Context: Evolution of Clinical Computing...

- Focus evolved from automating AR process in hospitals, to LAN-based SI of departmental systems, to enterprise clinical computing, toward region- and system-level process automation
- Changing healthcare organizational models and economic models ---> changing clinical computing priorities and designs
DATA REPOSITORY

Most Installations...

Evolutionary Stages

Phase 4: Enterprise-level Clinical Process Automation...

Different Imperatives, Different IT Strategies

Beyond the Enterprise: CHINs
Valid needs, No Buyer...
Community Health Information Needs

- Person Identification Server
- Membership Server
- Plan/Eligibility Server
- Health Record Server
- Resource Scheduling Server
- Care Protocol Server
- Commerce Server

Care Maps

- Defined and tested clinical methods
- Specific paths through map individualized
- Multiple opportunities for triage and adaptation
  - by planned choice
  - by risk
  - by availability
  - by length of trajectory
  - by cost
  - by new information

THE EYE OF THE CLINICIAN

- NLM Literature Databases
- External Knowledge
- Chart Review
- Edit Notes
- Personal Productivity
- Orders
- Decision Support
- Lab
- Meds
- Surgery
- Rad
- Dietary
- Nursing

“EMERGING FROM THE COMMON PROCESS MODEL”

- Infosystem Functional Specifications

IDENTIFY

- Enrollment
- CMPI
- MPI
- ADT
- Register
- Identify
- CPM

ASSESS

- Heuristic Plan Redesign
- Rule-Based Critiquing
- Focal AI
- Bibliographic
- Case-Based Learning
- History & Physical
- Assessment
- CPM
Summary...

- Diverse organizational models
- Healthcare IT emphasis moving from enterprise to health system focus
- Increased emphasis on patient’s role and MD risk
- An open, net-based commerce model fits in all organizational models

Definition...

- **Definition:**
  
  “Net-based clinical commerce” is the ability for patients, physicians, and other buyers of healthcare products and services to arrange, track, and pay for services, purchase healthcare commodity products from an electronic exchange, and manage health records on the Internet.

Context: Evolution of Commerce on the Internet...

- Commerce-enabled Web
  - banking
  - retail
- Extensive use in healthcare
  - research
  - communication
- Toward Clinical Commerce!!

Healthcare Commerce Entry Barriers

- Consumers purchase health *processes*
- Processes are ongoing with multiple points of contact with the provider
- Payment is by assignment of entitlement
- Payer exerts significant control over choice
- Payer exerts significant control over care
- Some goods and services require license
Patient Identification

- Example: Partners Health Care EMPI (Enterprise Master Patient Index)
  - Lisa Adragna, J AHIMA 1998 Oct;69(9):46-8, 50, 52
- Probabilistic Models for Matching

Partners EMPI Goals

- (Context) Partners Healthcare is affiliation of Brigham and Women’s, Mass. General, North-Shore Medical Center, Dana-Farber Cancer Institute, Spaulding Rehab, McLean, Partners Community Healthcare, …, formed in 1994.
- Eliminate need to renumber existing systems by creating Partners-wide unique ID
- Foundation for Clinical Data Repository
- Minimally invasive to current operations
- Rapid installation

Partners EMPI Matching

- Three methods of lookup:
  - Last and First names
  - First name, Date of Birth and Gender
  - Social Security Number
- If no matches, create new EMPI number
- If exact match on all known features (above), then link to existing EMPI entry
- If multiple possible matches, create new entry and flag as potential duplicate; manual intervention

Partners EMPI Experience (1997)

- 1M records entered from BWH & MGH
- 170,000 pair of records identified as potential duplicates
- 80% of new registrations match a unique EMPI entry; ~1800/day
- Sites made 3308 “within-entity” merges
- EMPI Admin staff (3 FTE) resolved over 200,000 potential duplicates.
Models of Patient Matching

- *Ad hoc* methods are weak, and difficult to improve
- Motivational observation:
  - Two records that share the last name Kowalczyk are more likely to refer to the same individual than two that share the name Smith.

Computing posterior odds of a match

- $O(\text{match}\mid \text{obs}) = O(\text{match})L(\text{obs}\mid \text{match})$
- $O(\text{match}) = 1/N$
- $L(\text{obs}\mid \text{match}) = P(\text{obs}\mid \text{match})/P(\text{obs}\mid \sim \text{match})$
- $P(\text{Smith},\text{Smith}\mid \text{match}) = P(\text{Smith})$
- $P(\text{Smith,Smith}\sim \text{match}) = P(\text{Smith})^2$
- $L(\text{Smith,Smith}) = 0.01006/0.01006^2 = 99.4$
- $L(\text{Kowalczyk,K…}) = 1/0.00001 = 100,000$

Multiple evidence accumulates

- Smith has same Soundex code as Smyth
- L partially matches Liz
- Same street address number, street name
- Ave. vs. Blvd.
- Different phone numbers
- …
- Assume conditional independence

Soundex

- Developed in 1800’s to represent phonetic similarity of names
  - E.g., Smith ➔ S-530; Smyth ➔ S-530
- Rules:
  - 1 B, F, P, V
  - 2 C, G, J, K, Q, S, X, Z
  - 3 D, T
  - 4 L
  - 5 M, N
  - 6 R
  - Disregard the letters A, E, I, O, U, H, W, and Y.
  - Special rules for doubled consonants, prefixes, …

Soundex examples

<table>
<thead>
<tr>
<th>Name</th>
<th>Soundex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Szolovits</td>
<td>S-413</td>
</tr>
<tr>
<td>Beethoven</td>
<td>B-315</td>
</tr>
<tr>
<td>Smyth</td>
<td>S-530</td>
</tr>
<tr>
<td>Kowalczyk</td>
<td>K-422</td>
</tr>
</tbody>
</table>

Typing/Transcription Error Model

- Consider probabilities of
  - Insertion
  - Deletion
  - Transposition
  - (Mutation) – varies with key layout
- Compute “distance” between mismatched features
- *Same idea in BLAST*
A Theory for Record Linkage

- Estimate $L(\text{obs} | \text{match})$ for all possible observations
  - Normally must make independence assumption to factor contributions of different aspects of $\text{obs}$
- Define decision policy
  - $P(\text{A+} | \text{obs})$, $P(\text{A-} | \text{obs})$, $P(\text{A?} | \text{obs})$, sum=1
  - Based on thresholds on $L$’s
- Yields errors
  - $P(\text{A+} | \sim \text{match})$ or $P(\text{A-} | \text{match})$, minimize $P(\text{A?})$

Blocking Sets

- Pair-by-pair comparison is expensive
  - E.g., 1M records at Partners $\rightarrow 10^12$ comparisons
- Therefore, compare only smaller subgroups
  - E.g., assume first letter of name is never misspelled, as Soundex does
  - E.g., assume Soundex pronunciation code is correct, even if spelling is not
- Increased chance of missing matches, much faster
### 4234 Soundex Classes

<table>
<thead>
<tr>
<th>Class</th>
<th>Name</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-530</td>
<td>Smith, Schmidt, Schmitt, Sneed, Schmid, Smoot, Snead, Shumate, Smyth, Sandy, Sand, Santo, Santoyo, …</td>
<td>1</td>
<td>1.118%</td>
</tr>
<tr>
<td>J-525</td>
<td>Johnson, Jenkins, Jensen, Jamison, Jameson, Jansen, Johansen, Janssen, …</td>
<td></td>
<td>0.998%</td>
</tr>
<tr>
<td>B-650</td>
<td>Brown, Bryan, Barron, Braun, Byrne, Bruno, Browne, Baron, Barney, Burney, …</td>
<td></td>
<td>0.779%</td>
</tr>
<tr>
<td>J-520</td>
<td>Jones, James, Johns, Jung, Janes, Jaynes, Jonas, Jenks, Jaimes, Jinks, Jahnke, Johannes, Janik, Janis, Jiang, …</td>
<td></td>
<td>0.775%</td>
</tr>
</tbody>
</table>

---

### Themes

- Growing levels of integration in health care
- Demand for integration of health IT systems raises challenging problems
- MPI is fundamental for record keeping
- Matching of records critical
  - Good theory, mixed practice
- Flip side: privacy concerns