Data Science: Exploring the Challenges and Opportunities for Nursing Research

Barbara Cochrane, PhD, RN, FAAN

Objectives

- Compare and contrast data science with other relevant analytic terms
- Describe analytic considerations involved in data science
- Identify opportunities for oncology nursing in data science
- The Women’s Health Initiative data and specimen resource as an exemplar

Overview

- Terminology
  - Big data
  - Data science
  - Predictive analytics
  - Precision medicine
Historical Evolution

- Types of data
  - Structured
  - Unstructured

- Hardware

- Software

Current Status

- Initiatives (e.g., NIH Big Data to Knowledge – BD2K; http://datascience.nih.gov/bd2k)

- Policies
  - HITECH Act
  - Triple Aim

- Challenges to address
  - Representativeness of data
  - Ethics, privacy, security
  - Data sharing/ownership
  - Workforce development

Expertise/Resources Needed

- Data scientists
- Data science team
- Partnerships
- Standardized terminology
- Access to data over time
Sources of Big Data

- Federal (e.g., Census, CMS)
- Research
- Electronic health records
- Social websites
- Quantification of self
- Internet of Things

Analytic Approaches/Strategies

- Consolidation of data (obtaining/cleaning)
  - Common data elements
  - Harmonization
  - Capturing data streams over time
- Exploring/visualizing/modeling data
- Interpreting/discovery of patterns

Opportunities for Oncology Nursing

- Contribute an understanding of the patient, setting, community
- Provide content expertise
- Integrative/inductive skills
- Develop data science expertise by working collaboratively on big data projects
  - WHI as an exemplar
WHI Recruitment & Enrollment

- Recruitment 1* by mass mailings
- Specific enrollment goals by:
  - Study component
    - Hormone Trials (HT)
    - Dietary Modification (DM)
    - Calcium-Vitamin D (CaD)
    - Observational Study (OS)
  - Age group:
    - 10% 50-54
    - 20% 55-59
    - 45% 60-69
    - 25% 70-79

Screening, Randomization/Enrollment

- 1-4 screening visits
- Clinical trial randomization:
  - 50:50, intervention:control in HT
  - 40:60, intervention:control in DM
  - 50:50, intervention:control in CaD
  (1 year after CT randomization)
- OS Enrollment: Women ineligible for or not interested in CT

WHI Components and Primary Outcomes

<table>
<thead>
<tr>
<th>Component</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hormone Therapy Trials: Coronary Heart Disease &amp; Fractures</td>
<td>27,347</td>
</tr>
<tr>
<td>Adverse effect for Breast Cancer? (16,608 E+P; 10,739 E-Alone)</td>
<td></td>
</tr>
<tr>
<td>Calcium-Vitamin D Trial: Fractures &amp; Colorectal Cancer</td>
<td>36,282</td>
</tr>
<tr>
<td>Dietary Modification Trial: Breast &amp; Colorectal Cancers &amp;</td>
<td>48,835</td>
</tr>
<tr>
<td>Coronary Heart Disease</td>
<td></td>
</tr>
<tr>
<td>Observational Study</td>
<td>93,676</td>
</tr>
</tbody>
</table>

161,808 women total
Women’s Health Initiative (WHI)

- Goal: Answer questions re. postmenopausal women’s health (cancer, heart disease, osteoporosis-related fractures)
- Vast scientific undertaking
  - 161,808 participants recruited from 1993-1998
  - 40 clinical centers (CCs) across the U.S.
    - 10 CCs: Recruitment goal to randomize 60% participants from racial/ethnic minority groups to achieve representation in same proportion as general population
    - 3 CCs collected bone density measures

The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health & Human Services (through contracts HHSN268201100046C, HSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C)

WHI Baseline Eligibility Criteria

- General criteria for inclusion:
  - Aged 50 to 79 years
  - Postmenopausal
  - Planning to reside in the area for at least 3 years
  - Able/willing to provide written informed consent

- Additional eligibility criteria specific to each study component, related to:
  - Safety
  - Competing risk
  - Adherence/retention
WHI Timeline
  - VCCs started in September 1993
  - NCCs started in February 1995
- Follow-up for 8-12 years
  - CT: contacts every 6 months
  - OS: annual mailings; 3-year clinic visit
- Close-out: 2004-2005
- WHI Extension Studies (long-term follow-up)
  - 2005-2010
  - 2010-2015 (includes home visit)
  - 2015-2020 (includes new clinical trials)

Data Collection (1993-2005)
- Self report forms:
  - Most were scannable
  - Most could be completed by interview or with assistance
- All participant materials, including self-report forms, available in Spanish
- Data collection protocols based on study component, study year, WHI priorities, and subsampling

Baseline Data Subsamples
- Functional status
  - Grip strength, timed walk, chair stand
  - 25% of CT participants 65 years at baseline
- Cognitive status
  - HT participants 65 years at baseline
- Bone density at 3 clinical centers (CT/OS)
  - Hip, spine, whole body
  - Urine collection
Clinical Trial  \( n = 68,132 \)
Dietary Modification  \( n = 48,835 \)
Hormone Trials  \( n = 27,347 \)
Calcium and Vitamin D (from CT at ~year 1)  \( n = 36,282 \)
WHI-Memory Study (WHIMS) (from HT \( \geq 65 \) yr)  \( n = 7,479 \)
Physical Performance (25% of CTs \( \geq 65 \) yr)  \( n = 6,800 \)
Blood Subsample (6% CT overall; 8.6% HT, 4.3% DM)  \( n = 4,000 \)

Baseline Self-Report Data (CT/OS)
- Medical history
- Reproductive history
- Family history
- Personal habits
- Hormone interview
- Current medications/supplements
- Psychosocial
- Dietary assessment

How Was Diet Measured?
Baseline
- Food Frequency Questionnaire All WHI
- 4-day food records DM Trial Only 4.6% + Cases
- 24 hr recalls 1% OS annually

Year 1
- Food Frequency Questionnaire All DM Trial
- 4-day food records DM Trial Only 4.6%
- Year 3,6,9
- 24 hr recalls DM Trial Only 4.6% of participants

Year 2-Close-out
- Food Frequency Questionnaire DM Trial: 1/3 annually OS: All at year 3
- Assess change in intake over time
- Assess intervention effects

Calibrate w/other diet measures
Baseline Clinical Data

- Resting pulse, blood pressure
- Waist/hip circumference
- Height, weight
- Blood collection
- Component-specific:
  - ECG (CT)
  - Clinical breast exam, mammogram (CT)
  - Pelvic exam and Pap smear (HT)
  - Endometrial aspiration or transvaginal uterine ultrasound (E+P)

Measures: Behavioral Outcomes

<table>
<thead>
<tr>
<th>Measure</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR-QOL</td>
<td>Rand 36-item survey</td>
</tr>
<tr>
<td>Symptoms</td>
<td>34 items used by PEPI, HWS, others</td>
</tr>
<tr>
<td>Sleep disturbance</td>
<td>10 items designed for WHI</td>
</tr>
<tr>
<td>Sexual function</td>
<td>5 items measuring activity and satisfaction</td>
</tr>
<tr>
<td>Cognitive function</td>
<td>3MSE</td>
</tr>
</tbody>
</table>

Measures: Social Environment

<table>
<thead>
<tr>
<th>Measure</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social support</td>
<td>9 items from the 15-item Medical Outcomes Study</td>
</tr>
<tr>
<td>Social integration</td>
<td>12 items from Alameda Co. Study, EPSE: religiosity, clubs, pets, living arrangements</td>
</tr>
<tr>
<td>Caregiver burden</td>
<td>2 items from Cardiovascular Health Study</td>
</tr>
<tr>
<td>Social strain</td>
<td>4 items from surveys on resources</td>
</tr>
<tr>
<td>Life events</td>
<td>11 events from Alameda Co., BHAT</td>
</tr>
</tbody>
</table>
Measures: Individual Differences

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Depression</strong></td>
<td>CES-D, short version</td>
</tr>
<tr>
<td><strong>Optimism</strong></td>
<td>Life Orientation Test—Revised</td>
</tr>
<tr>
<td><strong>Hostility</strong></td>
<td>Cynicism subscale of Cook-Medley</td>
</tr>
<tr>
<td><strong>Lack of expression</strong></td>
<td>7 items from Emotional Expressiveness Questionnaire</td>
</tr>
<tr>
<td><strong>Behavioral predictors</strong></td>
<td>Physical activity, smoking, alcohol use, caffeine, work experience</td>
</tr>
</tbody>
</table>

Follow-Up Data Collection

- **Clinical Trial**:
  - Semi-annual medical history update
  - Semi-annual management and safety interview (HT/CaD)
  - Annual visit for clinical measures & study forms
- **Observational Study**
  - Annual medical history and exposure updates
  - Year 3 visit for clinical measures & study forms

Outcome Adjudication

- Potential outcome confirmed or denied using study-specific adjudication criteria
- Confirmed outcomes documented:
  - Cardiovascular
  - Cancer
  - Fracture
  - Hospitalization
  - Report of death
  - Hysterectomy (HT only)
  - Venous thromboembolic disease (HT only)
Other Outcomes Ascertained by Self-Report

- Overnight hospitalizations (adjudicated for primary WHI outcomes)
- Hysterectomy (adjudicated in E+P)
- Hypertension (requiring medication)
- Other:
  - Falls
  - Glaucoma, cataracts
  - Osteoporosis, osteoarthritis, rheumatoid arthritis
  - Colon polyps
  - Gallbladder disease
  - Lupus
  - Health screenings

Specimens Available in the WHI

- Two blood draws on “entire” cohort –
  - Baseline and Year 03: HT, DM, Ca/D
- Blood drawn during Extension 2 home visit
- Plasma, serum, buffy coat, some extracted DNA
- Over time, results of multiple special study determinations should be available
- All other bio-specimens/ surgical tissue need new protocols
  - Tumor blocks for breast cancers being obtained

Core Analytes*

<table>
<thead>
<tr>
<th>Metabolic</th>
<th>Coagulation Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose</td>
<td>Factor VII Ag (antigen)</td>
</tr>
<tr>
<td>Insulin</td>
<td>Factor VII C (clotting activity)</td>
</tr>
<tr>
<td>Lipids</td>
<td>Fibrinogen</td>
</tr>
<tr>
<td>High-density lipoprotein-2</td>
<td>Micronutrients</td>
</tr>
<tr>
<td>High-density lipoprotein-3</td>
<td>Alpha-carotene</td>
</tr>
<tr>
<td>High-density lipoprotein cholesterol</td>
<td>Alpha-tocopherol</td>
</tr>
<tr>
<td>Low-density lipoprotein cholesterol</td>
<td>Beta-carotene</td>
</tr>
<tr>
<td>Apolipoprotein(a)</td>
<td>Beta-cryptoxanthine</td>
</tr>
<tr>
<td>Total cholesterol</td>
<td>Gamma-tocopherol</td>
</tr>
<tr>
<td>Triglyceride</td>
<td>Lycopene</td>
</tr>
</tbody>
</table>

*Measured prospectively in 6% CT subsample at baseline and years 1, 3, 6; in 1% OS subsample at baseline.
Storage and Distribution of Specimens

- NHLBI pays for specimen storage and retrieval
- Records of aliquoting and freeze/thawing available
- Good documentation of amount of various specimens remaining
- If interested in DNA, extracting from buffy coat may be necessary
- Exhaustion of specimens:
  - Biggest problem is for major disease outcome cases
  - Strong opportunities for “less traveled” diseases and conditions
  - For DNA, amplification possible

Is the WHI Big Data?
Other Resources Linked to the WHI

- National Death Index
- Centers for Medicare and Medicaid Services (CMS) data
- Biospecimen analyses from core and ancillary studies
- Other constructs/codes (e.g., geocoding)

WHI - Scientific Contributions

- Primary findings from clinical trials and their contribution to science and health
  - Hormone trials
  - Dietary modification
  - Calcium / vitamin D
- Identification of risks and health outcomes
- Development/testing of new hypotheses
- Launching of new trials
How can you access WHI resources?

- Paper proposal (in collaboration w/WHI investigators)
- Ancillary study (in collaboration w/WHI investigators; requires separate, non-WHI funding)
- Core study defined by WHI investigators (funded by WHI as subcontract with Clinical Coordinating Center)
- Limited access dataset (NHLBI)

Women’s Health Initiative Website (www.whi.org)

This site provides information on:

- The WHI program and major components
- WHI summary data (CT and OS) and data dictionaries
- Ancillary studies - process for proposing and conducting
- WHI publications - completed studies and published papers; process for proposing papers
- WHI investigators for potential collaboration

Summary/Conclusion

- Challenges
  - Big Data – constantly changing/growing
  - Expertise and resources
  - Evolving policies
- Opportunities
  - Promote standardization/interoperability
  - Identify/highlight nursing contributions to health and health care
  - Personalized medicine
  - Education/mentorship
Bibliography


