(Big) Data in Healthcare – A Clinician’s Perspective

International Symposium on Data-Driven Healthcare, 29 Aug 2017

A/Prof Low Cheng Ooi
Chief Medical Informatics Officer, MOH
Chief Clinical Informatics Officer, IHiS
What is Big Data?

“Big data generates a mixture of great promise, fantastic and delusion claims, impressive misunderstanding, and several early examples that are beginning to shape our understanding of it”

(Dan Ariely, Duke University Professor)

(John Glaser, Cener)
How did data become BIG data?

<table>
<thead>
<tr>
<th>Data Unit</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kilobyte (kB)</td>
<td>$10^3$</td>
</tr>
<tr>
<td>Megabyte (MB)</td>
<td>$10^6$</td>
</tr>
<tr>
<td>Gigabyte (GB)</td>
<td>$10^9$</td>
</tr>
<tr>
<td>Terabyte (TB)</td>
<td>$10^{12}$</td>
</tr>
<tr>
<td>Exabyte (EB)</td>
<td>$10^{18}$</td>
</tr>
<tr>
<td>Zettabyte (ZB)</td>
<td>$10^{21}$</td>
</tr>
<tr>
<td>Yottabyte (YB)</td>
<td>$10^{24}$</td>
</tr>
</tbody>
</table>

- 2012 worldwide digital healthcare data estimated to be around 500 petabytes and expected to reach 25,000 PB by 2020
- With the coming of sensor data from IoT it is widely expected that the healthcare data will reach Brontobyte in near future

- This will be our digital universe tomorrow
- This is our digital universe today
- 250 trillion of DVDs
- 1.3ZB of network traffic by 2016
- 1EB of data is created on internet each day
- 500TB of data is generated every day on Facebook data bases

* Source: https://twitter.com/hashtag/yottabyte
How Things Have Changed..
What is Big Data?

- Data sets whose size are beyond the ability of typical SQL database software to analyze.

- Characterized by 3 Vs: Volume, Velocity and Variety *(Gartner analyst Doug Laney, 2001)*

- Some include a 4th V: Veracity
  - Data uncertainty and reliability
  - This obscures the information in data *(IBM data scientists)*

- Some also include a 5th V: Value
  - Ability to turn data into value

* Source: [http://www.genalice.com/wp-content/uploads/2013/05/3Vs-of-Big-Data](http://www.genalice.com/wp-content/uploads/2013/05/3Vs-of-Big-Data)
"Big Data" in Healthcare Sector- Digitization of Healthcare

KEY STATISTICS

- Patients can generate different types of data from a plethora of sources in varied healthcare settings.
- Massive volume of information are generated with the digitization and adoption of EMR & Health IT technologies.

**Health Factors**

<table>
<thead>
<tr>
<th>Source Type</th>
<th>Percentage</th>
<th>Data Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exogeneous Sources</td>
<td>60%</td>
<td>1100 Terabytes</td>
</tr>
<tr>
<td>Genomics Sources</td>
<td>30%</td>
<td>6 Terabytes</td>
</tr>
<tr>
<td>Clinical Sources</td>
<td>10%</td>
<td>~1 Terabytes</td>
</tr>
</tbody>
</table>

**Data Generated**

- Per life time
  - Volume, Velocity, Variety, Veracity

**Statistics**

- **Today,** 80% of data is unstructured; with forms such as images, video, and email.
- **3GB** Human genome
- **16** Public Hospitals in Singapore; of which there are **8** Government restructured hospitals & **8** specialty centres.
- **120 MB** Mammogram
- **150 MB** MRI
- **250 MB** CT scan
- **20 MB** X-ray
- **0.55M** Total 2016 admissions in Singapore registered hospitals, according to Ministry of Health.
- **12.4%** Percentage of Singapore’s population aged 65 and above as of 2016.
- **At least 1000** more home palliative places by 2020.
- **At least 1000** more home palliative places by 2020.

**Exogeneous Sources**

- Genomics Sources
- Clinical Sources


© Copyright 2017 - Integrated Health Information Systems Pte Ltd
How did Healthcare generate data?

One Patient One Number
- NRIC/FIN
- Passport number
- Social security
- Demographic data
- Patient Master Index

One Encounter One Bill
- Registration/Encounter Number
- Bill number
- Procedure numbers
- Consumables, medications used
  - Itemised for billing
How did Healthcare generate data?

One Patient One Record
Patient Information

All on paper (no data as yet)

Laboratory results to laboratory systems and lab data and billing

Medications to Pharmacy systems

- Started as dispensing information, inventory control and billing, linked to orders and administration

Operating Theatre systems

- Scheduling, operation and procedure notes, operation tables - Billing
How did Healthcare generate data?

Patient Information

Hospital Discharge Summary
➢ Paper to digital (stored electronically, shared on print)

Radiology Films to PACS both high quality still images as well as video images

Electronic Orders
➢ Laboratory / radiology orders
➢ Medications orders
➢ Others
How did Healthcare generate data?

**Patient Information**

**Electronic Documentation**
- "electronic medical record"
- Free text vs coded; unstructured vs structured

**Electronic Referral**

**Acute care records to Patient Journey**
- National Electronic Health Records

**Disease Registries / Research databases**
How did Healthcare generate data?

Patient Information to Population Health

Consumables/Devices to Device Registries

Medication to CMIS (Critical Medical Information System) with Drug Allergy and Adverse Event data and Drug Registries

Pharmacy / Materials and Management systems to Group Procurement Office Database
Healthcare Big Data build with Trust

360° Profile of Population for targeted & personalised intervention

- National Registry of Diseases
- Lifestyle & Social Network Data
- Financial Data
- Geospatial Data
- Research/Edu Data
- Genome
- Shared IT Infra for Precision Medication (Aggregated Data)
- Shared IT Infra for collaboration (Aggregated Data)
- Household Means Test Social Programmes with Consent
- Various Programmes with Consent

Health Sector Data Standards Based

Shared IT Infra for Location Based Services (Aggregated Data)
How will Healthcare generate data in Future

*Sensing & Actuation* technologies in Healthcare will also be a big part of IR4.0

Robotic Nurse Assistant

Nanobots in Blood

"Head Transplants and Mind Transfer"

Robotic "Flight Simulator" Surgery

Electric assistance for Stroke recovery

Facebook team working on ways to read users’ thoughts

Apple Entering Glucose Monitoring Market With Apple Watch (AAPL)

Machine learning will replace human radiologists, pathologists, maybe soon
**Need for “Big Data” in Healthcare**

**Population Health**
- Increasing importance of population health management due to limited resources, an ageing population and an increasing prevalence of chronic diseases.
- “Big Data” could be utilized for population profiling, risk identification, resource allocation and forecasting purposes.

**Preventive Care**
- “Big data” could be used to identify risk factors in patients, and suggest lifestyle changes to prevent disease, monitoring patient’s progress and adjusting as necessary.

**Increase information available to doctors - reduce misdiagnoses**
- Leverage on collective knowledge of medical professionals.
- Compare and contrast effectiveness of different treatment models for patients with similar characteristics to prescribe the most effective treatment.
Admissions Prevention Predictive Model

Predictive System that identify discharge patient that likely to have multiple readmission was launch in April 2017

TRANSFORMING OUR HEALTHCARE SYSTEM TO MEET THE NEEDS OF SINGAPOREANS

To enable Singaporeans to receive appropriate care in the community and closer to home

HOSPITAL CARE

- Patient admitted to Hospital
- Risk prediction generated
- Based on past medical history predict the risk of multiple readmission
- Care Team assess High Risk Patient
- Clinical Team prepare the different tiered care plan for patient
- Provide intervention to suitable patient
- Patient Timely discharged & remain in community

COMMUNITY CARE

- Unplanned admission into the hospital
- Based on past medical history predict the risk of multiple readmission
- Clinical Team prepare the different tiered care plan for patient
- Provide intervention to suitable patient
- Patient Timely discharged & remain in community
- Community team review & report to Regional Health Systems
Facebook AI Creates Its Own Language

It comes after other AI developed in a similar pattern elsewhere, like when Google Translate invented its own language. The so-called ‘neutral network’ started translating phrases easier using it’s own language, according to the New Scientist.
Asking the right Question and hopefully getting the right answers
1st Dimension of Asking Questions

**FIRST DIMENSION**

**Specific and Exact Information**
- What is the latest potassium result?
- What and when was the last surgical procedure?

**Data needs little Processing**
- User interface to get to information
- +/- search engine
- Better if coded
- Setting of rules, parameters allow for simple alerts and reminders
- Not big data...
2\textsuperscript{nd} Dimensions of Asking Questions

\textbf{SECOND DIMENSION}

\begin{itemize}
  \item Specific and requires linked data (from other databases)
  \item Data needs link and needs little transformation
  \begin{itemize}
    \item Specialist checking on medication/procedures from another institution or general practitioner
    \item Administration asking for medication cost utilisation by physician
    \item Data linked at patient or disease level
    \item Structured, preferably coded or items mapped
    \item Software can call on more than 1 database and extraction
    \item Data mart or warehouse required with simple analytic engine
    \item Nearly big data....
  \end{itemize}
\end{itemize}
### 3rd Dimension of Asking Questions

#### THIRD DIMENSION

<table>
<thead>
<tr>
<th>Some level of uncertainty that requires Human judgement</th>
<th>Data needs a lot more processing, linking and transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Predictive model that provide Clinicians /care providers patient risk profile for targeted action</td>
<td>• Data non structured or structured, multiple formats, many databases, natural language processing, data modelling</td>
</tr>
<tr>
<td>• Population health management that support care coordination and prevention programmes through disease progression and cohort identification models</td>
<td>• Complex computing, more sophisticated tools</td>
</tr>
<tr>
<td></td>
<td>• Large data warehouse or linked multiple warehouses</td>
</tr>
<tr>
<td></td>
<td>• Advanced clinical decision support and alerts and advanced clinical risk profiling</td>
</tr>
</tbody>
</table>
Data needs a lot more processing, linking and transformation

Non-specific but do know the targeted business outcome

FOURTH DIMENSION

- Scientific algorithm coupled with high performance computing to detect patterns from the Big data
- For example, looking for unknown drivers for rising health cost

- Collaborative interactive analytics between Clinician, Administrator, Data scientist, Computing engineer and Business Analyst
- Powerful & efficient Big data processing engine
- A variety of techniques from statistic, predictive modeling, machine learning, and data mining
# DIFFERENT TYPES OF QUESTIONS

<table>
<thead>
<tr>
<th>1&lt;sup&gt;st&lt;/sup&gt; Dimension</th>
<th>2&lt;sup&gt;nd&lt;/sup&gt; Dimension</th>
<th>3&lt;sup&gt;rd&lt;/sup&gt; Dimension</th>
<th>4&lt;sup&gt;th&lt;/sup&gt; Dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple Query: Specific and exact information</td>
<td>Connected Qn: Specific and requires linked data</td>
<td>Complex Qn: Some level of uncertainty that requires human judgement</td>
<td>Explorative: Non-specific but do know targeted outcome</td>
</tr>
<tr>
<td>E.g. How many patients went for HBA1C tests this month?</td>
<td>E.g. What is the healthcare cost utilisation related to diabetes?</td>
<td>E.g. What is the survival rate of poorly controlled diabetes cohort in 2017?</td>
<td>E.g. What are the unknown key drivers of healthcare costs for diabetes cohort?</td>
</tr>
<tr>
<td>Data needs little processing</td>
<td>Data needs linkages and little transformation</td>
<td>Data needs a lot more processing, linking and transformation</td>
<td>Noisy dirty data; needs robust processing</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structured Data</td>
</tr>
<tr>
<td>Query &amp; Filtering</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data Warehouse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced Data</td>
</tr>
<tr>
<td>Advanced Query</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Advanced Queries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query &amp; Filtering</td>
</tr>
<tr>
<td>Advanced Query</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Big Data (Volume, Velocity, Variety)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parallel and In-Memory Processing</td>
</tr>
<tr>
<td>Statistical Modelling</td>
</tr>
</tbody>
</table>

| Source: Ayasdi |
A View on the Future

Precision

Personalized Treatment

Profiling

Diagnosis / Healthcare program

Population Segmentation

Timely Intervention

Targeted Intervention

Precision

Prescription

Prevention

Predictive

Personalized Treatment
… from A Physician’s Perspective

“ I am a doctor not a database …”

Star Trek Voyager “Doctor”

1995
End of Presentation
Thank you
eHealth Systems in Singapore in early 2000

**Multiple EMR Systems Multiple Lab & Radiology Systems**
- Public healthcare clusters
  - NHG and SingHealth hospitals, polyclinics and specialist clinics
- Ministry of Defence (MINDEF)

**Minimal EMR Systems**
- Private GPs (over 400 out of more than 2,000 have a CMS)
- Community Hospitals and other ILTC providers
… from A Physician’s Perspective

“ I am a doctor not a database …”

Star Trek Voyager “Doctor”

1995