Patient Linkage

Shaun Grannis, MD, MS
Research Scientist, Regenstrief Institute
Department of Family Medicine
Indiana University School of Medicine

Motivation

- Health care information is fragmented across many independent databases
- This situation makes record linkage crucial for such uses:
  - Clinical Care
  - Public Health Reporting
  - Research
  - Outcomes management
  - Vital status determination

Challenges

- Recording Errors
  - Phonetic
  - typographical
- Identifiers change
  - Last Name
  - Address
  - Phone
- Sharing Identifiers (SSN, etc.)

Potential Solutions

- National Patient Identifier
  - Recording errors
  - Sharing ID's
  - Lost/forgotten ID's
- Biometrics
  - Requires use of proprietary technology
  - Privacy concerns
- Linking algorithms

Linkage Methodologies

Deterministic/Heuristic	Probabilistic/Statistical
▪ Rapid Implementation	▪ Complex implementation
▪ Simple calculations	▪ Computationally intensive
▪ Relies on accurate data	▪ More forgiving of data errors
▪ May not function well with other data sets	▪ Algorithm is customized to data being linked

Previous Work

- Deterministic

  Any one of four combinations of identifiers including SSN, name and birth date.

- Results:
  - 90% Sensitive
  - 100% Specific

False-positive rate using SSN as the single linkage variable

	Hospital A	Hospital B
Indiana Subset	9.2%	4.7%
Full SSDMF	11.7%	5.7%

Causes of False Links with SSN

- 1/3 due to mix-up between beneficiary and guarantor:

Hospital data: **123-45-6789, DOYLE, ISAIAH, F, MALE, 12-20-1902**
SS death tape: **123-45-6789, DOYLE, ELLA, J, FEMALE, 03-04-1904**
SS death tape: **987-65-4321, DOYLE, ELLA, J, FEMALE, 03-04-1904**

- 1/3 from Typographical Error:

Hospital data: **123-45-6789, SMALL, FRANK, L, MALE, 10-21-1922**
SS death tape: **123-45-6789, JONES, PAT, D, MALE, 09-30-1943**
SS death tape: **123-46-6789, SMALL, FRANK, L, MALE, 10-21-1922**

- 1/3 from Unexplained collision:

Hospital data: **989-76-7654, BIGBY, TOMAS, Q, MALE, 11-07-1902**
SS death tape: **989-76-7654, WILSON, JAMES, D, MALE, 04-13-1913**

Hospital A linked to SSDMF: Name, Date of Birth, and Gender without SSN

Linkage Criteria	Indiana SSDMF		Full SSDMF	
	Sensitivity	Specificity	Sensitivity	Specificity
LN, FN, MB, DB, YB, G	70.1%	100%	70.0%	89.1%

Hospital B linked to SSDMF: Name, Date of Birth, and Gender without SSN

Linkage Criteria	Indiana SSDMF		Full SSDMF	
	Sensitivity	Specificity	Sensitivity	Specificity
LN, FN, MB, DB, YB, G	81.9%	100%	82.6%	89.5%

False Links: Top Name Frequencies from National SS Death Master File

1. William Smith (16,602)
2. James Smith (15,873)
3. John Smith (15,578)
4. Mary Smith (14,175)
5. Robert Smith (11,759)
6. William Johnson (10,254)
7. James Williams (10,059)

Replacing SSN with Zip code

Linkage Criteria	Hospital A		Hospital B	
	Sensitivity	Specificity	Sensitivity	Specificity
Combination of ZIP, FNY ,G, YB ZIP, FNY ,G, MB ZIP, FNY, G, DB	49.2%	100%	59.0%	100%

Current Work

- Compare probabilistic/statistical methods with deterministic/heuristic methods
- Outstanding Questions:
  How do probabilistic linkage algorithms compare with heuristic/deterministic methods?
- What are patient linkage best practices?

Probabilistic Linkage, Overview

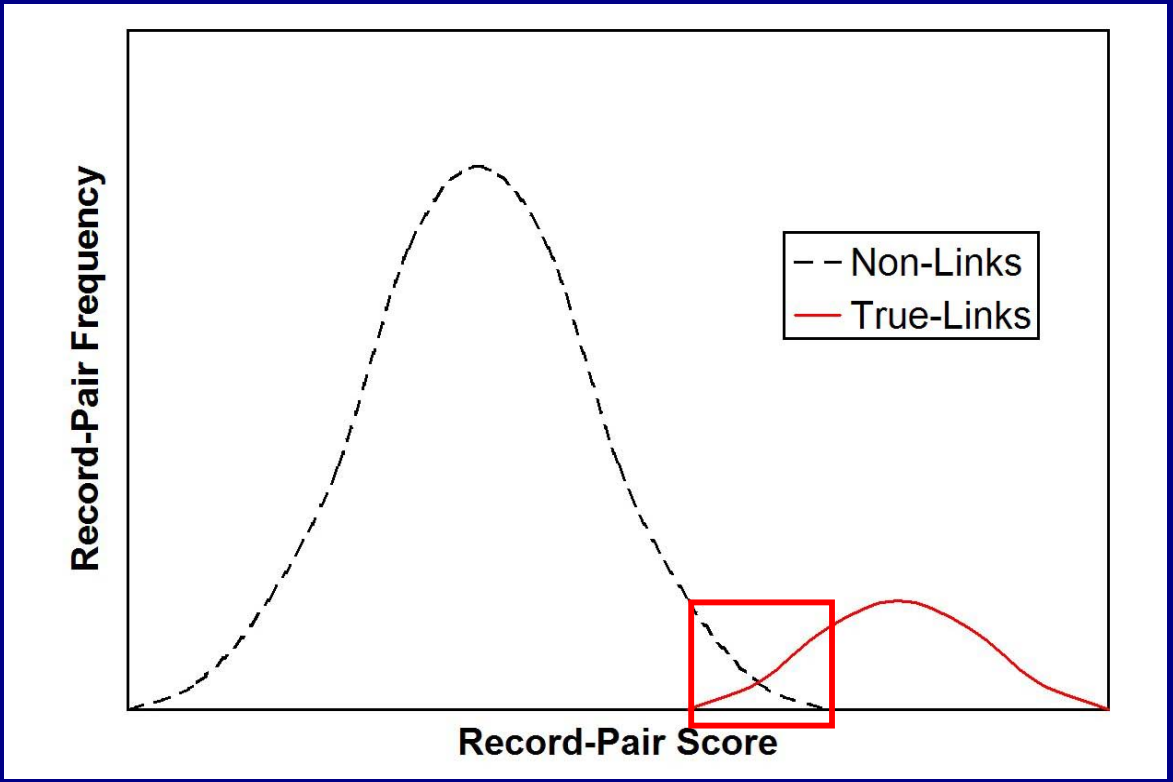
Generate Record Pairs: Each record pair is assigned a score:

File 1

- Record A
- Record B
- Record C

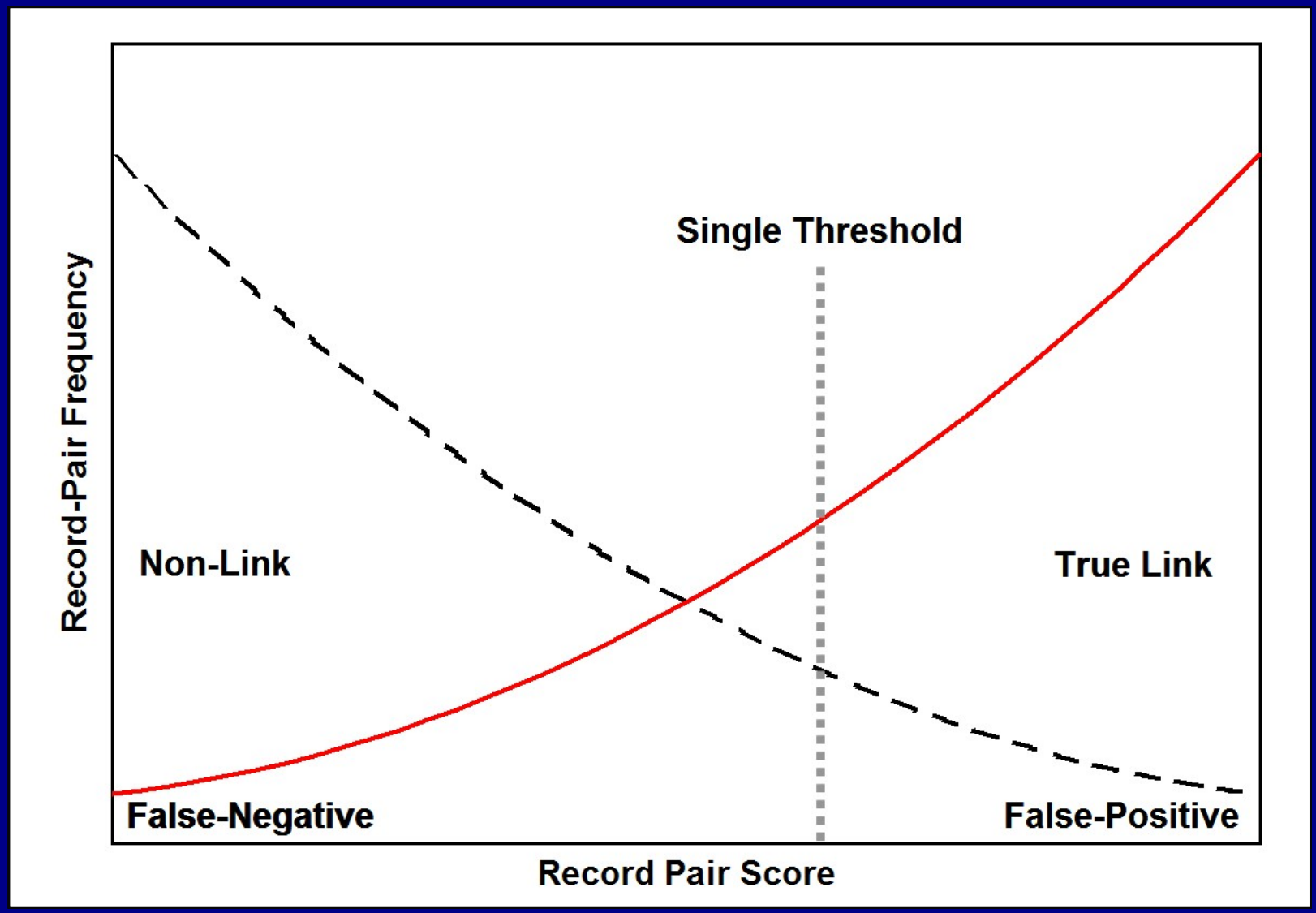
File 2

- Record X
- Record Y
- Record Z



Which are true links?

Probabilistic Linkage Overview: Human Review Thresholds



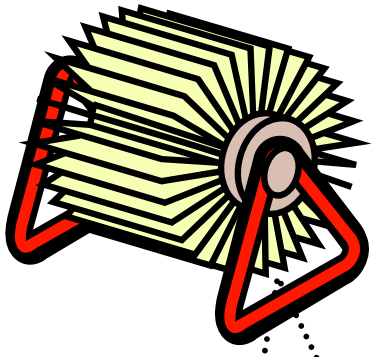
Accuracy and False Links

		Registry A	Registry B
Deterministic	Sensitivity	87.5	90.9
	Specificity	100	100
Probabilistic	Sensitivity	94.6	96.5
	Specificity	99.4	99.4

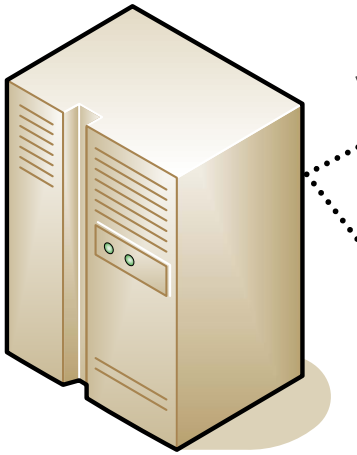
INPC's Linkage Algorithm

Immunization Registry

Patient ID: 123LMNOP
Name: Jane Doe
DOB: 01/01/04
SSN: N/A
Address: 555 Johnson Road
City: Indianapolis
State: Indiana
ZIP: 46202



Global Patient Index

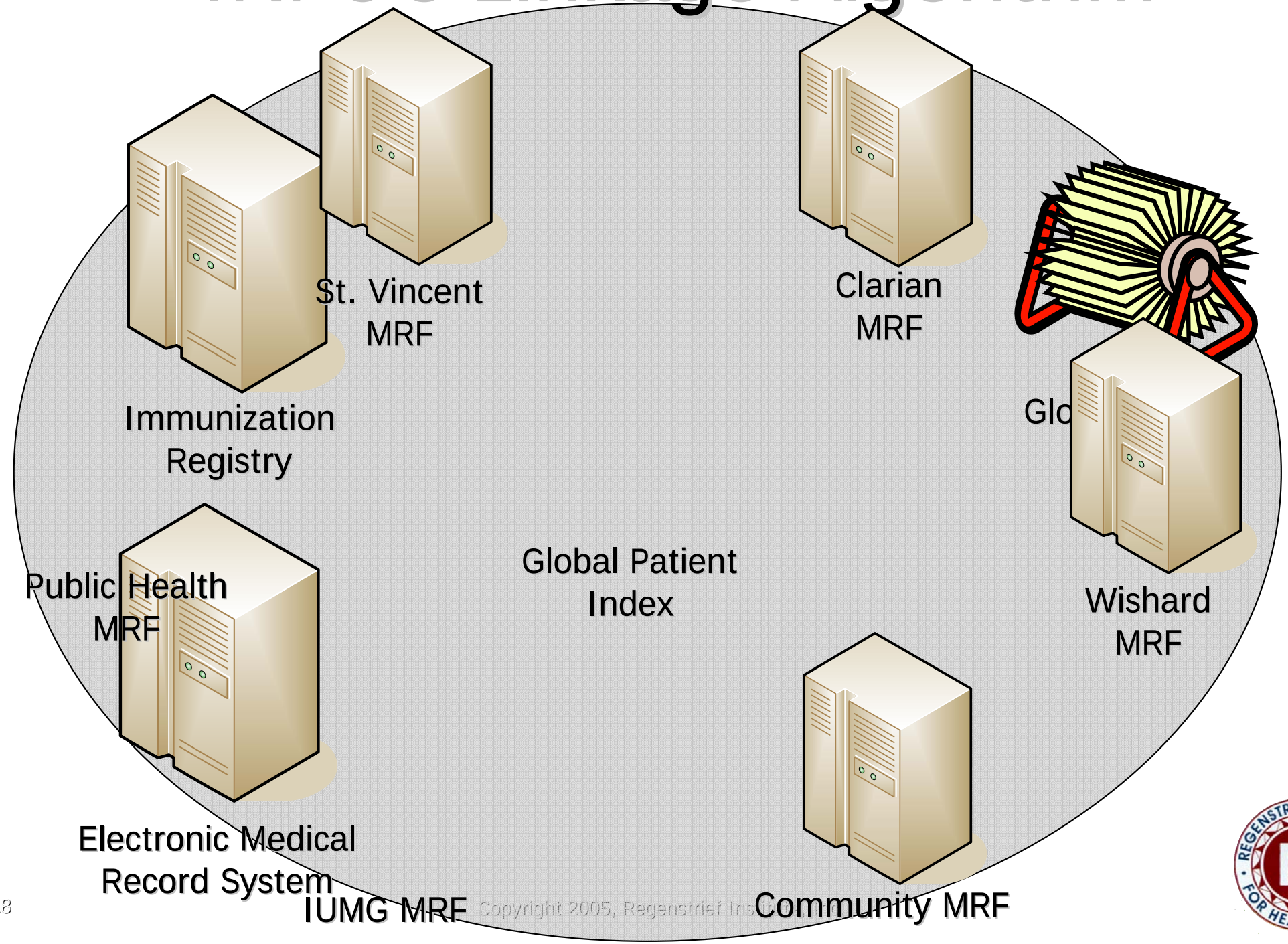


Electronic Medical Record System

Patient ID: 6789XYZ
Name: Jane Ellen Doe
DOB: 01/01/04
SSN:123-45-6789
Address: 555 Johnson Road
City: Indianapolis
State: Indiana
ZIP: 46202

Global ID:	45678
Name:	Jane Ellen Doe
Demographics..	
MRF1 ID:	OU81247
MRF2 ID:	4564356
PH MRF ID:	123LMNOP
MRF3 ID:	6789XYZ

The Indiana Network for Patient Care

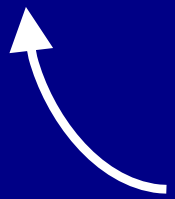


Global Patient Index Matching Algorithm Features

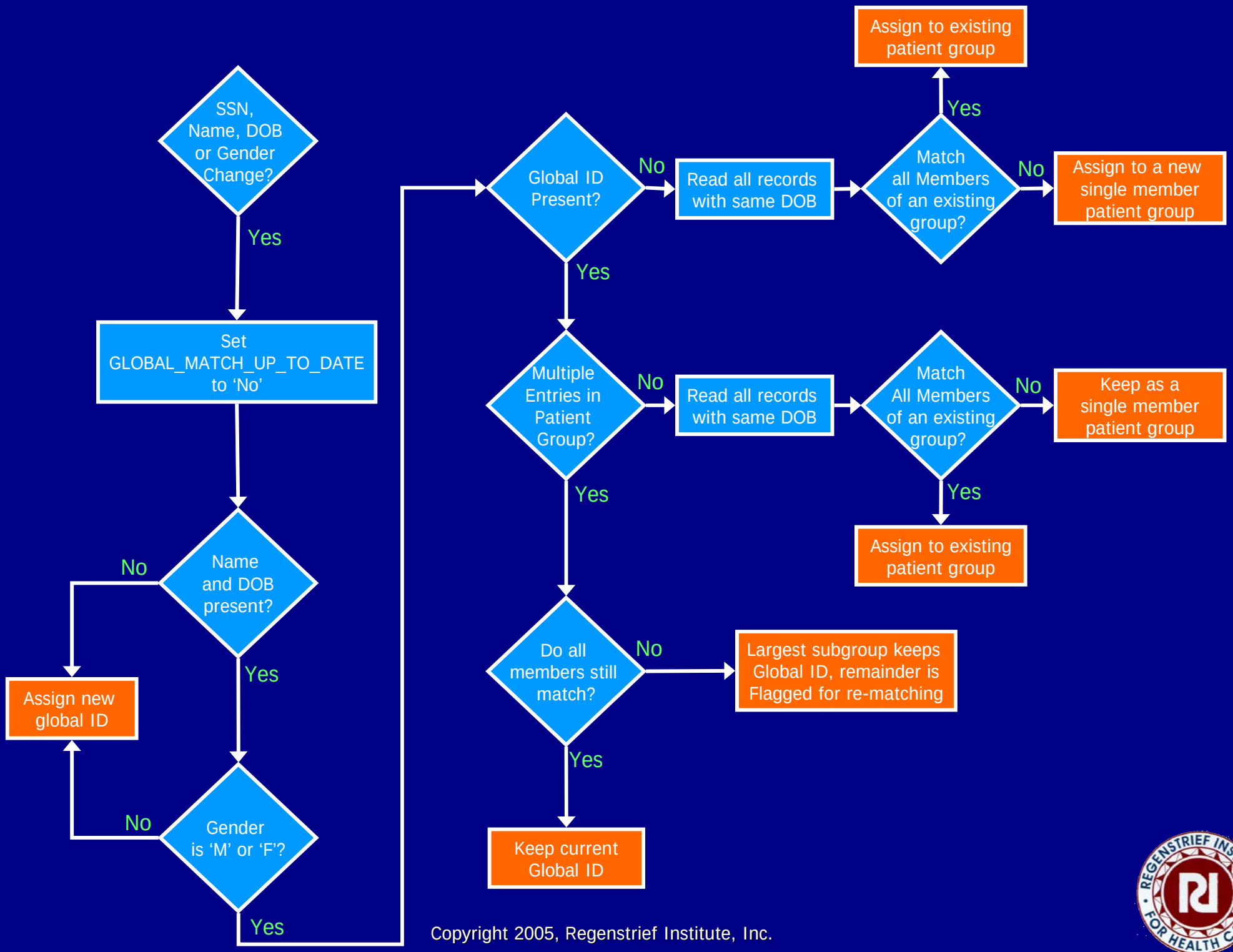
- One record per assigned medical record number per institution
- Implements the concept of a “patient group”: create logical links between each of these records
- Match using social security number, patient name, birth date and gender
- Use string comparator algorithm and phonetic transformations for near name matches
- Allow for minor transpositions in birth date

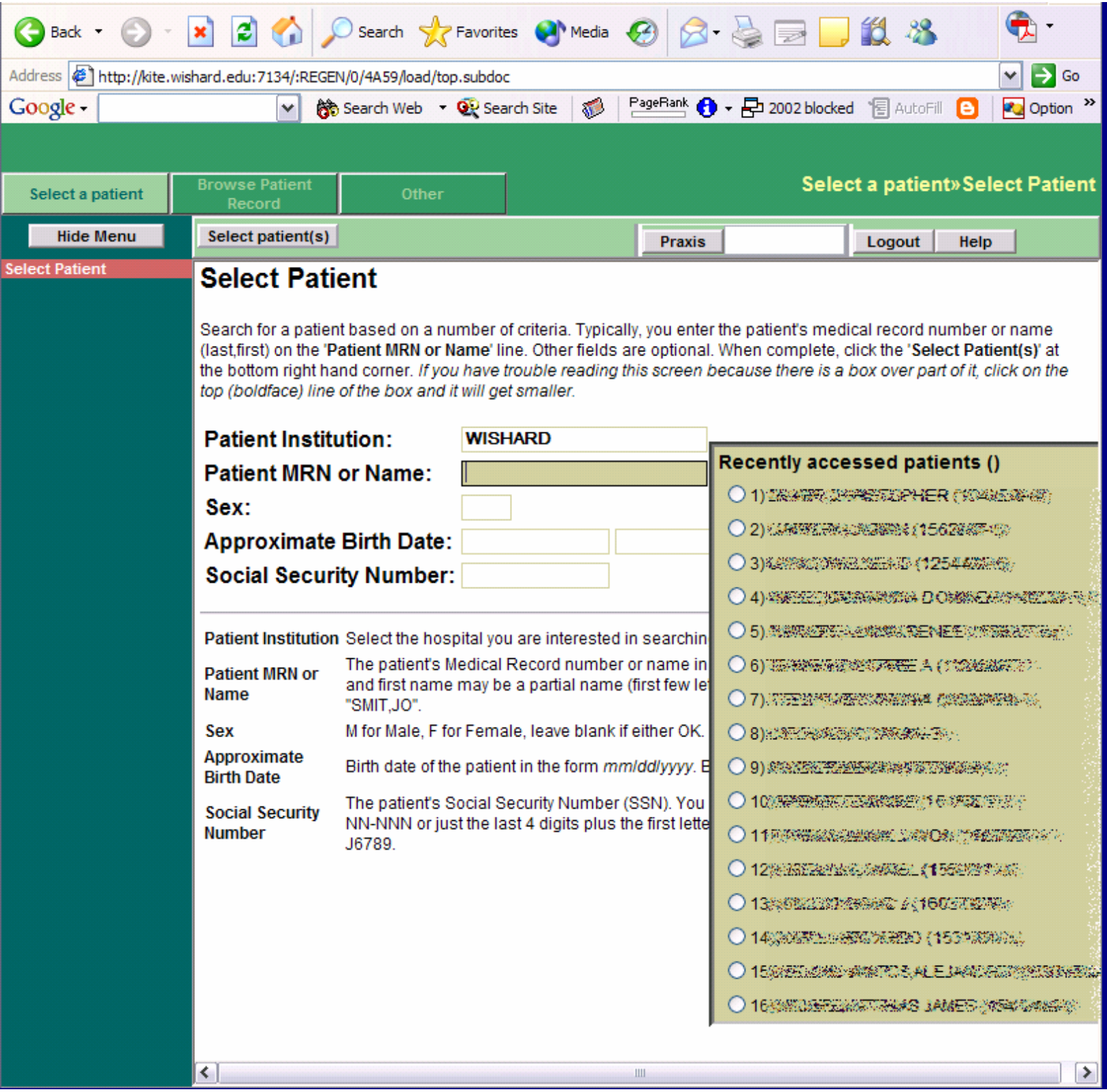
Global Patient Registry

Assigning Authority	Global #	Local Pat #	Patient Name	Birth date	Sex
Hospital A	99-1	231456	Sinkwell, Ralph J	12-2-59	M
Hospital B	123-0	A47239	Sinkwell, RJ	2-12-59	M
Hospital A	99-1	1032115	Sinkweil, Ralph	12-2-59	
Hospital C	101-0	A3276	Fredrick, Alice	4-14-78	F
Hospital A	101-0	2314590	Fredrick, Alyce	4-14-78	F

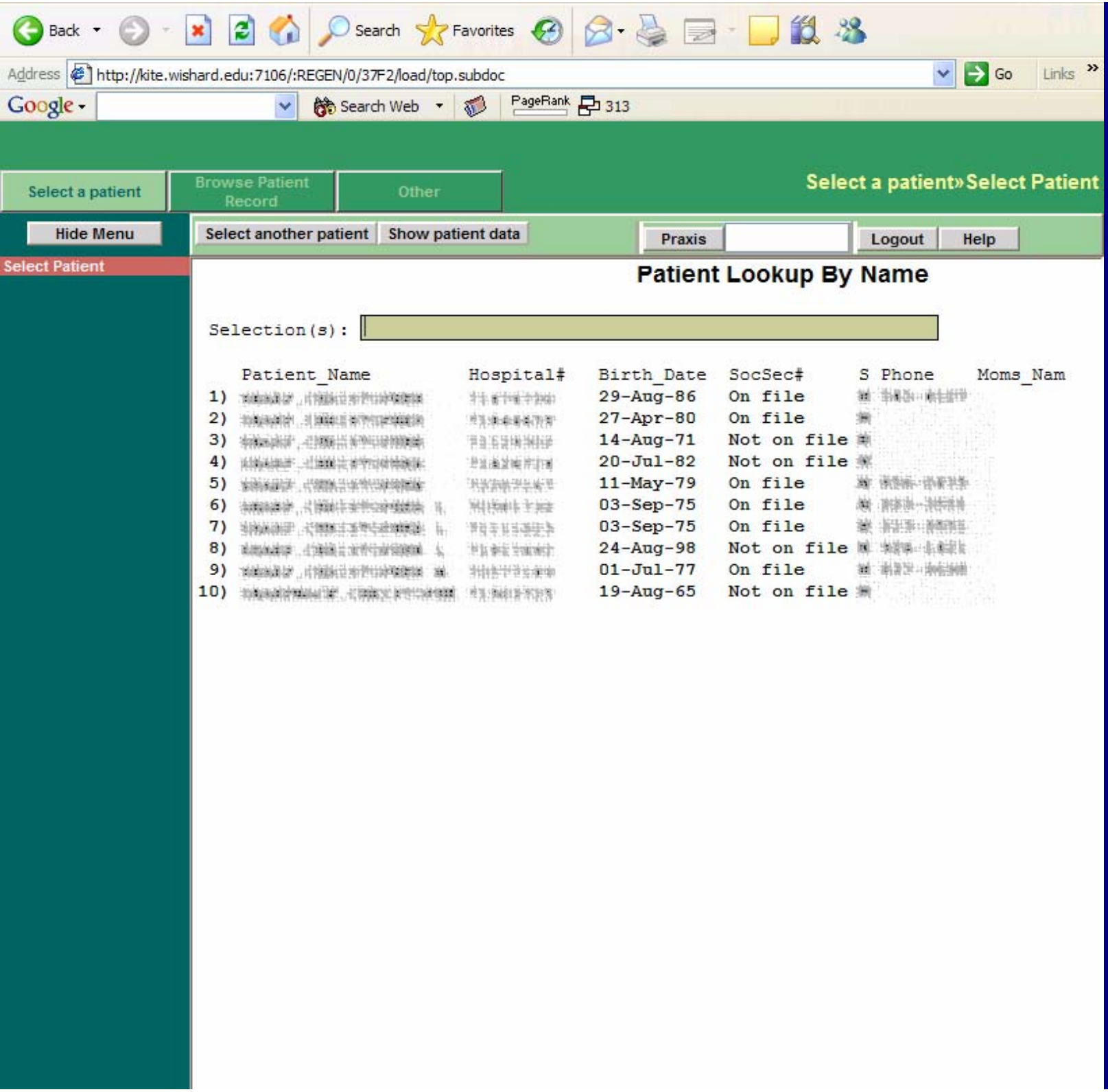


Global ID is for internal indexing only – not publicly exposed



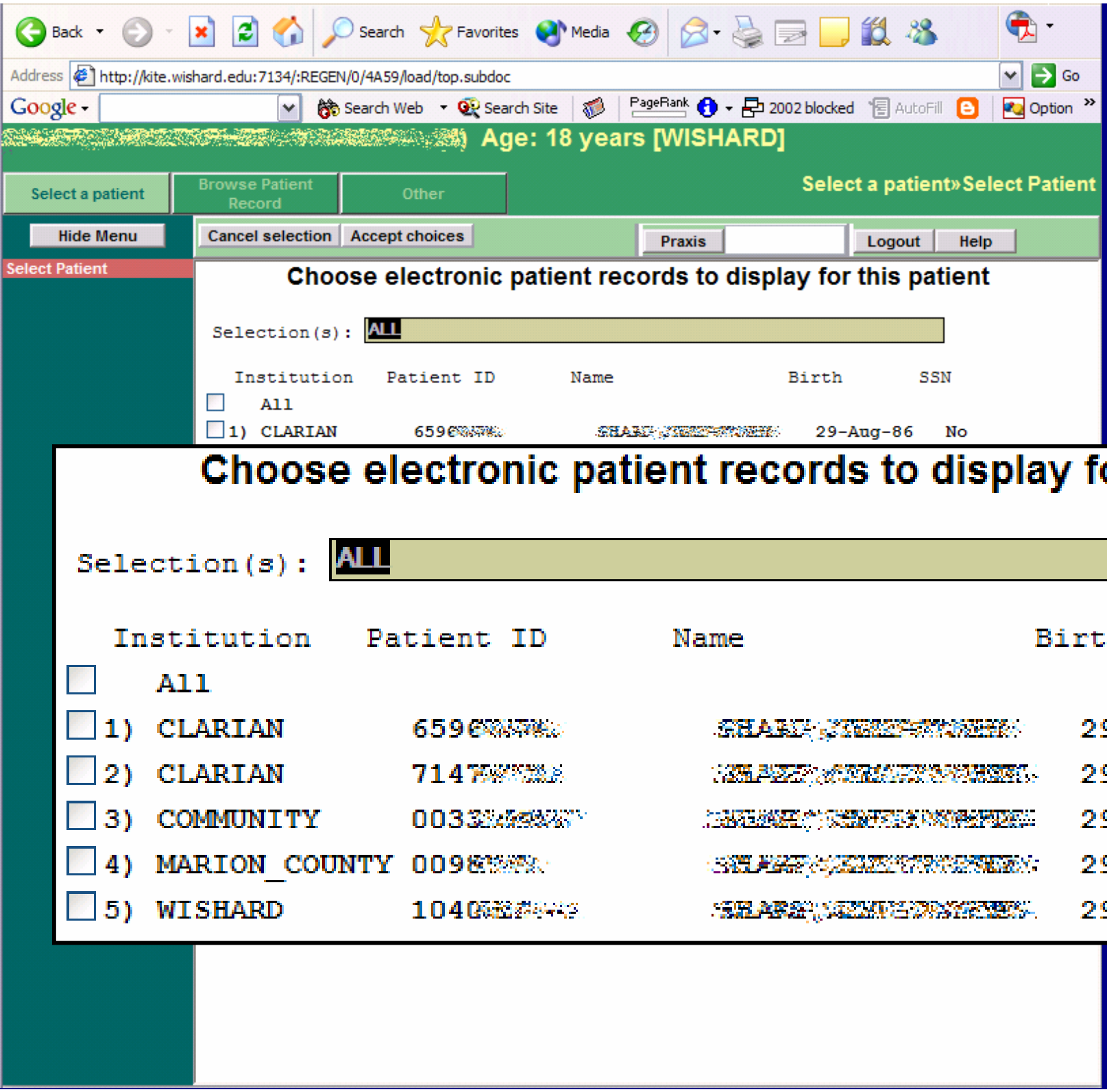


Results Retrieval: Patient Selection Dialog



Results Retrieval: Patient Selection Dialog

Results Retrieval: Global

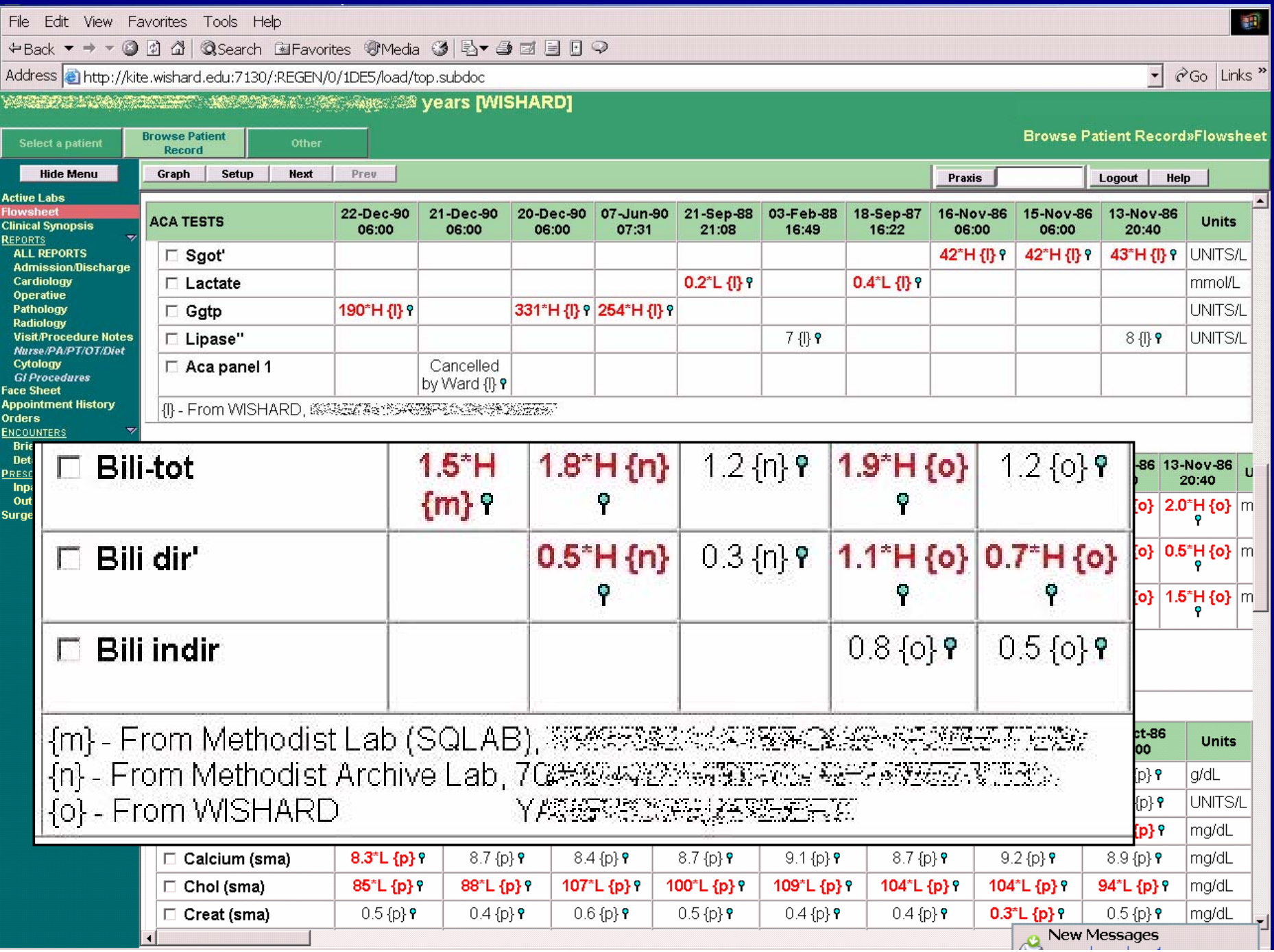


Choose electronic patient records to display for this patient

Selection(s): ALL

	Institution	Patient ID	Name	Birth	SSN
☐	All				
☐	1) CLARIAN	659		29-Aug-86	No
☐	2) CLARIAN	714		29-Aug-86	No
☐	3) COMMUNITY	0033		29-Aug-86	Yes
☐	4) MARION_COUNTY	009		29-Aug-86	No
☐	5) WISHARD	104		29-Aug-86	Yes

Results Retrieval Interface



Conclusions

- National Patient Identifier
  - An opt-in policy balances short-term and long-term expectations
  - NPI can be easily incorporated as another identifier in probabilistic linkage
  - Matching algorithms still necessary to locate patients who do not opt in, missing card, etc.
- Because accurate, consistent patient linkage is key to regional and national health care infrastructure:
  - We need to better understand linkage best-practices and establish standards, e.g. what fields to use under what circumstances
  - We need population gold-standard for testing and development

Patient Linkage

Shaun Grannis, MD, MS
Research Scientist, Regenstrief Institute
Assistant Professor, Family Medicine
Indiana University School of Medicine