From Data to Knowledge:
Data platform, information resource,
and predictive models

Yaorong Ge
October 10, 2018
Outline

• Data platform
  – Data integration
  – Data services

• Information resource
  – Semantic knowledge representation
  – Uncertainty in clinical knowledge

• Predictive models
  – Models for radiation therapy planning

• Toward an intelligent health data resource
Data Platform

Overarching Goals: Deliver to researchers

a. High quality data
b. Integrated knowledge, and
c. Effective tools to facilitate health research
i2b2 data warehouse platform

• i2b2 database
  – Star Schema

• i2b2 applications in a Hive Architecture
  – User management
  – Data analysis
  – Result visualization
Key Requirements

• Researchers must be able to explore data directly
Old Data Request Workflow

• Researcher
  – Seek IRB approval
  – Request data via an online request form (after IRB approval)

• TSI data control agents
  – Verify IRB protocol
  – Work with researcher to define search criteria – with extensive iterations
  – Fulfill request through EDW or TDW
New Data Request Workflow

• Researcher
  – Explore TDW for proper inclusion and exclusion criteria of a cohort
  – Seek IRB approval
  – Request data via an online request form (after IRB approval)

• TSI data control agents
  – Verify IRB protocol
  – Work with researcher to refine search criteria – with minimal iterations
  – Fulfill request through EDW or TDW
i2b2 Ontology-based Query Tool
Key Requirements

• Developers must be able to integrate heterogeneous data efficiently
i2b2 Star Schema

- **visit_dimension**
  - PK: encounter_num, patient_num
  - encounter_num: VARCHAR(10)
  - location_cd: VARCHAR(100)
  - location_path: VARCHAR(700)
  - start_date: DATETIME
  - end_date: DATETIME
  - visit_blob: TEXT(10)

- **observation_fact**
  - PK: encounter_num, concept_cd, provider_id, start_date
  - encounter_num: VARCHAR(20)
  - concept_cd: VARCHAR(20)
  - provider_id: VARCHAR(20)
  - start_date: DATETIME

- **patient_dimension**
  - PK: patient_num
  - patient_num: INTEGER
  - vital_status_cd: VARCHAR(10)
  - birth_date: DATETIME
  - death_date: DATETIME
  - sex_cd: VARCHAR(10)
  - age_in_years_num: INT
  - language_cd: VARCHAR(100)
  - race_cd: VARCHAR(100)
  - marital_status_cd: VARCHAR(100)
  - religion_cd: VARCHAR(100)
  - zip_cd: VARCHAR(20)
  - statecypZip_path: VARCHAR(200)
  - patient_blob: TEXT(10)

- **concept_dimension**
  - PK: concept_path
  - concept_path: VARCHAR(700)
  - concept_cd: VARCHAR(20)
  - name_char: VARCHAR(2000)
  - concept_blob: TEXT(10)

- **provider_dimension**
  - PK: provider_path
  - provider_path: VARCHAR(800)
  - provider_id: VARCHAR(20)
  - name_char: VARCHAR(2000)
  - provider_blob: TEXT(10)
Main Data Sources

• Clinical Data Repository (CDR) – until 9/12
  – Identified data from EMR, financial, surgery, labs, and ancillary systems (GE and others)

• Enterprise Data Warehouse (EDW)
  – Identified data from EPIC and other systems

• Clinical registries
  – Cancer registry
  – Cardiology registries
Cancer Registry (NAACCR)

• Categories include
  o Identification
  o Demographic
  o Recurrence/Death
  o Stage/Prognostic Factors
  o Textual Diagnosis
  o Textual Treatment
  o Pathology

* Encoded values from multiple standards
  (TNM,FORDS,NAACCR,SEER,WHO ICD-O)
Cardiology Registry

Categories of data include:
- Identification
- Demographics
- History and Risk Factors
- Clinical Evaluation
- Procedures
- Labs
- Coronary Anatomy
- Medications
- Post Procedure Events
Key Requirements

• The data platform must be a live resource
  – Should not be simply a passive process
  – Must be able to add data services to the platform easily
i2b2 Data Warehouse Platform
Wake Forest TDW Applications

• Clinical Trender
  – For a given cohort, plot trend charts of count of cohort and count and value of critical concepts (per month)
  – Basic elements of TDW Dash Board

• Radiology Report Analysis Tool
  – For a given cohort with radiology tests, search the reports for certain entity labels

• Phenotype Editor
  – For successful queries of certain cohorts, define them as new categories in the ontology after proper review
WF Clinical Trender

Identifies cohort of interest
Refines results
Specifies trend of interest
Trender conducts background queries
Trends updated with new data

Researcher analyzes trend graphs
Clinical Trender plugin
Trend of influenza patients

Shows seasonal uptick in number of patients treated
Trend of adverse effects of antibiotics in influenza patients

*Shows same seasonal uptick*
Trend of Vioxx (Rofecoxib) recipients

Taken off the market 9/20/2004 due to increased risk of heart attack
Trend of heart attacks among Vioxx (Rofecoxib) recipients

*Shows matching uptick over the period of Vioxx prescriptions*
WF Radiology Reports Analysis Tool

How many children complaining about nausea and vomiting are found to have hydrocephalus with a CT of the chest?
Location of terms identified and snippets produced

Affirmative cases can be expanded to see snippet from report text
Wake Forest vs. Boston Children’s
Figure 2: Proportions of various comorbidities in the two populations. Most comorbidities have similar rates, with the except of mental retardation (Wake Forest has many more).
Comorbidity Clusters in Autism Spectrum Disorders: An Electronic Health Record Time-Series Analysis

- 4927 BCH, 496 WFU, ICD9 codes in 0-15 years, Autism spectrum
- Group 1
  - Seizures (n = 120, seizure prevalence 77.5%).
- Group 2
  - Multi-system disorders (n = 197, GI disorder prevalence 24.3%, ear disorder prevalence 87.8%).
- Group 3
  - Psychiatric disorders (n = 212, psychiatric disorder prevalence 33.0%)
SCILHS – A PCORI CDRN
Data Platform

Overarching Goals: Deliver to researchers

a. High quality data
b. Integrated knowledge, and
c. Effective tools to facilitate health research
Outline

• Data platform
  – Data integration
  – Data services

• Information resource
  – Semantic knowledge representation
  – Uncertainty in clinical knowledge

• Predictive models
  – Models for radiation therapy planning

• Toward an intelligent health data resource
A Dataset Information Resource

• If you want a dataset that has more than 1,000,000 subjects to explore a deep learning algorithm
• If you want a dataset that is suitable for time series analysis
• If you are curious about existing analyses that have been applied to a known dataset
Representing and reasoning about dataset related knowledge
• Challenges of exploring datasets
   – Complexity, diversity, large scale, proprietary nature

• Limited help from existing resources
   – HealthData.gov, CEDAR, bioCADDIE

• Special challenges for entry-level data scientists
• The Dataset Information Resource (DIR) Framework
  – Integrate and represent knowledge of datasets
  – Be able to answer questions
• A framework designed for a target audience
  – Entry level data scientists
• A framework not to store datasets but to share dataset knowledge
Approach

- Question driven knowledge
- Standards based representation
• A paper

• Popular questions labeled as “dataset” on Quora and Stack Exchange

• Interview with health informatics novices
• Information about datasets
• Analytical methods that have been applied to the datasets
• Results of existing analyses
Fig. 3. Schema of extended W3C dataset description profile.
• RDF: Resource Description Framework
• Triple: subject + predicate + object
• SPARQL: SPARQL Protocol and RDF Query Language

• Existing framework example
  – Dbpedia
  – The Neuroscience Information Framework (NIF)
Fig. 1. Proposed architecture of DIR system.
Fig. 2. Infrastructure of DIR prototype.
<table>
<thead>
<tr>
<th><strong>MIMIC</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type declaration</strong></td>
</tr>
<tr>
<td><strong>Title</strong></td>
</tr>
<tr>
<td><strong>Description</strong></td>
</tr>
<tr>
<td><strong>Publisher</strong></td>
</tr>
<tr>
<td><strong>HTML page</strong></td>
</tr>
<tr>
<td><strong>Logo</strong></td>
</tr>
<tr>
<td><strong>Alternative titles</strong></td>
</tr>
<tr>
<td><strong>Keywords</strong></td>
</tr>
<tr>
<td><strong>License</strong></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
</tr>
<tr>
<td><strong>References</strong></td>
</tr>
<tr>
<td><strong>Citations</strong></td>
</tr>
<tr>
<td><strong>Preferred prefix</strong></td>
</tr>
<tr>
<td><strong>Version linking</strong></td>
</tr>
</tbody>
</table>
**MIMIC**

<table>
<thead>
<tr>
<th>Type declaration</th>
<th>dc:types:Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Medical Information Mart for Intensive Care</td>
</tr>
<tr>
<td>Description</td>
<td>MIMIC is an openly available dataset developed by the PhysioNet.</td>
</tr>
<tr>
<td>Publisher</td>
<td>PhysioNet</td>
</tr>
<tr>
<td>HTML page</td>
<td><a href="https://mimic.physionet.org/">https://mimic.physionet.org/</a></td>
</tr>
<tr>
<td>Logo</td>
<td><a href="https://mimic.physionet.org/img/logo.png">https://mimic.physionet.org/img/logo.png</a></td>
</tr>
<tr>
<td>Alternative titles</td>
<td>MIMIC, Multiparameter Intelligent Monitoring in Intensive Care</td>
</tr>
<tr>
<td>Keywords</td>
<td>ICU patient</td>
</tr>
<tr>
<td>License</td>
<td>GNU General Public License v2.0</td>
</tr>
<tr>
<td>Rights</td>
<td>MIMIC is made available largely through the work of researchers.</td>
</tr>
<tr>
<td>References</td>
<td><a href="https://mimic.physionet.org/about/mimic/">https://mimic.physionet.org/about/mimic/</a></td>
</tr>
<tr>
<td>Citations</td>
<td>Multiparameter Intelligent Monitoring in Intensive Care</td>
</tr>
<tr>
<td>Preferred prefix</td>
<td>MIMIC</td>
</tr>
<tr>
<td>Version linking</td>
<td>MIMIC-III v1.4</td>
</tr>
</tbody>
</table>

**MIMIC-III v1.4 db**

<table>
<thead>
<tr>
<th>Type declaration</th>
<th>dc:types:Distribution, dc:types:Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Medical Information Mart for Intensive Care v1.4</td>
</tr>
<tr>
<td>Description</td>
<td>The MIMIC-III (Medical Information Mart for Intensive Care) is a database that contains a variety of data related to patient care.</td>
</tr>
<tr>
<td>Creators</td>
<td>MIT Lab for Computational Physiology</td>
</tr>
<tr>
<td>Publisher</td>
<td>PhysioNet</td>
</tr>
<tr>
<td>License</td>
<td>GNU General Public License v2.0</td>
</tr>
<tr>
<td>File format</td>
<td><a href="http://www.iana.org/assignments/media-types/application/xsv-distribution">http://www.iana.org/assignments/media-types/application/xsv-distribution</a></td>
</tr>
<tr>
<td>Date of issue</td>
<td>2016-9-2</td>
</tr>
<tr>
<td>HTML page</td>
<td><a href="https://mimic.physionet.org/about/mimic/">https://mimic.physionet.org/about/mimic/</a></td>
</tr>
<tr>
<td>Language</td>
<td><a href="http://lexvo.org/id/iso639-3/eng">http://lexvo.org/id/iso639-3/eng</a></td>
</tr>
<tr>
<td>Version identifier</td>
<td>MIMIC-III v1.4</td>
</tr>
</tbody>
</table>

**Methods in publications**

- Logistic Regression (43.11%), Support Vector Machines (9.58%), Univariate Analysis (9.58%), Linear Algorithm (5.39%), Ensemble Learning (5.39%), Genet K-Means (2.40%), Survival Analysis (2.40%), SVAR McKean (1.80%), APRIORI (1.20%), Backward Stepwise Analysis (1.20%), Kernel SVM (1.20%), McNemar’s Test (1.20%), Multinomial Logistic (0.60%), Multiscale Entropy (0.60%), and more.

**Item listing**

- admissions, callout, caregivers, chartevents, chartevents_labitems, datetimesevents, diagnoses_icd, drgcode transactions
• Healthcare Cost and Utilization Project (HCUP)
• Truven Health MarketScan (MarketScan)
• Medical Information Mart for Intensive Care (MIMIC)
Extraction of Analytical Methods

Dictionary for rule-based NER

Publications using a dataset
- PubMed Central (PMC)
  - Query & Download
  - ~3000 publications in PDF format
  - Pre-processor
  - Preprocessed publications in text format

Dictionary for rule-based NER
- MethodOntology.owl
- MethodDictionary.txt

CLAMP

NER results

Methods in publications
### TABLE I. MOST FREQUENTLY-USED METHODS FOR HCUP, MARKETSCAN, AND MIMIC

<table>
<thead>
<tr>
<th>Rank</th>
<th>Method for HCUP</th>
<th>Method for MarketScan</th>
<th>Method for MIMIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>Logistic Regression (41.33%)</td>
<td>Chi-Squared Test (34.52%)</td>
<td>Logistic Regression (43.11%)</td>
</tr>
<tr>
<td>#2</td>
<td>Chi-Squared Test (38.24%)</td>
<td>Logistic Regression (33.10%)</td>
<td>Support Vector Machine (35.33%)</td>
</tr>
<tr>
<td>#3</td>
<td>Regression Model (21.44%)</td>
<td>Cox Regression (22.82%)</td>
<td>Cox Regression (18.56%)</td>
</tr>
<tr>
<td>#4</td>
<td>T-Test (29.90%)</td>
<td>T-Test (20.97%)</td>
<td>T-Test (17.96%)</td>
</tr>
<tr>
<td>#5</td>
<td>Multivariable Logistic Regression (19.69%)</td>
<td>Regression Model (14.84%)</td>
<td>Regression Model (17.37%)</td>
</tr>
</tbody>
</table>

**Methods in publications**
- Logistic Regression (43.11%)
- Analysis (9.58%)
- Univariate Linear Model (7.19%)
- Classification (5.39%)
- Ensemble (4.79%)
- Fisher's Exact Test (2.40%)
- K-Means (2.40%)
- Rank (1.80%)
- Linear Regression (1.32%)
- Correlation Based Method (1.20%)
- Linear Kernel Algorithm (0.60%)
- Chi-Square Test (0.60%)
- Simple Linear Regression (0.60%)

**Precision:** 93.82%, **Recall:** 90.53%
• Semantic Search by SPARQL-like Queries

```
[[Category:Summary Level]]
[[Dct:isVersionOf::<q>][[Category:Version Level]]
[[Subject number::>=1000000]]</q>]]
```

• Question Answering by 18 Parameterized Question Pages

**Run query: Which datasets have more than a specific number of subjects**

Question: Which datasets have more than [1000000] (number) subjects? (For example, type in "100000".)

(This is a "RunQuery" special page: you can see the form it uses at Form:Which datasets have more than a specific number of subjects, and the template it uses at Template:Which datasets have more than a specific number of subjects.)

Run query

Question: Which datasets have more than 1000000 subjects?

Answer: HCUP, MarketScan.
Outline

• Data platform
  – Data integration
  – Data services

• Information resource
  – Semantic knowledge representation
  – Uncertainty in clinical knowledge

• Predictive models
  – Models for radiation therapy planning

• Toward an intelligent health data resource
Radiation Treatment Planning
Why Need Treatment Planning?

In practice, dose given to tumor is limited by dose tolerance to healthy tissues and organs.

Goal: Maximize dose to tumor
Minimize dose to normal tissues and critical organs
Constraint Optimization Problem

- Number of beams
- Beam Orientations
- Each Beam’s Intensity Map
- Intensity levels
Central Question

• How do you know what dose levels to prescribe for the PTV and for each Organ at Risk (OAR)?
  – Must be as high and even as possible in PTV
  – Must be as low as possible in OARs
  – Must be achievable by the treatment system

• Guidelines are not optimal
  – Based on tolerance levels (at the high end)
  – Based on population outcomes
Patient Specific Dose Prediction

• Can we learn from existing database of high quality plans to achieve this?

Prior Plans Database

Patient Features           DVHs

Patient Features

Predictive Model

Dose
Dose Model

• Dose distribution: \( D = f(D_x, G, P, M) \)
• With feature selection: \( D_f = g(D_x, G_f, P, M) \)
• Assuming: \( P = \rho(D_x, G_f) \)

\[
D_f = g(D_x, G_f, \rho(D_x, G_f), M) = h(D_x, G_f, M)
\]
• Further assuming: \( M = M_0 \)

\[
D_f = h(D_x, G_f, M_0) = h'(D_x, G_f)
\]
• Extending to outcomes:

\[
E = \theta(D_f) = \theta(f(D_x, G_f, P, M)) = \varphi(D_x, G_f, P, M)
\]
Can we predict OAR Dose (DVH) from patient’s tumor and anatomical geometry?

\[ D_f = h'(D_x, G_f) \]
Predictive Model Development

- Feature engineering and selection
- Models
  - LR, SVR, ANN, CNN
  - Ensemble model
  - Model tree
- Training data
  - Outlier detection
  - Data partitioning
Inter- and Intra-Patient Variability
Dose Volume Histogram (DVH)

- Differential: Y% of volume has exactly X% of prescribed dose
- Cumulative: Y% of volume receives at least X% of prescribed dose

Target - ideal

OAR - ideal

Target - actual

OAR - actual
Distance To Target Histogram (DTH)

Dose in color map

Distance in color map

Prostate
rectum

DVH

DTH

Dose (%)

Distance (cm)

Dose in color map

Distance in color map

Prostate
rectum

DVH

DTH

Dose (%)

Distance (cm)
PCA of DVH and DTH

(a) bladder DVH
(b) rectum DVH
(c) bladder DTH
(d) rectum DTH

DVH/DTH dispersion for bladder/rectum in the database

(e) PCA of bladder DVH
(f) PCA of rectum DVH
(g) PCA of bladder DTH
(h) PCA of rectum DTH

2~3 number of PCs explain most of the variances (>95%)
DVH reconstruction by PCA

- DVH reconstructed from 1, 2, and 3 PCs
Important Features

• Investigation using stepwise regression
  – Input features
    • 3 PCA scores of DTH
    • OAR volumes
    • PTV volumes
    • OAR/PTV volume overlaps
  – Output features
    • 3 PCA scores of DVH
Dose Prediction for Prostate IMRT

• PTV
  – Prostate
• OARs
  – Rectum
  – Bladder
• 198 high quality prior plans (Duke University Radiation Oncology)
Validation: 14 Additional Plans

- SVR & MVNLR evaluate the achievable rectum DVH
Predictive Model Development

• Feature engineering and selection

• Models
  – LR, SVR, ANN, CNN
  – Ensemble model
  – Model tree

• Training data
  – Outlier detection
  – Data partitioning
Ensemble Model

- Stepwise
- Ridge
- LASSO
- Elastic net
For bladder prediction, the proposed ensemble method predicts significantly better than stepwise (p<0.001), ridge (p<0.001), lasso (p<0.001), and elastic net (p<0.001); for rectum prediction, the proposed ensemble method predicts significantly better than ridge (p<0.001), lasso (p<0.001), elastic net (p<0.001), and stepwise (p<0.001).
For bladder prediction, the proposed ensemble method predicts significantly better than stepwise (p=0.013), ridge (p<0.001), lasso (p=0.002), and elastic net (p<0.001); for rectum prediction, the proposed ensemble method predicts significantly better than ridge (p<0.001), lasso (p<0.001), elastic net (p<0.001), and performs similarly well as stepwise (p=0.210).
For bladder prediction, the proposed ensemble method predicts significantly better than stepwise ($p<0.001$), ridge ($p<0.001$), and performs similarly well as lasso ($p=0.753$), and elastic net ($p=0.841$). For rectum prediction, the proposed ensemble method predicts significantly better than stepwise ($p<0.001$) and ridge ($p<0.001$), and performs similarly well as lasso ($p=0.365$) and elastic net ($p=0.373$).
Trade-off Decisions

Group 1 (G1)
Only left parotid spared (14)

Group 2 (G2)
Only right parotid spared (14)

Group 3 (G3)
Both parotids spared (45)

Yellow: 100%
Dark Blue: 80%
Magenta: 50%
Light Green: 30%
Model Tree

- A combination of decision tree and linear regression
The global model has similar mean WSR as the baseline model.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Model</td>
<td>-0.016</td>
<td>0.036</td>
</tr>
<tr>
<td>Single model on bi-lateral cases</td>
<td>0.016</td>
<td>0.040</td>
</tr>
<tr>
<td>Single model on single-lateral cases</td>
<td>-0.039</td>
<td>0.037</td>
</tr>
<tr>
<td>Standard model on bi-lateral cases</td>
<td>0.010</td>
<td>0.070</td>
</tr>
<tr>
<td>Model tree on bi-lateral cases</td>
<td>0.007</td>
<td>0.034</td>
</tr>
<tr>
<td>Model tree on single-lateral cases</td>
<td>-0.014</td>
<td>0.042</td>
</tr>
<tr>
<td>Model tree on all cases</td>
<td>0.001</td>
<td>0.037</td>
</tr>
</tbody>
</table>
The global model has similar mean D50 for bi-lateral cases, and slightly worse D50 for the single-lateral cases.
Predictive Model Development

• Feature engineering and selection

• Models
  – LR, SVR, ANN, CNN
  – Ensemble model
  – Model tree

• Training data
  – Outlier detection
  – Data partitioning
• (a) and (b): Geometric novelty
  • (a): Extremely large bladder
  • (b): Large overlap between PTV and bladder
• (c) and (d): Dosimetric outlier
  • (c): Insufficient bladder sparing
  • (d): Insufficient bladder sparing
Method: Geometric Novelty Identification

Patient Features

Plan Features

1. Experiment on geometric novelty identification

- 37 prostate + lymph node (LN) plans
- 37 prostate bed plans
- 37 prostate + LN plans (inlier)
- Calculate leverage of one novelty and 37 inliers
- Leverage of prostate + LN (novelty) and prostate bed (inlier) plans
- ROC analysis

Leverage for novelty

Leverage for inlier

Population Mean

Treatment Site 2

G2

G3

Treatment Site 3
Studentized residual: 

$$r_i = \frac{e_i}{s(e_i)}$$

where $$e_i = y_i - \hat{y}_i$$, $$y_i$$ is the response variable for $$i$$ and $$\hat{y}_i$$ is the regression prediction for $$i$$. $$s(e_i)$$ is the standard deviation of prediction error.
• Novelty detection: One-class SVM.
  
  1. **Isometric mapping** was applied to the feature matrix (11D) to reduce the dimension to 2D (manifold learning); it preserves the topological relation between data points.
  
  2. **One-class SVM** was trained using all data as input using radial basis function (RBF) kernel. The penalty function was preset to 5% (percentage of novelty). The value of gamma is of our choice.

\[ k(x_i, x_k) = \exp(-\gamma \cdot \|x_i - x_k\|^2) \]
We are looking for a frontier in the *middle*.
Clinical partition vs. feature partition

- **Group 1 (G1):** prostate cases (a total of 126 cases).
- **Group 2 (G2):** prostate plus LN cases (a total of 126 cases).
- **Group 3 (G3):** prostate bed cases (a total of 103 cases).
Results: DVH Prediction Accuracy

S: Single-site model
Dose Prescription Based on Big Data

• Current dose prescription knowledge comes from clinical studies and experience
  – Small samples, biased population
• Can we use “bigger” data to derive personalized dose prescription?
  – Outcomes data
  – Biological data
An Intelligent Health Data Resource

• Integrate knowledge with data
  – Annotation
  – Phenotype
  – Prediction
  – Model

• Address uncertainty in knowledge
Acknowledgements

- Wake Forest
  - Jeff Carr, MD
  - Brian Ostawsieski
  - Michael Horvath
  - Shree Unde
  - William Scott
  - Julia Robertson
  - Lindsay Marion
  - Sheila Sanders, CIO IS
  - John Boehme, VP IS
  - IS Team lead by Linda Zimmerman
Acknowledgements

• UNCC
  – Jingyi Shi

• Havard
  – Finale Doshi-Velez, PhD
  – Isaac Kohane, MD PhD

• Duke
  – Jackie Wu, PhD
  – John Kirkpatrick, MD
  – Fangfang Yin, PhD
  – Sheng Yang, PhD
  – Jiahan Zhang, PhD
  – Lulin Yuan, PhD
Thanks!

Questions?