Record linkage – a method to generate an enriched dataset for the study of CPOE systems on test turnaround time (TAT)

Mary K Lam, Andrew Georgiou, Johanna I Westbrook

Health Informatics Research and Evaluation Unit,

Faculty of Health Sciences,

The University of Sydney

Outline

- What is record linkage
- Why link records to study the effect of CPOE on TAT
- The process of record linkage
- The results of record linkage
- Future work

What is record linkage

- Also known as data linkage

- The process of linking and aggregating records from two or more data sources

- Traditional techniques
  - Deterministic linkage
  - Probabilistic linkage

Why link records to study effect of CPOE

- CPOE systems
  - Enhance effectiveness and efficiency
  - Improve quality of care delivery

- Evidence based
  - Evaluation and monitoring

- Challenge
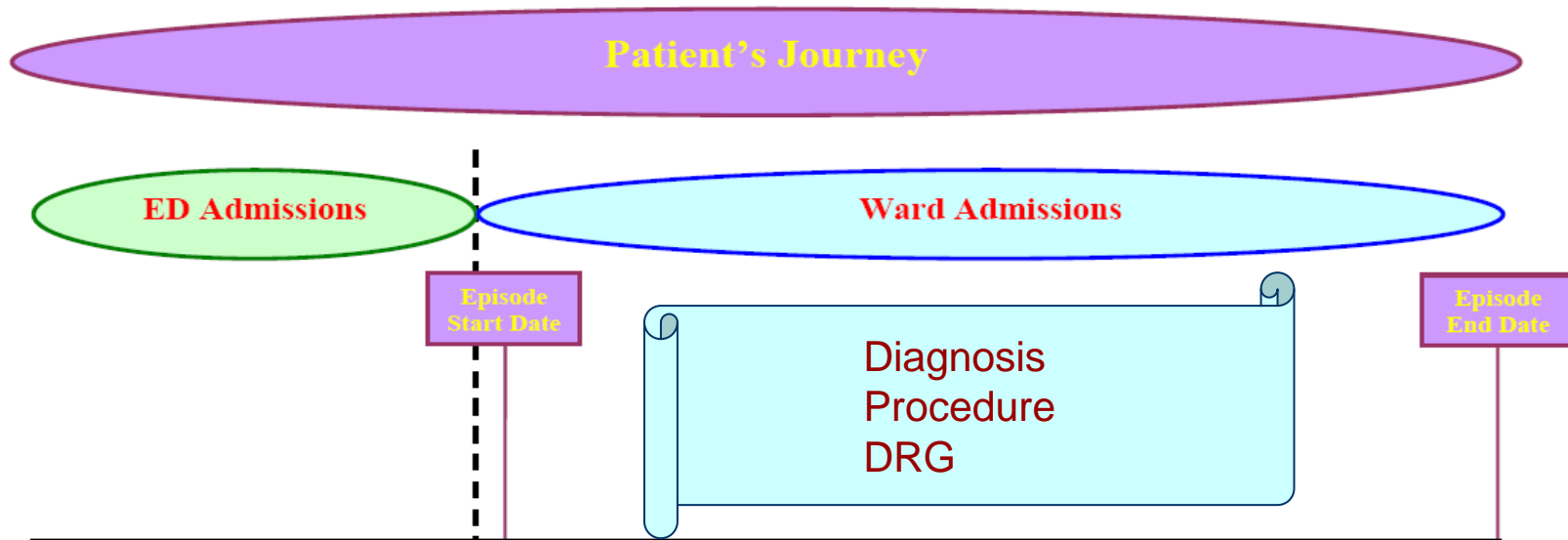 - Accessing the right information

Patient's Journey

ED Admissions

Ward Admissions

Episode Start Date

Episode End Date

Diagnosis
Procedure
DRG

Patient's Journey

ED Admissions

Ward Admissions

Episode
Start Date

Episode
End Date

Arrival
Date/Time

Triage
ED_LOS

Discharge
Date/Time

Diagnosis
Procedure
DRG

Patient's Journey

ED Admissions

Ward Admissions

Episode
Start Date

Episode
Start Date

Episode
End Date

Triage
ED_LOS

Diagnosis
Procedure
DRG

Arrival
Date/Time

Discharge
Date/Time

TAT
Test ordered
Urgency
Results

Admission
Date

Discharge
Date

Pathology Tests Ordered

Research Questions

- What is the effect of CPOE on TAT?

- What is the relationship between ED Triage category and Test TAT?

- What is the relationship between variables such as Triage category, TAT on ED LOS?

Patient's Journey

ED Admissions

Ward Admissions

Episode Start Date

Episode End Date

Triage
ED_LOS

Discharge Date/Time

Diagnosis
Procedure
DRG

Arrival Date/Time

Admission Date

Discharge Date

TAT
Test ordered
Urgency
Results

Pathology Tests Ordered

Methods

Dataset

- Data Extraction
  - August and September 2005
  - August and September 2006
- from
  - ED Information System
  - HosLab Information System

Datasets

ED Dataset	Usage
MRN	L
Arrival date time	L
DOB	V
Sex	V, IV
Triage Category	IV
Length of Stay in ED	DV
ED discharge date time	Cal
Plus 6 more variables	

HosLab Dataset	Usage
MRN	L
Admission Date	L
DOB	V
Sex	V, IV
Total TAT - Collection to completed	DV
Lab TAT - In Lab to completed	DV
Collection Year	IV
Test	IV
Urgency	IV
Abnormal Flag	IV
Office Hours	IV
Sample collection date time	Cal
Sample in lab date time	Cal
Sample complete date time	Cal
Plus 8 more variables	

L-Linkage Variable; V-Validation Variable; IV-Independent Variable; DV-Dependent Variable; Cal-Calculation

Data Linkage

- Records were linked using the deterministic linkage method
  - Linkage variables
    - MRN
    - Admission dates
  - Validation variables
    - DOB
    - Sex

Data Profiling

- Algorithm written in SPSS to check
  - Unique identifier – present
  - Missing values
  - Categorical variables - domain values correct
  - Date fields
    - correct format (dd/mm/yyyy)
    - compliant with logic rules
  - TAT variables – descriptive statistics
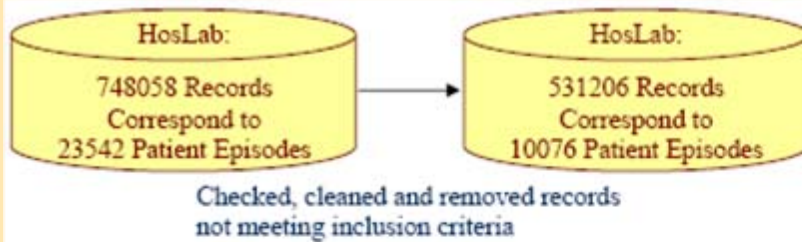 - Possible record duplications

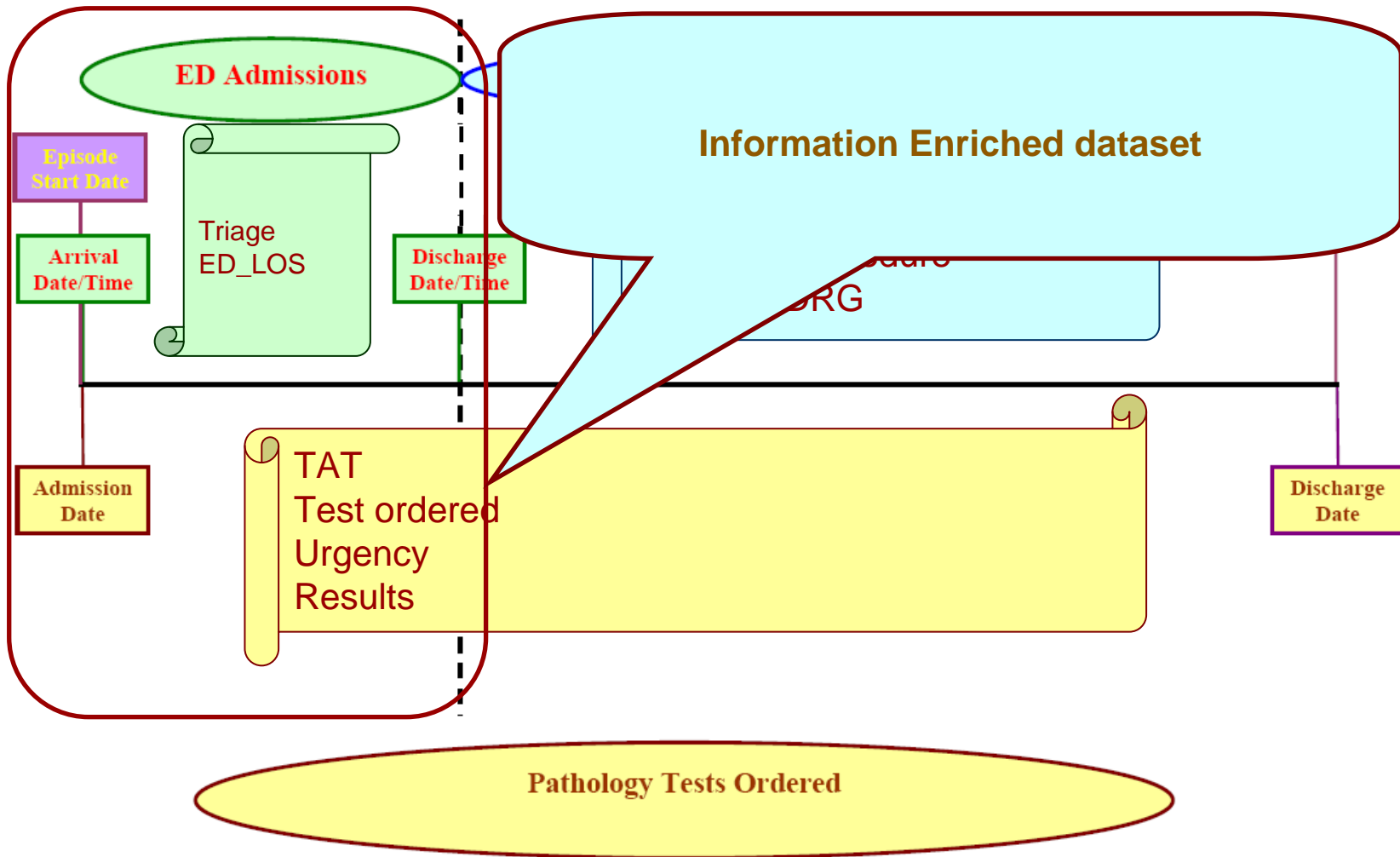| Data Quality	ED	HosLab	Actions (if required)
No. of variables	• 18	• 23	
Linkage variables and format validation	• MRN • Numeric	• MRN • numeric with 212 alphanumeric)	• Records in alphanumeric format removed
	• Arrival date time • dd/mm/yyyy hh:mm:ss	• Admission date • dd/mm/yyyy	• Divided date and time portion of Arrival date time into 2 variables: Admission date (dd/mm/yyyy) and Arrival time (hh:mm:ss)
Validation variables & format validation	• Sex (text)	• Sex (text)	
	• DOB (dd/mm/yy)	• DOB (dd/mm/yyyy)	• Converted DOB from ED set to dd/mm/yyyy
Missing values on linkage variables	• Nil	• Admission date (93631, 12.5%)	• Records removed
Duplicated records	• Nil	• Nil	
Variables with missing values	• 5	• 15	
Variables with incorrect domain values	• Nil	• Domain descriptions changed before and after CPOE implementation for 3 IVs	• Recodes domain names to match

Record inclusion criteria

No.	Criteria	Reasons
1	MRN and Admission data not missing	Linkage variables
2	TAT positive	Negative TAT – data collection error

Results

HosLab:

748058 Records
Correspond to
23542 Patient Episodes

HosLab:

531206 Records
Correspond to
10076 Patient Episodes

Checked, cleaned and removed records
not meeting inclusion criteria

Patient's Journey

ED Admissions

Information Enriched dataset

Episode
Start Date

Arrival
Date/Time

Triage
ED_LOS

Discharge
Date/Time

Procedure
DRG

Admission
Date

TAT
Test ordered
Urgency
Results

Discharge
Date

Pathology Tests Ordered

Future work

- Link ED – HosLab – PAS
  - DRG
  - Hospital LOS

Acknowledgements

- Supported by DoHA, QUPP grant

Thanks you