



UNIVERSITY
OF WOLLONGONG
AUSTRALIA

University of Wollongong
Research Online

Australian Health Services Research Institute

Faculty of Business

2011

Evaluation and selection instruments for outcomes evaluation: session 3

Janet E. Sansoni

University of Wollongong, jans@uow.edu.au

Nick Marosszeky

marossz@uow.edu.au

Publication Details

J. E. Sansoni & N. Marosszeky "Evaluation and selection instruments for outcomes evaluation: session 3", *Managing & Measuring Health Outcomes*, Menzies School of Health Research, Darwin, 8-11 March 2011, (2011)

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library:
research-pubs@uow.edu.au

Evaluation and selection instruments for outcomes evaluation: session 3

Abstract

Powerpoint presentation presented at Managing & Measuring Health Outcomes, Menzies School of Health Research, Darwin

Keywords

evaluation, selection, instruments, outcomes, 3, session

Publication Details

J. E. Sansoni & N. Marosszeky "Evaluation and selection instruments for outcomes evaluation: session 3", Managing & Measuring Health Outcomes, Menzies School of Health Research, Darwin, 8-11 March 2011, (2011)

Session 3 Part A

Evaluation and Selection of Instruments for Outcomes Evaluation

Jan Sansoni and Nick Marosszeky



University of Wollongong

Selecting Measures

Frame your research questions first by considering

- Do you want to compare the health status of your group (e.g. asthma) with those with other conditions? If so include a generic health status measure
- Do you want a detailed assessment of the symptoms for this condition – then use a disease specific measure
- Do you want to undertake cost utility analysis – then use a multi-attribute utility measure
- Do you want to assess functional skills- then use a functional measure
- Do you want to assess well-being or health related quality of life?
- Do you wish to make normative comparisons?

Other considerations

- Are the instruments appropriate to your group e.g. cultural appropriateness, translations available
- Do you want clinical ratings or self report forms or a mix of both
- Is the instrument appropriate for your mode of administration (e.g. telephone)
- Is it feasible to use these instruments in this setting – time constraints, respondent burden
- Remember ethics approvals and privacy issues

Psychometric Properties

- There are numerous instruments but better to choose ones with known and adequate psychometric properties (e.g. reliability, validity, responsiveness)
- DIY instruments are not a good idea – instrument design and validation is a lengthy process
- Even modifying or revising instruments requires revalidation
- Examine major texts such as McDowell, McDowell and Newell, Bowling etc

Narrowing the contenders

- Undertake electronic searches Medline, PsycInfo etc – see what others have used for similar studies. This may identify problems with particular instruments.
- Check the instrument texts to get some further review information about leading contenders
- Contact AHOC to get some ideas and they will also help re the grey literature

AHOC Instrument Review Sheet

- The AHOC instrument review sheet is used to undertake systematic reviews of instruments against standard criteria
- When selecting instruments review these criteria with reference to your planned instruments
- Some instrument reviews can be found at <http://chsd.uow.edu.au/ahoc>
- Also try www.qolid.org and NCHOD – <http://.phi.uhce.ox.ac.uk/phidb.html>
- and search the web

Instrument Review Protocol

- Title, author, publisher, supplier
- Costs and licensing (some instruments very expensive to use)
- Training requirements (self explanatory or requires accredited training?)
- Purpose
- Administration requirements (incl. time)
- Instrument type (self report, proxy report, rating scale; generic, disease specific etc)
- Instrument Structure (domains, subscales)

Instrument Review Protocol

- Scoring
- Normative Data – is their data for the gen. pop.?
- Clinical Reference Data (refer DOMS)
- Applications – where mostly used
- Reliability (internal consistency, test-retest, inter-rater)
- Validity (discriminatory power, correlation with other measures, construct and criterion validity)
- Responsiveness (sensitivity to change)

Instrument Review Protocol

- Cultural Applicability and Cultural adaptations (translations)
- Gender Appropriateness
- Age Appropriateness
- Summary
- References
- Reporter
- Date of report

AHOC Review Sheet Exercise

- Our instrument review sheet has recently been revised.
- Compare the original (P2 Session 3 Materials) with the revised version (P4)
- We revised this for the Dementia Outcomes Measurement Suite Project
- What are the differences?
- Use the revised review sheet for your review

Comparing Reviews

- Refer to the handout reviews of SF-36V1 and Katz ADL scale
- Which is the more comprehensive review?
- Is Katz in common use these days – what might this suggest about your search strategy?
- Compare the project team's reviews of the functional measures and discuss factors that might influence instrument selection
- If there are existing reviews (e.g. McDowell, Bowling) for instruments compare these as part of your review process

Instrument Review Assignment

- Choose an instrument* that has been used to assess patient outcomes and provide a systematic review using the AHOC criteria for assessment.
- Are the criteria equally important?
- How might you weight these criteria in order to choose the best instrument within a particular class (e.g. generic health status measures, disease specific measures, utility measures etc.)

* **Reviews of SF-12, SF-36 and Katz ADL excluded**

Session 3 Part B

The Evaluation and Selection of Instruments for Measurement Suites

Nick Marosszeky and Jan Sansoni



University of Wollongong

Measurement Suites

- The purpose of a measurement suite is to review all the instruments used to assess clients/patients or to measure outcomes for a particular target population/ group... could be to assess & assign services for the elderly, or to develop a standardized battery for outcomes monitoring for a particular condition (e.g. incontinence) or disease (e.g. asthma)
- Within a field the purpose of assessment or monitoring may vary across settings...primary/community care, in-patient, research, specialist practitioner... so need to take this into account
- Standardization of outcomes tools can also be a first step toward service benchmarking

Purpose of Measurement Suites

- Screening (ONI)
- More detailed/ tiered assessment (HACC)
- Standardizing outcomes measurement tools for a disease or condition (COMS)
- Routine outcomes monitoring leading to benchmarking (AROC)

Outcomes monitoring will require assessing clients 'before' the intervention and 'after' at appropriate time intervals (but it may be useful to consider building systems capable of both outcomes monitoring and assessment applications)

Example: Continence Outcomes Measurement Suite (COMS)

- To develop a set of recommendations concerning measurement tools that are appropriate for use by Australian clinicians, practitioners and researchers in the area of incontinence
- Recommendations have been developed for
 - Non-specialist primary care practitioners
 - Specialist Incontinence practitioners
 - Researchers

Identify Instrument Review Categories (COMS)

Clinical Measures

- Pad Tests
- Frequency Volume Charts and Bladder Diaries

Surveys and Questionnaires

- Urinary Incontinence Symptom Measures
- Faecal Incontinence Symptom measures
- Condition Specific Quality of Life Measures (Urinary, Faecal)
- Functional Measures
- Generic Health Status Measures/ Health Related Quality of Life Measures
- Multi Attribute Utility Instruments
- Patient Satisfaction (a separate consultancy)

Additional Elements

- Sometimes special symptom items or scales may also be included such as 'cognitive state' - when assessing the elderly (e.g. HACC Assessment Tool)
- Some standardized instruments may not be suited to all your target groups (e.g. children, demented, frail elderly)
- It is useful to have standard socio-demographic items in your measurement battery (e.g. ABS, other similar surveys)

Steps

- Review literature to determine possible instruments...and check related areas
- Get copies of all possible instruments
- Check databases such as QOLID or NCHOD, ACER, ring AHOC
- Review Texts – Bowling and McDowell & Newell very useful (see references)
- Set up expert panel (psychological measurement) and consult with user groups for feedback
- Use a standard form for instrument reviews (e.g. AHOC has one – refer to handouts)
- There will be many instruments – so what criteria might we use to decide which is best?

Types of Measures Considered

Generic HRQOL/ Health Status Measures – SF-36, NHP, SIP,
Generic QOL/ Well-Being Measures (not as relevant for continence outcome assessment) – an example would be COMQOL

Disease/ Symptom Specific – these usually are checklists of symptoms of a particular condition/ disease. These usually include symptom severity and symptom impact items. Examples for continence include Urogenital Distress Inventory and Incontinence Severity Index (for symptoms and severity) whereas Incontinence Impact Q (IIQ) or Incontinence QOL (IQOL) are really measuring the impact of symptoms on daily life

Associated Symptoms – Sometimes there will also be a single symptom measure (such as sexual or cognitive functioning) included in a battery. A particular issue for Dementia OMS.

Types of Measures Considered

Blends - Where a quality of life measure is combined with a disease specific or condition specific measure (e.g. FIQL). Some issues with these measures. Generally not preferred - although dementia may be an exception.

Generic Functional Status – FIM, Barthel, etc.

Health Utility Indexes – For economic evaluation particularly cost utility analysis - AQOL, EQ5D, HUI3 refer Hawthorne 2004.

Patient Satisfaction Measures: CSQ-8, SAPS – generic, and condition specific GUTTS. Global Indices of Satisfaction?

A number of these were then trialled in a community survey (refer session 5) to assess how well they performed.

Developing Criteria: Exercise

Examine the FIM, the Barthel and the Katz instruments for assessing functional skills

- What are some of the similarities and differences?
- What kind of items do they contain? (refer to the next slide and the coverage chart in handouts)
- How do you obtain the scores?
- Might some instruments be better used in some settings than others?
- Might you need more training to use some instruments?
- How long might they take to administer?
- What are some other factors we might consider?

Content of Functional Measures

Item	FIM	Barthel	Katz #
Bowels *	+	+	+
Bladder *	+	+	+
Grooming	+	+	
Toilet Use	+	+	+
Feeding	+	+	+
Transfer	+	+	+
Mobility	+	+	
Dressing	+	+	+
Stairs	+	+	
Bathing	+	+	+
Additional	Communication and Cognition		

* = refers to continence

some functions are combined

Criteria for selecting measures

Refresh these terms

- **Reliability**
- **Validity**
- **Responsiveness**
- **Normative Data / Clinical Data**
- **Type of Instrument**
- **Style of Instrument**

Criteria for Selection

- **Practical Utility**
- **Freedom from Confounding Factors**
- **Coverage/Relevance**
- **Mode of Administration**
- **Culture, Gender, Age Appropriateness**

COMS: How did we rate each measurement tool?

- We applied the following criteria to each measurement tool
 - Availability of comparison data/ usage
 - Length, ease and time to complete
 - Method of administration
 - Cost of using the instrument
 - Translations available
 - Ease of scoring
 - Sensitivity to change
 - Reliability evidence available
 - Validity evidence available
 - Adherence to psychometric/ utility axioms

COMS: How did we rate each measurement tool?

- The criteria had three point scales and were weighted according to an agreed index of importance
- Each tool was rated by an expert panel
- All but a very small number of ratings were the same across all raters, then majority rule prevailed
- The lowest score possible = 24, maximum score = 70...generally tools scoring below 47 were not recommended

Weights for Criteria

Criteria

Weight 1-3

- Availability of comparison data / usage
- Length, ease and time to complete
- Method of administration
- Cost of using the instrument
- Translations available
- Ease of scoring
- Sensitivity to change
- Reliability evidence available
- Validity evidence available
- Adherence to psychometric / utility axioms

Exercise: Rating these criteria

In small groups

- How might we rate these criteria...are some of these criteria more important than others?
- Assign your weights to these criteria
- Discussion – how do your 'weight' ratings match those used in COMS?

Weighting scheme used in ratings

For all categories a 3 point rating was used (refer to the sheet provided, but basically 1=inadequate, 2=satisfactory, 3=good. For the mode of administration 1=rating scale and 2=self report).

Criteria	Weight
Availability of comparison data / usage	3
Length, ease and time to complete	2
Method of administration	2
Cost of using the instrument	2
Translations available	1
Ease of scoring	2
Sensitivity to change	3
Reliability evidence available	3
Validity evidence available	3
Adherence to psychometric / utility axioms	3

Summary Table of Ratings for Functional Measures

Criteria	Weight	FIM	Barthel	Katz
Availability of comparison data / usage	3	3	2	1
Length, ease and time to complete	2	1	2	2
Method of administration	2	1	1.5	1
Cost of using the instrument	2	1	3	3
Translations available	1	2	2	2
Ease of scoring	2	1	3	2
Sensitivity to change	3	3	2	1
Reliability evidence available	3	3	2	1
Validity evidence available	3	3	2	1
Adherence to psychometric / utility axioms	3	3	2	1
Total		55	51	33

Note – multiply each rating by weight and then add to get the total score

Summary of Ratings for Social Isolation Instruments

Criteria	<i>Weight</i>	DJGLS	MSPSS	LSNS	MOS SSS	Friendship Scale^a
Theoretical basis	3	3	3	3	3	3
Availability of relevant comparison data	3	3	2.5	2	2.5	1.5
Length	2	3	3	3	2	3
Complexity of admin	2	2.5	3	3	2	3
Cultural Appropriateness	1	2	2	2	2	2
Ease obtain score	2	3	3	3	3	3
Sensitivity (Target Group)	3	2.5	2	2.5	2.5	2
Reliability	3	2.5	3	3	3	2
Validity	3	2.5	2.5	2.5	2.5	2
Cost-instrument	2	3	3	3	3	3
Cost-staff	2	3	3	3	3	3
Weighted Total		71.5	71	71	68.5	57.5

This is a new instrument with very few publications (including independent publications) as yet – but the limited evidence available is promising.
 Sansoni J et al. (2010) *Final Report: Effective Assessment of Social Isolation*. Centre for Health Service Development, University of Wollongong

Summary of Ratings for Cognitive Assessment Instruments

Criteria	Weight	MMSE (3MS)	SMMSE	RUDAS ^a	KICA-COG	KICA-CARER	IQCODE
Theoretical/empirical basis	3	3	3	3	3	3	3
Availability of comparison data	3	3	2	2	2	2	2
Length/feasibility of instrument for inclusion in battery	2	2	3	2	2	3	2
Complexity of administration/ cognitive burden	2	2	2	2	3	3	3
Cultural Appropriateness	1	2	2	3	3	3	2
Ease of obtaining score	2	2.5	2.5	3	3	3	3
Sensitivity	3	3	2.5	2	2	2	2.5
Reliability evidence	3	3	3	3	2.5 ^b	2 ^b	3
Validity evidence	3	3	3	2.5 ^b	2.5 ^b	2 ^b	3
Cost of the instrument	2	3	2 ^c	2	3	3	3
Cost of instrument administration	2	2	2	2	2	2	2
Weighted Total		70	65.5	62.5	65	64	68.5

Based on the DOMs review in 2008 – this review needs to be updated

b. Scored as 2 or 2.5 because of there being limited evidence/publications or independent publications but what there is indicates good sensitivity, validity and/or reliability.

c. Rated as 2 vs.1 as the costs are minimal and estimated at 12 cents per use

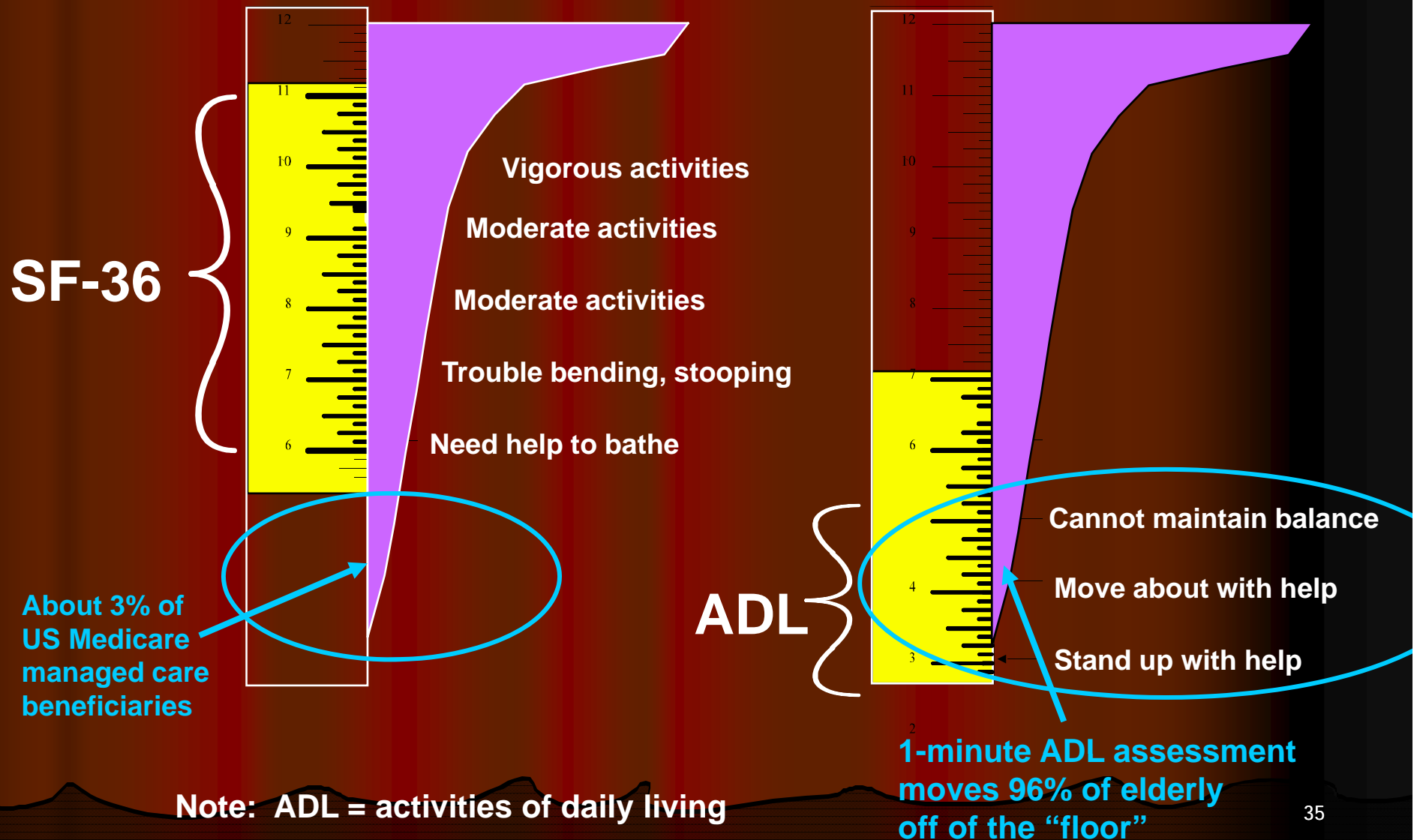
From Sansoni J et al.(2010) Selecting Tools for ACAT Assessment: A Report for the Aged Care Assessment Program (ACAP) Expert Clinical Reference Group. CHSD, University of Wollongong. Report for the Aged Care Assessment Program, Australian Government Department of Health and Ageing, Canberra

Discussion

Need to use instruments with good psychometric properties but need to consider the setting and the proposed applications

- FIM used for both assessment and outcomes monitoring in hospital/residential settings
- Barthel used more as an assessment or screening tool in community/primary care settings
- Techniques such as IRT will be helpful in refining and cross calibrating our measures

Combining and Refining Measures (Ware)



Some Useful References

For instruments and measures

Bowling, A. (1995) Measuring Disease, Open University Press

Bowling, A. (1997) Measuring Health, 2nd ED, Open University Press

McDowell, I. & Newell, C. (1996) Measuring Health, 2nd ED, Oxford University Press

Dittmar, S. S. & Gresham G.E. (1997) Functional Assessment and Outcome measures for the Rehabilitation Health Professional, Aspen Publications

Streiner, D.L. & Norman, G.R. (1995) Health Measurement Scales, 2nd ED, Oxford University Press

and refer to the health outcomes reading list provided.

Materials

- Reviews of SF-36V1 and Katz from COMS
- AHOC Instrument Review Sheet
- Copies of FIM, Barthel Katz instruments
- Copy Slide 30 for weight exercise.