Data Big and Small - The Meaning of Database Research

Adam Dzioryn, MD, PhD, FAAP
University of Rochester School of Medicine
Rochester, NY
Twitter: @adziorny
I have no relevant financial relationships with the manufacturer(s) of any commercial product(s) and/or provider(s) of commercial services discussed in this CME activity.

I **do not** intend to discuss an unapproved/investigative use of a commercial product/device in my presentation.
At the conclusion of the presentation, participants should be able to:

1. Define data science, machine learning, and their applications to pediatric critical care research
2. Identify example data science projects across the data spectrum
3. Understand the goals and challenges of “the last mile” implementation of decision support tools
‘There is an ethical imperative to capture “all of the data” — every heartbeat, every breath, to analyze for the benefit of our future patients.’

-- Randall Wetzel, “First Get the Data, Then Do the Science”, PCCM 2018
“The set of fundamental principles that support and guide the principled extraction of information and knowledge from data.”

(Sanchez-Pinto & Churpek, Chest, 2018)
How big is “Big Data”? 

Volume

Value

Veracity

Value

2,407 results
Example Applications

Comparative Effectiveness Research
• “Functional outcome after intracranial pressure monitoring for children with severe traumatic brain injury” – Bennett TD et al, JAMA Peds, 2017

Predictive Modeling
• “Multicenter development and validation of a risk stratification tool for ward patients” – Churpek MM et al, AJRCCM 2014

Clustering and Phenotyping
• “Derivation and validation of novel phenotypes of multiple organ dysfunction syndrome in critically ill children” – Sanchez-Pinto LN et al, JAMA Network Open, 2020

Natural Language Processing

Physiologic Waveform Analysis
• “Development of a Heart Rate Variability Risk Score to Predict Organ Dysfunction and Death in Critically Ill Children” – Badke C et al, PCCM, 2021
Start Local, Think Bigger

- Local QI example: Co-oximetry & VBG
- Local research example: Simultaneous Hgb
- Local validation: PEDSnet – VPS
- Multi-center federated: Pediatric CDS
- Multi-center centralized: PICU Data Collaborative
Local QI example: VBG vs CoOX

\[ y = 1.016x \]
\[ R^2 = 0.9993 \]

\( (\text{Dziorny, Fitzgerald, Weiss. Soc Crit Care Med, 2018}) \)
Local research example: *Simultaneous Hgb*

**Objective:** Measure analytic & clinical accuracy of paired Hgb results

- **N = 3,636,797**
  - Patients = 12,005
  - Visits = 18,268

- **N = 295,154**
  - Patients = 11,005
  - Visits = 16,312

- **N = 37,521**
  - PICU: 25,448
  - CICU: 12,073
  - Patients = 5,791
  - Visits = 7,562
Objective: Measure analytic & clinical accuracy of paired Hgb results

Local research example: Simultaneous Hgb

(Dziorny, Wolfe, Srinivasan; Soc Crit Care Med, 2019 & PCCM [Under Review])

N = 37,521
PICU: 25,448    CICU: 12,073
Local validation: **PEDSnet – VPS**

- Manually abstracted database of consecutive PICU admissions from over 132 hospitals
- Low clinical detail

- Electronic Health Record (EHR) abstracted data from 8 children’s hospitals
- Granular but suffers cohort identification
Local validation: *PEDSnet – VPS*

**Goal**: Find the most likely matched subjects between datasets by **feature matching**.
Local validation: PEDSnet – VPS

(Dziorny, Lindell, Bennett et al. PCCM, 2020)

Sensitivity: 98.5%
Specificity: 99.3%
PPV: 99.9%
Local validation --> Multi-site: PEDSnet – VPS

<table>
<thead>
<tr>
<th>Site ID</th>
<th>N, Total</th>
<th>N, Above Cutoff (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>11,602</td>
<td>11,089 (95.6)</td>
</tr>
<tr>
<td>B</td>
<td>22,250</td>
<td>20,720 (93.1)</td>
</tr>
<tr>
<td>C</td>
<td>15,290</td>
<td>15,194 (99.4)</td>
</tr>
<tr>
<td>D</td>
<td>15,842</td>
<td>15,595 (98.4)</td>
</tr>
<tr>
<td>E</td>
<td>10,133</td>
<td>9,136 (90.1)</td>
</tr>
<tr>
<td>F</td>
<td>7,340</td>
<td>7,122 (97.0)</td>
</tr>
<tr>
<td>All Sites</td>
<td>82,457</td>
<td>78,856 (95.6)</td>
</tr>
</tbody>
</table>

(Brennan, ..., Dziorny, Soc Crit Care Med, 2022)
Objective: Measure interruptive CDS alert burden across pediatric health systems using multiple burden metrics
Multi-center (federated): Pediatric CDS

Queries developed at a single site, validated at a second site

Queries and interface shared with all sites

Aggregation of “limited dataset” (dates) row-level data among sites

(Mandl & Kohane, Nature Biotechnology, 2015)
Multi-center (federated): Pediatric CDS

Alerts per 100 Orders

<table>
<thead>
<tr>
<th>Site</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>FD/ICU</td>
<td>0.00</td>
<td>0.13</td>
<td>0.36</td>
<td>0.06</td>
<td>0.47</td>
<td>0.17</td>
<td>0.20</td>
</tr>
<tr>
<td>ICU</td>
<td>1.45</td>
<td>13.24</td>
<td>22.52</td>
<td>2.29</td>
<td>6.99</td>
<td>50.19</td>
<td>16.11</td>
</tr>
<tr>
<td>IP - Non-ICU</td>
<td>0.73</td>
<td>1.42</td>
<td>2.49</td>
<td>0.90</td>
<td>3.31</td>
<td>2.27</td>
<td>1.85</td>
</tr>
<tr>
<td>Perioperative</td>
<td>0.02</td>
<td>0.24</td>
<td>0.58</td>
<td>0.20</td>
<td>0.48</td>
<td>0.28</td>
<td>0.30</td>
</tr>
<tr>
<td>Ambulatory</td>
<td>0.04</td>
<td>0.03</td>
<td>0.18</td>
<td>0.02</td>
<td>0.03</td>
<td>0.74</td>
<td>0.17</td>
</tr>
<tr>
<td>HOD</td>
<td>0.21</td>
<td>0.14</td>
<td>0.13</td>
<td>0.41</td>
<td>0.02</td>
<td>0.11</td>
<td>0.17</td>
</tr>
<tr>
<td>Ancillary</td>
<td>0.00</td>
<td>0.00</td>
<td>0.09</td>
<td>0.03</td>
<td>0.13</td>
<td>0.01</td>
<td>0.05</td>
</tr>
<tr>
<td>Average</td>
<td>0.35</td>
<td>2.54</td>
<td>3.76</td>
<td>0.56</td>
<td>1.63</td>
<td>7.68</td>
<td>2.76</td>
</tr>
</tbody>
</table>

Alerts per Encounter

| Site | A  | B  | C  | D  | E  | F  | Average |

Alerts per IP Day

| Site | A  | B  | C  | D  | E  | F  | Average |

Alerts per Clinician Day

http://pediatriccds.org/

Orenstein EW, Kandaswamy S, Muthu N et al. JAMIA (2021)
Multi-center (centralized): PICU Data Collaborative

Raw, multi-site EHR data

Harmonization

Quality assurance

Enhancement

Research datasets

(A Adapted with permission from Nelson Sanchez-Pinto, 2022)

<table>
<thead>
<tr>
<th>Site</th>
<th>Patients</th>
<th>Encounters</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>16,415</td>
<td>22,540</td>
</tr>
<tr>
<td>B</td>
<td>12,970</td>
<td>17,810</td>
</tr>
<tr>
<td>C</td>
<td>19,624</td>
<td>30,681</td>
</tr>
<tr>
<td>D</td>
<td>10,365</td>
<td>16,233</td>
</tr>
<tr>
<td>E</td>
<td>12,972</td>
<td>18,120</td>
</tr>
<tr>
<td>F</td>
<td>8,243</td>
<td>11,265</td>
</tr>
<tr>
<td>G</td>
<td>16,979</td>
<td>21,857</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td><strong>97,568</strong></td>
<td><strong>138,506</strong></td>
</tr>
</tbody>
</table>

https://pedsdata.org/
What is Clinical Informatics?

(William Hersh, 2011)

(Charles Friedman, JAMIA, 2009)
Clinical Decision Support

“Knowledge and person-specific information, intelligently filtered or presented at appropriate times, to enhance health and health care"

-- Osheroff, 2007

Right information
Right people
Right channel
Right format
Right time
The application of ML/Al to healthcare has failed to produce meaningful changes to patient care.

Single-center ML models are rarely extended to multi-center validation or shared implementation.

(Chen & Asch, NEJM 2017)
Not All Implementations Are Effective

Computerised clinical decision support systems and absolute improvements in care: meta-analysis of controlled clinical trials

Janice L Kwan,^1,2 Lisha Lo,^3 Jacob Ferguson,^4 Hanna Goldberg,^4 Juan Pablo Diaz-Martinez,^5 George Tomlinson,^5 Jeremy M Grimshaw,^6 Kaveh G Shojania^2,3,7

Unexpected Increased Mortality After Implementation of Sold Computerized Physician Order Entry Alerts

Yong Y. Han, MD^‡; Joseph A. Carcillo, MD^‡‡; Shekhar T. Venkatraman, MD^†; Robert S.B. Clark, MD^‡‡; R. Scott Watson, MD, MPH^‡‡‡; Trung C. Nguyen, MD and Richard A. Orr, MD^‡‡

Unintended Effects of a Computerized Physician Order Entry Nearly Hard-Stop Alert to Prevent a Drug Interaction

A Randomized Controlled Trial

Brian L. Strom, MD, MPH; Rita Schinnar, MPA; Faten Aberra, MD, MSCE; Warren Bilker, PhD; Sean Hennessy, PharmD, PhD; Charles E. Leonard, PharmD; Eric Pifer, MD
“Discovery is a Constant Process”

Stages

1. Knowledge Capture
   - Observation notes
   - Collected artifacts
   - Interview & focus group transcripts

2. Cognitive Analysis & Representation
   - Work domain demands, barriers & facilitators
   - Practitioner skills & strategies
   - Cognitive support requirements
   - Leverage points

3. Conceptual Design
   - Use case development
   - High level visualization design
   - User feedback

4. Prototype Development
   - Detailed visualizations
   - Refined cognitive support requirements
   - User feedback

5. Work Oriented Evaluation
   - User assessment of usability & usefulness
   - Objective performance
   - Recommendations for improvement
   - Forward-looking opportunities

(Hettinger, Roth, Bisantz; J Biomed Inform, 2017)
This is your machine learning system?

Yup! You pour the data into this big pile of linear algebra, then collect the answers on the other side.

What if the answers are wrong?

Just stir the pile until they start looking right.
Training & Resources

• Online training programs

• Degree-granting programs (e.g. Certificates, Master’s Degrees)

• Medical fellowship programs (e.g. Clinical Informatics)

• Informal (experiential) learning
Acknowledgements

University of Rochester
• Pediatrics (Pat Brophy & Jill Cholette)
• Biomedical Engineering (Diane Dalecki)
• UR Health Lab (David Mitton)

Children’s Hospital of Philadelphia
• DBHI (Bryan Wolfe)
  • Charles Bailey, Bob Grundmeier, Tony Luberti
• Critical Care & Anesthesiology (Bob Berg)
  • Akira Nishisaki, Brad Lindell, Heather Wolfe, Vijay Srinivasan, Julie Fitzgerald

KL2 Research Mentors
• Martin Zand (URMC), Jim McMahon (URMC), Naveen Muthu (Emory), Nelson Sanchez-Pinto (Lurie)

PALISI
• Mela Bambea, SC Chair

AAP SOCC
• Ellie Hirshberg

Peds CDS Collaborative
• Evan Orenstein, Eric Kirkendall

PICU Data Collaborative
• Randall Wetzel, Reid Farris, Nelson Sanchez-Pinto

Funding Agencies
• American Medical Association
• Assoc of Ped Program Directors
• CHOP Foerderer Program
• URMC CTSI (KL2)