Seminar 4 Lecture

"CTS Research Project

Outcomes Measures and Controls"

Year 1 (2012-13)

Clinical and Translational Science (CTS) Initiative

University of New England

29 November 2012

Robert H. Ross, Ph.D., CTS Initiative Director and Research Associate Professor,
College of Graduate Studies-Public Health, University of New England
Address: McDougal Hall 102, 716 Stevens Avenue, Portland ME.
Email rross1@une.edu. Tel 207-221-4569. Cell 978-505-8892

Outline

1. Recap Lecture 3

2. CTS Study Outcomes Measures and Controls (Kane and Radosevich chap 5-7, 9)
   - **Chapter 5**: (Ross 29Nov12) "Generic Health Outcomes Measures"
   - **Chapter 6**: (You 6Dec12) "Health-Related Quality of Life
   - **Chapter 7**: (Ross 29Nov12) "Condition-Specific Measures"
   - **Chapter 9**: (You 6Dec12) "Demographic, Psychologic, and Social Factors"

1. Recap Lecture 3

In Lecture 3 of this year's Faculty seminar we discussed CTS Study Implementation, highlighting Hulley and Cummings

- Chapter 14: Addressing Ethical Issues
- Chapter 15: Designing Questionnaires and Interviews
- Chapter 16: Data Management
- Chapter 17: Implementing the Study and Quality Control
- Chapter 18: Community and International Studies

To recap, here is a slide or two per section of Lecture 3:

Recap Lecture 3

H&C Chap 14: Addressing Ethical Issues

- Investigators must assure that projects observe the ethical principles of respect for persons, beneficence, and justice; research meets applicable federal requirements, above all Informed consent and IRB review; informed consent covers nature of the project, potential risks, benefits, alternatives
- Vulnerable populations need added protection: children, prisoners, pregnant, cognitive deficiency, social disadvantage
- Scientific misconduct: fabrication, falsification, plagiarism
- Issues for researchers: conflicts of interest, authorship
- Musts with RCTs: intervention arms in equipose; control group receives appropriate intervention; trial closed if one intervention shown safer, more effective; confidentiality gets added attention w/ previously collected specimens and data

Recap Lecture 3

H&C Chap 15: Design Questionnaires Interviews

- Clinical research quality depends on the quality and appropriateness of Q's and I's; instruments must be as valid and reproducible as possible before study start
- Open-ended questions: answers w/o Investigator limitations
- Closed-ended questions: easier to answer, and analyze; response options need to be exhaustive, mutually exhaustive
- Questions need to clear, simple, neutral, appropriate for the study population; Investigators need to eliminate ambiguous terms, double-barreled questions, hidden assumptions, and answer options that do not fit the question
- Instruments need be easily read; interview questions easily read aloud; format spacious, uncluttered, suited to electronic data entry method chosen, with instructions, directions.

Recap Lecture 3

H&C Chap 15: Design Questionnaires Interviews

- Multi-item scales combine questions, measure abstract constructs, e.g. attitudes, health status; questions measure single characteristics, responses need be internally consistent
- Existing instruments may give more valid, reliable results; for new instruments, use existing measures as models, for ideas
- Before study start: pretest, time, adapt whole instrument set
- Pretest new instrument to improve question, instruction clarity and to refine instrument range, reproducibility, validity
- Self-administered Qs are more economical than interviews, more standardiz-able, added privacy may enhance validity
- Interviews give completer responses, may enhance validity
- Computer-assisted Telephone interviewing (CATI), e-mail, and Internet can increase instrument admin efficiency

Recap Lecture 3

H&C Chap 16: Data Management

- Database = tables: rows = records or entities e.g. participants, columns = fields or attributes e.g. measurements
  - Data dictionary gives names, data type, description, range of allowed values for all fields of the database
- Data entry system is used to populate tables; transcription from paper requires double data entry to ensure fidelity
  - Electronic data captured by on-screen forms or web page eliminates transcription step
- Spreadsheet okay for simple database but relational database using management software required for complex databases
- Database queries sort, filter data, calculate values; monitor data entry, report study progress, format results for analysis
- Databases w/ personal identifiers need secure servers, restricted access, and auditing

Recap Lecture 3

H&C Chap 17: Implement Study, Quality Control

- Study start-up first assemble resources: space, staff, budget
- Finalize protocol then by pretesting and piloting recruitment, measurement, interventions, outcomes assessment plans: minimizes need for subsequent in-study protocol revisions
  - Minor protocol revision, e.g. add Questionnaire item, then easy but IRB may need to approve and data analysis may be affected
  - Major protocol revision, e.g. change intervention or primary outcome, have big implications, should be done only w/ approval of such key bodies as DSMB, IRB, funder
- Study conduct then is systematic and Quality controlled
  - Clinical practice QC: OP manual, staff training/certif, perform review
  - Lab procedures QC: blinding, labeling specimens taken, using standard pools, blinded duplicates
  - Data management QC: oversees completeness, accuracy, integrity of data collection, editing, entry, and analysis

Recap Lecture 3

H&C Chap 18: Community, International Studies

- Object: to discover regional differences, e.g. in disease epi or cultural factors that determine intervention effectiveness
- Local participation in clinical research can have secondary benefits to region, e.g. enhanced scholarship, self-sufficiency
- Practical issues more difficult in these settings, e.g. funding, mentoring; thus start small, i.d. local advantages, network
- Collaboration bet AMC and community investigators can be top-down (community I's conduct study originating from MC) or bottom-up (MC I's help community I's do own research)
- International research adds challenges: communication and language, cultural differences, funding, power differences, financial and admin practices, ethics. Rewards can include helping needy people, big PH impact, rich x-cult experiences

2. CTS Study Outcomes Measures and Controls

CTS Study Outcomes Measures and Controls (Kane and Radosevich chap 5-7, 9)

- **Chapter 5**: (Ross 29Nov12) "Generic Health Outcomes Measures"
- **Chapter 6**: (You 6Dec12) "Health-Related Quality of Life
- **Chapter 7**: (Ross 29Nov12) "Condition-Specific Measures"
- **Chapter 9**: (You 6Dec12) "Demographic, Psychologic, and Social Factors"

Generic Health Outcomes Measures (K&R Chapter 5)

Condition-specific outcomes measures (chap 7) assess signs, symptoms of given medical condition, effects on a person's life

Generic measures (chap 5) trade broad coverage against greater responsiveness, discrimination (choice should reflect purpose of the measure); apply and enable comparison across diseases, treatments, demographics; assess single or multiple domains of heath-related functioning in daily life; summarize wide spectrum of health concepts that apply to individuals and populations in different health states, e.g. as do the **Sickness Impact Profile** (SIP, Bergner et al. 1981) and the **Short Form Health Survey** (SF-36. Ware et al. 1992, 1993, 1994, the most widely used generic measure today, translated into > 120 languages; see "QualityMetric" website)

Generic Health Outcomes Measures (K&R Chapter 5)

Generic measures

- capture elements that transcend single diseases, thus may be used to compare treatments across diseases and populations
- assess the physical, psychological, and social dimensions of health consistent with WHO (1948) definition of health as "a state of complete physical, mental and social well-being and not merely the absence of disease"
- go beyond the conventional measures of health used in epi and clinical research (mortality, morbidity, life expectancy, as derived from vital records including census counts, birth and death records) to reflect the importance people attach to their health and functioning
- serve clinicians as bottom-line indicators of treatment effects

Generic Health Outcomes Measures (K&R Chapter 5)

Generic measures

- assess health along entire range from well-being to disability
- augment clinical data and provider perceptions that focus on signs and symptoms of disease
- track the "natural history" of a patient's perceived health status and QoL, e.g. (K&R chap 6) using the (Dartmouth) COOP Charts (Nelson et al. 1990, 1996) or the SF-36 (Ware and Sherbourne 1992)
- register patients' own assessment, own experience of physical, emotional, and social health
- measure HRQL (K&R chap 6), complement to morbidity and mortality measures in studies of disease and treatment impact on patients

Generic Health Outcomes Measures (K&R Chapter 5)

Generic measures

- enable calculating QALYs (denominator of cost effectiveness ratios) when the generic measure is constructed to assign relative values to different health states, e.g. as do the
  - Health Utilities Index Mark 3 (Feeney et al. 2002)
  - EuroQol or EQ-5D (EuroQol Group, 1990, Johnson et al. 1998)
  - Quality of Well-Being Scale (Kaplan and Anderson 1988
- each is referred to as a "health utility" or "health preference" measure, which assigns values to health states that reflect patient preferences for being in a given health state (e.g. death); the EQ-5D can identify 972,000 possible health statuses with 8 questions (Feeney et al 2002)
- such measures 1) assess treatment effects in terms most relevant to patient and 2) enable cost effectiveness analysis

Generic Health Outcomes Measures (K&R Chapter 5)

Generic measures

Two further uses:

- Risk adjustment, where a generic measure can be more useful predictors of cost variation than diagnosis or pharmacy-based measures of patient risk (Pope et al. 1998), e.g. as when
  - the Duke Health Profile was more predictive of primary care charges than diagnoses or provider severity measures (Parkerson et al. 2005)
- HC organization profiling, e.g. hospitals or nursing homes, where by means of patient or residents' completing such measures these institutions may be well differentiated on quality (Kane et al. 2004)

For summary of "Advantages and Disadvantages of Generic Measures" see K&R **Table 5-1**, for "Criteria for Choosing a Generic Measure" K&R **Table 5-2**

Generic Health Outcomes Measures (K&R Chapter 5)

Health Outcomes Domains. Generic measures assess single (unidimensional) or multiple health (multidimensional) domains.

- HRQL (chap 6) measures are multidimensional and assess function, social activity, cognition, emotional well-being, sleep and rest patterns, energy and vitality, perceived health, life satisfaction
- scope of generic outcomes measures: the 6 D's: death, disease, disability, discomfort, dissatisfaction, dollars expended for HC services
  - Includes the traditional two (mort/morb) plus added domains which provide a door into a patient/person's health-related experience
- For "Seven Generic Domains of Health" see K&R **Table 5-3** and "Selected Unidimensional Measures" K&R **Table 5-4**

Generic Health Outcomes Measures (K&R Chapter 5)

Health Outcomes Domains. Comprehensive reviews of generic measures: Kane and Kane (2000) and McDowell (2006), MAPI Research Trust (http://www.mapi-trust.org/). Seven domains are

1) Physical functioning (PF), or the range of a person's mobility and independence in three types of physical ability: fitness, or physiological health; basic self-care activities (ADLs); advanced, integrated independent living activities (IADLs). The choice of PF measure depends on the population studied

2) Psychological health (PH), or a person's range of positive and negative emotions; often focused on anxiety, depression. The generic measure SF-36 includes a five-item PH scale, the MHI-5, which assesses depression, anxiety, mood (Berwick et al. 1991)

Generic Health Outcomes Measures (K&R Chapter 5)

Health Outcomes Domains

3) Social Functioning (SF), or social interaction, interdependence of a person within social environs in four ways: social roles (e.g. job, parenting); engagement in community, neighborhood (social integration); closeness of interpersonal relationships (quality of social network); social support (e.g. emotional, physical.

4) Pain, besides mental health, pain is the most frequently reported reason for physician visit (or "chief complaint" AHRQ 2008); measures of pain assess the degree of physical discomfort in terms of intensity, duration, frequency, timing, precipitating and alleviating factors, all critical to developing a history of present illness.

Generic Health Outcomes Measures (K&R Chapter 5)

Health Outcomes Domains.

5) Cognitive functioning (CF), or the range of a person's intellectual ability, measured in three ways: memory (e.g. significant dates, events), reasoning ability (e.g. computational tasks), orientation (to person, place, time in current surrounding)

6) Vitality, in two constructs: energy and sleep and rest (basic needs besides food and shelter); e.g. the Pittsburg Sleep Quality Index (**PSQI**). Instruments assess sleep as a complex phenomenon characterized by sleep quality, bedtime routine, wake-up time, sleep latency, and duration. The PSQI assesses the partner as well. Both the SF-36 and SIP dedicate questions to available energy and fatigue.

Generic Health Outcomes Measures (K&R Chapter 5)

Health Outcomes Domains.

7) Overall well-being (OWB), or life satisfaction, global measure (combining physical, psychological, social dimensions) of a person's sense of contentment, or health status and happiness, e.g. "In general, would you say your health is excellent, very good, good, fair, or poor?", found to be a good predictor of mortality (Idler and Kasl 1991) Combined with the other six domains, assessment of OWB gives a complete picture of HRQL.

Generic Health Outcomes Measures (K&R Chapter 5)

Practical Considerations

- Measure at baseline and follow-up to track treatment impact
- Serial assessment of generic measures is the core of health outcomes research
- Include health transition item, which asks person to compare present to previous health state (Feinstein 1987), with the generic measure used: enhances interpretation of the measure, adds useful information per se (Fischer et al. 1999)
- Use existing generic measure instead of creating new one: the development work is extensive, dwarfs most applied studies

Settle the following practical issues, but only after the conceptual model and psychometric issues have been resolved

Generic Health Outcomes Measures (K&R Chapter 5)

Settle these Practical Considerations

- Length of time needed to administer, complete questionnaire
- Appropriate format for survey: teleph, face-to-face, self-admin
- Use of proxy respondents or not
- Cost of administration: data collection and entry
- Complexity of measurement and scoring methods
- Acceptability of survey to patients/respondents and clinicians
- Expected format for presenting results, hint: policy-makers, clinicians often find single, summary values more likable than separate domain scores
- Treat scores not as indisputable, objective indicators of underlying health but as present findings

All these criteria cannot always be optimally met: keep going

Generic Health Outcomes Measures (K&R Chapter 5)

Choosing a measure. Investigators have an array of sophisticated patient-reported generic measures available now to complement or substitute for three traditional indicators of generic health (death, disease, disability, utilization of HC services) reported from medical records, vital statistics, and hospital charts.

Utilization is sometimes used as a proxy for health status but it is difficult to interpret as a measure of health because of access-to-services differences and other factors relation to a population's utilization including cultural and economic factors which may disable comparison within and among populations of the relationship between health and health services utilization

Generic Health Outcomes Measures (K&R Chapter 5)

Conclusion

1) Think which health domains are salient to your problem, then choose a measure which best captures (measures vary on which domains and combinations of domains are included)

2) Generic measures are the best way to capture multidimensional aspects of health, i.e. cross-domain (physical, mental, social, cognitive, pain, vitality, well-being); use a generic measure if overall health is the desired outcome

3) Measurement should be made at baseline to indicate where a person's health course began: improvement or worsening of health can be established only by comparing before and after; include a simple health transition item

4) The more easily understood the measure, the more useful it is; generic measures anchored to both real life (lived experience) and to clinical context (e.g. treatment for a condition) are more readily interpreted and interpretable, as when a numeric value readily expresses a health state.

Condition-Specific Measures (K&R Chapter 7)

Generic measures (GMs, chap 5) measure outcomes in variety of settings, but for this breadth (across diseases, treatments, demographics) they trade depth (responsiveness, discrimination)

Condition-specific measures (CSMs) are more responsive to changes in health status because they are more sensitive. CSMs

- are available across many different diseases and conditions
- assess specific diagnostic groups or patient populations' "clinically important changes"
- measure changes in most salient aspects of specific condition
- reflect aspects of functioning that are closely tied to condition
- respond to small treatment effects which generic measures can miss

Condition-Specific Measures (K&R Chapter 7)

Two types of Condition-specific measures

- Clinical: primarily measures of signs, symptoms, and tests
- Experiential: assesses impact of the disease or condition on patient, hence evaluates health in ways like those of a generic measures, only more fine-tuned to that disease or condition

So why not assess just with Generic measures (GMs)? Granted, GMs cover many facets of disease outcomes, the SF-36 health survey (chap 6) for example which

- taps eight distinct aspects of functional health status including (Table 6-6) Physical functioning, Role limitations-physical, Bodily pain, General health, Social functioning, Vitality, Role limitations-emotional, and Mental health, as well as
- two composite summaries: physical and mental health

Condition-Specific Measures (K&R Chapter 7)

A GM like the SF-36 thus works across dimensions, e.g.

- a drug may be found to improve Physical functioning but cause fatigue (reduced Vitality) as a side effect
- the contrasting outcomes on these two dimensions go into a summary score across all eight of the SF-36's subscales
- the problem is that, with just this summary score, clinically important findings could be missed (Patrick and Deyo 1989)
- thus in intervention research, by not isolating the dimension/s of greatest interest, a true treatment effect could be masked
- a CSM designed to assess fatigue, and more sensitively, could accurately detect the drug's clinically significant side effect
- Bottom-line: CSMs hone in on what is especially salient, a gain on clinical sensitivity (depth) albeit a loss on breadth

Condition-Specific Measures (K&R Chapter 7)

GMs may miss clinically significant treatment effects due to floor (or ceiling) effects, e.g.

- a GM normed to distinguish healthy from unhealthy persons may not be able to distinguish unhealthy from very unhealthy persons because the entire sample may bunch at the lower end of the scale before and after an intervention
- a CSM in contrast is often designed/aimed at a particular segment of the distribution, e.g. those from the ill to very ill

A GM may likewise omit a necessary dimension of health, e.g.

- positive effects of treatment for hypertension may be missed using the SF-36 because improved BP may go unnoticed to patients, despite its profound long-term consequences for health

Condition-Specific Measures (K&R Chapter 7)

But CSMs too have drawbacks, including that

- to measure a condition more precisely, a CSM measures more narrowly than a GM, which may miss important intervention effects
- CSMs cannot be used to compare among conditions, e.g. improvements in diabetes care cannot thereby be readily compared to decrease in arthritis symptoms
- many conditions have a plethora of available scales, thus if an investigator is interested not only in her/his treatment effect/s but also in comparing these effects to those of other studies, using a CSM may be a handicap, in contrast to using GMs like the SF-36 or EuroQual 5D which support such comparisons

Condition-Specific Measures (K&R Chapter 7)

The Combined GM/CSM approach

- CSMs are intuitively appealing to clinicians, thus GM/CSM combinations are employed that tap the strengths of each
- Example 1: in Damiano et al. (1995), a study of cataract patients which used both a GM, the Sickness Impact Profile (SIP), and a CSM, the Vision-specific measure VF-14, to evaluate surgical outcomes,
  - the VF-14 found post-op improvement in visual acuity unrelated to SIP score
  - but the SIP provided insights that would have been missed using the VF-14 alone, notably that behaviors not expected to be related to vision, including "I act irritable and impatient," were found to be highly correlated to better visual acuity pre-surgery

Condition-Specific Measures (K&R Chapter 7)

The Combined GM/CSM approach

- Example 2: in Bombardier et al. (1995, Bellamy et al. 1988), a study of pain and physical functioning after knee surgery which used both a GM, the SF-36, and a CSM, the knee pain-specific WOMAC (Western Ontario and McMaster Universities Osteoarthritis Index), to evaluate surgical outcomes,
  - among patients who reported knee pain the GM (SF-36) was not but the CSM (WOMAC) was able to distinguish patients in need of surgery
  - but after surgery, patients were often recovered enough that the WOMAC was unable to distinguish among patients, although some were extremely disabled, while the SF-36 was able to do so

Condition-Specific Measures (K&R Chapter 7)

Here are some Alternatives to GM or CSM only and GM/CSMs combination approaches (Table 7-2 and *passim*)

Approach	Discussion
Modify a generic measure for a specific condition	But once a scale has been modified and re-weighted, it is no more comparable to the original than a completely unrelated scale. And the advantage of a GM is lost, the ability to compare to other studies that have used the same measure.
Attach a condition-specific supplement to general measure	Goal is to have the condition-specific supplement not overlap measurement of the domains of the GM but expand it into domains of added interest. This retains advantage of the GM, comparability, and taps domains of interest in the supplement.
Use battery of condition specific measures	Addresses the narrowness and specificity of single CSMs by expanding the number of domains measured. But it may be easier, cheaper, and more thorough to use GM/CSM combinations. And a battery of CSMs does not have the comparability of a GM, alone or in combination with CSMs.

Condition-Specific Measures (K&R Chapter 7)

Choosing a Condition-specific measure

- First step is "to understand the natural history of the disease and to construct a theory regarding precisely how the intervention will impact the condition and when, during the course of the natural history, the measurement occurs."
- Then, "With that model in place, available condition-specific measures can be evaluated to find one that taps the exact domain/s, along with when and where the intervention is expected to have an impact."
- Bottom line: in order to select an appropriate CSM, the investigator must know what s/he wishes to measure.

Condition-Specific Measures (K&R Chapter 7)

Choosing a Condition-specific measure

- Choosing appropriate domains is the key to selecting the right CSM/s: the choice is informed by whether the CSM selected is meant to target a symptom, sign, test, or function (Table 7-3)
- NB: It is **not clear**, for example, that symptoms reported by patients are less reliable than other types of CSMs.
- Each type of measure has weaknesses.
- Each may tap different domains measuring impact of the condition on the patient.
- Some domains, e.g. pain, may be measured by each of the four methods (see Kane's discourse on rheumatoid arthritis, pp. 147-48).

Condition-Specific Measures (K&R Chapter 7)

Types of Condition-specific measures (Table 7-3 and *passim*)

	Definition	Example
Symptoms	Reported, but not confirmed by other means	Pain, shortness of breath
Signs	Results reported by medical profession after direct exam, an "opinion" or report	Heart murmur
Test	Objective, reproducible finding by a medical professional, e.g. a lab test, requiring an interpretation	Blood pressure, blood glucose level
Function test	Measurement of item related to the condition, but not the condition itself, e.g. ADL/s	Test of a patient's ability to walk up stairs

Condition-Specific Measures (K&R Chapter 7)

Generic measures: to include or not to include with a CSM?

- **GMs complement CSMs**: "An intervention's primary impact may be in one domain, but the intervention may also … have secondary impacts in several different domains."
- **Be clear** why a GM is included: Is it to capture unexpected results (i.e. trolling without an underlying model, typical of early research on a new treatment)? or to test hypotheses (which produces stronger conclusions)?
- **Answer**: Instead of "throwing a battery of tests at a problem, without any underlying conceptual model," do the conceptual work, frame hypotheses, and incorporate a GM as warranted.
- **E.g.** knee surgery is expected to affect mobility, but mightn't that mobility also reduce depression caused by isolation?

Condition-Specific Measures (K&R Chapter 7)

Generic measures: to include or not to include with a CSM?

- **Likewise**, overall health itself, measured by a GM, may differentially affect the intervention's impact on the main (condition-specific) outcome measure, e.g.
  - in back surgery for patients with back pain (measured pre/post by the CSM Roland-Morris Low Back Pain and Disability Rating Scale) the success of the of the intervention may depend in part on overall health status inasmuch as the more ailments the patient the less relief from surgery even if the surgery worked perfectly.
- A GM could reveal this, a finding which could then be used to target treatment at (sub) populations more or most likely to benefit from treatment.

Condition-Specific Measures (K&R Chapter 7)

Finally, choosing to create Condition-specific measure

- the work (time, cost) to create a new measure is daunting: do not do it unless there is no acceptable measure available.
  - acceptance of study results using new measure will depend on acceptance of the measure, which acceptance will require the I to provide strong evidence of reliability, internal consistency, and validity
  - results from using new measure will be hard to compare with others'
  - therefore start with CSM developed, validated by other investigators
  - usually many validated CSMs per condition, e.g. arthritis has at least five (Patrick and Deyo 1989); see McDowell, I (2006). *Measuring health: A guide to rating scales and questionnaires* (3rd ed.). New York: Oxford University Press for ~100 measures for common conditions, e.g. pain, mental status, depression, and physical disability
  - Ovid database interface (at Medline) gives access to the "Health and Psychosocial Instruments" database, with >15,000 articles/references

Condition-Specific Measures (K&R Chapter 7)

Last slide!

- **MAPI Research Trust** has an extensive battery of self-reported condition-specific measures in its "Patient Reported Outcome and Quality of Life Instruments Database" (PROQLID: MAPI Research Trust, 2009): particular use for cross-cultural and translations of measures (Acquadro et al. 2004)
- Be guided by these considerations in choosing a CSM:
  - statistically, a measure needs be reliable, valid, responsive, unbiased, and precise in the range where effects are expected; choose measures appropriate for your population, which don't bunch at floor or ceiling
  - practically, it should not be burdensome for investigator or participant to undertake, should have a track record (to facilitate comparison)
  - theoretically, it should cover domains of greatest interest, based on the disease or condition model first elaborated; incorporate item periods (e.g. "In the past four weeks …") suited to intervention and condition
  - analytically, measure/s should square with anticipated method/s of analysis, appropriate statistical tests, and power analysis/sample size