Building a Perioperative Data Warehouse

Matthew A Levin MD
Assistant Professor
Department of Anesthesiology
Division of Cardiothoracic Anesthesia
Icahn School of Medicine at Mount Sinai
Disclosures

• None
Outline

• Overview and general considerations
• Sinai experience
• Vanderbilt experience
• Summary and wrapup
Define: Data Warehouse

data·ta ware·house

noun COMPUTING

noun: data warehouse; plural noun: data warehouses.

1 a large store of data accumulated from a wide range of sources within a company and used to guide management decisions.

https://www.google.com/search?num=20&site=&source=hp&q=define%3A+data+warehouse&oq=define%3A
Define: Perioperative data warehouse

1 a large store of data accumulated from a wide range of sources within a health care organization and used to:

- support/enable PQRS reporting
- support/enable departmental or hospital PI/QA
- support/enable research
- archive data from legacy AIMS systems
- support Perioperative Surgical Home?
Tools for building a periop DW

- People
- Support
- Hardware
- Software
Challenges

• Getting data out of AIMS
• Integrating hospital data
• Integrating non-hospital data
• Validation/cleaning/phenotyping
• Ongoing HW/SW support
• Who will pay for all of this??
**Dimensional vs Normalized Approach**

- **Dimensional aka “star schema”**
  - single “fact” table, many reference “dimension” tables

- **Normalized**
  - like a regular relational database

- **Pluses and Minuses to each**
Is my AIMS a periop DW?

• AIMS is a tool for intraop charting
• AIMS usually only includes limited periop data
• AIMS reporting may be limited to pre-built reports
• AIMS cannot archive from other AIMS systems (e.g., legacy)
DW as an archive

• What to do with old data when migrating to a new AIMS?
  • Keep an old machine to access old data?
  • Declare data bankruptcy and start anew?
  • Build a warehouse and populate with old data?

• Many issues with schema mapping, patient identifier mapping, etc.

• Build a feed that adds new data from new AIMS
Can Epic be a Periop DW?

• Not really
• Hierarchal db, can’t query across patients/encounters
• Restricted access
  • need to go to Verona, WI
    (pop. 10,632 including cows)
Can Epic be a Periop DW?

- Yes, maybe, sortof
- Reporting workbench - runs directly off Cache
- Clarity - Relational reporting extract from Cache
- Cogito - Star schema extract from Clarity

---

Epic Clarity Is Not a Data Warehouse

Posted on July 27, 2012 by edgewatertech

It’s not even the reporting tool for which your clinicians have been asking!

- [https://uccsc.ucsf.edu/session/cogito-data-warehouse-technical-impressions](https://uccsc.ucsf.edu/session/cogito-data-warehouse-technical-impressions)
Getting data out of Epic

• Reporting (via Clarity)
• Realtime - email alerts fired by BPA’s
• Realtime - Custom feed
  • Web Services (Epic as provider, SOAP protocol)
  • External calls (Epic as consumer)
  • Message Passing (bi-directional SOA model, XML/XSD schemas)
• $$$$
Epic data - Caveats

- No real-time intraoperative vitals
- Will have to collect vitals directly*

- No realtime MAR (Medication Administration Reconciliation) data
Can a Cloud-based AIMS be a Periop DW

- There are now several “cloud” based AIMS
- They all store case data on remote servers
- They all provide facilities for reporting (dashboards, analytics) and for sending PQRS data to AQI
- Unclear how they integrate periop data
Can AQI be a periop DW?

• Qualified yes

• AQI is a Patient Safety Organization (PSO) and a Qualified Clinical Data Registry (QCDR)

• They’ve built a robust data mining infrastructure (see poster #29)

• Whatever data you give them, they can mine for you - if you ask

• But, not realtime, and de-identified
  • No dashboard, no way to link back to individual cases in your data warehouse
The Sinai Experience - ORDW
ORDW

q30 sec; 1-2 min latency

Physiologic data

q15 sec

event- and data-driven notifications

bi-weekly

monthly

manually*

MSDW

Reports

Reporting Server

Billing Server

Social Security Death Index

RxNORM

Research

*not actively updated

Saturday, January 10, 15
ORDW external data

• Social Security death index via automated loader
  • kept in separate db to limit access

• Other data loaded manually
  • CPT codes, HCUP CCS, personnel lists...
Hardware & software

- 16 core machine
- 144 GB RAM, 3 TB disk
- Ubuntu 14.04, MySQL 5.6 (Percona Server)
Schema

- Main tables mimic CompuRecord schema
- Separate db for Periop data, quasi-relational
  - Already abstracted so won’t have to change
- Separate databases for ssn, summary table, realtime vs historical data
- Continuously updated summary table
Perioperative data schema

- Custom feed, designed with and implemented by MSDW team
- Separate database, separate schema, quasi-normalized
  - 2 week lag (matches lag in MSDW Epic feed)
- Rich set of perioperative data
Summary table

- Key for DW use
- Wide (300 column) table which aggregates data from many base CR tables
- Many calculated fields and flags
- Table definition abstracted
  - programmatic schema update and table regeneration (perl script)
- Primary initial query table for:
  - Research
  - Billing
  - Reporting and compliance
<table>
<thead>
<tr>
<th>Columns</th>
<th>Type</th>
<th>Description/Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>birth_date</td>
<td>date</td>
<td>PHI</td>
</tr>
<tr>
<td>service_date</td>
<td>date</td>
<td>PHI; Date of service. If doesn’t exist, use date of anesthesia_start</td>
</tr>
<tr>
<td>service_week</td>
<td>smallint</td>
<td>Week of service date</td>
</tr>
<tr>
<td>service_year</td>
<td>smallint</td>
<td>Year of service date</td>
</tr>
<tr>
<td>age</td>
<td>smallint</td>
<td>datediff of birth_date and service_date</td>
</tr>
<tr>
<td>is_age_pediatric</td>
<td>bool</td>
<td>True if age &lt; 19</td>
</tr>
<tr>
<td>gender</td>
<td></td>
<td>Can be Male, Female, Undetermined</td>
</tr>
<tr>
<td>height_cm</td>
<td>double</td>
<td>Height in centimeters</td>
</tr>
<tr>
<td>weight_kg</td>
<td>double</td>
<td>Weight in kilograms</td>
</tr>
<tr>
<td>bmi</td>
<td>double</td>
<td>BMI based on height and weight</td>
</tr>
<tr>
<td>bsa</td>
<td>double</td>
<td>BSA using Mosteller formula: (cm*kg/3600)</td>
</tr>
<tr>
<td>ibw</td>
<td>double</td>
<td>IBW using Devine for males (50 + 2.3 * (in - 60)) or Robinson for females (45.5 + 2.3 * (in-60))</td>
</tr>
<tr>
<td>patient_class</td>
<td></td>
<td>Combo ItemID=43; Restricted to Inpatient, DAS, Ambulatory in decreasing order of precedence; Office if location is anesfluoro</td>
</tr>
<tr>
<td>asa_status</td>
<td></td>
<td>ASA Physical Status (1 through 6)</td>
</tr>
<tr>
<td>is_pregnant</td>
<td>bool</td>
<td>6402=N/A, 6403=Not pregnant, 6404=Pregnant, 0=Other</td>
</tr>
<tr>
<td>is_emergency</td>
<td>bool</td>
<td>True if emergency flag set</td>
</tr>
<tr>
<td>mallampati</td>
<td></td>
<td>Mallampati score</td>
</tr>
<tr>
<td>primary_cpt</td>
<td></td>
<td>The &quot;primary&quot; CPT; usually the CPT with the highest number of anesthesia RVUs</td>
</tr>
<tr>
<td>proc_cpt_list</td>
<td></td>
<td>list of all CPTs, comma separated; ordered by CPT; excludes 99100</td>
</tr>
<tr>
<td>cpt1</td>
<td></td>
<td>First CPT listed; exclude 99100</td>
</tr>
<tr>
<td>cpt2</td>
<td></td>
<td>Second CPT</td>
</tr>
<tr>
<td>cpt3</td>
<td></td>
<td>Third CPT</td>
</tr>
</tbody>
</table>
Size

• ~550,000 cases
  • 300-400 new cases added daily, from 4 hospitals
  • no periop data from 3 new sites

• ~770 GB

• Vitalsigns table: ~600 GB, 5+ billion rows
## OR Occupancy for 12/29/2014 through 1/3/2015

<table>
<thead>
<tr>
<th>Time</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Friday</th>
<th>Wednesday</th>
<th>Thursday</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:00</td>
<td>24</td>
<td>29</td>
<td>16</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>08:30</td>
<td>27</td>
<td>31</td>
<td>20</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>09:00</td>
<td>26</td>
<td>31</td>
<td>19</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>09:30</td>
<td>27</td>
<td>31</td>
<td>19</td>
<td>19</td>
<td>2</td>
</tr>
<tr>
<td>10:00</td>
<td>26</td>
<td>31</td>
<td>18</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>10:30</td>
<td>27</td>
<td>29</td>
<td>17</td>
<td>21</td>
<td>1</td>
</tr>
<tr>
<td>11:00</td>
<td>23</td>
<td>29</td>
<td>17</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>11:30</td>
<td>24</td>
<td>26</td>
<td>17</td>
<td>18</td>
<td>2</td>
</tr>
<tr>
<td>12:00</td>
<td>24</td>
<td>26</td>
<td>19</td>
<td>20</td>
<td>2</td>
</tr>
</tbody>
</table>
ORDW and Epic

• MSDW acts as intermediary
  • Daily export from Cache to Clarity
  • From Clarity to MSDW every 2 weeks
    - feed designed and maintained by MSDW team in collaboration with Epic reporting team

• From MSDW to us every 2 weeks

• From us to Epic q15 min

• Result: we get what we want and don’t really have to deal with Epic
Anesthesia Summary Report

Performed Procedure

- CORONARY ARTERY BYPASS, USING VENOUS GRAFT(S) AND ARTERIAL GRAFT(S); TWO VENOUS GRAFTS; CORONARY ARTERY BYPASS, USING ARTERIAL GRAFT; SINGLE ARTERIAL GRAFT; LV ANEURYSMECTOMY WITH THROMBUS EVACUATION AND GORETEX PATCH; LIGATION OF LA APPENDAGE

Anesthetic Technique Primary: General
Anesthetic Technique Secondary:
Position: Supine

Anesthesia Start Date: 2014-12-29 07:37:00
Procedure Start Date: 2014-12-29 08:48:00
Procedure End Date: 2014-12-29 13:10:00
Anesthesia End Date: 2014-12-29 13:45:00
Anesthesia Care Team: ACUNAJ
ASA Status: 4
Emergency: No
OP Room: anesngpor04
Prophylaxis: Antibiotic(s) given as per protocol
ORDW - what we don’t have

- Full text of notes
  - 90% of note content is redundant and useless anyway
- Imaging reports, or image data
- Allergies (Epic reported)
- All Epic flowsheet data
  - For example, vent data for ICU patients
    - not contained in any existing Clarity reports
    - would have to have Epic reporting team build a new report
      - low priority for reporting team
    - MSDW then has to modify their Clarity feed (and test/validate)
    - MSDW then has to modify our anesthesia feed (and test/validate)
The Vanderbilt* experience

• SQL Server based
• Batch loads of both AIMS and external periop data
  • AIMS data loaded from backup files q8h
  • Periop data loaded from Enterprise Data Warehouse daily
• Cleaning/modeling using T-SQL MERGE statements
• Phenotyping - risk score calculation, etc.
• Reporting via Tableau software
  • Custom one-off queries for research requests
• 680,000+ cases
• 5 billion+ vitalsigns

*Thanks to Jon Wanderer MD for this information
ORDW - future directions

• Enable distributed queries for improved performance
• Integration with other local data warehouses - genomics!
• Periop data from other Mount Sinai health system sites
• More summarization and phenotyping
• Visualization
Summary

- Building a periop DW is a major effort
- Reward is a rich dataset that can be used to answer deep research questions
- Most groups will not need a periop DW or want to build one. Built in reporting facilities will be good enough.
- Epic is not a data warehouse
Thank you