

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

- Rapid Implementation
- Simple calculations
- Relies on accurate data
- May not function well with other data sets

Probabilistic/Statistical

- Complex implementation
- Computationally intensive
- More forgiving of data errors
- 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

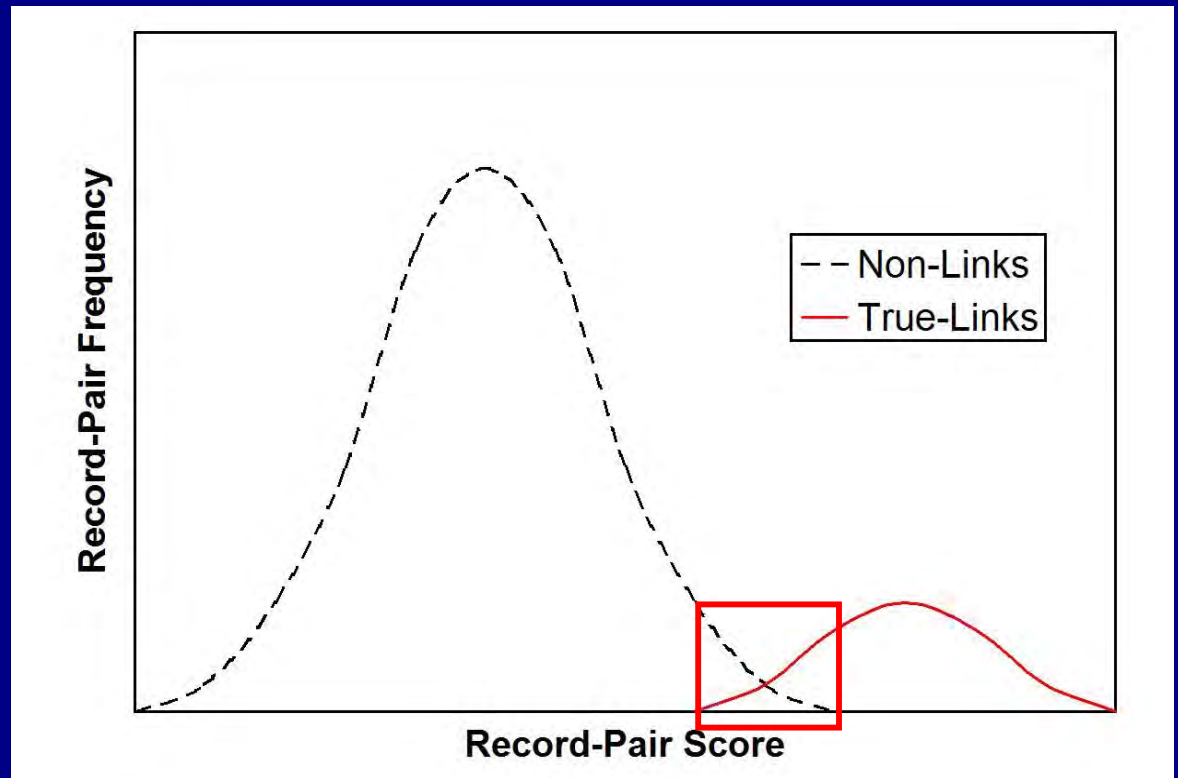
- Compare probabilistic/statistical methods with deterministic/heuristic methods
- Outstanding Questions:
How do probabilistic linkage algorithms compare with heuristic/deterministic methods?
- What are patient linkage best practices?



Probabilistic Linkage, Overview

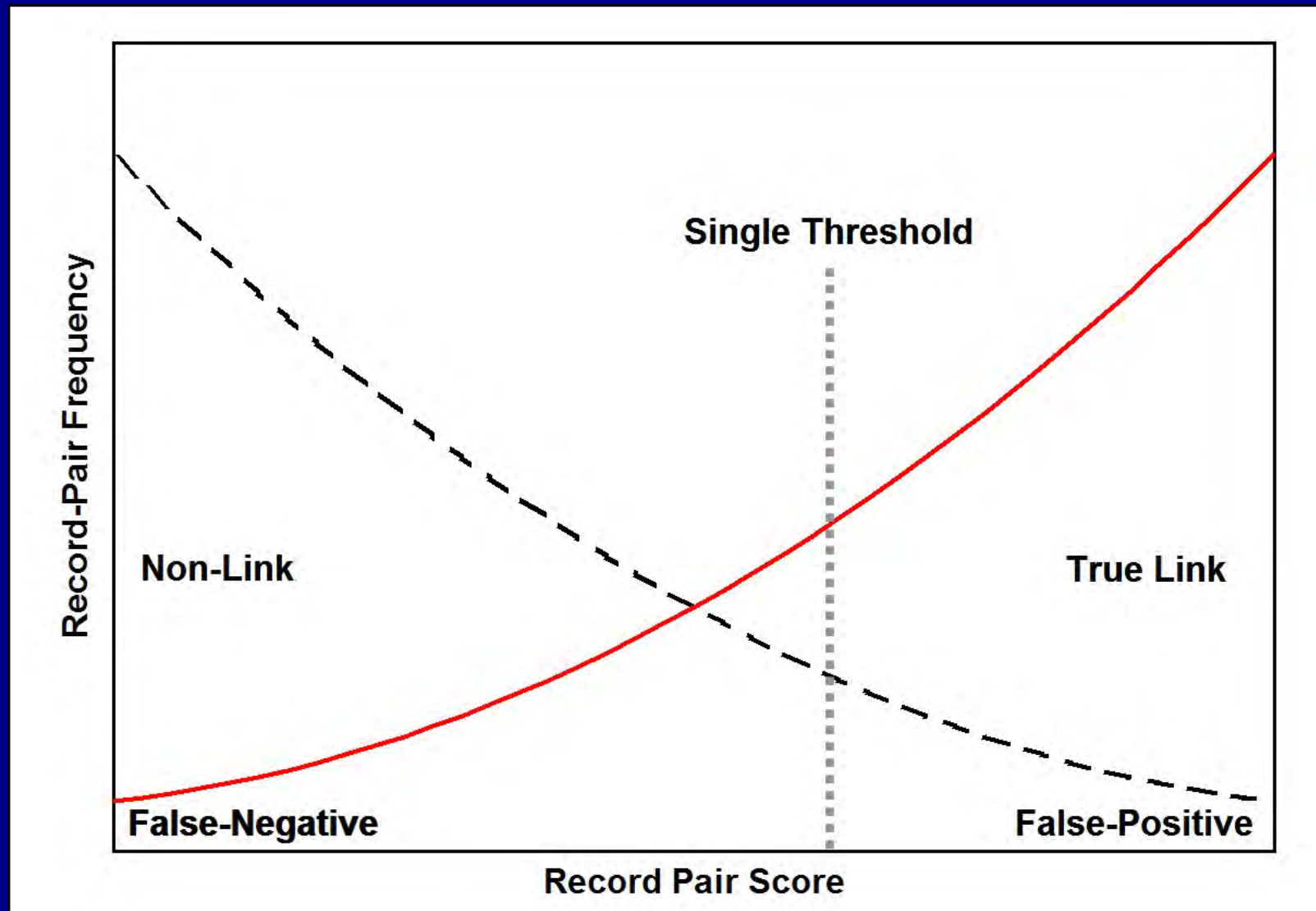
Generate Record Pairs. Each record pair is assigned a score:

File 1	File 2
Record A	Record X
Record B	Record Y
Record C	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



Electronic Medical Record System

Patient ID: 6789XYZ

Name: Jane Ellen Doe

DOB: 01/01/04

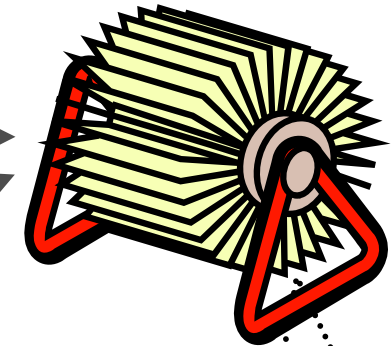
SSN: 123-45-6789

Address: 555 Johnson Road

City: Indianapolis

State: Indiana

ZIP: 46202

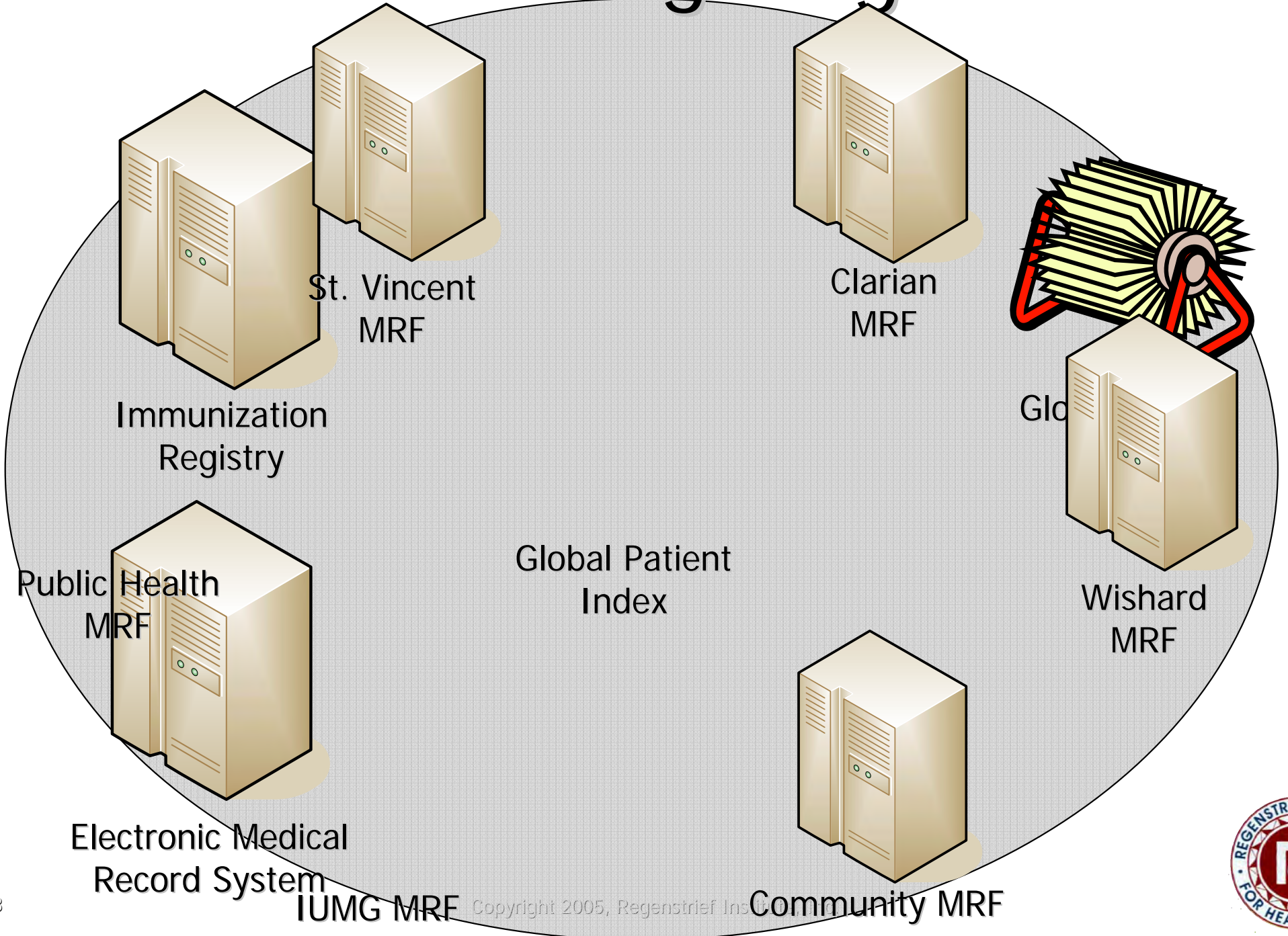


Global Patient Index

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



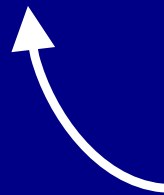
Global Patient Index Matching Algorithm Features

- 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



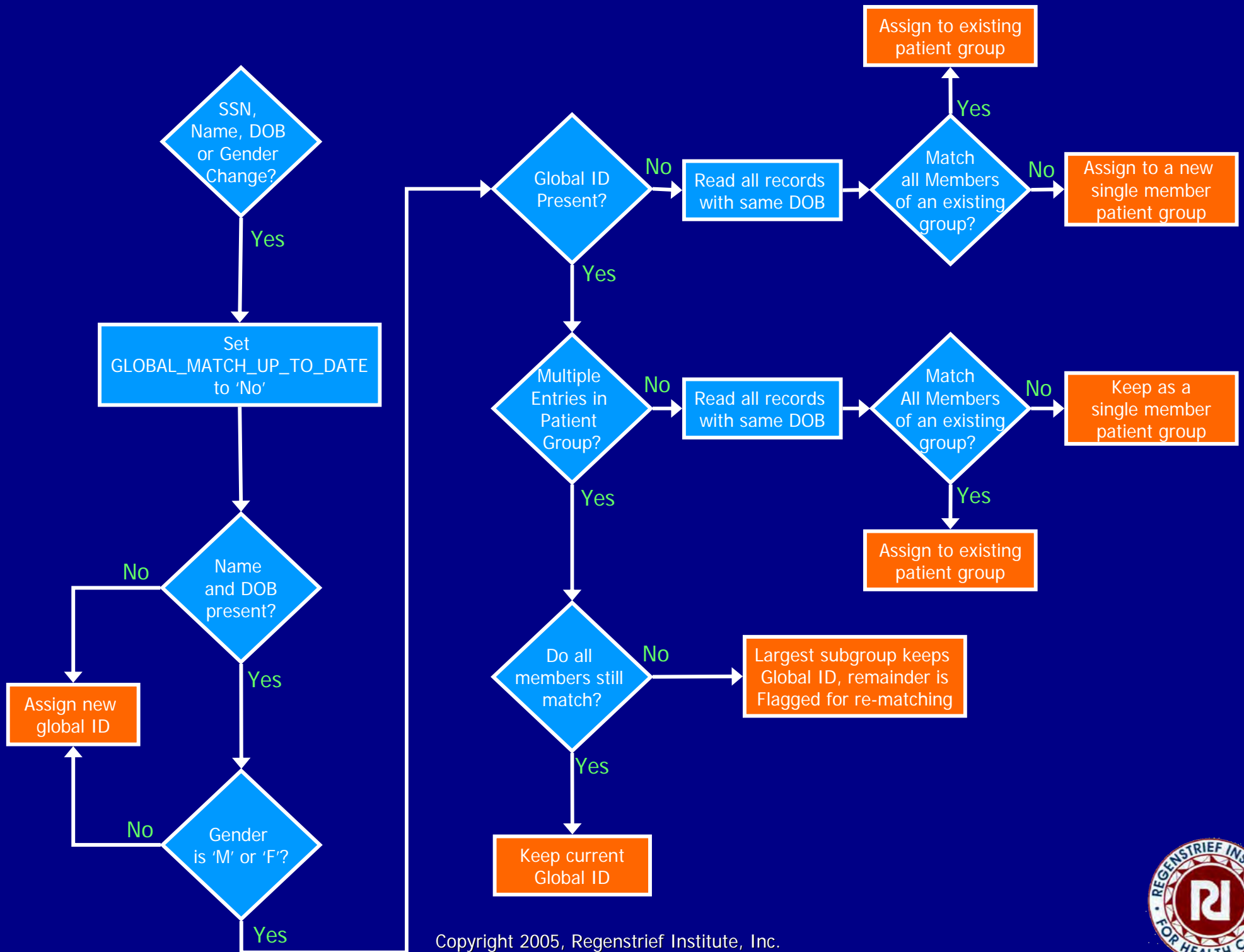
Global Patient Registry

<u>Assigning Authority</u>	<u>Global #</u>	<u>Local Pat #</u>	<u>Patient Name</u>	<u>Birth date</u>	<u>Sex</u>
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





Select Patient

Search for a patient based on a number of criteria. Typically, you enter the patient's medical record number or name (last,first) on the 'Patient MRN or Name' line. Other fields are optional. When complete, click the 'Select Patient(s)' at the bottom right hand corner. *If you have trouble reading this screen because there is a box over part of it, click on the top (boldface) line of the box and it will get smaller.*

Patient Institution:

Patient MRN or Name:

Sex:

Approximate Birth Date:

Social Security Number:

Patient Institution Select the hospital you are interested in searching

Patient MRN or Name The patient's Medical Record number or name in and first name may be a partial name (first few letters) "SMIT,JO".

Sex M for Male, F for Female, leave blank if either OK.

Approximate Birth Date Birth date of the patient in the form mm/dd/yyyy. E

Social Security Number The patient's Social Security Number (SSN). You can enter NN-NNN or just the last 4 digits plus the first letter J6789.

- Recently accessed patients ()**
- 1) ~~SMITH, CHRISTOPHER (1046827)~~
 - 2) ~~SMITH, CHRISTOPHER (1562827)~~
 - 3) ~~SMITH, CHRISTOPHER (1254427)~~
 - 4) ~~SMITH, CHRISTOPHER (1046827)~~
 - 5) ~~SMITH, CHRISTOPHER (1046827)~~
 - 6) ~~SMITH, CHRISTOPHER (1046827)~~
 - 7) ~~SMITH, CHRISTOPHER (1046827)~~
 - 8) ~~SMITH, CHRISTOPHER (1046827)~~
 - 9) ~~SMITH, CHRISTOPHER (1046827)~~
 - 10) ~~SMITH, CHRISTOPHER (1046827)~~
 - 11) ~~SMITH, CHRISTOPHER (1046827)~~
 - 12) ~~SMITH, CHRISTOPHER (1562827)~~
 - 13) ~~SMITH, CHRISTOPHER (1603227)~~
 - 14) ~~SMITH, CHRISTOPHER (1537227)~~
 - 15) ~~SMITH, CHRISTOPHER (1537227)~~
 - 16) ~~SMITH, CHRISTOPHER (1537227)~~

Results Retrieval: Patient Selection Dialog



Patient Lookup By Name

Selection(s):

	Patient_Name	Hospital#	Birth_Date	SocSec#	S Phone	Moms_Nam
1)			29-Aug-86	On file		
2)			27-Apr-80	On file		
3)			14-Aug-71	Not on file		
4)			20-Jul-82	Not on file		
5)			11-May-79	On file		
6)			03-Sep-75	On file		
7)			03-Sep-75	On file		
8)			24-Aug-98	Not on file		
9)			01-Jul-77	On file		
10)			19-Aug-65	Not on file		

Results Retrieval: Patient Selection Dialog



Age: 18 years [WISHARD]

Select a patient Browse Patient Record Other Select a patient» Select Patient

Hide Menu Cancel selection Accept choices Praxis Logout Help

Select Patient

Choose electronic patient records to display for this patient

Selection(s): ALL

Institution	Patient ID	Name	Birth	SSN
<input type="checkbox"/> All				
<input type="checkbox"/> 1) CLARIAN	6596	SHAWN STEVENSON	29-Aug-86	No

Results Retrieval: Global

Choose electronic patient records to display for this patient

Selection(s): ALL

Institution	Patient ID	Name	Birth	SSN
<input type="checkbox"/> All				
<input type="checkbox"/> 1) CLARIAN	6596	SHAWN STEVENSON	29-Aug-86	No
<input type="checkbox"/> 2) CLARIAN	7147	COLLEEN CHRISTOPHERSON	29-Aug-86	No
<input type="checkbox"/> 3) COMMUNITY	0033	KEVIN JAMES WILSON	29-Aug-86	Yes
<input type="checkbox"/> 4) MARION_COUNTY	0096	SHAWN STEVENSON	29-Aug-86	No
<input type="checkbox"/> 5) WISHARD	1040	SHAWN STEVENSON	29-Aug-86	Yes



Results Retrieval Interface

File Edit View Favorites Tools Help

Address http://kite.wishard.edu:7130/REGEN/0/IDE5/load/top.subdoc

years [WISHARD]

Select a patient Browse Patient Record Other

Graph Setup Next Prev Praxis Logout Help

Active Labs
 Flowsheet
 Clinical Synopsis
 REPORTS
 ALL REPORTS
 Admission/Discharge
 Cardiology
 Operative
 Pathology
 Radiology
 Visit/Procedure Notes
 Nurse/PA/PT/OT/Diet
 Cytology
 GI Procedures
 Face Sheet
 Appointment History
 Orders
 ENCOUNTERS

ACA TESTS	22-Dec-90 06:00	21-Dec-90 06:00	20-Dec-90 06:00	07-Jun-90 07:31	21-Sep-88 21:08	03-Feb-88 16:49	18-Sep-87 16:22	16-Nov-86 06:00	15-Nov-86 06:00	13-Nov-86 20:40	Units
<input type="checkbox"/> Sgot'								42*H {f} †	42*H {f} †	43*H {f} †	UNITS/L
<input type="checkbox"/> Lactate					0.2*L {f} †		0.4*L {f} †				mmol/L
<input type="checkbox"/> Ggtp	190*H {f} †		331*H {f} †	254*H {f} †							UNITS/L
<input type="checkbox"/> Lipase''						7 {f} †				8 {f} †	UNITS/L
<input type="checkbox"/> Aca panel 1		Cancelled by Ward {f} †									

{f} - From WISHARD, [REDACTED]

<input type="checkbox"/> Bili-tot	1.5*H {m} †	1.8*H {n} †	1.2 {n} †	1.9*H {o} †	1.2 {o} †						
<input type="checkbox"/> Bili dir'		0.5*H {n} †	0.3 {n} †	1.1*H {o} †	0.7*H {o} †						
<input type="checkbox"/> Bili indir					0.8 {o} †		0.5 {o} †				

{m} - From Methodist Lab (SQLAB), [REDACTED]
 {n} - From Methodist Archive Lab, 70 [REDACTED]
 {o} - From WISHARD

<input type="checkbox"/> Calcium (sma)	8.3*L {p} †	8.7 {p} †	8.4 {p} †	8.7 {p} †	9.1 {p} †	8.7 {p} †	9.2 {p} †	8.9 {p} †			mg/dL
<input type="checkbox"/> Chol (sma)	85*L {p} †	88*L {p} †	107*L {p} †	100*L {p} †	109*L {p} †	104*L {p} †	104*L {p} †	94*L {p} †			mg/dL
<input type="checkbox"/> Creat (sma)	0.5 {p} †	0.4 {p} †	0.6 {p} †	0.5 {p} †	0.4 {p} †	0.4 {p} †	0.3*L {p} †	0.5 {p} †			mg/dL

New Messages



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

