Big Data Project in CISE/Health IT: From DiabeticLink to SilverLink

Dr. Hsinchun Chen, Ph. D.
Regents’ Professor, Thomas R. Brown Chair
Director, AI Lab, University of Arizona
Expert, Smart & Connected Health Program, NSF/CISE/IIS

Dr. Randall Brown, MD, University of Arizona Health Network

Acknowledgement: NSF, NIH, NLM, NCI
Outline

• Health Big Data Analytics
• Electronic Health Record (EHR) & Patient Social Media Analytics Research
• DiabeticLink Development
• SilverLink Design (Dr. Chen)

• Why do data scientists need medical experts? (Dr. Brown)
Health Big Data Analytics

- Health IT Landscape
- NSF Smart & Connected Health (SCH)
Health IT: The Perfect Storm

- US, Obama Care, HIT meaningful use; Healthcare.gov troubles; aging baby boomers

- China, $120B healthcare overhaul; one-child policy; reverse pyramid (4-2-1)

- Taiwan, National Health Insurance (NHI) policy; NHI and EHR databases
Health Expenditure as a Share of GDP, 2000-12, G7

Data Source: OECD Health Statistics 2014
U.S. Hospitals’ Adoption Of Electronic Health Record (EHR) Systems, 2008–13 (Obamacare/Affordable Care Act, since March 23, 2010)

Health Big Data Landscape

• Health Big Data:
  – genomics (sequences, proteins, bioinformatics; 4 TBs per person)
  – health care (EHR, patient social media, sensors/devices, health informatics; Keiser, VA, PatientsLikeMe)

• Smart health, Health 2.0 (social) & 3.0 (health analytics)
• Health analytics: EHR analytics (Columbia, Vanderbilt, Utah, OHU, Harvard, IBM + Watson); patient social media analytics (UIUC, ASU, PLM, DailyStrength)
• AMIA (NLM, 2000+ participants), ACM HIT, IEEE ICHI, Springer ICSH China ➔ special issues, ACM TMIS, IEEE IS (Chen et al.)

• NSF SCH $80M, 2011-2014; infrastructure, data mining, patient empowerment, sensors/devices
• NLM $40M; NIH Reporter Search, $250M with EHR; NIH Big Data To Knowledge (BD2K) Initiative, $100M/year
Smart and Connected Health: From Medicine to Health

ECG
EEG
Pulmonary Function
Posture
Gait
Balance
Step Size
GPS
SpO₂
ECG
Blood Pressure
Step Height
Training
Chronic Care
Social Networks
Health Information
Performance Prediction
Early Detection

(Source: Dr. Howard Wactlar, IEEE IS, 2012; NSF)
From Traditional Health to Smart Health: NSF Perspective

<table>
<thead>
<tr>
<th>Traditional Health</th>
<th>Smart Health</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EPISODIC, REACTIVE</strong></td>
<td><strong>PROACTIVE and PREVENTIVE</strong></td>
</tr>
<tr>
<td><strong>FOCUS ON DISEASE</strong></td>
<td><strong>FOCUS ON WELLBEING</strong></td>
</tr>
<tr>
<td><strong>HOSPITAL-CENTRIC</strong></td>
<td><strong>QUALITY OF LIFE</strong></td>
</tr>
<tr>
<td><strong>TRAINING &amp; EXPERIENCE BASED</strong></td>
<td><strong>PATIENT-CENTRIC, HOME-BASED</strong></td>
</tr>
<tr>
<td><strong>FRAGMENTED, LOCAL DATA</strong></td>
<td><strong>MORE EVIDENCE – BASED, DATA-DRIVEN DECISION SUPPORT</strong></td>
</tr>
<tr>
<td><strong>NAÏVE, PASSIVE, PATIENTS</strong></td>
<td><strong>INTEROPERABLE, AVAILABLE ANYWHERE, ANYTIME</strong></td>
</tr>
<tr>
<td></td>
<td><strong>EMPOWERED, ENAGAGED, INFORMED, PARTICIPATING</strong></td>
</tr>
</tbody>
</table>
Smart and Connected Health: People, Technology, Process

Subjective
- Concerns
- Patient Reported Outcomes

Objective
- Clinical measures
- Laboratory findings
- Sensor data

Assessment
- Diagnosis
- Categorical reporting
- Prognosis/Trajectory

Plan
- Treatment planning
- Self-care planning
- Post treatment
- Surveillance

Decision Support Needs
- Risk modeling
- Diagnostic support
- Treatment selection
- Guideline adherence
- Error detection/correction

Information Exchange

Clinic-based EHR Data

Patient-based Health Data

Medical Team

Hospital System

Medical Researcher

Patient & Family

- Situational awareness
- Population health
- Continuity of care
- Identify side effects
- Inform discovery
# Smart and Connected Health Research Areas

| **Digital Health Infrastructure** | • Integration of EHR, clinical and patient data  
|                                 | • Access to information, data harmonization  
|                                 | • Semantic representation, fusion, visualization |
| **Informatics and Infrastructure** | |

| **Data to Knowledge to Decision** | • Data-mining and machine learning  
|                                 | • Inference, cognitive decision support system  
|                                 | • Bring raw image data to clinical practice |
| **Reasoning under uncertainty** | |

| **Empowered Individuals** | • Systems for empowering patient  
|                         | • Models of readiness to change  
|                         | • State assessment from images video |
| **Energized, enabled, educated** | |

| **Sensors, Devices, and Robotics** | • Assistive technologies embodying computational intelligence  
|                                  | • Medical devices, co-robots, cognitive orthotics, rehab coaches |
| **Sensor-based actuation** | |
SCH Sample Projects
SCH: EXP: Intelligent Clinical Decision Support using Probabilistic and Temporal EHR Modeling

**Motivation:**
- Decision-support tools have potential to exploit the wealth of clinical data in EHRs as well as expert recommendations
- Balance costs vs outcomes, reason with uncertainty
- New AI techniques needed to calculate patient-specific, temporal, statistically-justified treatment plans

**Technical Approach:**
- Advance AI state of the art: Partially-Observable Markov Decision Processes, Statistical Relational Learning
- Prototype clinical decision support system (CDSS) that integrates SRL disease progression models into a POMDP decision making model
- Three disease scenarios: stroke, cardiology, ER readmission

**Contacts:**
- Kris Hauser, Sriraam Natarajan, Shaun Grannis
  Indiana University
- Marshfield Clinic, Vanderbilt, Regenstrief Institute, Advanced Health Logic
- [http://www.iu.edu/~motion/decisionssupport](http://www.iu.edu/~motion/decisionssupport)
Mapping the Cardiac Acousteome: Biosensing and Computational Modeling Applied to Smart Diagnosis and Monitoring of Heart Conditions

Motivation
• Cardiac auscultation is the science and art of diagnosis of heart conditions via the stethoscope – this diagnostic modality is essentially unchanged in over a 100 years.
• New sensors and signal processing algorithms combined with high-fidelity bioacoustic modeling, and rigorous testing and validation, could provide the impetus needed to usher this modality into the 21st century.

Contacts
• Rajat Mittal, (Ph.D.), Mechanical Engineering, JHU (mittal@jhu.edu),
• Andreas Andreou, (Ph.D.), Electrical Engineering, JHU
• William R. Thompson, (M.D.), Pediatric Cardiology, JHU
• Theodore Abraham (M.D.), Cardiology, JHU.

Technical Approach
• High-fidelity hemoacoustic modeling and simulation to understand generation and scattering of heart sounds in thorax.
• Bioinspired approach to sound measurement and localization with a new compact acoustic sensor array (the “StethoVest”).
• Generative (model based) statistical pattern recognition algorithms for abnormal heart sounds.

Computational (in-silico), experimental
Fittle+: Theory and Models for Smartphone Ecological Momentary Intervention

Motivation:
• Smartphone platforms can project behavioral-change techniques into everyday life, to improve diet and fitness
• Intensify and prolong interventions and impact
• Develop fine-grained predictive models to address rich, fine-grained data collected on daily behavior change dynamics.

Contacts:
• PIs: Peter Pirolli & Michael Youngblood, PARC
• Collaborators: Robert Kraut (CMU), Joy Zhang (CMU-SV), Mark Wilson (UC Berkeley)

Technical Approach:
• Intelligent, personalized coaching that plans and delivers personalized interventions in the right contexts
• Online, mobile social interaction to support motivation, commitment, and learning
• Integrated with mobile sensing and wearables (Fitbit)
• Testbed; Weight loss programs at a large Health Maintenance Organization - Kaiser Permanente Hawaii (KPH)
Some of My Prior Security & Medical Informatics Projects
COPLINK System: Crime Data Mining
COPLINK Identity Resolution and Criminal Network Analysis (DHS)

Cross-jurisdictional Information Sharing/Collaboration

Arizona IDMatcher

Law-enforcement Data

CAN Visualizer

Border Crossing Data (AZ, CA, TX)

Vehicles

People

Identity Resolution

Detect false and deceptive identities across jurisdictions using a probabilistic naïve-Bayes based resolution system.

Criminal Network Analysis

High-risk Vehicle Identification

Identify high-risk vehicles using association techniques like mutual information using border crossing and law enforcement data.

Criminal Link Prediction

Predict interaction between individuals and vehicles using link prediction techniques to identify high-risk border crossers.

Suspect Traffic Burst Detection

Detect real-time anomalies and threats in border traffic using Markov switching and other models.

* Only the grayed datasets are available to the AI Lab

Techniques: Entity resolution, anomaly detection, link prediction, spatial-temporal clustering, burst detection

Partners: TPD, PPD, DHS, CBP; criminology, public policy
COPLINK: Crime Data Mining

The New York Times, November 2, 2002
COPLINK assisted in DC sniper investigation

ABC News April 15, 2003

Google for Cops: Coplink software helps police search for cyber clues to bust criminals

A computerized way for police to coordinate crime databases

Washington Post, March 6, 2008
National dragnet is a click away! COPLINK in use in 3,500 police agencies in US!

Commercialized in 2000; Merged i2 in 2009; acquired by IBM in 2011 for $500M
Dark Web: Terrorism Informatics

- Dark Web: Terrorists’ and cyber criminals’ use of the Internet
- Collection: Web sites, forums, blogs, YouTube, etc.
- Analytics: Authorship analysis, multilingual affect analysis & visualization, social infection modeling, dark network analysis
- 20 TBs in size, with close to 10B pages/files/messages (the entire LOC collection: 15 TBs)

Partners: LOC, terrorism study, M Sageman, West Point CTC, OKC MIPT, Israel ICT; CIA, FBI, NSA
Security Informatics & Dark Web

Chen, 2006

Chen, 2012
Funding: NSF, NIH, NLM, NCI ($3M); Digital Library Program
Publications: ISR, JAMIA, JBI, IEEE TITB
Impact: medical knowledge mapping & visualization ➔ health informatics

Medical Informatics
Cancer Map: 2M CancerLit articles, 1500 maps (OOHAY, DLI)
BioPortal: Infectious Disease Tracking and Visualization, SARS, WNV, FMD (ISR, 2009)
From EHR to Patient Intelligence: A Case Study
## EHR Data Mining: Summary of Prior Work

<table>
<thead>
<tr>
<th>Study</th>
<th>Predicting Event</th>
<th>Data Source</th>
<th># of Features</th>
<th>Feature Selection</th>
<th>Handle Missing Values</th>
<th>Predictive Models</th>
<th>Model Evaluation</th>
<th>Metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lin et al. (2014)</td>
<td>Hospitalization among diabetes patients</td>
<td>EHR</td>
<td>79</td>
<td>EKB</td>
<td>Y</td>
<td>CM</td>
<td>SF</td>
<td>AUC</td>
</tr>
<tr>
<td>Sun et al. (2013)</td>
<td>Transition points in hypertension control</td>
<td>EHR</td>
<td>NA</td>
<td>SMLB</td>
<td>Y</td>
<td>LR, NB, RF</td>
<td>CV</td>
<td>A, AUC</td>
</tr>
<tr>
<td>Tammemägi et al. (2013)</td>
<td>Risk of lung cancer</td>
<td>CST</td>
<td>11</td>
<td>EKB</td>
<td>N</td>
<td>CM, LR</td>
<td>EVD</td>
<td>AUC</td>
</tr>
<tr>
<td>Sesen et al. (2012)</td>
<td>Lung cancer one-year-survival</td>
<td>CST</td>
<td>9</td>
<td>EKB</td>
<td>N</td>
<td>BN, NB</td>
<td>CV</td>
<td>AUC</td>
</tr>
<tr>
<td>Sun et al. (2012)</td>
<td>Risk of heart failure</td>
<td>EHR</td>
<td>500+</td>
<td>EKB+SMLB</td>
<td>NA</td>
<td>SOR</td>
<td>CV</td>
<td>AUC</td>
</tr>
<tr>
<td>Yeh et al. (2011)</td>
<td>Hospitalization among hemodialysis patients</td>
<td>CST</td>
<td>26</td>
<td>EKB</td>
<td>Y</td>
<td>ARM, DT</td>
<td>CV</td>
<td>AC</td>
</tr>
<tr>
<td>Khosla et al. (2010)</td>
<td>Stroke risk in 5 years</td>
<td>CST</td>
<td>796</td>
<td>SMLB</td>
<td>Y</td>
<td>CM, SVM</td>
<td>CV</td>
<td>AUC, CI</td>
</tr>
<tr>
<td>Wu et al. (2010)</td>
<td>Risk of heart failure</td>
<td>EHR</td>
<td>179</td>
<td>SMLB</td>
<td>Y</td>
<td>LR, SVM</td>
<td>CV</td>
<td>AUC</td>
</tr>
<tr>
<td>Lewis et al. (2009)</td>
<td>Risk of heart failure</td>
<td>CST</td>
<td>17</td>
<td>EKB</td>
<td>N</td>
<td>CM</td>
<td>CV</td>
<td>CI</td>
</tr>
<tr>
<td>Cho et al. (2008)</td>
<td>Onset of diabetic nephropathy</td>
<td>EHR</td>
<td>184</td>
<td>SMLB</td>
<td>Y</td>
<td>LR, SVM</td>
<td>CV</td>
<td>AUC</td>
</tr>
<tr>
<td>D’Agostino et al. (2008)</td>
<td>Risk of cardiovascular events</td>
<td>CST</td>
<td>8</td>
<td>EKB</td>
<td>N</td>
<td>CM</td>
<td>BR</td>
<td>CI</td>
</tr>
<tr>
<td>Levy et al. (2006)</td>
<td>Survival time of heart failure patients</td>
<td>CST</td>
<td>24</td>
<td>EKB</td>
<td>Y</td>
<td>CM</td>
<td>EVD</td>
<td>AUC</td>
</tr>
</tbody>
</table>

**Note.** A=accuracy; ARM=association rule mining; AUC=area under the receiver-operating-characteristic curve; B=Boosting; BN=Bayesian network; BR=bootstrap resample; CI=concordance index; CM=Cox model; CST=cohort studies or trials; CV=cross validation; DT=decision trees; EKB=evidence or knowledge based; EHR=electronic health records; EVD=external validation data; LR=logistic regression; NA=not available in the paper; NB=naive bayes; RF=random forest; SMLB=statistical or machine learning based; SOR=scalable orthogonal regression; SVM=support vector machine.
Clinical Process and EHR Data: 1M Patients, 100M records, 10 years (HIPAA, IRB approved)

Patient background:
- CHART (病歷基本資料檔)
- AGEGROUP1 (年齡屬主檔)
- AGEGROUP2 (年齡分齡表身檔)
- PTTYPE (身份代碼檔)

Outpatient Modules
- PTER (急診掛號檔)
- PTOPD (門診掛病患檔)
- CODINGOPDA (門診疾病分類診斷檔)
- CODINGOPD (門診疾病分類資料檔)
- CODINGOPDP (門診疾病分類處置檔)
- ORDAOPD (門診病患診斷檔)
- HORDERA (門診病患診斷檔)
- ORDSOOPD (門診S.O.檔)
- ACNTOPD (門診病患醫令明細檔)
- PRICE (收費標準檔)
- HRECOPD1 (歷史門診收據檔(表頭))
- HRECOPD2C (歷史門診收據檔(表身))
- HRECOPD2D (歷史門診收據檔(借方))

Registration → Diagnosis (Symptoms and Diseases) → Treatment (Procedures and Orders) → Transaction / Receipt

Inpatient Modules

Hospital:
- BED (床位主檔)
- BEDGRADE (床位等級代碼檔(表頭))
- BEDGRADE2 (床位等級代碼檔(表身))
- BEDSTATUS (床位狀態檔)
- DEPT (部門代碼檔)
- DIV (科別代碼档)
- DOCTOR (醫師代碼檔)
- HOSPITAL (醫院代碼檔)

Disease:
- APDRGD (DRG疾病代碼對照檔)
- DRGICD9 (DRG疾病代碼對照檔)
- ICDGROUP1 (ICD分類主檔)
- ICDGROUP2 (ICD分類表身檔)
- ICD (國際疾病分類代碼檔)

Operation:
- PTOR (手術病人主檔)
- PTORDRPT (手術病人報告記錄明細檔)
- ORSAMPLE (手術內容模組主檔)
- PTORALLLOG (手術異常記錄LOG檔)
- PTORDIAG (手術病人術後診斷)
- PTORLOG (手術病人報告記錄明細檔)
- PTORSTAFF (手術參與人員檔)

LIS:
- LABITEM1 (檢驗項目主檔(表頭))
- LABITEM2 (檢驗項目主檔(表身))
- LABGROUP (檢驗組別代碼檔)
- LABP1 (檢驗病理表頭檔)
- LABP2 (檢驗病理表身檔)
- LABS1 (檢驗病理組織學1檔)
- LABS2 (檢驗病理組織學2檔)
- LABX1 (檢驗異動表頭檔)
- LABX2 (檢驗異動表身檔)

PACS:
- EXAMITEM (檢查項目檔)
- PTEXAM (申請單主檔)
- PTEXAMINDEX (申請單表頭檔)
- PTEXAMITEM (申請單檢查項目檔)
- PTEXAMRPT (申請單報告檔)

Note: Tables with underlines contain free-text data.
Research Framework

- **Guideline-based feature selection**: obtain clinically meaningful features
- **Temporal Regularization**: handle irregularly spaced data
- **Data abstraction**: reduce data dimensionality and bring out semantics
- **Multiple imputation**: handle missing data
- **Extended Cox models**: time-to-event modeling with time-dependent covariates
Guideline-based Feature Selection (cont.)

- We arrange the concepts in three dimensions: evaluations, diagnoses, and treatments.
  - About one hundred concepts are extracted and encoded from the AACE guidelines. The table shows a subset of the instances. We then manually map these concepts to the corresponding items in EHRs, resulting in about 400 ICD9 diagnosis codes, 150 unique treatments, and 20 lab tests and physical evaluations.

<table>
<thead>
<tr>
<th>Category</th>
<th>Evaluations</th>
<th>Diagnoses</th>
<th>Treatments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diabetes</strong></td>
<td>• HbA1c</td>
<td>• Polyuria</td>
<td>• Insulin regimen</td>
</tr>
<tr>
<td></td>
<td>• Fasting glucose</td>
<td>• Polydipsia</td>
<td>• DPP-4 inhibitors</td>
</tr>
<tr>
<td></td>
<td>• 2 hours after meal glucose</td>
<td>• Polyphagia</td>
<td>• GLP-1 agonists</td>
</tr>
<tr>
<td></td>
<td>• 75 g oral glucose tolerance test</td>
<td>• Unexplained weight loss</td>
<td>• Metformin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Hypoglycemia</td>
<td>• Sulfonylurea</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Hyperglycemia</td>
<td>• Thiazolidinedione</td>
</tr>
<tr>
<td><strong>Cardiovascular</strong></td>
<td>• Blood pressure</td>
<td>• Hypertensive diseases</td>
<td>• Antihypertensive therapy</td>
</tr>
<tr>
<td><strong>diseases</strong></td>
<td>• LDL cholesterol</td>
<td>• Acute coronary syndromes</td>
<td>• Antiplatelet therapy</td>
</tr>
<tr>
<td></td>
<td>• HDL cholesterol</td>
<td>• Coronary heart diseases</td>
<td>• Lipid lowering therapy</td>
</tr>
<tr>
<td></td>
<td>• Triglycerides</td>
<td>• Disorders of lipid metabolism</td>
<td></td>
</tr>
</tbody>
</table>
Data Abstraction (cont.)

- **Concept abstraction**
  - **Diagnosis:** ICD9 codes are mapped to a higher order concept by using the Clinical Classifications Software (Elixhauser et al. 2013)
  - **Treatment:** medications are categorized by their family/class names
The Final Feature Set

<table>
<thead>
<tr>
<th>Category</th>
<th>Category code</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>Baseline</td>
<td>Sex, age, smoking, HbA1c, fasting glucose, LDL cholesterol, triglyceride, systolic blood pressure, BUN, creatinine, body weight</td>
</tr>
<tr>
<td>Concept abstracted treatment features</td>
<td>TX</td>
<td>ACE inhibitors, ARBs, amputation, antihypertensive therapy, antiplatelet therapy, DPP4 inhibitors, insulin, lipid lowering therapy, metformin, sulfonylureas, thiazolidinediones</td>
</tr>
<tr>
<td>Temporal abstracted baseline features</td>
<td>TA</td>
<td>States and trends of baseline features (HbA1c, glucose AC, LDL cholesterol, triglyceride, BUN, creatinine, body weight, systolic blood pressure)</td>
</tr>
</tbody>
</table>

Note: Some lab tests, e.g., HDL cholesterol or glomerular filtration rate, were not included in our final feature set because less than 50% of our patients receive these tests.
Extended Cox Model

• Cox proportional hazards model is a popular tool for time-to-event analysis.
  – Proportional hazards: \( h(t, x_1) = \theta h_0(t, x_1') \)

• A Cox model (Cox 1972) is given by
  \[
  h(t, X) = h_0(t) \exp \left( \sum_{i=1}^{P_1} \beta_i X_i \right)
  \]

• An extended Cox model allows covariates to change over time, enabling a more flexible modeling framework (Fisher and Lin 1999). The extended model is given by
  \[
  h(t, X(t)) = h_0(t) \exp \left( \sum_{i=1}^{P_1} \beta_i X_i + \sum_{j=1}^{P_2} \delta_j X_j(t) \right)
  \]

where \( h(t, X(t)) \) is the hazard value at time \( t \),
  \( h_0(t) \) is an arbitrary baseline hazard function,
  \( X \) is a covariate matrix, containing \( P_1 \) time independent covariates and \( P_2 \) time dependent covariates.
Results: Diabetes

- Bootstrap analyses: Full Extended Cox Model achieves 89% accuracy.
Statistically Significant Covariates

<table>
<thead>
<tr>
<th>Risk factors</th>
<th>Feature category</th>
<th>Hazard ratio</th>
<th>Lower CI bound</th>
<th>Upper CI bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open wounds of extremities (CCS 236)</td>
<td>DX</td>
<td>12.898</td>
<td>1.495</td>
<td>121.187</td>
</tr>
<tr>
<td>Acute and unspecified renal failure (CCS 157)</td>
<td>DX</td>
<td>11.243</td>
<td>1.569</td>
<td>81.705</td>
</tr>
<tr>
<td>Insulin treatment</td>
<td>TX</td>
<td>6.082</td>
<td>3.780</td>
<td>9.787</td>
</tr>
<tr>
<td>Smoking</td>
<td>Baseline</td>
<td>2.750</td>
<td>1.745</td>
<td>4.336</td>
</tr>
<tr>
<td>Antiplatelet therapy</td>
<td>TX</td>
<td>2.145</td>
<td>1.200</td>
<td>3.841</td>
</tr>
<tr>
<td>Upward trend of body weight</td>
<td>TA</td>
<td>1.788</td>
<td>1.238</td>
<td>2.582</td>
</tr>
<tr>
<td>Upward trend of fasting glucose</td>
<td>TA</td>
<td>1.642</td>
<td>1.098</td>
<td>2.459</td>
</tr>
<tr>
<td>Diabetes mellitus without complication (CCS 49)</td>
<td>DX</td>
<td>1.489</td>
<td>0.965</td>
<td>2.298</td>
</tr>
<tr>
<td>HbA1c</td>
<td>Baseline</td>
<td>1.112</td>
<td>1.004</td>
<td>1.233</td>
</tr>
<tr>
<td>LDL cholesterol</td>
<td>Baseline</td>
<td>1.007</td>
<td>1.000</td>
<td>1.013</td>
</tr>
<tr>
<td>Fasting glucose</td>
<td>Baseline</td>
<td>1.003</td>
<td>1.002</td>
<td>1.004</td>
</tr>
<tr>
<td>Sulfonylurea treatment</td>
<td>TX</td>
<td>0.663</td>
<td>0.442</td>
<td>0.994</td>
</tr>
<tr>
<td>Low level state of fasting glucose</td>
<td>TA</td>
<td>0.545</td>
<td>0.304</td>
<td>0.976</td>
</tr>
</tbody>
</table>

Note 1: p-value ≤ .05 in two or more imputed data sets
Note 2: CI=95% confidence interval
Note 3: When a hazard ratio is significantly greater than one, the risk factor is deemed positively associated with the event. On the other hand, if a hazard ratio is significantly below one, the risk factor is negatively associated with the event.
DiabeticLink Risk Prediction Engine

Compare to average patients

Your risk of getting a stroke is **2.59** times higher than average patients in your age.

What-if analysis

If you control your LDL Cholesterol to the level of 130, your risk of stroke is **62% lower** than your current status.

Hemoglobin A1C: (%) 5% - 7.0%
LDL Cholesterol: (mg/dl) 50 - 130
Systolic Blood Pressure: (mmHg) 90 - 120

Risk changes: **62%**

Estimate again

Stroke time prediction

You have 50% change get a stroke **in 2 years**.
You have 90% chance get a stroke **in 4 years**.
DiabeticLink Patient Portal

DiabeticLink US and Taiwan teams
<table>
<thead>
<tr>
<th>Feature</th>
<th>DiabeticLink</th>
<th>dLife</th>
<th>American Diabetes Association</th>
<th>Diabetes.co.uk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregated Forums</td>
<td>✔️</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Adverse Drug Reactions</td>
<td>✔️</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>EHR Search tool</td>
<td>✔️</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Portal Tracking app integrated with Mobile</td>
<td>✔️</td>
<td>No</td>
<td>No</td>
<td>(Portal tracking, but no integration)</td>
</tr>
<tr>
<td>Mobile Tracking app (Glucose, weight, A1c, medications, food log, etc)</td>
<td>✔️</td>
<td>(Glucose only; search recipes; Q&amp;A)</td>
<td>(ADA journals only)</td>
<td>No</td>
</tr>
<tr>
<td>Community Forums; Social Connectivity; Meal Plans; Food &amp; Fitness Guides; Diabetes Guides and Health Information</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
</tbody>
</table>
Homepage
Tracking Dashboard

Diabetes Tracker

Average Glucose Per Category

Average Glucose for Last 10 Days

Average Carbs for Last 10 Days

Average Carbs Per Meal

My Account

Dashboard
Settings
View Entered Data

Entries Overview

- Glucose
- Food
- Insulin
- Blood Pressure
- Weight
- HbA1c

Glucose:
- All Categories
- All Symptoms

View Glucose Records Inc. mmol/L, mg/dL

Food:
- All Categories

Insulin:
- All Categories

Graph showing blood glucose levels (mg/dL) with time (8:00 AM to 8:00 AM over Oct 8 to Oct 9).
Search Drug: Avandamet
Compare Two Drugs: Avandia and Avandamet
<table>
<thead>
<tr>
<th>Basics</th>
<th>Lifestyle Factors</th>
<th>Complications</th>
<th>Treatments</th>
<th>Hba1c Graph</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient 1</td>
<td>None</td>
<td>None</td>
<td>Medications: 9, Hospitalizations: 1, Dialysis: 0</td>
<td>Stable</td>
</tr>
<tr>
<td>Patient 2</td>
<td>None</td>
<td>None</td>
<td>Medications: 11, Hospitalizations: 2, Dialysis: 0</td>
<td>Stable</td>
</tr>
<tr>
<td>Patient 3</td>
<td>None</td>
<td>None</td>
<td>Medications: 4, Hospitalizations: 2, Dialysis: 0</td>
<td>Improving</td>
</tr>
<tr>
<td>Patient 4</td>
<td>None</td>
<td>Kidney Problems</td>
<td>Medications: 20, Hospitalizations: 5, Dialysis: 0</td>
<td>Improving</td>
</tr>
<tr>
<td>Patient 5</td>
<td>None</td>
<td>Eye Problems</td>
<td>Medications: 6, Hospitalizations: 0, Dialysis: 0</td>
<td>Stable</td>
</tr>
<tr>
<td>Patient 6</td>
<td>Overweight</td>
<td>Eye Problems</td>
<td>Medications: 11, Hospitalizations: 1, Dialysis: 0</td>
<td>Improving</td>
</tr>
</tbody>
</table>
View Patient Summary Chart
DiabeticLink Risk Engine

Compare to average patients

Your risk of getting a stroke is **2.59** times higher than average patients in your age.

What-if analysis

If you control your LDL Cholesterol to the level of 130, your risk of stroke is **62% lower** than your current status.

- Hemoglobin A1C: (%)
  - Current: 7.0
  - Target: 5%
- LDL Cholesterol: (mg/dl)
  - Current: 120
  - Target: 130
- Systolic Blood Pressure:(mmHg)
  - Current: 120
  - Target: 130

**Risk changes: 62%**

Stroke time prediction

You have 50% chance get a stroke **in 2 years**.
You have 90% chance get a stroke **in 4 years**.
DiabeticLink
Taiwan
• We consolidated the modules and grouped the modules that have similar features together.

• The 飲食地圖 includes the original recipe and restaurant modules.

• The 健康報馬仔 is the module which allows the users to search useful information such as clinic address, open hours; NHI stats; and ADE.
New Mobile Layout

文章列表

文章呈現方式1(開頭)

文章呈現方式2(文中圖片)
SilverLink Design

US, Taiwan, China teams
SilverLink: Supporting Senior Citizens Living Independently

- Elderly who stay alone
- Elderly with a chronic condition
- Elderly with worried children

Age group: 55-90 years living independently

Remote family support: email, video, chart

Chronic conditions:
- Diabetes
- Dementia
- Chronic heart conditions
SL Hardware Requirement

• **Activity Sensors**
  – Keychain sensor
    • in and out of the house
  – Other sensors can be attached to pillboxes, refrigerator door, bathroom or shower door, front door, slippers, chairs, etc.
    • Each sensor is labeled with a one-digit number
  – Wearable sensor / SOS wristband
    • Detect person’s movement and fall
    • emergency alarm

• **Embedded BLE and accelerometer for all sensors**

• **Home Gateway with built-in 3G service**
  – Placed at a central location of the house for receiving sensor data and transmitting data to SL SmartCloud
SL Software Requirements

• SL SmartCloud
  – Database
    • Store the raw data collected from the activity sensors
  – Analytics Engines
    • Advanced pattern recognition and abnormal pattern detection algorithm
  – Alerts and notifications dispatcher

• SL web portal

• SL mobile app
SL SmartCloud

- Data collection API
- Activity log and portal databases
- Analytics engines
- Notification service

Data flow

Data source

- Remote sensors
- Data Collection API
- Raw sensor data

Data transformation and integration

- Noise reduction
- Sanitization
- Sanitized activity data

Analytics approaches

- Pattern recognition
- Signal detection
- User Activity Data

Notification service

Client

- Portal
- SMS
- Email
SL Web Portal - Dashboard

**Notifications**

10/13/2014 5:30PM
Alice has not left the house today

10/13/2014 12:30PM
Alice, did you forget to eat lunch?

10/13/2014 1:30PM
Alice, did you forget to take the morning pill?

**System Health**

- **Sensor #1**: Kitchen
  - Status: Good

- **Sensor #2**: Front Door
  - Status: Good

- **Sensor #3**: Pillbox
  - Status: Error

- **Sensor #4**: Chair
  - Status: Low Battery

- **Keychain**
  - Status: Good

- **Gateway**
  - Status: Good

**Sensor Settings**
SilverMail

- SilverMail will act as an incentive:
  - Providing monthly health summary and reminders.
  - Providing photographs, messages and videos from family every month.

Goal:
- Motivate the elderly to keep using the tracking features and better manage their health.
- Bridge the gap between the user and their loved ones.
## Competition: Feature comparison
(US/China testing: June-December 2015)

<table>
<thead>
<tr>
<th>Competitors/Features</th>
<th>SOS button</th>
<th>Motion Sensors for home monitoring</th>
<th>Analytics engine (fall, anomaly detection)</th>
<th>Emergency response services</th>
<th>Web portal (sensor status, etc.)</th>
<th>Glucose, BP Tracking using tablet</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SilverLink</strong></td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Alert 1</td>
<td>✔</td>
<td>✗</td>
<td>✗</td>
<td>✔</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td><strong>MyLively</strong></td>
<td>✔</td>
<td>✔</td>
<td>✗</td>
<td>✔</td>
<td>✔</td>
<td>✗</td>
</tr>
<tr>
<td>Phillips medical alert system</td>
<td>✔</td>
<td>✗</td>
<td>✗</td>
<td>✔</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td><strong>Bay Medical Systems</strong></td>
<td>✔</td>
<td>✗</td>
<td>✗</td>
<td>✔</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Medical Guardian</td>
<td>✔</td>
<td>✗</td>
<td>✗</td>
<td>✔</td>
<td>✗</td>
<td>✗</td>
</tr>
</tbody>
</table>
Health Big Data Research: Summary

- What type of data do your consumers want? By what mechanisms do you determine this? ➔ health contents; patients, physicians, etc.
- What makes this type of data “big” (e.g., volume, velocity, variety, veracity)? ➔ integrated EHR, patient social media, genomics
- What infrastructure, funding, and policies are needed to generate this data? ➔ health government funding, industry interest
- Do you confront issues of data standards and interoperability? ➔ Yes, HIE (health information exchange)
- Do you confront issues of privacy/security/ethics? ➔ Yes, HIPAA, IRB; international dimension
- What types of methods are used to analyze your data by producers? By consumers? ➔ Data/text/web mining, from database to analytics
- Do you confront issues of limited current capacity in the data sciences? ➔ Yes, domain knowledge, inter-disciplinary teams, showing impacts
- Does your project involve partnerships or other types of sustained organizational relationships? ➔ Yes, governments, hospitals, etc.
- How has your big data work changed your field? ➔ Perfect Storm!
- What advice do you have for others running big data projects? ➔ training data scientists (4Vs, DB, statistics, ML), forming inter-disciplinary teams, prioritizing goals (high-impact problems), demonstrating impacts/values

- Big Data in Education? ➔ http://ai.arizona.edu/mis510/ Web Computing and Mining (my experience)
For questions and comments

hchen@eller.Arizona.edu
http://ai.Arizona.edu