Opportunities for Big Data in Medical Records

Mike Conlon, PhD

mconlon@ufl.edu

A Medical Record is …

- A legal document. MD accountable for contents. Data belongs to the patient
- Protected Health Information. Security, Privacy
- Records of visits
  - Patient identification and dates – who was seen when
  - Notes – transcribed text from MD dictation
  - Lab reports
  - Pointers to images
  - Billing – diagnostic codes, procedure codes, prescriptions
- Organized for care – rapid access to historical information for individual patients

Problems with Medical Records

- Incomplete. Focused on provider, not patient. One patient may have many providers
- MD bias. Data elements reported as needed for care
- Notes -> text, not structured data. Few common data elements
- Immature, nonstandard use of vocabulary for common elements. Procedures (CPT), diagnoses (ICD9/10), medications (RxNorm), UMLS (100 vocabularies, 1 million concepts, 5 million concept names)

What's Happening – New Data

- Patient reported outcomes. Quality of Life, other self-reported measures
- Genetics – 3 billion base pairs. $7K, under $1K in 2 years. 1 Gig per person.
- Metabolomics – small molecule distributions at the cellular level. Time, disease, location, function varying. PubChem 19 million small organic molecules
- Microbiome – gene sequencing of symbiotic organisms (10 times the number of human cells (100 trillion)). Time varying.

What's Happening – Pooling Data

- Collecting data across providers
- Creating a patient-centric record. Personal Health records. Microsoft HealthVault. Google Health discontinued.
- Some large collections
  - VA – 8M patients per year. Common medical record
  - Insurance/HMO claims data – United Health Care (70M), WellPoint (68M), Kaiser Permanente (9M)
  - Harvard Hospitals (20). SHRINE system
  - UF and partners. Potential for 5M patients

What's Happening – UF

- Epic electronic medical record for UF Health
- Integrated Data Repository to provide data for research organized for query and analysis. Since June 2011: 350K pts, 2M visits, 11M procedures, 14M diagnoses
- Under a consent process, addition of scoped genetic data (1,500 pts/yr; 256 SNPs)
- Under IRB approval, pooling de-identified data across systems in Florida (Gainesville, Jacksonville, Orlando)

Improving Care With Big Data

- Show me patients like my current patient (disease, condition, history, genome, metabolome, microbiome)
- Show me treatment alternatives those patients received
  - n_1 patients received t_1
  - n_2 patients received t_2
  - n_3 patients received t_3
- Show outcomes from these alternatives
  - Patients receiving t_1 had these outcomes ….
  - Patients receiving t_2 had these outcomes ….
  - Patients receiving t_3 had these outcomes ….
- Generate recommendation based on patient's concerns for risk/reward

Reducing Time to New Therapy

- Clinical trials for new therapies may take 10-15 years to complete and cost $1B for one new drug
- Using molecular data (genetic, metabolomic and microbiome), data mining techniques may discover patterns of response (positive and negative, efficacy and side effects) to therapy associated with molecular profiles allowing:
  - Definitive trial results obtained faster
  - Trial results associated with molecular profiles to guide recommendations for favorable and unfavorable scenarios for therapy