
Tap to Read More
or
Both Breast (Bi-lateral) Removal
Surgery to remove your breast that does not...
Tap to Read More
No Reconstruction
You may choose not to have reconstruction after a mastectomy.
Tap to Read More
or
Reconstruction
There are several ways to remake the look and feel of a natural breast. See options below.
Tissue
Tissue from your back, belly or legs is used to recreate your breast.
Tap to Read More
or
Implants
Silicone or saline implants are placed under the remaining chest tissue.
Tap to Read More
Radiation
Radiation therapy uses radiation to treat cancer.
Tap to Read More
Chemotherapy
Chemotherapy is medication that can treat cancer.
Tap to Read More
Hormone / Endocrine
Endocrine therapy may shrink cancers that are...
Tap to Read More
HER2
HER2-directed therapy uses medication to target the HER2 protein.
Tap to Read More

Multi-center/registry level

- NSQIP=surgical quality (American College of Surgeons)
- SEER= cancer statistics (US Federal)
- HCUP= trends about hospital stays, discharges (US Federal)
- MEPS=medical expenditures (US Federal)
- STS Database (Society of Thoracic Surgeons)

Advantages of using registries

- Readily accessible
- Detailed subgroup analyses
- Collection of low prevalence events
- Longitudinal designs
- Healthcare delivery designs

Disadvantages of using registries

- Hospitals
  - Costs (participation, trained abstractor)
  - Sample of patients (e.g. 20% for NSQIP)
  - Time to abstract (e.g. 20 min for 135 fields in NSQIP)
- Researchers:
  - Data lag (12 months?)
  - Less details (“data depth”)
  - Understand limitations and strengths of these databases

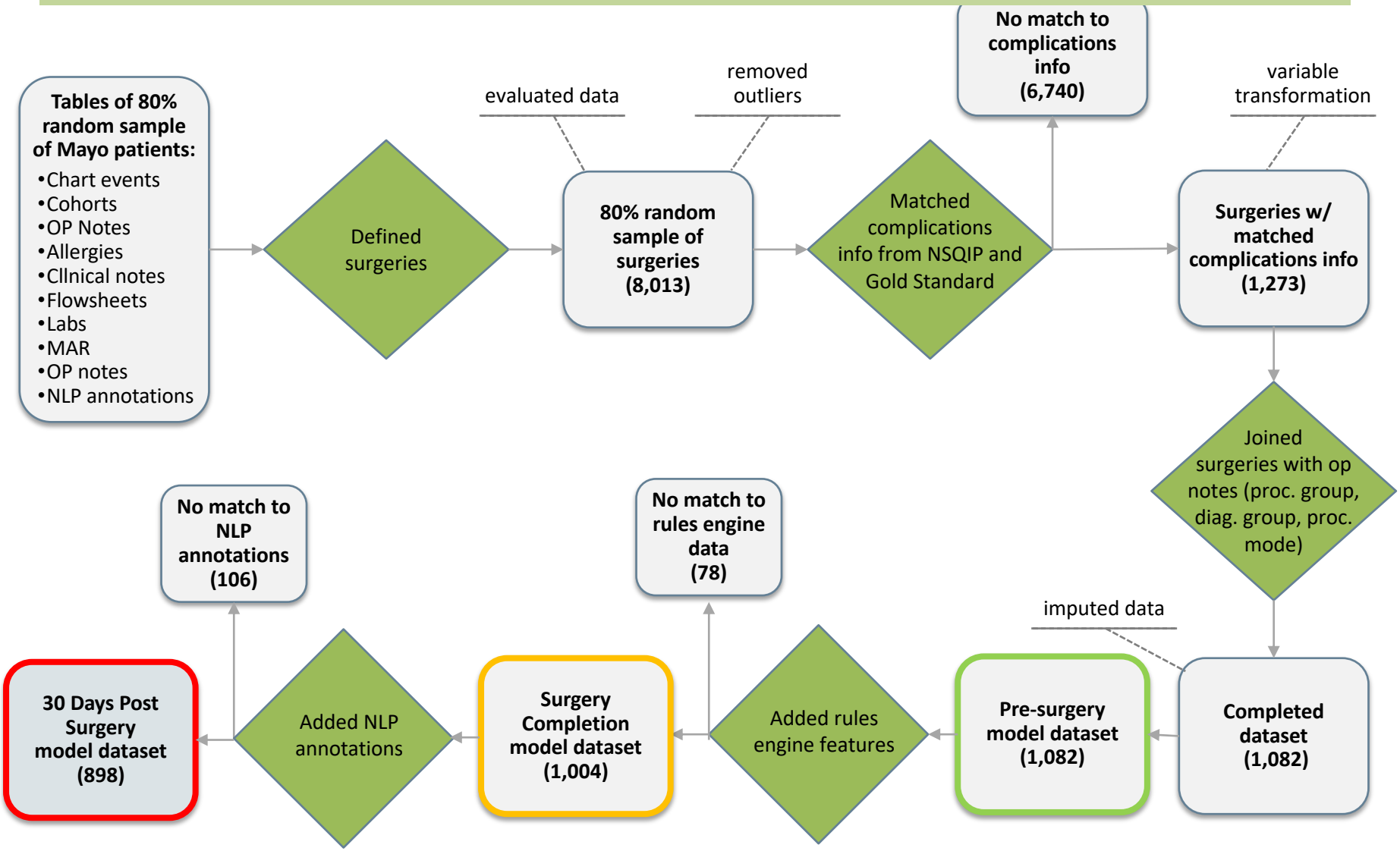
Discordance: administrative and registry

	Administrative (EHR)	Clinical registry, e.g NSQIP
Purpose	Billing	Quality of care
Level	Hospital	National
Input	Providers; multiple formats	Trained staff in abstracting; standard formats
Events	Entire course of hospitalizations	Set of complications within 30 days
Comp rate*	1.5%	0.6%

Discordance: 43% due to definitions/criteria
Coding error (5% both, 37%admin, 15% NSQIP)

** Etzioni et al. Ann. Surg 2018. n=16,559*

Data to researchers



Weller et al. Stat Meth Med Res 2017

Many data sources for researchers

- Electronic health records
- Integrating survey data, images, reports,…
- Compliance with clinical protocols
- Treatment choices confounded with patient's values
- Information from wearable health devices
- Multiple centers, different countries (language, formats)
- Reading data from the web, e.g. online health social networks

EHR for statisticians: Lessons learned

1. EHR development:
   - Statisticians need to be involved in all steps of the EHR development: “Otherwise it’s a mess…”
   - How was the data acquired, managed, and curated?
   - Understand limitations and strengths of data sources
2. Extract data for analysis
   - Ethics approval may be required from multiple institutions/agencies
   - Identify inclusion/exclusion criteria
   - Selection vs confounding bias

EHR for statisticians: Lessons learned

3. Prepare data for analysis
   - Integration of different data sources
   - Natural language processing
   - Data visualization
   - Bigger or complex data => MORE data cleaning
   - Evaluate many confounding factors