

The Office of the National Coordinator for
Health Information Technology



Component 24:

Health Care

Data Analytics

Component Guide

Health IT Workforce Curriculum
Version 4.0/Spring 2016

This material (Comp 24) was developed by Oregon Health & Science University and The University of Texas Health Science Center at Houston, funded by the Department of Health and Human Services, Office of the National Coordinator for Health Information Technology under Award Number 90WT0001 and 90WT0006.

This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/>

Component Number: 24

Component Title:

Health Care Data Analytics

Component Description:

Data analytics has been defined as the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions. Although the focus of health IT in recent years has been on electronic health record (EHR) implementation and capturing and sharing of data, work in the future will shift to putting that data and information to use improving individual health and health care delivery. As the quantity and complexity of health care data grow through EHR data capture, genomics, and other sources, the number of facts per clinical decision will increase, requiring increasing support for those making decisions. This component presents data analytics focused on the key needs of health care.

Component Objectives:

At the completion of this component, the student will be able to:

1. Describe different types of data generated in health care
2. Communicate data analysis results
3. Describe different approaches to identity management
4. Discuss the value and approaches of machine learning and natural language processing
5. Describe the application of risk adjustment and predictive modeling
6. Select a secondary use (re-use) of clinical data and describe its goals and limitations
7. Conduct basic data analyses for a specified purpose
8. Describe the applications of data analytics in clinical and patient-oriented settings
9. Define the learning health system and describe its operations
10. Apply the principles of usability to data capture, analysis, and usage

Component Files:

Each unit within the component includes the following files:

- Lectures (voiceover PowerPoint in .mp4 format); PowerPoint slides (Microsoft PowerPoint format), lecture transcripts (Microsoft Word format); and audio files (.mp3 format) for each lecture.
- Application activities (discussion questions, assignments, or projects) with answer keys.
- Self-assessment questions with answer keys based on identified learning objectives.
- Some units may also include additional materials as noted in this document

Component Units with Objectives and Topics

Unit 1: Introduction to Health Care Data Analytics

Description:

This unit introduces the basics of working with health care data for the novice. The different types of data are explored, as well as the array of technology and tools available for working with data. Big data is defined and the special challenges related to working with data are discussed.

Objectives:

1. Categorize data into the different types.
2. Select an appropriate technology or tool for working with different data types.
3. Determine whether data fits the definition of Big Data.
4. Summarize the challenges faced when working with Big Data.
5. Define or apply common terms used in data analysis, such as sample, paired, t-test, histogram, population, correlation vs. causation, descriptive, etc.

Lectures:

- a. Overview of Health Care Data Analytics (21:12)
 1. Introduction to working with data
- b. Understanding the Data (22:37)
 1. Selection of technology tools – overview of technology tools for data
 2. Big Data – definitions and challenges

Unit 2: Working with Data

Description:

This unit is a laboratory exercise using publicly available data sets and Microsoft Excel®. The activities focus on preparing data for analysis and performing simple analyses.

Objectives:

1. Describe reasons why data need to be cleaned or modified before analysis
2. Demonstrate ability to identify and correct basic errors in data
3. Demonstrate ability to perform descriptive statistics
4. Demonstrate ability to use pivot tables
5. Demonstrate ability to create histograms
6. Describe the relationship between a database in an HIT system and data analysis tools

Lecture:

- a. Tools used in Data Analysis (15:32)
 1. Reasons for cleaning or modifying data before analysis
 2. Identifying and correcting basic errors in data
 3. Performing descriptive statistics
 4. Using pivot tables
 5. The relationship between a database in an HIT system and data analysis tools

Additional Files:

The exercise uses four Excel spreadsheets, comp24_unit2_dataset_chronic_conditions.xlsx, comp24_unit2_dataset_healthcare_associated_infections_state.xlsx, comp24_unit2_dataset_motor_vehicle_occupant_death_rate_by_age_and_gender.xlsx, and comp24_unit2_dataset_prevalence_and_trends_data_tobacco_use.xlsx, which is included with the other files for this unit.

Unit 3: Secondary Use of Clinical Data**Description:**

This unit describes the major re-uses (or secondary uses) of clinical data, mainly from the electronic health record.

Objectives:

1. Describe the secondary uses of clinical data, including the electronic health record.
2. Discuss the limitations and challenges of re-using clinical data.
3. Conduct a data re-use analysis for health care quality measurement utilizing a sample data set.

Lectures:

- a. Opportunities for Secondary Use of Clinical Data (21:17)
- b. Limitations and Challenges for Reusing Clinical Data (27:23)

Additional Files:

This unit contains both a discussion and an exercise. The exercise uses an Excel spreadsheet, comp24_unit3_exercise.xls, which is included with the other files for this unit.

Unit 4: Communicating Data Analysis Results

Description:

This topic introduces the basics of communication of data analysis results, written, visual, and oral. The choice of communication mode, given the analysis goals and results will be explored. Challenges for communicating technical data, as well as solutions will be discussed. The best use of interactive visual representations of data to amplify human cognition will be studied.

Objectives:

1. Select the best data communication mode, given the analysis goals and results.
2. Interpret data analysis results.
3. Present solutions for a variety of technical data communication challenges.
4. Prepare a simple data visualization using open-source tools.
5. Participate in the design and development of a complex data visualization.

Lectures:

- c. Data Analysis Communication and Visualization (14:43)
 1. Effectively communicate analyses and their interpretation
 2. Effectively communicate technical issue
 3. Apply concepts of data visualization to results of data analysis

Unit 5: Patient Identification

Description:

The goal of this unit is to describe how patients are identified for data analytical purposes. After describing the key attributes of patient identifiers, the unit will discuss standard identifiers and methods to link records.

Objectives:

1. Define the key attributes of patient identifiers.
2. Describe the challenges of duplicate and overlaid records.
3. Discuss the pros and cons of standard identifiers vs. linking records.
4. Describe the methods used for patient record-matching.
5. Match a sample set of patient records.
6. Discuss the benefits and limitations of de-identified data.

Lectures:

- a. Key Attributes of Patient Identifiers (14:45)
- b. Benefits and Limitations (8:48)

Unit 6: Machine Learning and Natural Language Processing

Description:

The goal of this unit is to introduce the topics of machine learning and natural language processing in the context of health care data analytics.

Objectives:

1. Describe the major tasks for which machine learning is used.
2. Compare and contrast the major approaches for machine learning.
3. Describe the major tasks for which natural language processing is used.
4. Describe the major approaches and challenges in processing clinical narratives.

Lectures:

- a. What is Machine Learning? (08:58)
- b. What is Natural Language Processing? (18:08)
- c. Clinical Natural Language Processing (NLP) Approaches, Projects, and Future Directions (16:35)

Unit 7: Data Analytics in Clinical Settings

Description:

Data analytics in clinical settings will describe the current and potential future use of analytics in health care. It will highlight examples from value-based payment systems, from patient safety, and from health care quality. Tools and approaches to maximize analytics capabilities will be described, including algorithm creation and implementation. Current investments in analytics will be described, with a focus on health system governance and operations.

Objectives:

1. Describe the current state of data analytics in clinical settings.
2. Identify key tools and approaches to improve analytics capabilities in clinical settings.
3. Describe different governance and operations strategies in analytics in clinical settings.
4. Discuss value-based payment systems and the role of data analytics in achieving their potential
5. Analyze data used in population management and value-based care systems.

Lectures:

- a. Measurement Issues: Current State (Specifications, Integration, and Examples) (21:18)
- b. Measurement Issues: Tools, Approaches, and Governance (Specifications, Integration, and Examples) (17:55)
- c. Population Management and Value-Based Care (13:42)

Additional Files:

The exercise in this unit uses an Excel spreadsheet, comp24_unit7_data_exercise.xls, along with an exercise key, comp24_unit7_data_exercise_key.xls, which are included with the other files for this unit.

Unit 8: Learning Health Systems

Description:

Learning health systems use data and information about current practices, combined with knowledge, to discover new knowledge and opportunities for improvement in health and well-being. This unit will describe learning health systems, highlight models, and help learners evaluate learning health system capabilities.

Objectives:

1. Define a learning health system.
2. Compare the current state of health systems versus the promise of learning health systems.
3. Discuss different models of learning health systems in theory and practice.
4. Evaluate the capabilities of learning health systems.
5. Characterize the data and data systems necessary for a learning health system.

Lecture:

- a. Learning Health Systems (13:25)

Unit 9: Usability

Description:

This unit focuses on the convergence of usability and data analytics. Specifically, the analytics that are required to assist in determining the usability of health information systems will be examined.

Objectives:

1. Discuss the different threats to HIS usability.
2. Determine a plausible analysis given a usability concern.

Lectures:

- a. Health Information Systems and Usability (17:13)
 1. Usability issues in Data Analytics

Unit 10: Risk Adjustment and Predictive Modeling**Description:**

Risk adjustment and predictive modeling is a common component of population management and value-based care systems. This unit defines risk adjustment, predictive modeling, and validation. The unit provides analytical concepts and examples of how they are used in health and health care.

Objectives:

1. Define risk adjustment, predictive modeling, and validations of models in health care.
2. Identify the health care and other data needed to perform risk adjustment and predictive modeling.
3. Relate risk adjustment and population segmentation to allocation of health care resources and health care redesign.
4. Discuss uses of risk adjustment and modeling in value-based models of care.
5. Delineate the use of health information technology in the creation, delivery, and evaluation of prediction models.
6. Describe ethical considerations in risk adjustment in population management.

Lectures:

- a. What Are Risk Adjustment and Predictive Modeling? (17:28)
- b. Applications of Risk Adjustment and Predictive Modeling (8:09)
- c. What is the Future of Risk Adjustment and Predictive Modeling? (18:33)

Additional Files:

The exercise in this unit uses an Excel spreadsheet, comp24_unit10_data_example.xls, which is included with the other files for this unit.

Component Authors

Component Originally Developed by:

Assigned Institutions:

Oregon Health & Science University and the University of Texas Health Science Center at Houston

Team Leads:

David Dorr, MD, MS, OHSU

Susan H. Fenton, PhD, RHIA, FAHIMA, UTH

Peter Graven, PhD, OHSU

William Hersh, MD, OHSU

Primary Contributing Authors:

Susan H. Fenton, PhD, RHIA, FAHIMA, UTH

William Hersh, MD, OHSU

Lecture Narration:

Voiceover Talent

Kristen Hagopian

Digital One, Portland, OR, <http://digone.com/>

Susan H. Fenton, PhD, RHIA, FAHIMA, UTH

Team Members:

David Dorr, MD, MS, OHSU

Susan H. Fenton, PhD, RHIA, FAHIMA, UTH

Peter Graven, PhD, OHSU

William Hersh, MD, OHSU

Kerri F. Nussbaum, MS, OHSU

Megan Crossan, BS, UTH

Kimberly Smith, PhD, MT(ASCP), UTH

Creative Commons



This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/>.

DETAILS of the CC-BY NC SA 4.0 International license:

You are free to:

Share — to copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material

Under the following conditions:

Attribution — you must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use:
Courtesy of (name of university that created the work) and the ONC Health IT program.

NonCommercial — You may not use the material for commercial purposes.
Note: Use of these materials is considered “non-commercial” for all educational institutions, for educational purposes, including tuition-based courses, continuing education courses, and fee-based courses. The selling of these materials is not permitted. Charging tuition for a course shall not be considered commercial use.

ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Notices:

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation.

No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material

To view the Legal Code of the full license, go to the CC BY NonCommercial ShareAlike 4.0 International web page (<https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode>).

Disclaimer

These materials were prepared under the sponsorship of an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Likewise, the above also applies to the Curriculum Development Centers (including Columbia University, Duke University, Johns Hopkins University, Oregon Health & Science University, University of Alabama at Birmingham, and their affiliated entities) and Workforce Training Programs (including Bellevue College, Columbia University, Johns Hopkins University, Normandale Community College, Oregon Health & Science University, University of Alabama at Birmingham, University of Texas Health Science Center at Houston, and their affiliated entities).

The information contained in the Health IT Workforce Curriculum materials is intended to be accessible to all. To help make this possible, the materials are provided in a variety of file formats. Some individuals may not find the PowerPoint slides fully accessible and should instead utilize the PDF version of the slides together with the .mp3 audio file and/or Word transcript to access the lectures. For more information, please visit the website of the ONC Workforce Development Programs at <https://www.healthit.gov/providers-professionals/workforce-development-programs> to view the full accessibility statement.