Preparing Data for Analysis

Sankalp KHANNA

CSIRO Australian e-Health Research Centre, Brisbane, Australia
Improving public hospital performance through efficiency improvements

Creating an evidence base to support policy and decision making

Patient Flow @ AEHRC

1. Linking ambulance, ED and admissions data
2. Disease surveillance
3. ED Length of stay performance
4. Better bed demand prediction
5. Patient flow visualisation
6. Patient flow and hospital occupancy
7. Bed configuration
8. Adverse event analysis
9. Early discharge strategies
10. Readmission prediction (frequent-flyers)

http://www.csiro.au/patientflow
A Typical Data Analysis Project

Plan and Access Data → Link → Transform → Cleanse → Profile → Analyse

Preparing Data for Analysis
Best Practice in Data Prep

• Check data formats
• Verify data types
• Graph your data
• Verify data accuracy
• Identify outliers
• Deal with missing values
• Check your assumptions about how the data is distributed
• Back up and document everything you do
Pre-Prep – Planning and Accessing Data

• Collaborative planning
• Identify the data you need – all of it
• Make sure it is all there
• Obtain regulatory approvals
• Verify the sample extract
• Obtain the data
Where Best Practice fails Health Data Prep

• Lag between data collection and processing/availability

• Data quality and consistency issues

• Poor “link-ability”

• Insufficient domain knowledge
Avoiding common roadblocks

Consider:
- Nuances of data
- Domain terminology and KPIs
- Impact and path to impact
- Clinical engagement

If you will do a lot of it, consider:
- Performance
- Scalability
- Reusability
Thank you

Australian e-Health Research Centre, CSIRO

Sankalp Khanna
Research Scientist

+61 7 3253 3629
Sankalp.Khanna@csiro.au
www.aehrc.com