Patient Reported Outcomes: Word from the front

Bill Plummer, PhD
Professor Emeritus of Computer Science
University of Missouri
plummerb@missouri.edu

Chief Scientist
Universal Research Solutions LLC
bill@oberd.com

A(nother) taxonomy of physicians

Type I: PROs are subjective, unscientific, and useless

Type II: My main problem with PROs is that I am too busy already, but it looks like I'm stuck

Type III: This stuff may really improve my practice

This talk will try to show how Type III's can help change the face of health care, how Type II's can sharply reduce their pain, and how Type I's can moderate into the role of "loyal opposition"

News from Lake Woebegone (and the rest of Minnesota)

In 2008, Minnesota passed the "State Health Reform Act"

Wide-ranging quality improvement initiatives

Significant reporting and process requirements, finalized 2011

"Basket of Care" Concept:

Organize thinking around an entire process of patient care

MN Dept of Health required to develop 7 initial baskets

Implementation Approach

DOH turned to existing players in the state healthcare reform arena

Institute for Clinical Systems Improvement (ICSI)

Founded 1993 by Mayo, Park Nicollet, and Health Partners

Memebership now includes 55 major practices, and 85% of physician in MN

Sponsorship includes the 6 large non-profit insurers

Minnesota Community Measurement (MNCM)

Basket of Care

Total Knee Replacement was one basket chosen

TKR Subcommittee made serious commitment to PRO

Oxford Knee Scale - functional outcomes

EQ-5D - quality of life outcomes

Required pre-op and 3,6, and 12 month post-op administration

National Agenda

At the national level, the bet was placed on information technology as the weak link in US healthcare

Only 2% of physicians were using an EMR in 2006
Fiscal incentives were offered
Evidence-based practice was called for
Data gathering and exchange was mandated

HIT will be the enabler!

Are we getting carried away?

["*Every problem has a simple solution--neat, plausible, and wrong.*"--*Mencken*]

In Lake Woebegone they know the Devil's In the details

Collect data (what data…by whom…quality… compliance….

Store data (what medium…by whom…safety…privacy…HIPAA…

What equipment (computers….staff…cost….upkeep….

Reports (analysis…submissions…follow-ups…by whom…

Start up time (mandates... deadlines....lose my shirt.....

WHERE AM I GOING AND WHAT AM I DOING IN THIS CAREBASKET ???

The data collection problem...

Can be addressed for modest cost and modest person-hours

While controlling your personal time commitment using solution options that range from roll-your-own to turnkey.

By building an infrastructure that makes your practice better, and meets the requirements as a natural consequence

... if you are careful

["His solution was not right. It was not even wrong." -- Heisenberg]

A model architecture for data collection

Employs a fully-hosted computer environment in which

The computers, data storage, and software are remotely remotely located

Only internet access is required to use the system

You retain ownership and control of all data which you submit

Professional standards of security, privacy and disaster recovery are provided

Model Architecture

Side benefits of hosted environment

Patients can fill forms from home, or in the office on a touch-screen device (e.g. iPAD)

Research collaborators can be located anywhere

The iPAD can be used by provider staff throughout the work flow, all the way into the OR

Reminders and other communication with patient can be automated, via internet

Model architecture

Makes a wide choice of outcome instruments available

Scoring and analysis tools available

Interfaces developed by human-computer interaction specialists to offer pleasant user experience and adaptive features to address patient physical limitations

Operates in both clinical and research settings

Model Architecture

Side benefits

Custom outcome forms may be added

Patient history and other intake forms can be collected from home or on an iPAD in the office

Patient satisfaction instruments can be automated

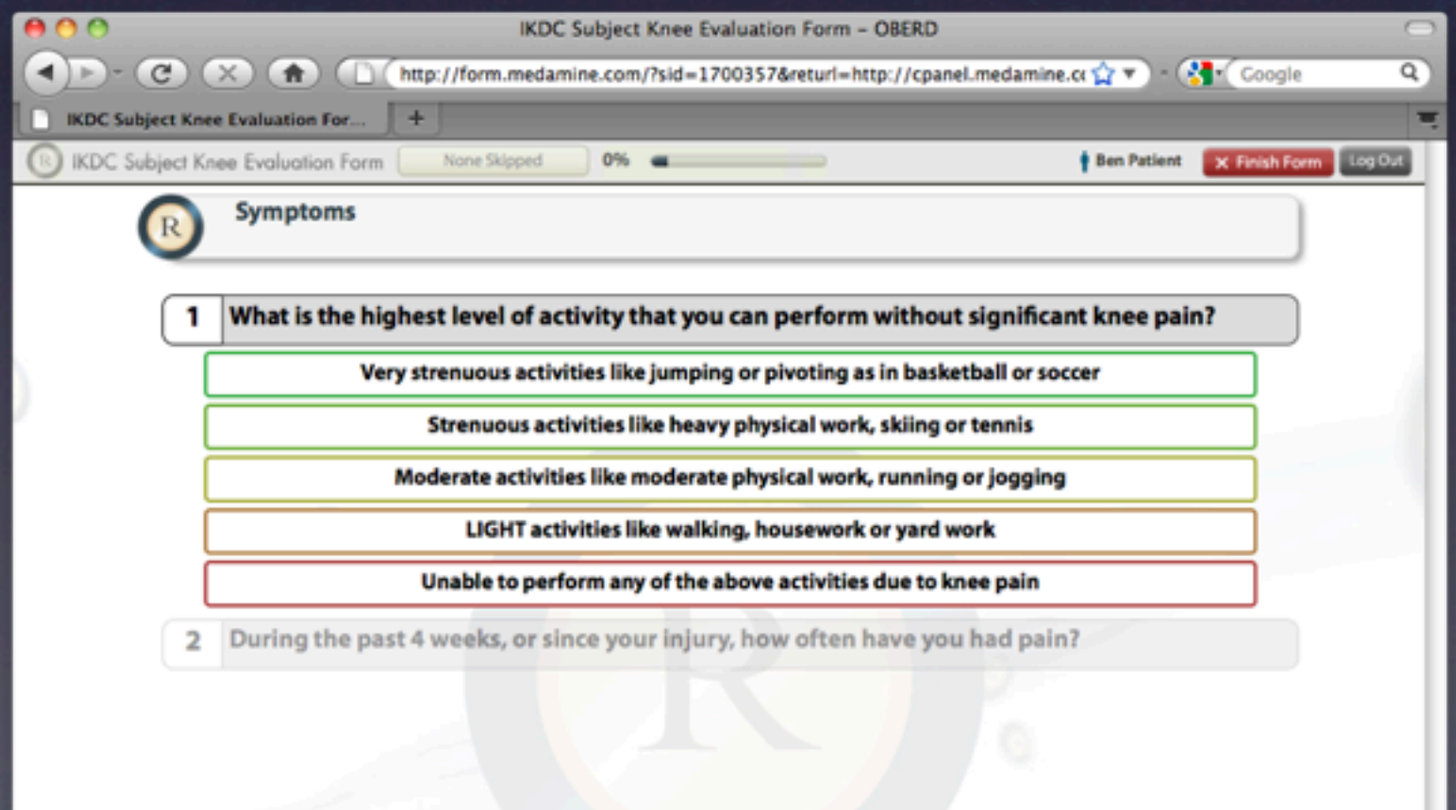
Patient education materials can be communicated

Paper reports may be set up to meet agency requirements, deal with referrals, etc

Model Architecture

Side Benefits

- Replace "paper under glass format"
- Use computer power to adjust question flow based on answers



Model Architecture

Is cognizant of emerging communications protocols and software certifications

Permits the system to be a part of the wider health information environment

Meaningful use certification of any relevant modules insures that any data collected will be shareable and secure

Model architecture

Side benefits:

The system that deals with your PRO can potentially integrate with other medical software in the practice.

Patient communication can be driven by the appointment system

EMR and research data can be collected once and shared when useful

CPT codes benefit both billing and efficacy studies

System building strategies

Don’t build one: Systems that address the components you need may already exist

Combine services from several vendors

Open Source software exists that can provide excellent building blocks (or even an MU-certified EMR)

Some care is in order

[“When committed to software it transcended wrong, and entered the realm of the malevolent” -- Plummer]

How to think about PRO Instruments

PRO is fashionable these days and therefore may seem to be settled science or just obvious

Actually there are some subtleties, some controversies, and some pitfalls.

There is even controversy about the word "outcomes" : the strongest proponents really want to emphasize follow-up as the prime motivation

Functions of questionnaires

Elicit opinions *("Did your physician behave in a professional manner?")*

Elicit observable facts that are most easily obtained from the patient *("Can you walk a mile without pause?")*

Elicit information about "latent traits" *("What is your anxiety level today?")*

Purposes of questionnaires

Opinion

Understand patient motivations better

Factual

Judge patient function

Latent trait

Measure patient condition as perceived or experienced

Pitfalls of questionnaires

Opinion

Unknown basis, highly changeable

Factual

May require estimates or trained judgement

Latent traits

When can subjective judgements be regarded as valid scientific data?

Scales and scores

Likert scale: multiple choice, choices ordered by severity, described by familiar adjective, usually numbered

Example: How hard is it for you to do house work?

1 Easy 2 Somewhat difficult 3 very difficult

Numbers put the answers in order and are easy to record

ISSUE: It is tempting to take differences, ratios, averages etc. of these numbers, but is it meaningful?

Scales and Scores

Visual Analog Scale (VAS)

Line, with end-points labelled

Pt is free to indicate answer by a mark anywhere on the line

Score is distance from the "good" end.

Hope to get "direct" right brain info.

Example: How hard is is for you to do housework?

Score recorded = 63

ISSUES: What does the "score" mean? Is it consistent between different people? Is it just another ordering method? Is a linear scale appropriate?

Validation

Most PROs aggregate responses in some way and assign a numerical score

The process of spelling out the meaning of the score is called "validation"

Does the score qualify as a measurement in any sense that resembles measurement in the physical sciences?

It is generally agreed that a measure must be repeatable, within some limits of reliability

Classical Validation Process

Addresses the question "Is the instrument valid?"by examining various characteristics deemed essential

Is the construct to be measured clearly defined? Based on a theory? Unidimensional? Context clear? ("Construct validity")

Are the questions clear and meaningful to the target population? ("Content validity")

Does the score predict anything ("Predictive Validity")

Etc.

“Modern” validation process

Believes that the whole idea of constructs and latent traits runs aground on the concept of meaning of something you cannot observe

Holds that the meaning of the of the score is the instrument itself (e.g. intelligence is defined as the score on a Stanford-Benet test)

The game is to use one or more such instruments to make predictions that are observable

Implies that any change in a questionnaire in principle defines a new quantity.

Current validity thinking

Need to keep constructs--that's where theories live.

Focus on the fact that validity resides in the combination of instrument and its application--it is the study that is either valid or not

Validity is not independent of the intended use of the study or the population under study

Validity and the construct can evolve over time as more data comes in

The FDA weighs in (2009)

Permits PRO findings to be an experimental end point and even a "claim" for drug labeling

Developed in-house expertise on PROs and validity of studies

Has issued industry guidance for PRO based claims

Requires a dossier that deals with the construct, the logic behind the instrument for drawing a conclusion about the construct, and the justification for application to the test population

De facto position of medical literature

Clings to the notion of validated instrument

Mostly uses classical model and vocabulary

Very few validations would pass FDA scrutiny

An interesting 2005 JBJS editorial by Bartram Zarins, MD entitled "Are Validated Questionnaires Valid?" and the lively discussion it provoked puts the matter in an orthopaedic context

The Oxford Knee Score

Discussions with 20 patients to see how they actually described their knee status led to an initial 20 question instrument

Administered to 2nd group, comments were obtained, and the questions refined

Two more refinement stages led to a 12 question form, with answers numbered 1 to 5 in order of severity of knee disability

Two more refinement stages led to the 12 question form, with answers numbered 1 to 5 in order of severity of knee disability

Oxford Knee Score

During the past four weeks

1) How would you describe the pain you usually have from your knee?

2) Have you had any trouble with washing and drying yourself (all over) because of your knee?

3) Have you had any trouble getting in and out of a car or using public transport because of your knee? (whichever you tend to use)

4) For how long have you been able to walk before the pain from your knee becomes severe? (with or without a stick)

Oxford Knee Score

5) After a meal (sat at a table), how painful has it been for you to stand up from a chair because of your knee?

6) Have you been limping when walking, because of your knee?

7) Could you kneel down and get up again afterwards?

8) Have you been troubled by pain from your knee in bed at night?

Oxford Knee Score

9) How much has pain from your knee interfered with your usual work (including housework)?

10) Have you felt that your knee might suddenly “give way” or let you down?

11) Could you do the household shopping on your own?

12) Could you walk down a flight of stairs?

Each question has 5 answer choices, numbered 1-5

Most use the word "moderate" for 3

Exceptions were questions 1, 6, 8 which seem somewhat unsymmetrical:

PAIN: 1 None 2 Very mild 3 Mild 4 Moderate 5 Severe

LIMP: 1 Rarely/never 2 Sometimes or just at first 3 Often, not just at first 4 Most of the time 5 All of the time

NIGHT PAIN: 1 No nights 2 Only 1 or 2 nights 3 Some nights 4 Most nights 5 Every night

The score for a question was the same as the answer number, and the 12 question scores were summed

The psychometric properties were studied by pre-op and post-op administration to 117 consecutive patients

The classical validity concepts were satisfied

Of particular value were test-retest results establishing that the minimum statistically significant difference between two scores was 6.45 points for this sample

As is typical with Likert scale scoring, there was no showing that the numbers assigned had any meaning except to order the answers

No evidence that a unit difference (between 3 and 4, say) was any sense equal to a unit difference between other values (say 2 and 3)

No evidence that the questions were of equal importance for assessing overall knee function

EQ-5D (EuroQoL Group)

Five questions to assess five factors deemed key components of health-based quality of life status

Visual analog scale for patient assessment of overall health

Original version presented three answer choices for each question (EQ-5D-3L)

New version released in 2011 offers five answer choices (EQ-5D-5L)

The questions are meant to seek factual information; the VAS is intended to measure a latent construct

EuroQoL emphasizes that the numerical labels refer only to ordering and that they may not be used arithmetically

Health status is to be reported by a 5-tuple consisting of the 5 scores

11111 is the best possible status
55555 is the worst status for the 5L version
12145 shows a mix of good and bad functioning

There are a total of 3125 possible states

EuroQoL seeks to provide a number in the range 0 to 1 that reflects the value of each state as perceived by the general population (anchored by 11111 at 1)

One technique is to base the value on the VAS scores reported in a national survey

Another uses "time trade-off" questions ("Would would you prefer 10 years at 33333 or 5 years at 11111")

Surveys have been done in 20 or so countries, typically with 2000 or more respondents

This provides a quality-of-life weighting that can be applied to a given health improvement for purposes of cost-benefit calculations, directions for research, etc.

Summary

Use (carefully) information technology to deal with PROs

A "Model Architecture" that deals with PROs and much more can be implemented today with minimal on-site infrastructure

PRO instruments alone cannot insure "validity" absent consderation of the type of information, the underlying constructs, the target population, and intended purpose of the data

Disclosures

Universal Research Solutions LLC offers a product which implements a number of features of the Model Architecture