

**The Science of
Health Outcomes Measurement:
Facts, fictions and future challenges**

Neil K. Aaronson

Advances in Health Outcomes Measurement

Bethesda, MD

June 23-25, 2004

The way to capture an audience's attention is with a demonstration where there is a possibility the speaker may die.

Jearl Walker, Cleveland State University

A note on terms

- Health outcomes
- Health status
- Quality of life
- Health-related quality of life
- Patient-reported outcomes (PROs)

fact

n. a thing known to be true || a statement about something which has occurred, *he got the facts distorted* || (*law*, in certain phrases only) a crime as a matter of fact, in point of fact, the fact of the matter is... (introductory phrases used to emphasize an explanation or confession) to tell you the truth in fact (usually in contradistinction to some supposed state of affairs) in truth, actually (fr. L. *factum*, a thing done)

The New International Webster Dictionary of the English Language, 1995

fiction

n. A literature consisting of invented narrative, esp. the novel and short story || **a falsehood** (e.g., that there exists a ‘man in the street’) **conventionally accepted as true because it is useful to make the assumption**

The New International Webster Dictionary of the English Language, 1995

fact or fiction?

The term “(health-related) quality of life,”
is well defined and widely understood.

Fact – if you keep things simple

Fiction – if you dig deeper

"Quality of life is a vague and ethereal entity, something that many people talk about, but which nobody clearly knows what to do about." Campbell et al., 1976

"The idea has become a kind of umbrella under which are placed many different indexes dealing with whatever the user wants to focus on." Feinstein, 1987

"Quality of life is an ill-defined term...it means different things to different people, and takes on different meanings according to the area of application." Fayers & Machin, 2000



4 criteria for evaluating clinical effectiveness of chemotherapeutic agents in lung cancer

D.A. Karnofsky et al., *Cancer* 1:634 , 1948

- subjective improvement
- objective improvement
- performance status
- length of survival

Subjective improvement

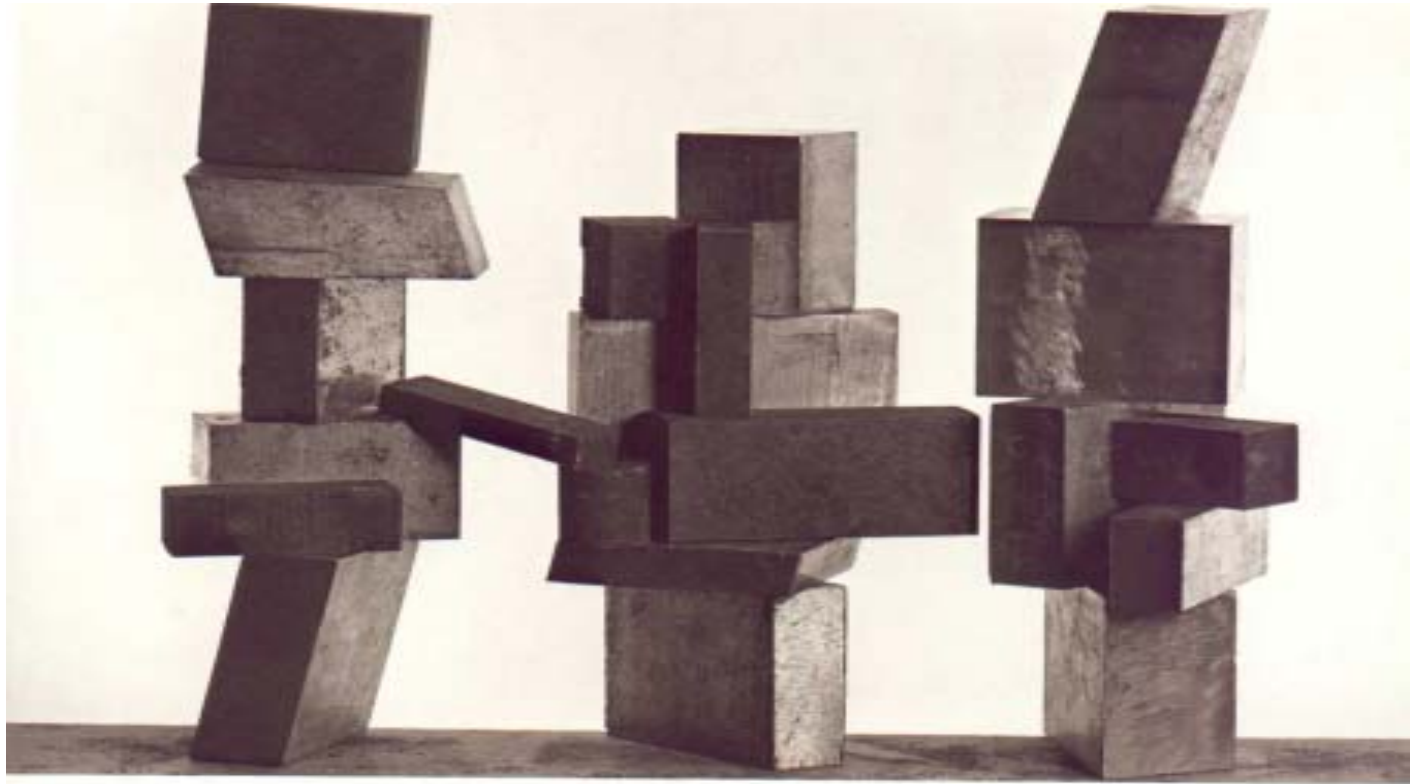
- improvement in mood and attitude
- general feeling of well-being
- activity, appetite, and the alleviation of distressing symptoms such as pain, weakness, and dyspnea

WHO definition of health, 1948

“A state of complete physical, mental and social well-being, and not merely the absence of disease or infirmity.”

Key dimensions of quality of life as defined by ASCO, 1995

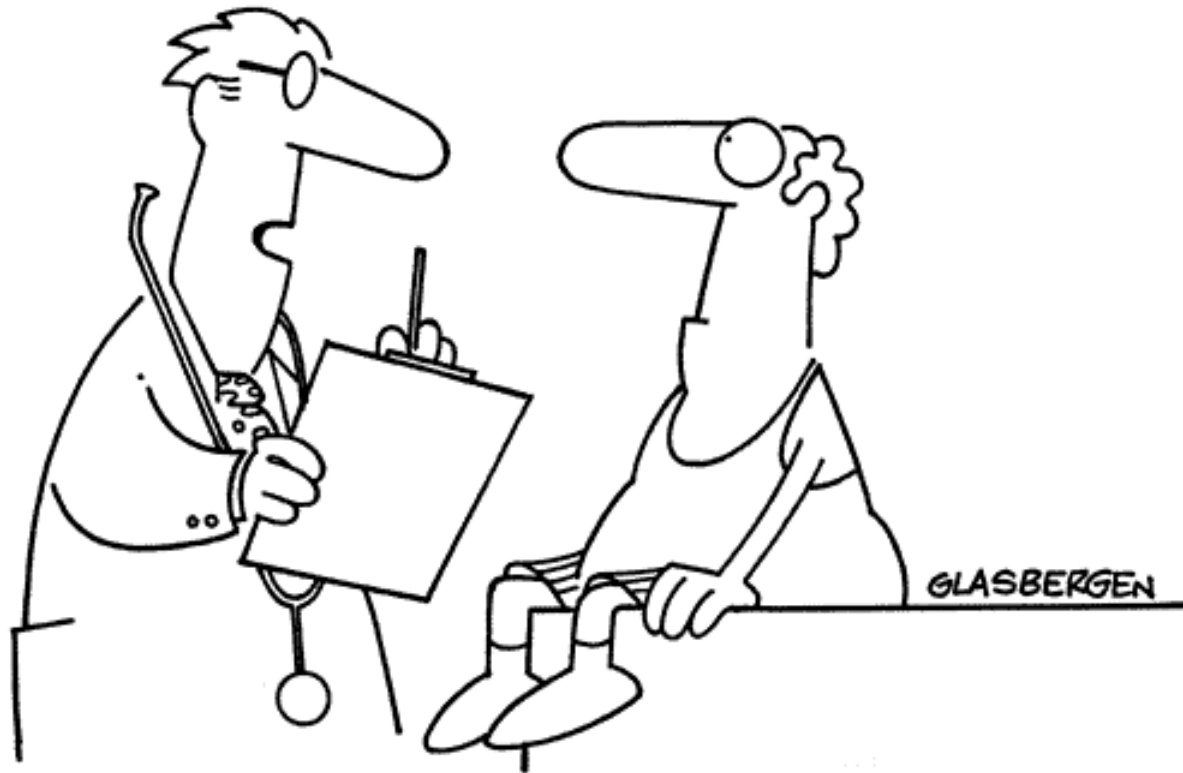
- Physical** Symptoms commonly caused by cancer and the toxicities of treatment
- Psychologic** Effects of cancer and its treatment on cognitive function and emotional state
- Social** Effects of cancer and its treatment on interpersonal relationships, school, work and recreation



Attributes of QL definitions

- Non-specific vs. health-related
- Health states (or status) versus personal evaluation of those states (e.g., expectations, discrepancies, satisfaction)
- Scope of concerns (e.g., spirituality or existential issues)
- Polarity of concerns (well-being vs. dysfunction and its resolution)

Copyright 2002 by Randy Glasbergen.
www.glasbergen.com



“We can’t find anything wrong with you, so we’re going to treat you for Symptom Deficit Disorder.”

Does it matter?

- Yes, because the content of QL questionnaires reflects the underlying definition.
- It may be less important in clinical trials, where group comparisons will be internally valid, regardless of the definition used.
- It is more important in comparing results across trials and in observational (e.g., prevalence) studies.

Examples of QL definitions

“The difference between the hopes and expectations of the individual and the individual’s present experience.”

Calman, 1987

“The functional effect of an illness and its consequent therapy upon a patient, as perceived by the patient.”

Schipper et al. 1996

Covinsky et al. Am J Med 1999; 106:435-440

- Study of 493 elderly patients
- 43% of those with worst physical functioning and 47% with highest levels of psychological distress rated their QL as “good or excellent”
- QL was rated as poor by 15% of those with the best physical functioning and 21% with the lowest levels of psychological distress

Is there a role for IRT/CAT in defining HRQL?

- Probably not
- To the contrary, explicit HRQL definitions should guide IRT/CAT development

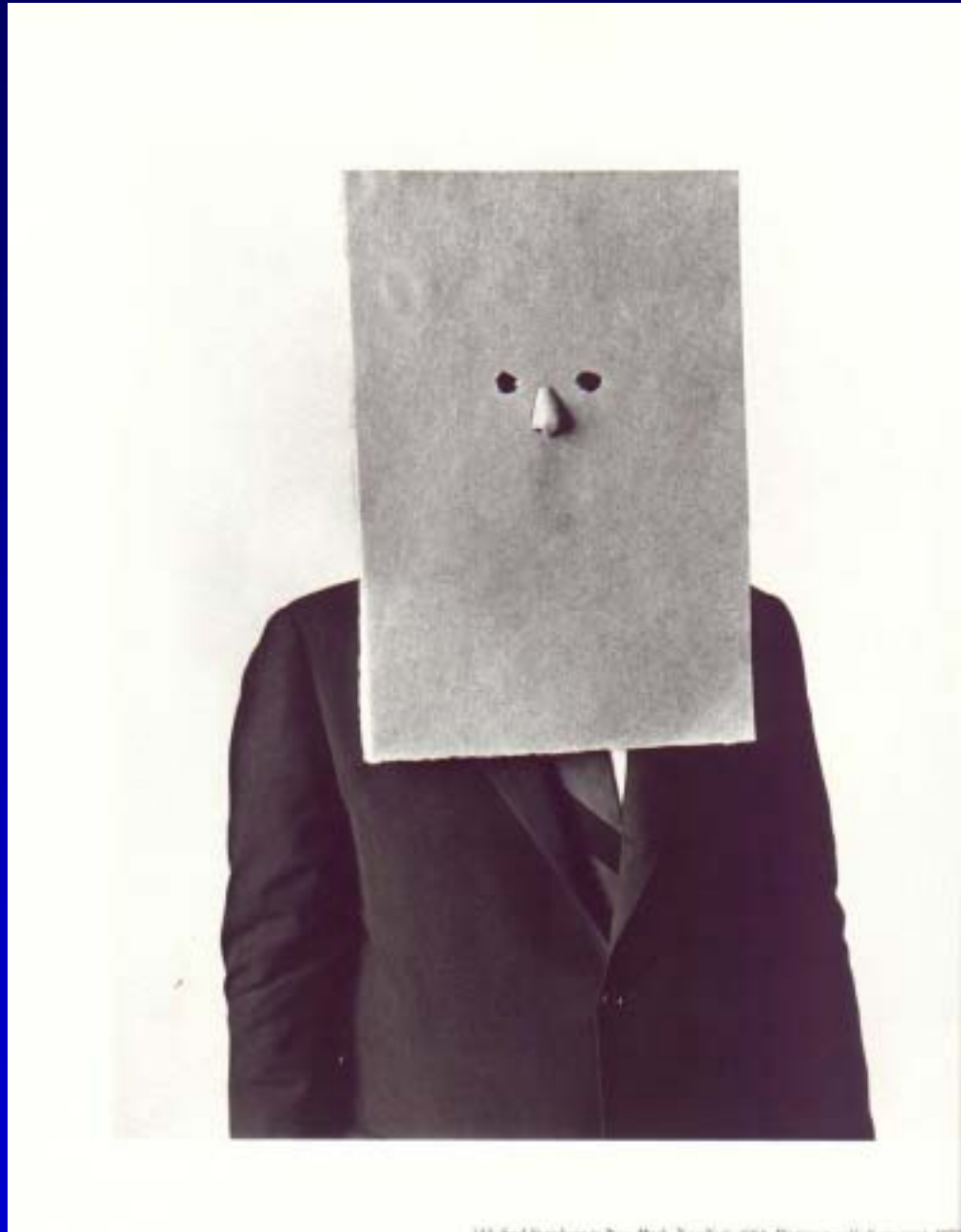
fact or fiction?

The patient is the sole legitimate source of information about his/her QL. Other “proxy” raters (e.g., family members, health care providers) are, at best, poor substitutes.

(partial) fiction



Slide prepared by Neil K. Aaronson, Ph.D.



Slide prepared by Neil K. Aaronson, Ph.D.

Self-report can be limited by:

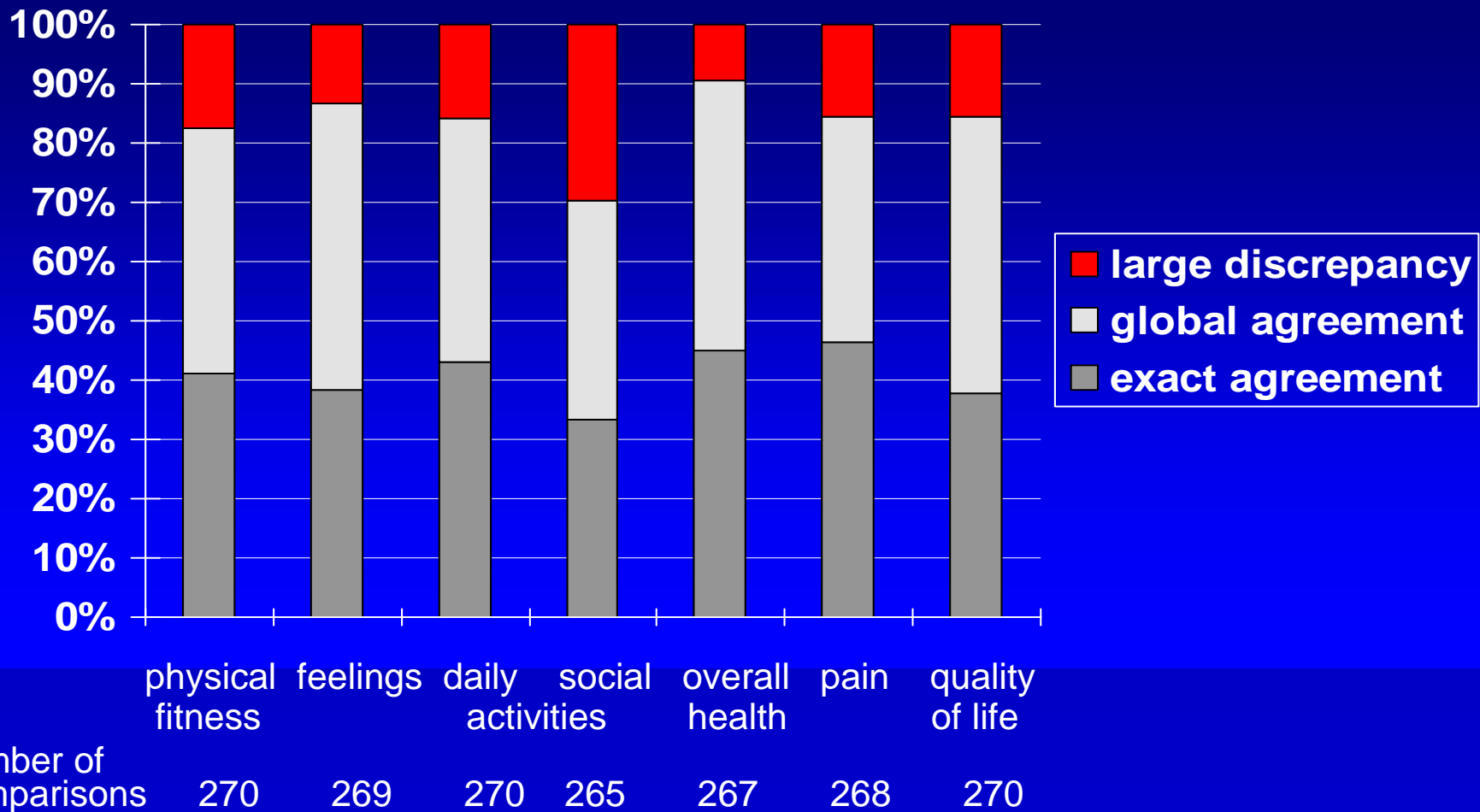
- age (very young or old)
- cognitive impairment
- communication problems
- symptom distress
- physical disability
- emotional distress

Exclusion of highly relevant subgroup of patients can result in biased study outcomes

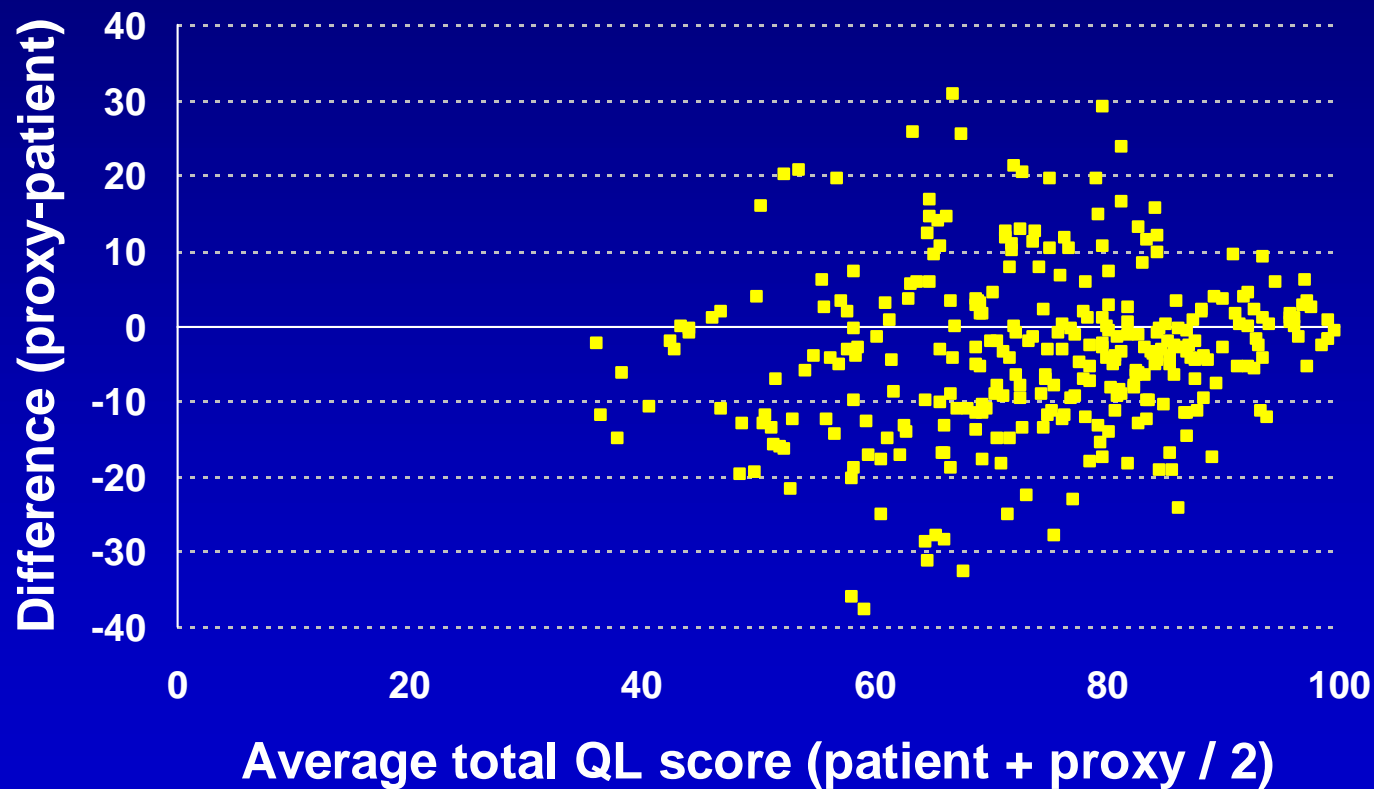
The role of health care providers and significant others in evaluating the QL of patients with chronic disease
Sneeuw KCA et al. 2002; J Clin Epidemiol 55:1130-43

- 23 studies published between 1991 - 2000
- Moderate/high patient – proxy agreement
- Proxies tended to rate patients as having more problems than did patients themselves
- Magnitude of differences was small (median standardized difference 0.20)

Proportion of agreement by COOP/WONCA QL domain



Bland-Altman plot for total QL score



Is there a role for IRT/CAT in proxy HRQL ratings?

- We simply don't know
- Need to examine empirically whether IRT-based and/or CAT-versions of proxy HRQL assessments yield the same promising results as have been seen to date

fact or fiction?

Although there are many HRQL questionnaires from which to choose, the dust is settling and a “best bet” can be identified based on a comparison of psychometric characteristics and performance.

fiction

Generic HRQL instruments

- Sickness Impact Profile (SIP)
- Nottingham Health Profile (NHP)
- Spitzer QL Index
- COOP/WONCA Charts
- MOS 36-Item Health Survey (SF-36)
- World Health Organization (WHOQoL)

Generic QL instruments

- Sickness Impact Profile (SIP)
- Nottingham Health Profile (NHP)
- Spitzer QL Index
- COOP/WONCA Charts
- **MOS 36-Item Health Survey (SF-36)**
- World Health Organization (WHOQoL)

Cancer-specific HRQL questionnaires

- Functional Living Index – Cancer (FLIC)
- Cancer Rehabilitation Evaluation System (CARES)
- Rotterdam Symptom Checklist (RSCL)
- EORTC QLQ-C30
- Functional Assessment of Cancer Therapy (FACT-G)

Cancer-specific QL questionnaires

- Functional Living Index – Cancer (FLIC)
- Cancer Rehabilitation Evaluation System (CARES)
- Rotterdam Symptom Checklist (RSCL)
- EORTC QLQ-C30
- Functional Assessment of Cancer Therapy (FACT-G)

Key psychometric attributes of HRQL instruments

- measurement model
- reliability
- validity
- responsiveness
- interpretability
- cultural adaptability
- burden

Is there a role for IRT/CAT in choosing among extant HRQL questionnaires?

- Yes, definitely:
 - Provide insight into the cultural (in)variability of questionnaires (DIF)
 - calibration of instruments for direct comparison and interpretation of results
 - BUT.....

Choice of HRQL instrument should be driven by:

- the research question(s) to be addressed
- the population under study
- the conceptual basis of candidate questionnaires
- the specific content and wording of candidate questionnaires

Negative affect items

SF-36

“Have you felt so down in the dumps that nothing could cheer you up?”

“Have you felt downhearted and blue?”

FACT-G

“I feel sad”

QLQ-C30

“Did you feel depressed?”

Future perspective items

SF-36 “I expect my health to get worse.”

FACT-G “I worry about dying.”

CARES-SF “I worry about whether the cancer will progress.”

QLQ-C30 --

fact or fiction?

Given the plethora of HRQL questionnaires currently available, there is little or no need for continued efforts at instrument development.

fiction

- Condition-specific questionnaires tend to be more sensitive to group differences and responsive to inter- and intra-individual changes over time

supplemental modules/scales

- combine “core” instrument with condition-specific modules/scales
 - EORTC “modules”
 - FACT subscales
 - NCIC symptom checklists

advantages of core + module approach to QL assessment

- facilitates comparison of results across studies
- provides sufficient flexibility to address questions specific to a given patient population or treatment

Is there a role for IRT/CAT in developing new HRQL measures?

BINGO

- improve reliability/precision
- improve validity
- increase efficiency/decrease burden

fact or fiction?

Drug regulatory agencies and key clinical oncology groups are increasingly open to and supportive of the use of HRQL outcomes in clinical trials.

In theory, factual –

In practice, fiction?

U.S. FDA 1985

“...Survival and quality of life are the key efficacy parameters.”

Johnson and Temple, Cancer Trt Rep, 1985

U.S. FDA 1996

“The Oncologic Drugs Advisory Committee has recommended that beneficial effects on QoL and/or survival be the basis for approval of new anticancer drugs...”

Beitz, Gnecco & Justice, JNCI Monographs, 1996

Endpoints in U.S. F.D.A. approval of oncology drugs: 1990 - 2002

- Marketing approval given to 57 drugs via standard procedures
- Basis of approval:
 - Survival – 32%
 - Tumor response – 46%
 - Tumor-specific symptoms – 23%
 - Other – 16%
 - Quality of life – 0%

Johnson JR et al. J Clin Oncol 2003; 21:1404-11

A question of miscommunication?

Williams G, Pazdur R, Temple R (FDA Division of Oncology Drug Products), *J Biopharm Stat* 2004

- “...in the late 1970’s and early 1980’s, FDA determined that acceptable endpoints for cancer drug approval were survival or an improvement in the quality of a patient’s life, e.g., an improvement in tumor-related symptoms.”

A question of miscommunication?

Williams G, Pazdur R, Temple R (FDA Division of Oncology Drug Products), *J Biopharm Stat* 2004

- “...in the late 1970’s and early 1980’s, FDA determined that acceptable endpoints for cancer drug approval were survival or an improvement in the quality of a patient’s life, *i.e.*, an improvement in tumor-related symptoms.”

- “Drug sponsors are encouraged to include symptom assessments in cancer clinical trials and to perform further research to improve symptom-assessment methods.”

Williams G, Pazdur R, Temple R
J Biopharm Stat 2004

“In theory there is no difference between theory and practice. In practice there is.”

Yogi Berra

Can IRT/CAT play a role in these FDA-driven developments

- Yes, BUT
- The complexity of IRT/CAT may be challenging to (and challenged by) those within the FDA (and elsewhere)
- IRT may be of limited use if one chooses to assess many symptoms in a simple way (e.g., symptom checklists)

fact or fiction?

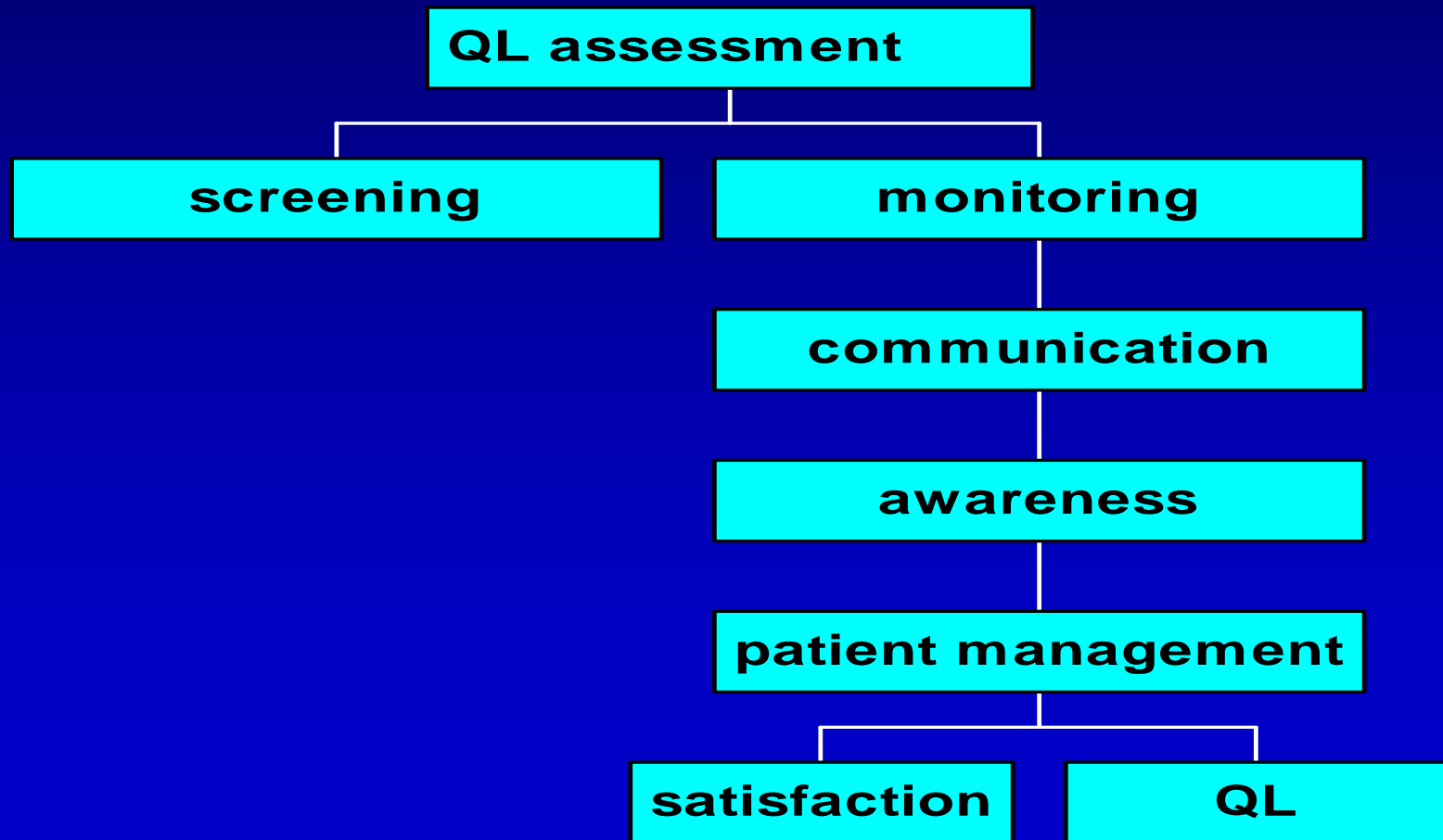
QL assessment is ready for prime time as a tool in daily clinical practice.

“faction”

Brodman K. et al. The Cornell Medical Index:
An adjunct to medical interview JAMA 1949;
140:531-34

- 195 item self-administered questionnaire on physical and psychological symptoms and medical history
- completed prior to office visit in 10-30 minutes; high compliance rates
- Elicited information not found in medical records

Modeling the use of QL assessment in clinical practice



QL assessment in daily clinical practice: Feasibility

- Self-administered questionnaires can be completed quickly in office-based practice
- Computer-assisted (e.g., touchscreen) administration is acceptable and efficient
- Does not interfere with normal clinic routine or lengthen average visit time



Slide prepared by Neil K. Aaronson, Ph.D.

QL assessment in daily clinical practice:

16 randomized studies published 1987-2004

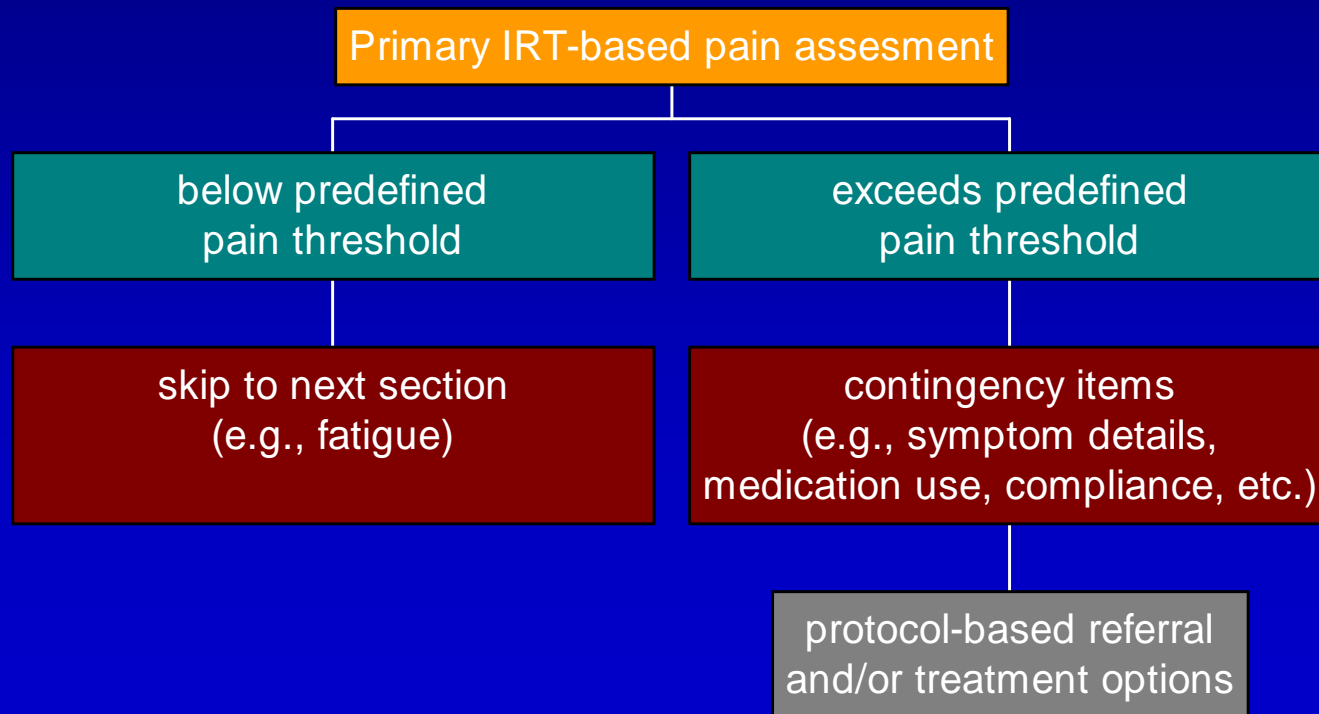
**4 of which were in oncology setting: Taenzer et al. 2000;
McLachlan et al. 2001; Detmar et al. 2002; Velikova et al. 2004**

- communication +
- awareness +
- patient management +/-
- satisfaction -
- health outcomes +/-

Is there a role for IRT/CAT in
HRQL assessment in daily
clinical practice?
BINGO AGAIN

- improve reliability/precision
- improve validity
- increase efficiency/decrease burden

CAT + contingency approach to HRQL assessment in daily clinical practice



Summarizing the potential role of IRT/CAT in:

- Defining HRQL
- Proxy assessment
- Selecting among measures
- Developing new measures
- Informing regulatory guidelines
- Assessing HRQL in daily clinical practice
- None
- Unknown
- Moderate
- Major
- Moderate
- Major

“The best way to predict the future is to invent it.”

Alan Kay