Designing and implementing measurement suites: screening, assessment, outcomes evaluation and service benchmarking

Kathy Eagar
University of Wollongong, keagar@uow.edu.au
Designing and implementing measurement suites: screening, assessment, outcomes evaluation and service benchmarking

Abstract
Workshop overview

- A common language for the workshop - what is an outcome?
- Some starting points based on what we've learned
- PCOC and AROC as examples
- Performance measurement using routinely collected data
- Interpreting routinely collected data
- Open discussion

Keywords
implementing, measurement, suites, service, designing, screening, benchmarking, assessment, outcomes, evaluation

Publication Details
Designing and Implementing Measurement Suites: Screening, Assessment, Outcomes Evaluation and Service Benchmarking

Professor Kathy Eagar
Director, Centre for Health Service Development

Shaping Up in Health: How does Australia become the world’s best?
7-9 October 2009, Hobart
Workshop overview

- A common language for the workshop - what is an outcome?
- Some starting points based on what we’ve learned
- PCOC and AROC as examples
- Performance measurement using routinely collected data
- Interpreting routinely collected data
- Open discussion
How do you assess an ‘outcome’?
Whose assessment counts?

**The Person**
- Survive
- Have friends
- Have things to do
- Come to terms with loss
- Be happy
- Function as independently as possible
- Have maximum confidence and control
- Get better

**The Provider**
- Maximum improvement
- Minimum carer burden
- Minimum burden on the health system

**The Payer**
- Maximum improvement at minimum cost?
- Minimum burden on society?
Health Outcome

A change in an individual or group of individuals that can be attributed (at least in part) to an intervention or series of interventions

3 key ideas:
- change
- attribution
- intervention

Health Outcome ≠ Health status
Before and after

- Health outcome = difference in health status 'before and after' intervention.
  - grounded in an acute care paradigm in which sick patients receive treatment and, as a result, get better.
  - the way that clinicians (and consumers) typically judge the success of most health care interventions.

- Of limited value in measuring the outcomes for people with protracted and chronic illnesses.
  - Some people have conditions that last a lifetime.
Outcomes: Before and After

The difference before and after the intervention

Outcome = 40 point improvement
With and without

- Health outcome = the difference between the person's quality of life and health status if they had received no intervention (or another type of intervention) and that person's expected quality of life and health status with the intervention.

- Includes outcomes for both consumers and carers.
Outcomes: With and Without

The expected difference with and without an intervention

Outcome with this intervention is now either -20, 20, 40 or 50 points improvement, depending on what might have happened with no intervention or another type of intervention.
Outcomes have to be linked to the goal of the intervention

No change, or an arrest in the rate of decline, can be a good outcome in some cases
## A Matrix of Outcomes

<table>
<thead>
<tr>
<th>Length of Life</th>
<th>Improve</th>
<th>Same</th>
<th>Worse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Longer</td>
<td>++++++</td>
<td>+</td>
<td>--</td>
</tr>
<tr>
<td>Same</td>
<td>+++</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Shorter</td>
<td>++</td>
<td>-</td>
<td>-----</td>
</tr>
</tbody>
</table>
Outcomes assessment can’t be a one-off event

- Need reassessment, based on a protocol:
  - clinical criteria (e.g., diagnosis, pall care phase)
  - pre-agreed time periods (e.g., each 90 days) or
  - natural bookends (e.g., hospital discharge)

- Types of outcomes at these points:
  - alive or dead (level 1)
  - better or worse (level 2)
  - better or worse than expected (level 3)
  - value for money (level 4)
Some starting points for the workshop

Based on our experiences
The Australian health system cannot afford to collect data for only one purpose.

Good reasons to collect data:
- immediate use with a consumer - screen, assess, diagnose etc
  - help consumers to get the right services at the right time
- information sharing - common language (including with consumers) and referral
- priority setting - eg, waiting list management
- pay and accounting for health care funding
Possible system level uses of data

- Outcome measurement and evaluation
  - not sustainable purpose in its own right

- Benchmarking
  - not sustainable purpose in its own right

- Accountability and reporting
  - regarded in the field as just more paperwork
  - can be fudged if not a by-product of information collected for other purposes
If you want data for outcome evaluation and benchmarking:

- Start by designing measurement suites that are useful for other purposes:
  - immediate use with a consumer - screen, assess, diagnose etc
  - help consumers to get the right services at the right time
  - information sharing - common language and referral
  - priority setting - eg, waiting list management
  - paying for health care - funding, payment etc
Outcomes occur at different levels

And can be evaluated at different levels
Outcomes and evaluation hierarchy

- 'Process, Impact and Outcome' not enough
- Level 1: Impact on, and outcomes for, consumers
  - patients, families, friends, communities
- Level 2: Impact on, and outcomes for, providers
  - professionals, organisations
- Level 3: Impact on, and outcomes for, the system
  - structures and processes, networks, relationships
Hierarchy of measurement

u Level 1: Impact on, and outcomes for, consumers
  - measured at the person-level and the organisational level
  - capacity to benchmark at the organisational level

u Level 2: Impact on, and outcomes for, providers
  - some measurement possible (eg, workforce competency, availability, satisfaction, turn-over)
  - but little or no systematic benchmarking

u Level 3: Impact on, and outcomes for, the system
  - benchmarking ideas not currently at this level (eg, sustainable systems)
CHSD evaluation framework

- What did you do?
  - PROJECT DELIVERY

- How did it go?
  - PROJECT IMPACT

- What’s been learned?
  - CAPACITY BUILDING

- Will it keep going?
  - SUSTAINABILITY

- Are your lessons useful for someone else?
  - GENERALISABILITY

- Who did you tell?
  - DISSEMINATION

Focus of workshop is on delivery and impact at both person and organisational level
<table>
<thead>
<tr>
<th></th>
<th>What did you do?</th>
<th>How did it go?</th>
<th>Can you keep going?</th>
<th>What has been learnt?</th>
<th>Are your lessons useful for someone else?</th>
<th>Who did you tell?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level 1</strong></td>
<td>Impact on, and outcomes for, consumers (consumers, families, carers, friends, communities)</td>
<td>Impact on consumers&lt;br&gt;Impact on communities</td>
<td>Sustainability assessment</td>
<td>Capacity building assessment</td>
<td>Generalisability assessment</td>
<td>Dissemination</td>
</tr>
<tr>
<td>Direct care delivery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professional development</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Level 2</strong></td>
<td>Impact on, and outcomes for, providers (professionals, volunteers, organisations)</td>
<td>GPs&lt;br&gt;Associated health professionals&lt;br&gt;Nurses&lt;br&gt;Others</td>
<td>Sustainability assessment</td>
<td>Capacity building assessment</td>
<td>Generalisability assessment</td>
<td>Dissemination</td>
</tr>
<tr>
<td>Governance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct care</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professional development</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Level 3</strong></td>
<td>Impact on, and outcomes for, the system (structures and processes, networks, relationships)</td>
<td>System level impacts&lt;br&gt;External relationships</td>
<td>Sustainability assessment</td>
<td>Capacity building assessment</td>
<td>Generalisability assessment</td>
<td>Dissemination</td>
</tr>
<tr>
<td>Governance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct care</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professional development</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The major challenges are cultural and practical

Not technical
Common and routine tools and systems are possible, but...

Implementation is hard work and made more difficult when the policy environment and rationale appears unclear.

Training is a crucial investment domain.

Paperwork burdens are a major limitation.

Culture change is hard and requires time and ongoing support.

No, I don't care what you say. I'm not giving up my forms!
A development cycle for outcomes assessment and benchmarking

But it’s a bit more chaotic in practice!
One off studies

Outcome studies → Culture Change
Routine systems

Outcome studies → Routine outcome measures → Routine outcome systems (training, data collection protocols & processes)

Culture Change
Feedback

Outcome studies -> Routine outcome measures -> Routine outcome systems (training, data collection protocols & processes) -> Culture Change

Feedback -> Performance measurement
Benchmarking

Outcome studies → Routine outcome measures → Routine outcome systems
(training, data collection protocols & processes)

Culture Change

Benchmark (use the data to identify best practices and then implement them)

Feedback ← Performance measurement
The benchmarking cycle

Outcome studies → Routine outcome measures → Routine outcome systems (training, data collection protocols & processes)

Evaluate & refine (measures & systems) → Benchmark (use the data to identify best practices and then implement them) → Feedback → Performance measurement

Culture Change
Exercise 1

**Scenario:** A quality health service **provider** (ie. a champion), knows of other quality health service providers in Australia, and they all decide that they want to **demonstrate** their effectiveness in improving health outcomes.

How should they go about doing this? Outline the steps required.

**Scenario:** A quality health service **funder** (ie. a champion), knows of other funders in Australia, and they all decide that they want to **know** whether their clients / members etc are achieving the health outcomes they should.

How should they go about doing this? Outline the steps required.
Palliative Care Outcomes Collaboration (PCOC) as an example of a routine outcomes systems
The Palliative Care Outcomes Collaboration (PCOC)

A national initiative funded by the Department of Health & Ageing to introduce routine assessment of palliative care quality and outcomes across Australia

PCOC:
- Supports continuous quality improvement of palliative care
- Benchmarks service to improve practice
- Measures outcomes (service and patient/carer)
- Standardises palliative care assessment
- Develops a “common language” for clinicians including primary care
The ultimate measure of the quality of health care is the outcomes that patients and carers achieve.
PCOC is a collaboration

- Centre for Health Service Development, University of Wollongong (PCOC Central)
  - Professor Kathy Eagar

- Department of Palliative and Supportive Services, Flinders University (PCOC South)
  - Professor David Currow

- Western Australian Centre for Cancer and Palliative Care, Curtin University of Technology (PCOC West)
  - Professor Samar Aoun

- Institute of Health & Biomedical Innovation Queensland University of Technology (PCOC North)
  - Professor Patsy Yates
PCOC Staffing

Team at University of Wollongong:
- Manager
- Quality and Education Manager
- Data and IT support
- Statisticians
- Administrative support

Quality Improvement Facilitators (QIFs) based in Brisbane, Melbourne, Adelaide, Perth and Wollongong
How PCOC works

- Work with services to incorporate the PCOC data collection into routine practice
- Provide ongoing support through training and assistance with IT
- Analyse the data and provide feedback on the results to individual services - reports every 6 months
- Facilitate benchmarking with other services
- Assist services with practice quality changes
Overview of Progress (1)

- 111 palliative care services (of about 160 in Australia) have agreed to join PCOC in last 3.5 years, with 86 submitting data for last PCOC Report.
- Majority are large metropolitan services.
- Estimate is that these services represent more than 80% of specialist palliative care episodes.
- All other specialist PC services across Australia are at various stages of follow up, with most expected to join.
Overview of Progress (2)

- Seven national reports
  - Report #7 covers 1 Jan to 31 Jul 2009
- Annual national patient and carer surveys
- Over 2,500 clinicians trained
- Three national benchmarking workshops in 2009
- Early stage planning for V3 dataset of the patient outcomes data set has started
The program logic for PCOC data

Information to be collected at different levels
PCOC information architecture

- **Level 1 Patient**
  - eg, age, sex, diagnosis, postcode

- **Level 2 Episode of palliative care**
  - eg, referral source, time between referral & 1st assessment, episode type, accommodation at start & end, level of support at start & end, place of death

- **Level 3 Phase**
  - eg, Phase (stable, unstable, deteriorating, terminal, bereaved), function at start & end, symptoms at start & end, reason for phase end

Phase - the level at which outcomes are measured

- Phase of care - stage of illness
  - stable, unstable, deteriorating, terminal, bereaved

- For each phase that the patient goes through:
  - Provider type (eg, multidisciplinary, nursing only)
  - Model of care (eg, direct, shared care, consultation-liaison)
  - Start and end dates
  - Reason for phase change
  - Symptom scores at start and end
  - Functional scores at start and end

Quality and outcome measures - 1

- Phase movements
- Change in function
  - RUG-ADL and Kamcfsky
- Change in problem severity
  - PC problem severity scale and SAS
- How episodes start and end
- ALOS (days seen) x phase
- Place of death x level of support
Quality and outcome measures - 2

- Access measures
  - Postcode
  - ATSI
  - Language / country of birth
- Time between referral and assessment
- Diagnostic group
- Model of care planned / provided
- (Consultative services)

3 initial benchmark measures

- Time between referral and 1st contact
- Change in pain from beginning to end of phase
- Time in unstable phase
- Next step is to introduce 3-4 additional measures. Under consideration are:
  - psychological/spiritual problems - PCPSS (Palliative Care Problem Severity Score)
  - carer problems - PCPSS
  - nausea - SAS (Symptom Assessment Score)
  - fatigue - SAS
  - dyspnoea - SAS

A constant theme - unexplained variation

No matter what the measure, we find significant variations between services that we are working to understand and reduce

Some examples...
Variability among inpatient units

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>SD</th>
<th>Difference (X fold)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average length of stay</td>
<td>14.0</td>
<td>6.2</td>
<td>18.6</td>
<td>3.3</td>
<td>3</td>
</tr>
<tr>
<td>Discharge to community</td>
<td>25.1%</td>
<td>12.1%</td>
<td>64.2%</td>
<td>15.4</td>
<td>5</td>
</tr>
<tr>
<td>Stable after unstable</td>
<td>25.8%</td>
<td>4.2%</td>
<td>51.5%</td>
<td>16.2</td>
<td>12</td>
</tr>
<tr>
<td>Function better</td>
<td>9.8%</td>
<td>3.6%</td>
<td>15.7%</td>
<td>3.6</td>
<td>4</td>
</tr>
<tr>
<td>Symptoms improve</td>
<td>22.4%</td>
<td>5.9%</td>
<td>40.6%</td>
<td>11.2</td>
<td>7</td>
</tr>
</tbody>
</table>

The picture is no different for community and consultation services.

Pain at phase end for patients with moderate or severe pain at start (SAS)

Patients self-reported pain in last 3 days (Patient Outcome Scale V2)
Patients self-reported other symptoms in last 3 days (POS-2)

Patients self-reported depression in last 3 days (POS-2)

Carers - Have you had someone to help you with practical tasks?
Carers - Information on Carer Payment or Allowance?

The PCOC approach

Measure outcomes routinely using validated assessment tools

Improve outcomes by consulting the evidence base (CareSearch) & assessing against standards (NSAP)

Compare the outcomes a service achieves with peers, the national average & PCOC benchmarks

The PCOC cycle

CareSearch to inform benchmark development

PCOC Quality Improvement Facilitators (QIFs) work with each service to identify areas where service has room for improvement, does not meet benchmark

Service submits PCOC data each 6 months

Feedback:
- 6 month reports
- benchmarking workshops

CareSearch helps QIFs & service assemble best available evidence on how to improve

If necessary, QIFs help service review itself against relevant NSAP standards to identify potential process improvements

Service changes practices
An increasingly sophisticated evidence-based sector

- Early days - ‘We don’t need to measure outcomes, our patients and carers are really satisfied with the care we provide’
- Then - ‘The data must be wrong’
- Now - ‘We now have information we’ve never had before. What does this mean for the way we provide care? How can we improve the way we organise our service?’

Exercise 2

Scenario: You are the director of a service that is participating in PCOC. The outcomes your service is achieving seem well below those of other comparable services.

What, if anything, would you do about this? What steps would you take? What would be the main challenges? How would you deal with them?

Exercise 3

Scenario: Minister Roxon announces that she wishes to introduce public report cards for all services, including palliative care. She wants to know whether PCOC reports that identify each service should be posted on the web. She also wants to know whether to introduce ‘Paying for Performance’ and pay more to services that achieve the best outcomes.

What advice would you give her?
Now you’ve got the data, how do you interpret / use it?

AROC

- AROC = Australasian Rehabilitation Outcomes Centre
- A joint initiative of the Australian rehabilitation sector (providers, payers, regulators and consumers).
- Established in 2002 by the AFRM on behalf of its industry partners.
- CHSD as data manager.
- Is a not-for-profit self-funding centre with own management board but attached to CHSD.
5 Roles

- National 'benchmarking centre'.
- National 'data bureau' that receives and manages data on rehabilitation services in Australia.
- Education and training in outcome measurement.
- Certification centre for the FIM.
- R & D centre that develops R & D proposals and seeks external funding for its research agenda.

AROC Data Collection

- 180 facilities in Australia and New Zealand submit data to AROC (public and private sectors)
- More than 50,000 episodes are submitted per year
- Database now has over 400,000 episodes of care

AROC Data Set

- Demographic items such as
  - Date of Birth,
  - Sex,
  - Postcode,
  - Country of Birth,
  - Usual accommodation,
  - Living with on admission,
  - Discharge destination, etc
AROC Data Set

- Clinical items such as
  - Impairment Code
  - FIM scores on admission
  - FIM scores on discharge
  - Interruption days
  - Date of Onset
  - Date of FIM assessments

The FIM

<table>
<thead>
<tr>
<th>Episode FIM Score</th>
<th>Motor Subscale</th>
<th>Start</th>
<th>Unit</th>
<th>FIM Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AROC Uses the Data to...

- Provide reports for information and comparison
  - ...for providers and funders

- Provide baseline data for benchmarking workshops
  - ...to start the discussion around how services are provided
Overall Rehabilitation Outcomes
Summary - change in measures 2000-2006

<table>
<thead>
<tr>
<th></th>
<th>Difference from 2000 data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Age (years)</td>
<td>72.3</td>
</tr>
<tr>
<td>Length of stay (days)</td>
<td>21.8</td>
</tr>
<tr>
<td>FIM admission score</td>
<td>88.7</td>
</tr>
<tr>
<td>FIM discharge score</td>
<td>101.3</td>
</tr>
<tr>
<td>FIM change (admit to div)</td>
<td>14.4</td>
</tr>
<tr>
<td>FIM change (per week)</td>
<td>4.7</td>
</tr>
</tbody>
</table>

Dear old Mabel next door has to go to hospital for some rehab.

She must choose which hospital.

She asks you for advice.

Where should Mabel go for her rehab?

The four options

| Hosp 1  | 14 |
| Hosp 2  | 15.6 |
| Hosp 3  | 18.4 |
| Hosp 4  | 18.4 |
| National | 18.8 |
### The four options

<table>
<thead>
<tr>
<th></th>
<th>Average change in FIM score</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hosp 1</td>
<td>14</td>
<td>(12.0 - 16.0)</td>
</tr>
<tr>
<td>Hosp 2</td>
<td>15.6</td>
<td>(13.6 - 17.7)</td>
</tr>
<tr>
<td>Hosp 3</td>
<td>18.4</td>
<td>(16.3 - 20.5)</td>
</tr>
<tr>
<td>Hosp 4</td>
<td>18.4</td>
<td>(16.0 - 20.7)</td>
</tr>
<tr>
<td>National</td>
<td>18.8</td>
<td></td>
</tr>
</tbody>
</table>

### The four options

<table>
<thead>
<tr>
<th></th>
<th>Average change in FIM score</th>
<th>95% CI</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hosp 1</td>
<td>14</td>
<td>(12.0 - 16.0)</td>
<td>18</td>
</tr>
<tr>
<td>Hosp 2</td>
<td>15.6</td>
<td>(13.6 - 17.7)</td>
<td>20</td>
</tr>
<tr>
<td>Hosp 3</td>
<td>18.4</td>
<td>(16.3 - 20.5)</td>
<td>22</td>
</tr>
<tr>
<td>Hosp 4</td>
<td>18.4</td>
<td>(16.0 - 20.7)</td>
<td>22</td>
</tr>
<tr>
<td>National</td>
<td>18.8</td>
<td></td>
<td>20</td>
</tr>
</tbody>
</table>

### Or, from another perspective...

<table>
<thead>
<tr>
<th></th>
<th>Cost per episode</th>
<th>Average change in FIM score</th>
<th>Cost per FIM point improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hosp 1</td>
<td>$22,379</td>
<td>14</td>
<td>$1,599</td>
</tr>
<tr>
<td>Hosp 2</td>
<td>$14,134</td>
<td>15.6</td>
<td>$906</td>
</tr>
<tr>
<td>Hosp 3</td>
<td>$14,870</td>
<td>18.4</td>
<td>$808</td>
</tr>
<tr>
<td>Hosp 4</td>
<td>$16,375</td>
<td>18.4</td>
<td>$890</td>
</tr>
</tbody>
</table>
But...

- Outcomes vary because there are differences between hospitals.
- Outcomes also vary because there are differences between patients within hospitals (the hospital’s casemix).
- We need to control for casemix to help understand the differences in outcomes between hospitals.

Control for casemix???

- AN-SNAP is a casemix classification
  - a method of grouping episodes of care based on consumer attributes that best explain the cost of care
  - iso-resource - consumers in the same class cost about the same amount to treat
  - clinically sensible
  - the right number of classes

The AN-SNAP Version 1 Rehabilitation Classification

- All Rehabilitation
- Overnight Rehabilitation 32 classes
- Ambulatory Rehabilitation 15 classes
Structure of the overnight rehabilitation branch

- Assessment only: Class 201
- FIM Motor = 13
- FIM Motor > 13
- 2 classes: Classes 202 and 203
- 10 major impairment groups
- most of which are split on FIM Motor, FIM Cognition and Age
- Classes 204 - 232

eg, 5 Stroke Classes:

- Class 204 - Motor 63-91, cognition 20-35
- Class 205 - Motor 63-91, cognition 5-19
- Class 206 - Motor 47-62
- Class 207 - Motor 14-46, age >= 75
- Class 208 - Motor 14-46, age <= 74

4 classes for brain dysfunction

Class 209  Motor 71-91
Class 210  Motor 29-70, age >= 55
Class 211  Motor 29-70, age <= 54
Class 212  Motor 14-28
Controlling for differences between patients

- Assign episodes to a ‘casemix class’.
  - Similar consumers in the same class
  - Different consumers in different classes
- When outcomes results are standardised to take account of the mix of consumers, any remaining differences can be attributed to differences between the hospitals.
  - Similar to standardising for age and sex

Casemix adjusted relative mean improvement (CARMi)

- For each episode, calculate the change in FIM score
- For each episode, calculate the difference between the FIM change and the average change for the relevant casemix class.
- Average across the hospital to produce the hospital’s CARMi score.

To interpret a CARMi score

- CARMi for your hospital > 0
  - on average, your patients’ FIM scores improved more than similar patients in the national database.
- CARMi for your hospital = 0
  - your patients achieved about the same level of improvement as similar patients in the database.
- CARMi for your hospital < 0
  - your patients achieved less improvement than similar patients in the database.
What's that mean for Mabel?

### CARMI (FIM)

<table>
<thead>
<tr>
<th></th>
<th>Average Change in FIM score</th>
<th>CARMI (FIM)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hosp 1</td>
<td>14</td>
<td>0.1</td>
<td>(-1.8 - 1.9)</td>
</tr>
<tr>
<td>Hosp 2</td>
<td>15.6</td>
<td>1.9</td>
<td>(0.1 - 3.7)</td>
</tr>
<tr>
<td>Hosp 3</td>
<td>18.4</td>
<td>3.8</td>
<td>(1.6 - 5.3)</td>
</tr>
<tr>
<td>Hosp 4</td>
<td>18.4</td>
<td>1.8</td>
<td>(-0.5 - 4.1)</td>
</tr>
<tr>
<td>National</td>
<td>18.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Other measures

<table>
<thead>
<tr>
<th></th>
<th>CARMI (LCS)</th>
<th>95% CI</th>
<th>% Home</th>
<th>% Readmitted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hosp 1</td>
<td>1.1</td>
<td>(-1.1 - 3.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hosp 2</td>
<td>5.8</td>
<td>(3.9 - 9.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hosp 3</td>
<td>-2</td>
<td>(-3.7 - -0.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hosp 4</td>
<td>-4.4</td>
<td>(-6.3 - -2.5)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
And financially

<table>
<thead>
<tr>
<th></th>
<th>Cost per episode (raw)</th>
<th>Average case complexity</th>
<th>CW=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hosp 1</td>
<td>$22,370</td>
<td>1.43</td>
<td>$15,633</td>
</tr>
<tr>
<td>Hosp 2</td>
<td>$14,134</td>
<td>1.57</td>
<td>$8,997</td>
</tr>
<tr>
<td>Hosp 3</td>
<td>$14,870</td>
<td>1.64</td>
<td>$9,655</td>
</tr>
<tr>
<td>Hosp 4</td>
<td>$16,375</td>
<td>1.27</td>
<td>$12,901</td>
</tr>
</tbody>
</table>

Example 2

- New Zealand Mental Health Classification and Outcomes Study (NZ-CAOS)
  - To develop the first version of a national casemix classification for specialist mental health services in NZ
  - To trial the introduction of outcome measurement into routine clinical practice
- 8 participating District Health Boards (DHBs)

Variables used in the classification

- Length of stay (Complete vs Ongoing inpatients)
- Age
- Ethnicity (adults)
- HoNOS start scores (adult inpatient)
- Diagnosis (child/youth inpatient)
- HoNOSCA start scores (child/youth)
- Legal status (adults)
- Focus of Care (adults)
Average HoNOS improvement by DHB

Why the differences?

- DHB 1 either:
  - provides the best clinical care and support and therefore gets the best outcomes or/and
  - has a mix of consumers that happen to be the most likely group to improve

- Need to standardise for the consumer mix (i.e., the casemix) to make the comparison meaningful

Types of variation

1. Variation due to differences in the ways that health services treat patients

2. Variation due to differences in the kinds of patients treated
Controlling for differences in the mix of consumers

- The casemix classification is the measurement tool
- Assign episodes to a 'casemix class'
- Standardise outcome measures to take into account the casemix within the DHB.

Improvement on the HoNOS by inpatient class

Improvement on the HoNOS by community class
Improvement on the HoNOSCA by community class

HoNOSCA CARMIC

Mental Health ethnicity results (NZ)

<table>
<thead>
<tr>
<th>Ethnicity Grouping</th>
<th>Cost per episode (raw)</th>
<th>Average case complexity</th>
<th>CW=1</th>
<th>Diff.</th>
<th>% diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maori</td>
<td>$7,032</td>
<td>1.49</td>
<td>$4,719</td>
<td>S39</td>
<td>0.8%</td>
</tr>
<tr>
<td>Pacific Island</td>
<td>$9,235</td>
<td>1.93</td>
<td>$4,792</td>
<td>$111</td>
<td>2.4%</td>
</tr>
<tr>
<td>Euro</td>
<td>$3,776</td>
<td>0.81</td>
<td>$4,662</td>
<td>-$19</td>
<td>-0.4%</td>
</tr>
<tr>
<td>All</td>
<td>$4,961</td>
<td>1.00</td>
<td>$4,681</td>
<td>$0</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

Diff. = difference from NZ casemix-adjusted average (the average cost per episode that is unexplained by the classification)
Back to PCOC

Same issue - the need for casemix adjustment - pain control as an example

Change in pain depends on where you start

<table>
<thead>
<tr>
<th>Phase</th>
<th>SAS Pain at start</th>
<th>Average change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>-1.26</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-0.74</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>-0.31</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>-0.50</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.34</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>0.56</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>1.24</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>1.84</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>2.20</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>2.60</td>
</tr>
</tbody>
</table>

So, PCOC needs a composite measure to control for both phase and start score

Negative= pain gets worse
Positive = pain gets better

Pain Casemix-Adjusted Score

- **P-CAS for your service > 0**
  - on average, your patients' change in pain was better than similar patients in the national database.

- **P-CAS for your service = 0**
  - your patients' pain scores changed about the same as similar patients in the database.

- **P-CAS for your service < 0**
  - your patients' change in pain scores was worse than similar patients in the database.
Exercise 4

Scenario: You are the manager of a health service provider organisation, or the manager of a government health program, and you have to implement a ready-made system of routine outcomes measurement.

How would you go about doing this? What would be the main challenges? How would you deal with them?
Some issues worth discussing

Open discussion

- The balance between collecting data for accountability and reporting purposes versus quality and outcome improvement
- If casemix-adjusted outcome measures are possible in palliative care, rehabilitation and mental health, why aren't they in use in acute care?
- Public report cards - incentives and issues
- Action to improve things - conceptual, service-level, system-level
- Finding the right balance between realism and rigour
- Managing culture change
  - I don't care what you say, I'm not giving up my forms
  - We're different (more complex, important, have less resources etc etc)
- Integrating new measures into, and replacing, routine practices
- The relationship between research and implementation
- Training and other required investments, including the clever use of IT