INTERNSHIP PRESENTATION

By
Dr. Deepthi Dintyala, MD, MPH,
PSM-BMHI
BACKGROUND

- Professional Background – Physician, MPH, PSM – BMHI
- Interests – Data Analytics, HEOR, Precision Medicine, Medical technologies
- Faculty advisor – Dr. Fei Yu
INTERNSHIP EXPERIENCES

Master's Internship
  - Alva10
    - Breast cancer
  - Atopic Dermatitis
  - ENABLE
    - HiDAV
    - DataAware
ALVA10 – COMPANY BACKGROUND

Who are they?
- Boston based consulting firm
- Disinterested Middle organizations

What do they do?
- Develop evidence-based scientific models to support sound economic decision making
- Develop toggle tools and software solutions to help clients in precision medicine efforts

Who are their clients?
- Payers
- Diagnostic technology companies
- Researchers
Title – Evaluating clinical and economic burden of breast cancer, and drive precision efforts in disease management

- Purpose
- Background
- Methodology
- Deliverables
- Recommendations
PURPOSE - CASE STUDY

Background – After the advent of insurance companies into precision medicine space, there is greater motivation for payers to adopt efforts and foster collaboration efforts with diagnostic companies with the purpose of reducing healthcare costs.

Annual economic burden of breast cancer in U.S is over $20 Billion.

<table>
<thead>
<tr>
<th>Who is the client?</th>
<th>What is the goal?</th>
</tr>
</thead>
</table>
| A leading payer company (Hypothetical) | • To reduce healthcare expenditure  
• To identify opportunities in breast cancer care pathway to implement precision efforts |

<table>
<thead>
<tr>
<th>What are the assets and challenges of client?</th>
<th>Solution?</th>
</tr>
</thead>
</table>
| • Large payer datasets  
• Challenges in developing models backed by scientific evidence | Alva10 Breast Cancer Precision Medicine optimization platform |

Business impact?  
Data driven decision making, backed by scientific evidence
Breast Cancer
Disease of the breast tissue characterized by uncontrolled growth

Symptoms
Asymptomatic to painful lump in breast and/or axillary lymph node

Risk factors
Age, family history, parity, body mass index

Diagnosis
Screening by mammography, biopsy, radiological examination

Management
Surgery, chemotherapy, radiotherapy, hormone replacement therapy, combination strategies

Staging
Follows AJCC staging – Stages 0 – Stage IV
Disease burden

- **Incidence**  estimated 330,000 new cases – 81% invasive

- **Risk** - 1 in 8 (13%) of all women have lifetime risk of developing Invasive Breast Cancer

- **Mortality** – estimated annual deaths 43,600 women – 2nd most common cancer death

- **Morbidity** – 3.8million women live with a previous diagnosis of breast cancer

- **Recurrence**
  - Local recurrence – 11%
  - Distant recurrence/malignancy – 12%

---

**Economic burden**

- U.S Total expenditure - $20.5 billion per annum

- Per patient Annual healthcare costs:
  - Minimum $80,000 in Stage 0
  - Maximum of $180,000 in Stage IV
METHODOLOGY

Current status
- Study guidelines
- Current care pathway

Literature review
- Gather facts and numbers from published literature
- Identify gaps and pain points

Modelling
- Develop business impact model

Validation
- Validate published data with payer data insights and vice versa

Recommendations
- Develop recommendations and a revised schema for care
BUSINESS IMPACT MODEL

• Breast cancer care was studied and modelled into three phases; Screening, Diagnosis, and Treatment

• The model can be used by payers to plug in their data, and get economic and patient impact data
FINDINGS - PAIN POINTS

- Sub-optimal screening compliance - <63% over 44 years of age
- Stage at diagnosis – less than 46% of non-screening complaint women have stage I diagnosis vs. 63% among compliant women.
- Delayed diagnosis estimated in >10% of all breast cancer diagnosis
- False positive Recall rates – upto 16%
- Annual national cost due to false positive results – $2.8 billion
- Gene expression profiling – only 54% prescription
  - GEP can show upto 40% change in chemotherapy prescription
RECOMMENDATION – OPPORTUNITIES TO REDUCE COSTS

SCREENING COMPLIANCE

- $$ - LOWER BARRIERS TO SCREENING THROUGH DX

SCREENING TECHNOLOGY

- $$ - REDUCED OVER/UNDER-SCREENING

SCREENING OUTCOMES

- $$ - REDUCED FALSE POSITIVES/RECALLS/ MISSED Dx

STAGE STRATIFICATION

- $$ - EARLIER STAGE DETECTION

TREATMENT Utilization

- $$ - EFFICIENT TREATMENT

ECONOMIC OUTCOMES

- $$ - REDUCED COSTS

Complete Snapshot of your breast cancer care model stratified by patient demographics, risk profiles, probabilities
Gap analysis in resource utilization

Confidential & Proprietary
LIMITATIONS

- Limited data available on certain necessary variables such as compliance rates, demographic data at a population level, and the prescription data at provider level – possibly error due to generalization and extrapolation
- Payer data that was used by our partner organization in analysis was a Medicare data set, meaning it only had data for ages 65 years and greater
- With limited time on the project, we couldn’t get deeper into phases 2 and 3 and the research was predominantly restricted to screening and diagnostics strategies
CHALLENGES AND LESSONS

- Introduction to economic aspects of healthcare – A significant new learning
- Vast extent of disease area – although overwhelming, helped gain a deeper understanding
- Difficulty finding data as required – required significant extrapolation and logical thinking
ENABLE stands for Extensible Network-Accessible Biomedical and Health Informatics Lifelong learning Environment

- An extension of CHIP programs

- Introduction to basic Data Science concepts
  - Open for everyone
  - Free of cost
SUMMER BOOTCAMPs

- HiDAV - Health Informatics Data Analytics and Visualization
  - **Target students** - undergraduate students from underrepresented backgrounds
  - Provides didactic introduction to programming, data analysis, data visualization, research methodologies with a mentored hand-on project.
  - **Purpose** – Introduction to opportunities and pathways in Health informatics and spur interest with help of real-time projects
  - **Perks** – Opportunity to advance career by meeting with experts and employers in the field
  - **Other** – Provide professional and leadership development by participating in journal clubs, poster presentations, and presentations.
  - **Additional** – Career counseling during and after the program
SUMMER BOOTCAMPS

- **DataAware** – Data Analytics for Teen Advancement
  - **Target students** – High school students from the triangle area
  - Provides didactic introduction to programming, data analysis, data visualization, research methodologies with a mentored hand-on project by an expert in the field or a faculty mentor
- **Purpose** – Introduce students to basic concepts of data analytics, and encourage them to gain collaborative research experience
- **Perks** – engage with experts in the field, participate in various university related activities
- **Other** – engage in various educative and fun activities organized by the teaching team
ROLES AND RESPONSIBILITIES

- Instruction
  - Synchronous classes – on biostatistics, epidemiology, research methodology
  - Data Mining and Text Mining modules
    - Didactic teaching or Data Mining modules and coding to exemplify each step

- Mentorship
  - First mentor in HiDAV, second mentor in DataAware
  - Oversee journal club
  - Provide one-on-one guidance
ROLES AND RESPONSIBILITIES...

- Participate in quality assurance of the asynchronous modules and Jupyter notebooks

- Make dynamic decisions on curriculum in collaboration with program coordinators – Shikha and Claire

- Provide example code to students to implement their analysis
  - Following is an overview of steps taken to conduct the analysis – it acted as an example from which students picked codes for their group projects

---

1.2 How to run this module

Throughout this module you will encounter both text and code cells. Please run each cell in this Notebook by clicking "Run" button in the Toolbar or by pushing Shift+Enter keys.

The cell above is an example of a code cell. You will be running numerous code cells like the one below throughout the case. Select the cell and select the run button above.

```python
In [ ]:
# This is an example of a code cell
print('Congratulations!')
print('You've run your first code cell.')
```
SAMPLE PROJECT – STROKE PREDICTION

**Purpose** – To predict stroke in a given population by given characteristics

**Dataset** – Stroke prediction dataset from Kaggle

Source type- Open source, CSV file

- **Stroke prediction dataset**

**Language** – R

**Notebook** – Kaggle
The methodology included a series of steps that educated students on the algorithms, libraries, and codes as required to perform data analysis:

Data analysis methods – Secondary data analysis of patient data. The steps included:
1. **Prepare notebook** – Identify and add libraries as appropriate to create the right environment
2. **Loading CSV file** of dataset – using the URL pathway
3. **Selecting variables** appropriate to study
4. **Data preparation** – assigning appropriate variable types
5. **Data cleaning** – cleaning up missing values, wrong values, empty fields by appropriate imputation methods such as replacing missing values with means/medians
6. **Univariate analysis** – study the nature and properties of dataset population by running statistical summarization and data visualization techniques
7. **Bivariate analysis** – correlational studies by running algorithms and finding Pearson correlational coefficients, Chi-square tests, paired t-tests, ANOVA tests; creating visualizations, and interpretation.
CHALLENGES AND LESSONS

- Notebook related – Owing to minimum experience working on Jupyter notebooks, there were some challenges initially in setting up the environment for myself.

- Language – As someone experienced with concepts of Data analysis and with working on other languages and platforms such as STATA and Python, there was an initial learning curve in getting used to applying the same principles in a different language; R.

- Challenges working in a remote environment, and handling young students.
OVERALL EXPERIENCE

- Data analysis on datasets in ENABLE with leadership and project management principles
- Interpretation of published literature and contextualizing findings in an economic and patient outcomes perspective

- A whole circle – Rich and wholesome experience
SKILLS

- Programming
- Data Analytics
- Biostatistics and epidemiology
- Health systems in United States
- Business analytics
- Data management
- Population Health Management
ACKNOWLEDGEMENTS

- Dr. Javed Mostafa – Director, CHIP programs, UNC
- Mike Fried – Director of Data Analytics, ALVA10
- Dr. Blake Long – Chief Medical Officer, ALVA10
- Dr. Fei Yu - Assistant Professor, SILS, UNC
- Shikha Yadav and Claire Paulson, ENABLE programs, CHIP, UNC
- Mika Wang – PhD student, CHIP, UNC
REFERENCES


- Jagar Jasem, Arya Amini, Rachel Rabinovitch, Virginia F. Borges, Anthony Elias, Christine M. Fisher, and Peter Kabos. 21-Gene Recurrence Score Assay As a Predictor of Adjuvant Chemotherapy Administration for Early-Stage Breast Cancer: An Analysis of Use, Therapeutic Implications, and Disparity Profile. Published online ahead of print at www.jco.org on March 21, 2016. by American Society of Clinical Oncology 0732-183X/16/3417w-1895w/$20.00 DOI: 10.1200/JCO.2015.65.0887
THANK YOU

Questions?