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

Copyright 2005, Regenstrief Institute, Inc.

Challenges

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

Copyright 2005, Regenstrief Institute, Inc.

Potential Solutions

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

Copyright 2005, Regenstrief Institute, Inc.

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

Copyright 2005, Regenstrief Institute, Inc.

Previous Work

- Deterministic

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

- Results:
  - 90% Sensitive
  - 100% Specific

Copyright 2005, Regenstrief Institute, Inc.

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%

Copyright 2005, Regenstrief Institute, Inc.

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**

Copyright 2005, Regenstrief Institute, Inc.

Hospital A linked to SSDMF: Name, Date of Birth, and Gender <u>without</u> SSN

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

Copyright 2005, Regenstrief Institute, Inc.

Hospital B linked to SSDMF: Name, Date of Birth, and Gender <u>without</u> SSN

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

Copyright 2005, Regenstrief Institute, Inc.

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)

Copyright 2005, Regenstrief Institute, Inc.

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%

Copyright 2005, Regenstrief Institute, Inc.

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?

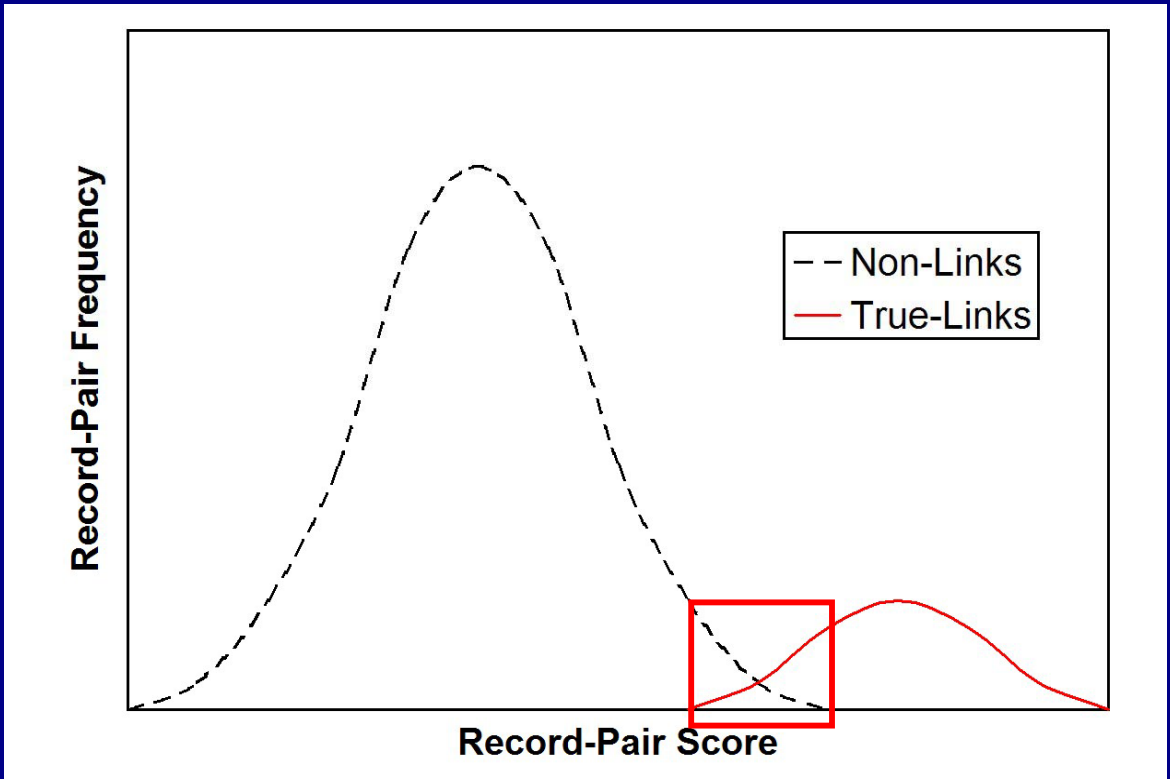
Copyright 2005, Regenstrief Institute, Inc.

Probabilistic Linkage, Overview

Generate Record Pairs:

Each record pair is assigned a score:

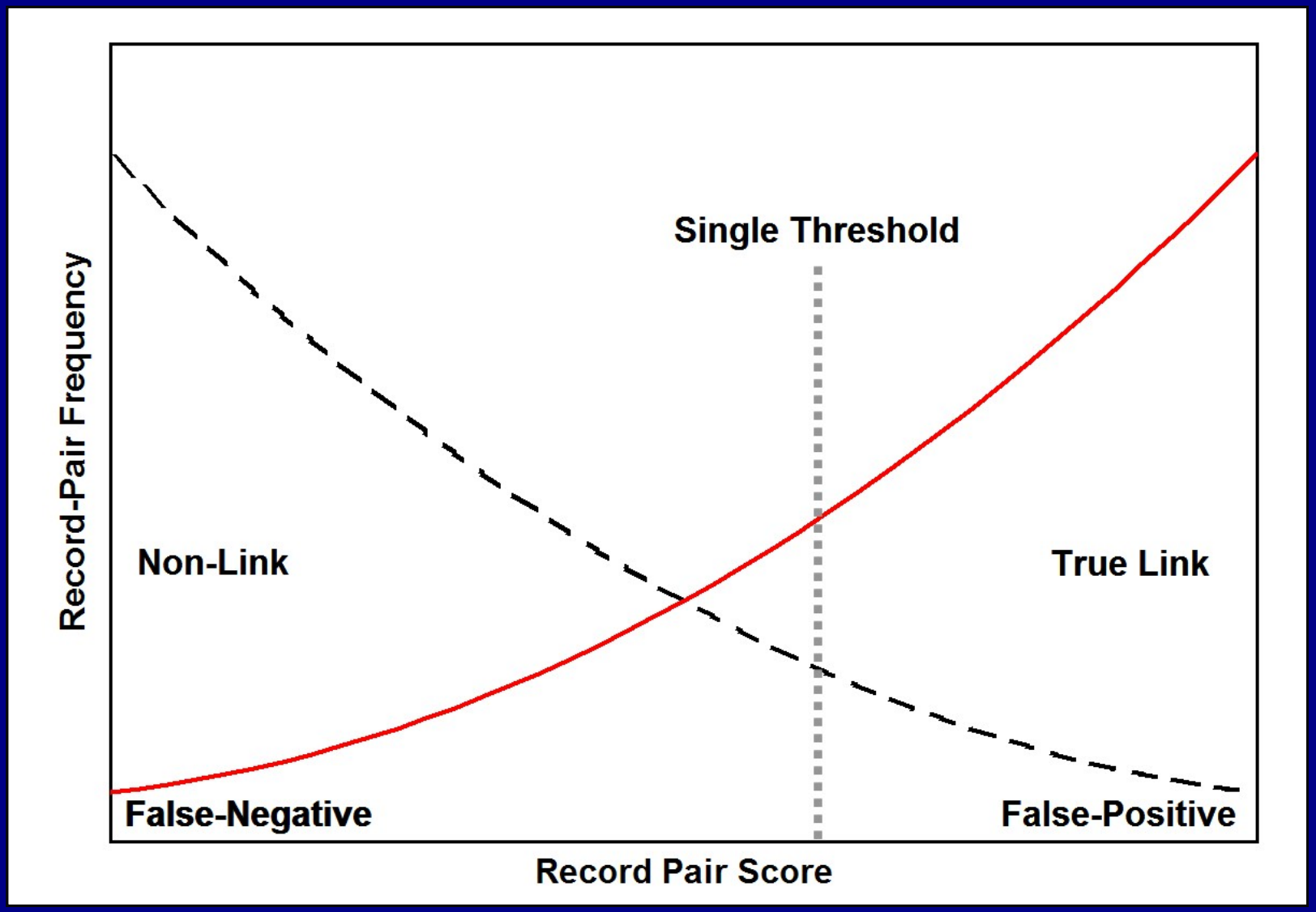
File 1	File 2
Record A	Record X
Record B	Record Y
Record C	Record Z



Which are true links?

Copyright 2005, Regenstrief Institute, Inc.

Probabilistic Linkage Overview: Human Review Thresholds



Copyright 2005, Regenstrief Institute, Inc.

Accuracy and False Links

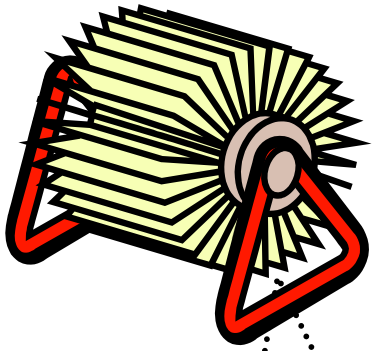
		Registry A	Registry B
Deterministic	Sensitivity	87.5	90.9
	Specificity	100	100
Probabilistic	Sensitivity	94.6	96.5
	Specificity	99.4	99.4

Copyright 2005, Regenstrief Institute, Inc.

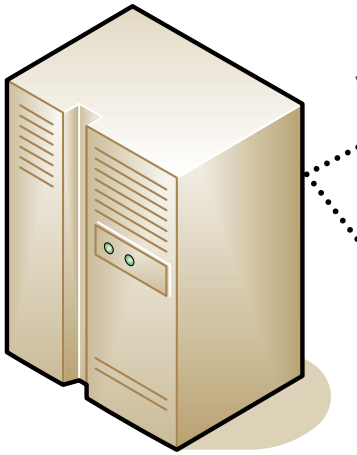
INPC's Linkage Algorithm

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



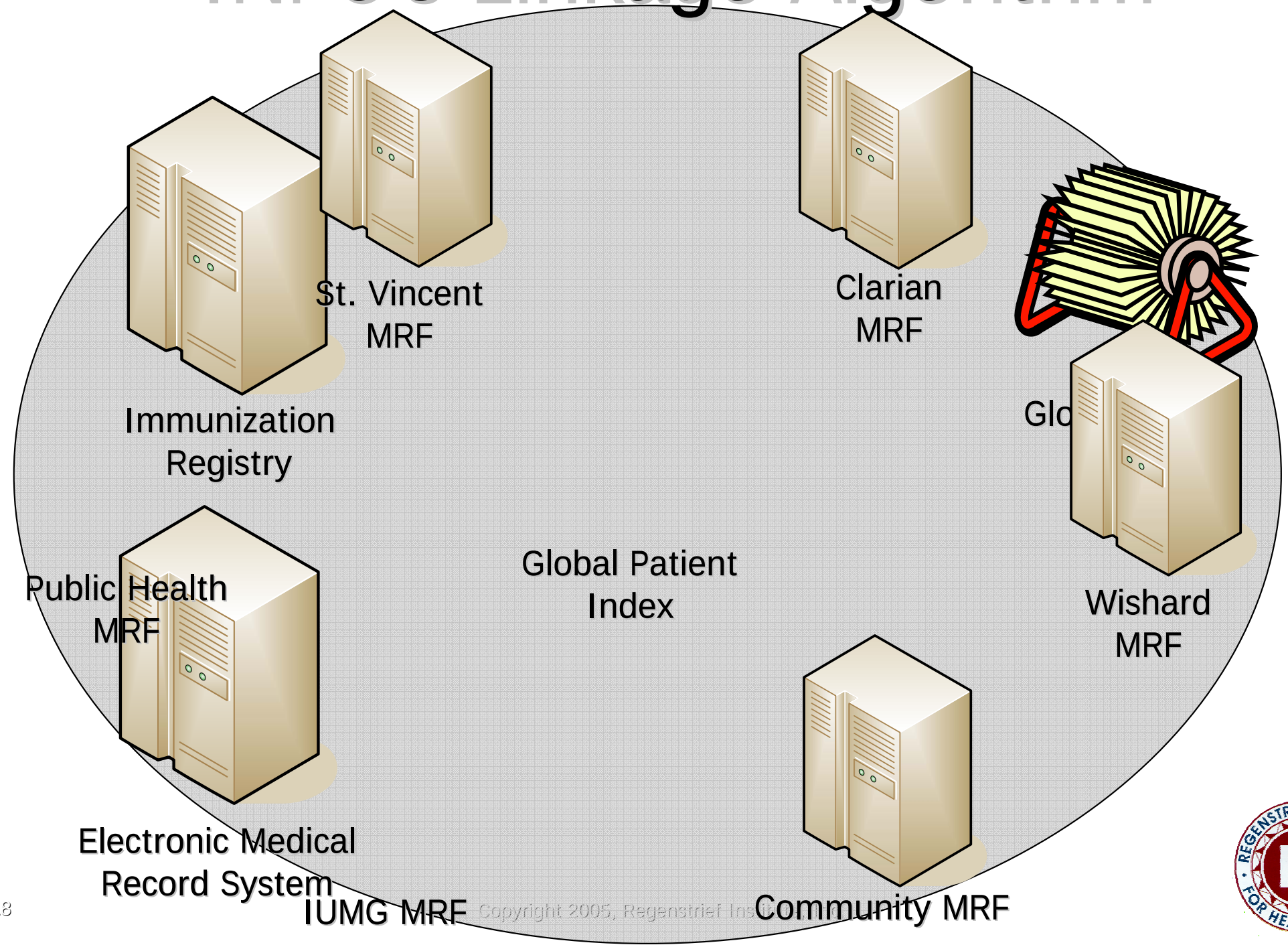
Electronic Medical Record System

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

Copyright 2005, Regenstrief Institute, Inc.

The Indiana Network for Patient Care



Copyright 2005, Regenstrief Institute

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

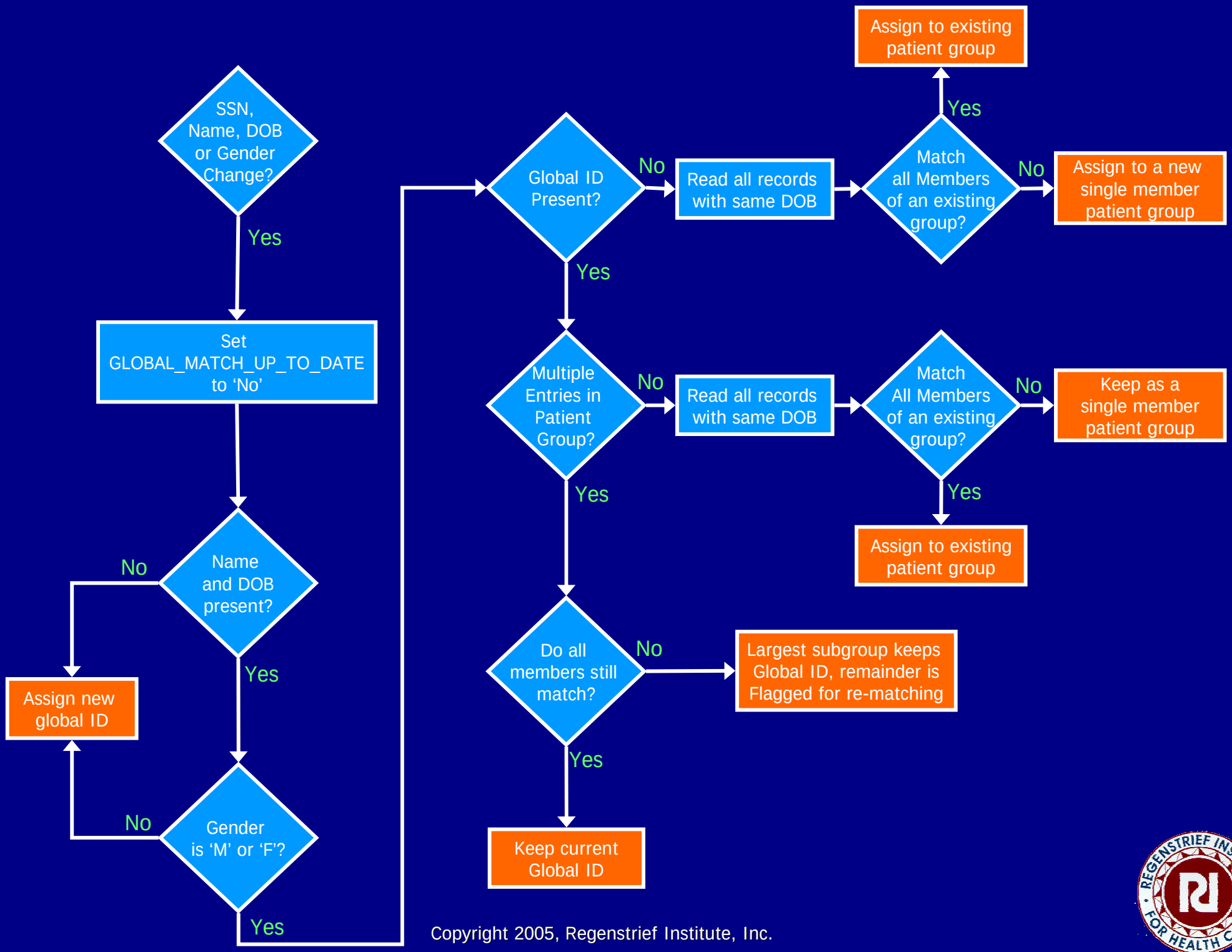
Copyright 2005, Regenstrief Institute, Inc.

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

Copyright 2005, Regenstrief Institute, Inc.



Copyright 2005, Regenstrief Institute, Inc.

Results Retrieval: Patient Selection Dialog

Back Search Favorites Media

Address http://kite.wishard.edu:7134/:REGEN/0/4A59/load/top.subdoc Go

Google Search Web Search Site PageRank 2002 blocked AutoFill Option

Select a patient | Browse Patient Record | Other

Select a patient»Select Patient

Hide Menu | Select patient(s) | Praxis | Logout | Help

Select Patient

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: WISHARD

Patient MRN or Name:

Sex:

Approximate Birth Date:

Social Security Number:

Patient Institution	Select the hospital you are interested in searchin
Patient MRN or Name	The patient's Medical Record number or name in and first name may be a partial name (first few le "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 NN-NNN or just the last 4 digits plus the first lette J6789.

Recently accessed patients ()

1) [illegible]
2) [illegible]
3) [illegible]
4) [illegible]
5) [illegible]
6) [illegible]
7) [illegible]
8) [illegible]
9) [illegible]
10) [illegible]
11) [illegible]
12) [illegible]
13) [illegible]
14) [illegible]
15) [illegible]
16) [illegible]

REGENSTRIEF INSTITUTE FOR HEALTH CARE

Results Retrieval: Patient Selection Dialog

Back Search Favorites

Address http://kite.wishard.edu:7106/:REGEN/0/37F2/load/top.subdoc Go Links

Google Search Web PageRank 313

Select a patient | Browse Patient Record | Other

Select a patient»Select Patient

Hide Menu | Select another patient | Show patient data | Praxis | Logout | Help

Select Patient

Patient Lookup By Name

Selection(s):

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

REGENSTRIEF INSTITUTE FOR HEALTH CARE

Results Retrieval: Global

Back Search Favorites Media

Address http://kite.wishard.edu:7134/:REGEN/0/4A59/load/top.subdoc Go

Google Search Web Search Site PageRank 2002 blocked AutoFill Option

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
All				
1) CLARIAN	659[illegible]	[illegible]	29-Aug-86	No
2) CLARIAN	714[illegible]	[illegible]	29-Aug-86	No
3) COMMUNITY	0033[illegible]	[illegible]	29-Aug-86	Yes
4) MARION_COUNTY	009[illegible]	[illegible]	29-Aug-86	No
5) WISHARD	104[illegible]	[illegible]	29-Aug-86	Yes

Choose electronic patient records to display for this patient

Selection(s): ALL

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

REGENSTRIEF INSTITUTE FOR HEALTH CARE

Results Retrieval Interface

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://kite.wishard.edu:7130/:REGEN/0/1DE5/load/top.subdoc Go Links

years [WISHARD]

Select a patient | Browse Patient Record | Other

Browse Patient Record»Flowsheet

Hide Menu | 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
Sgot'								42*H {l}	42*H {l}	43*H {l}	UNITS/L
Lactate					0.2*L {l}		0.4*L {l}				mmol/L
Ggtp	190*H {l}		331*H {l}	254*H {l}							UNITS/L
Lipase''						7 {l}				8 {l}	UNITS/L
Aca panel 1		Cancelled by Ward {l}									

{l} - From WISHARD, [redacted]

Bili-tot	1.5*H {m}	1.8*H {n}	1.2 {n}	1.9*H {o}	1.2 {o}						
Bili dir'		0.5*H {n}	0.3 {n}	1.1*H {o}	0.7*H {o}						
Bili indir				0.8 {o}	0.5 {o}						

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

									Units
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
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
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

Copyright 2005, Regenstrief Institute, Inc.

Patient Linkage

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