Pitfalls in Measurement

Brian Feldman MD, MSc, FRCPC
The Hospital for Sick Children
University of Toronto

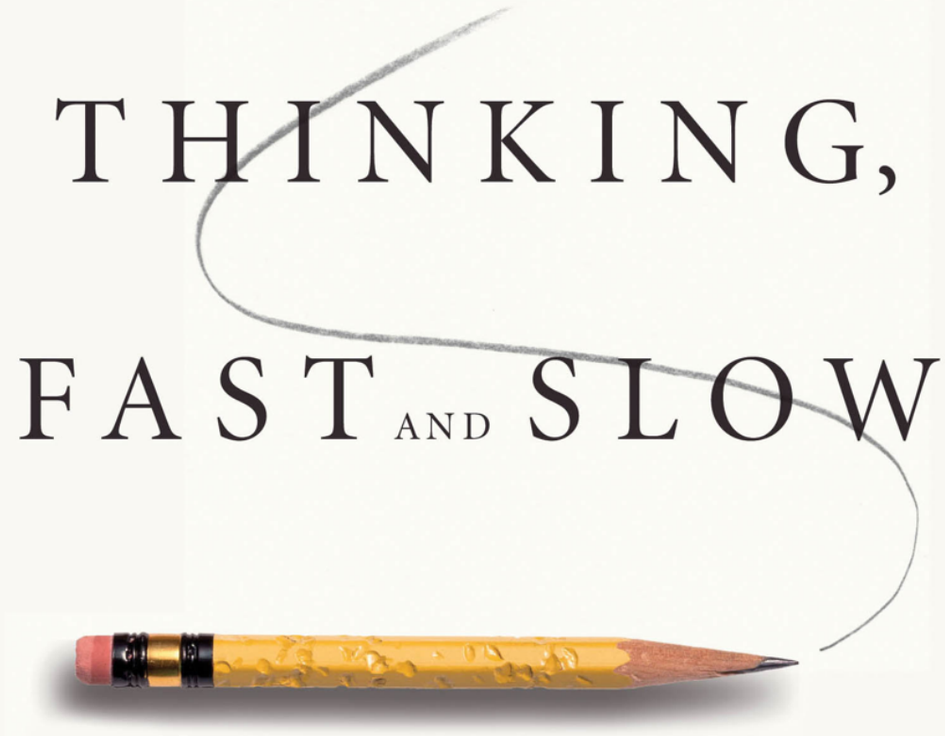
‘you can’t manage what you don’t measure’

Why do we need to measure things?

> ...when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind...
>
> —Lord Kelvin

Cognitive biases

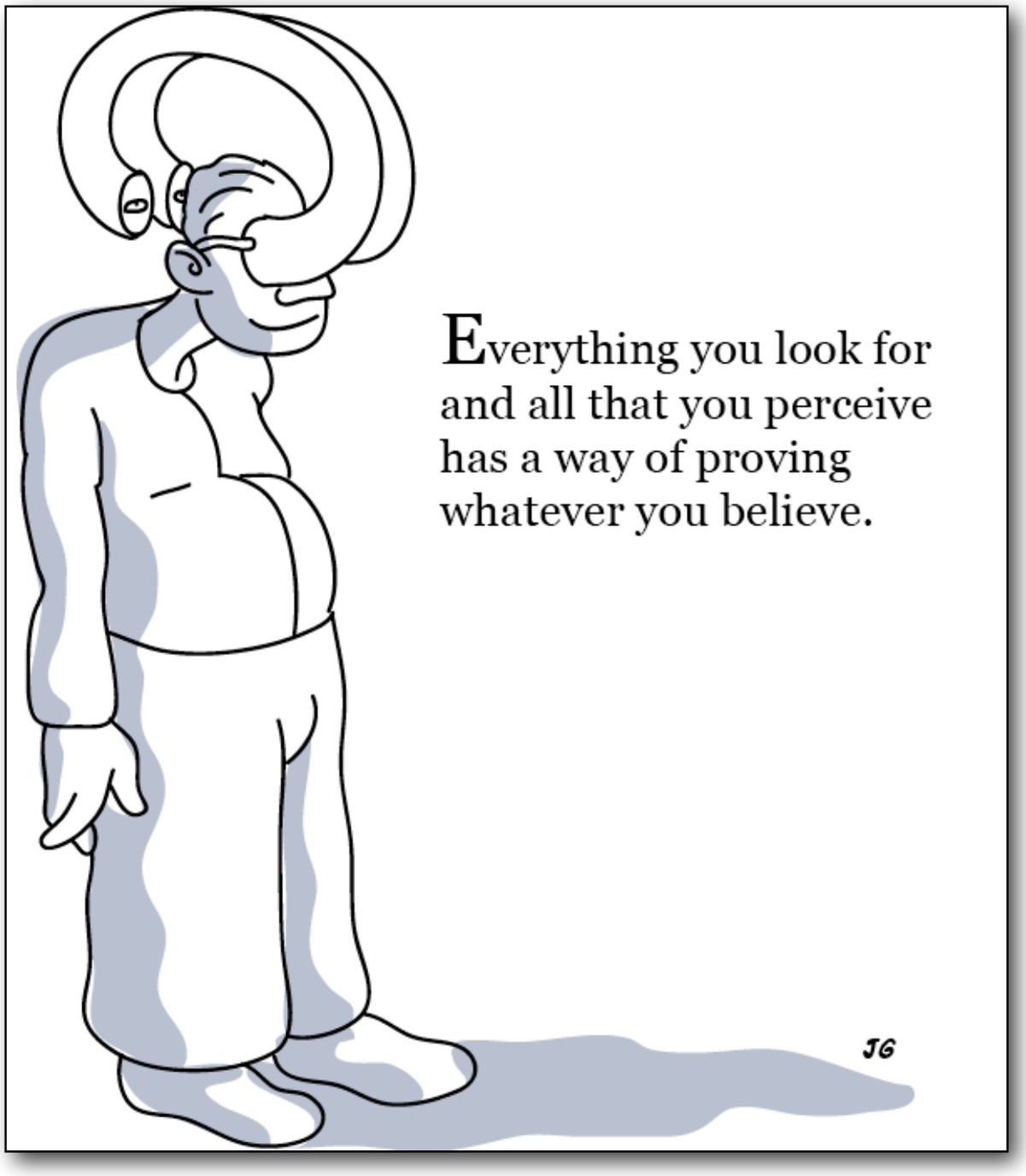
THINKING, FAST AND SLOW



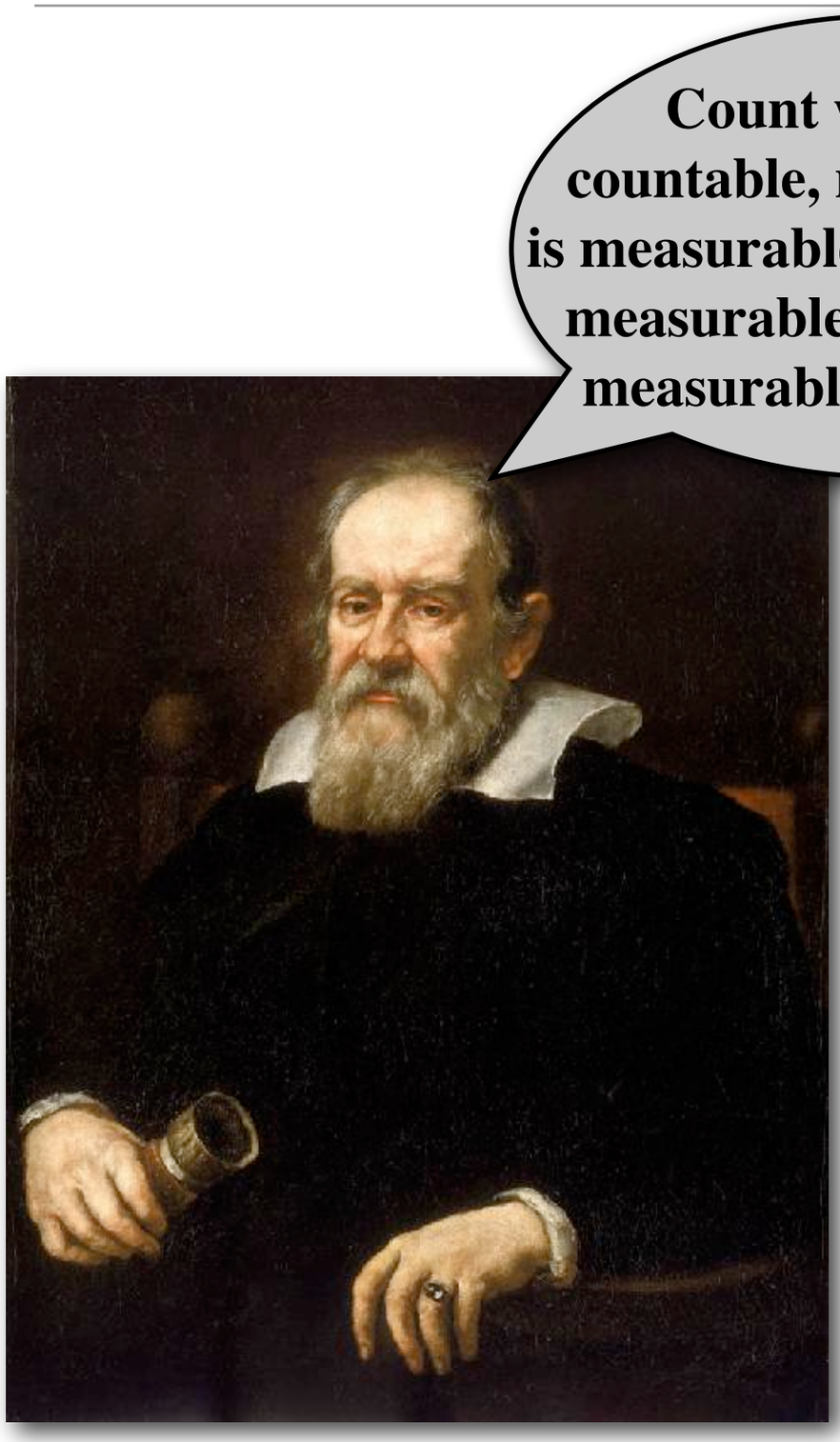
DANIEL KAHNEMAN

WINNER OF THE NOBEL PRIZE IN ECONOMICS

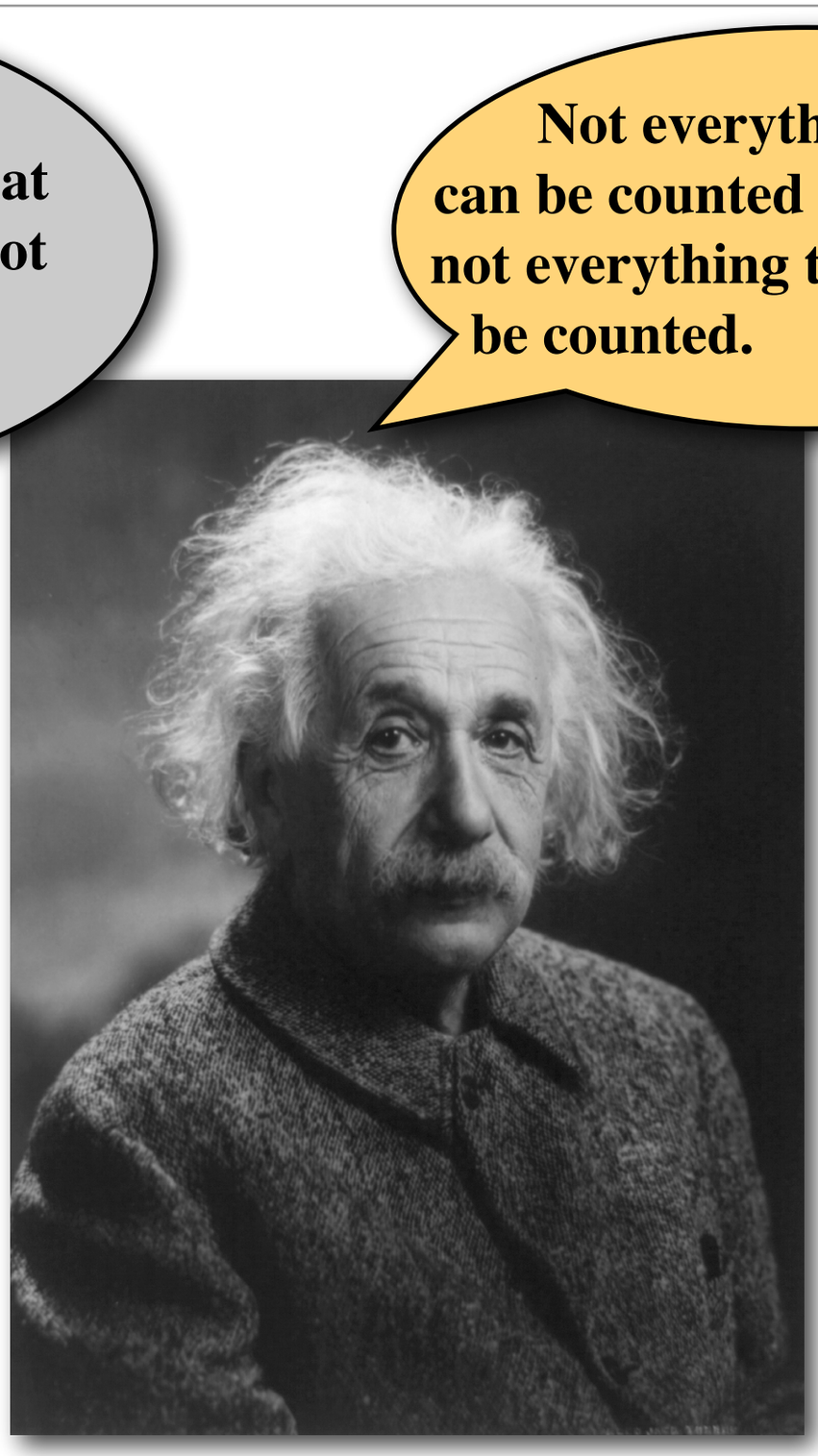
Confirmation bias



You can’t manage what you don’t measure



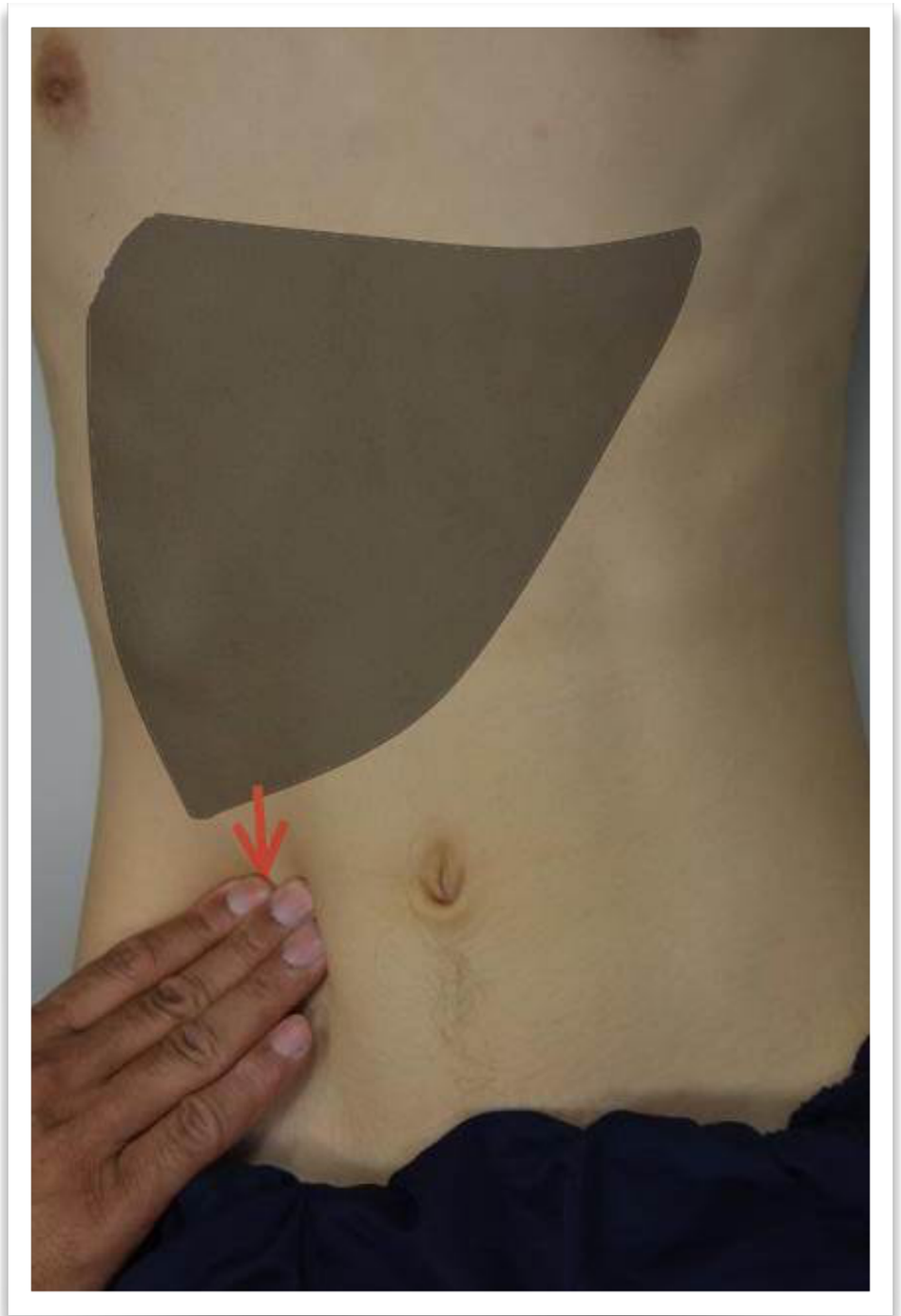
Count what is countable, measure what is measurable. What is not measurable, make measurable.



Not everything that can be counted counts, and not everything that counts can be counted.

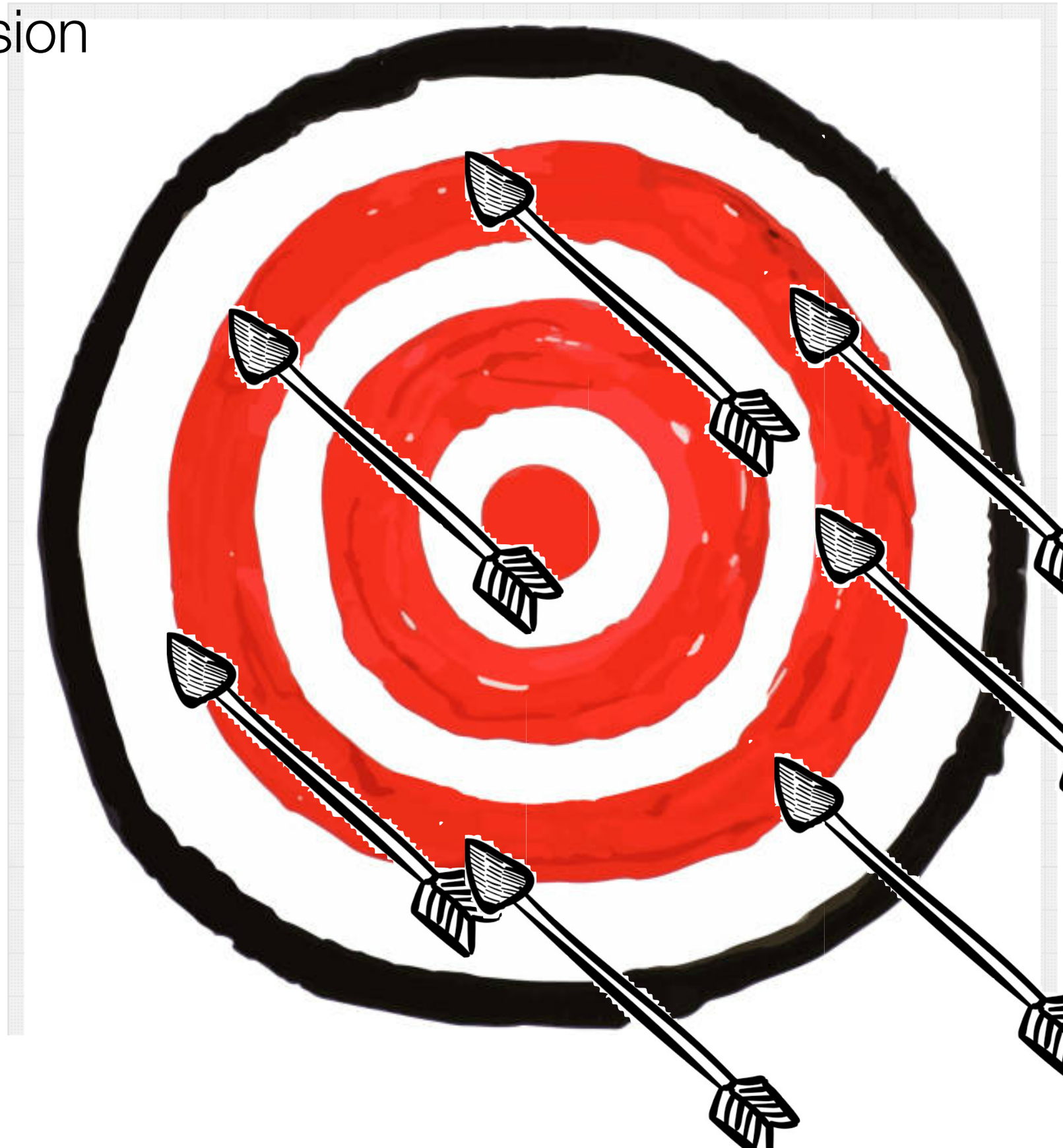
1. You can’t manage what you don’t measure

We want to know the TRUE size of the liver

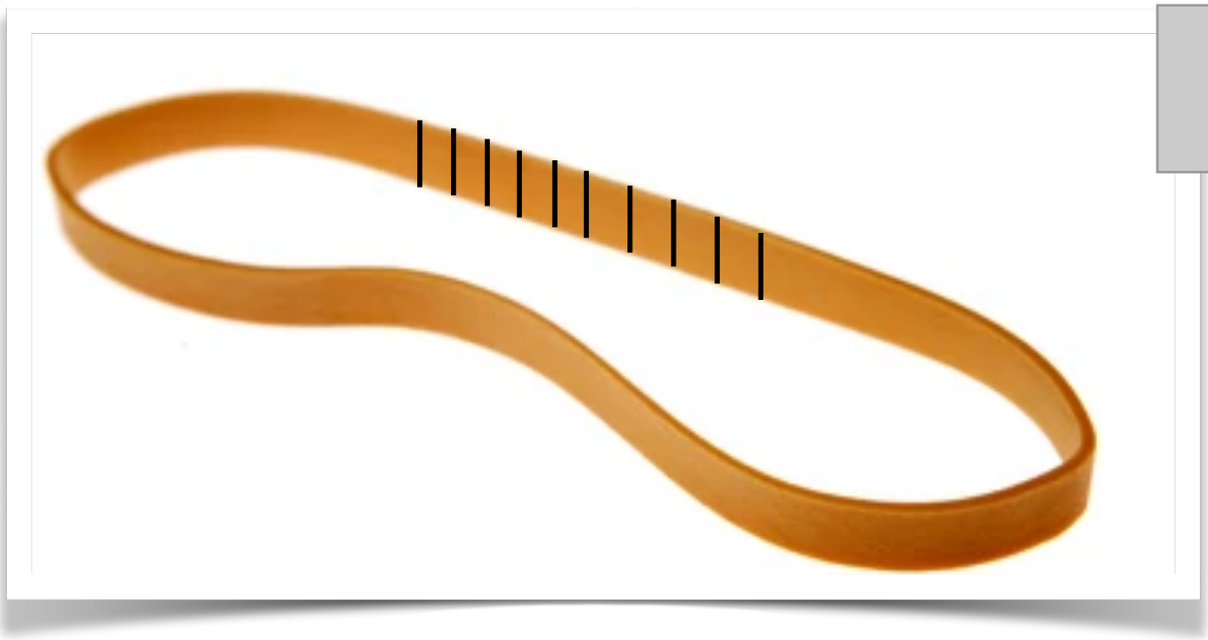


https://clinipedia.wordpress.com/2013/07/11/liver-method-of-examination/

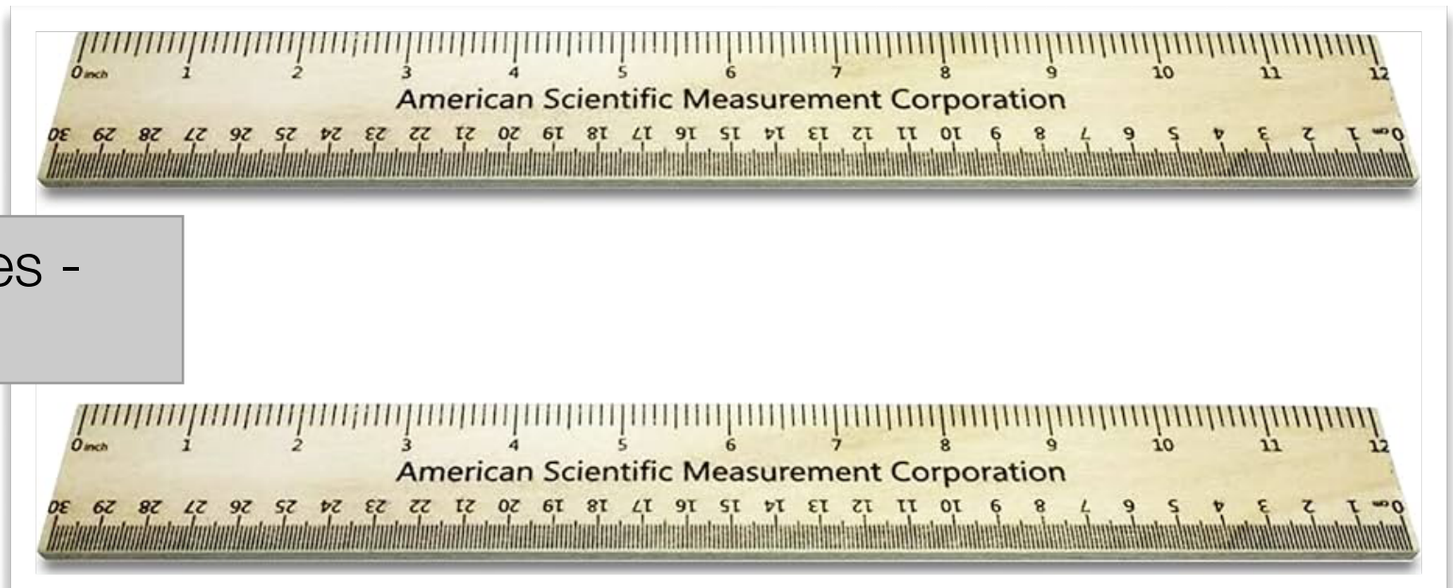
Imprecision



Reliability (precision)

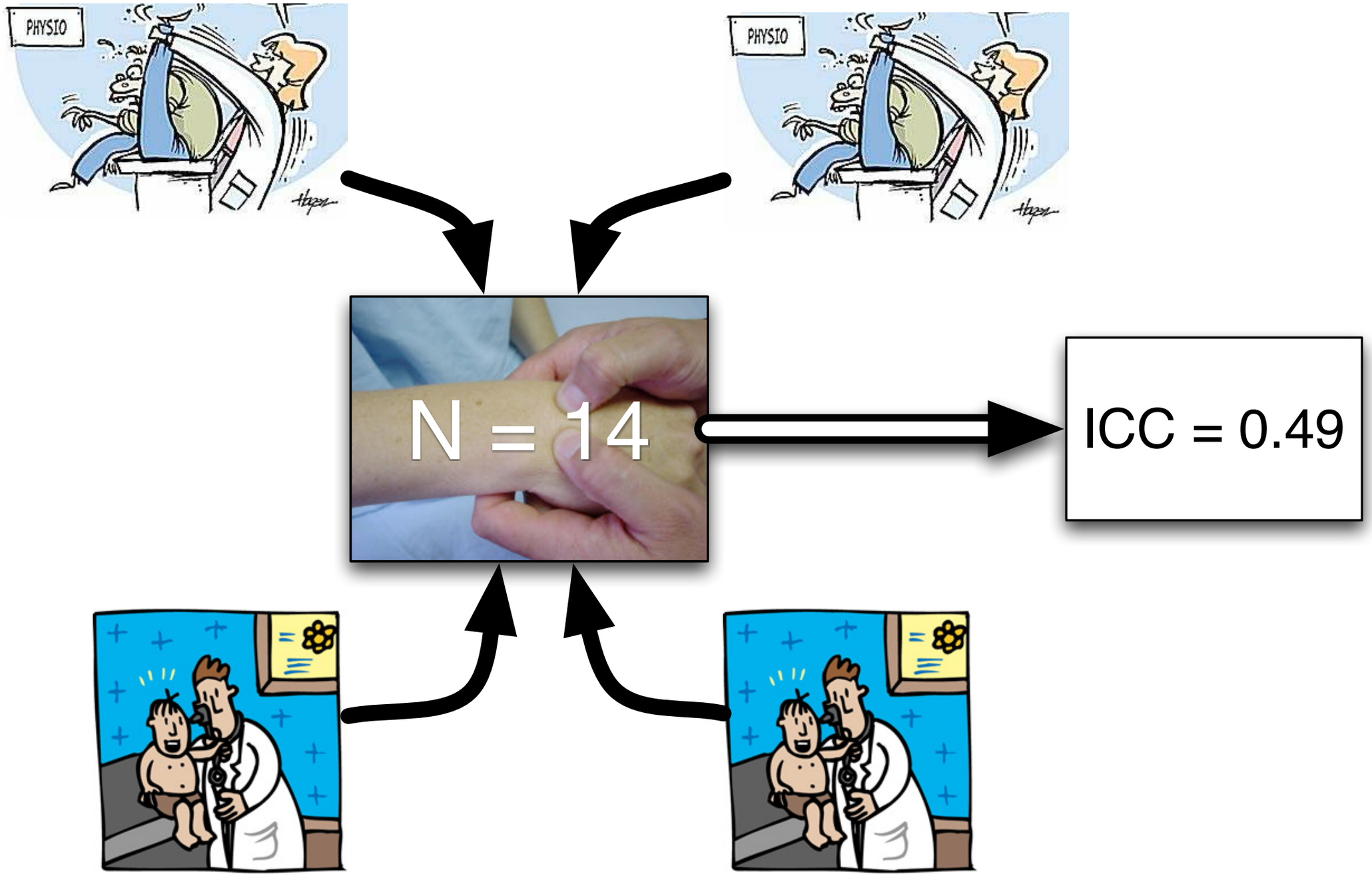


Each time you measure - INTRA rater



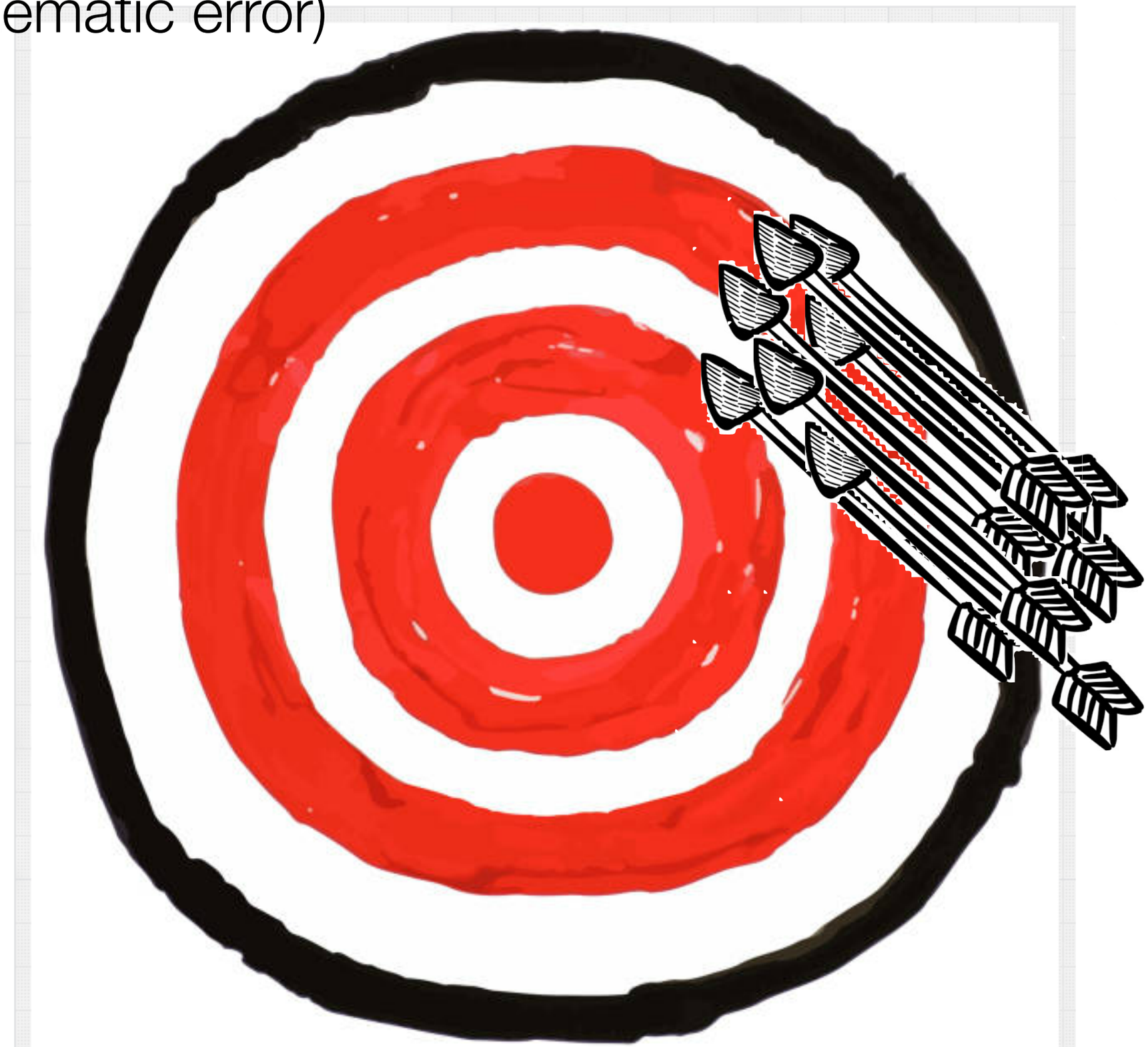
Each time someone else measures - INTER rater

Joint counts (Structure and Function)

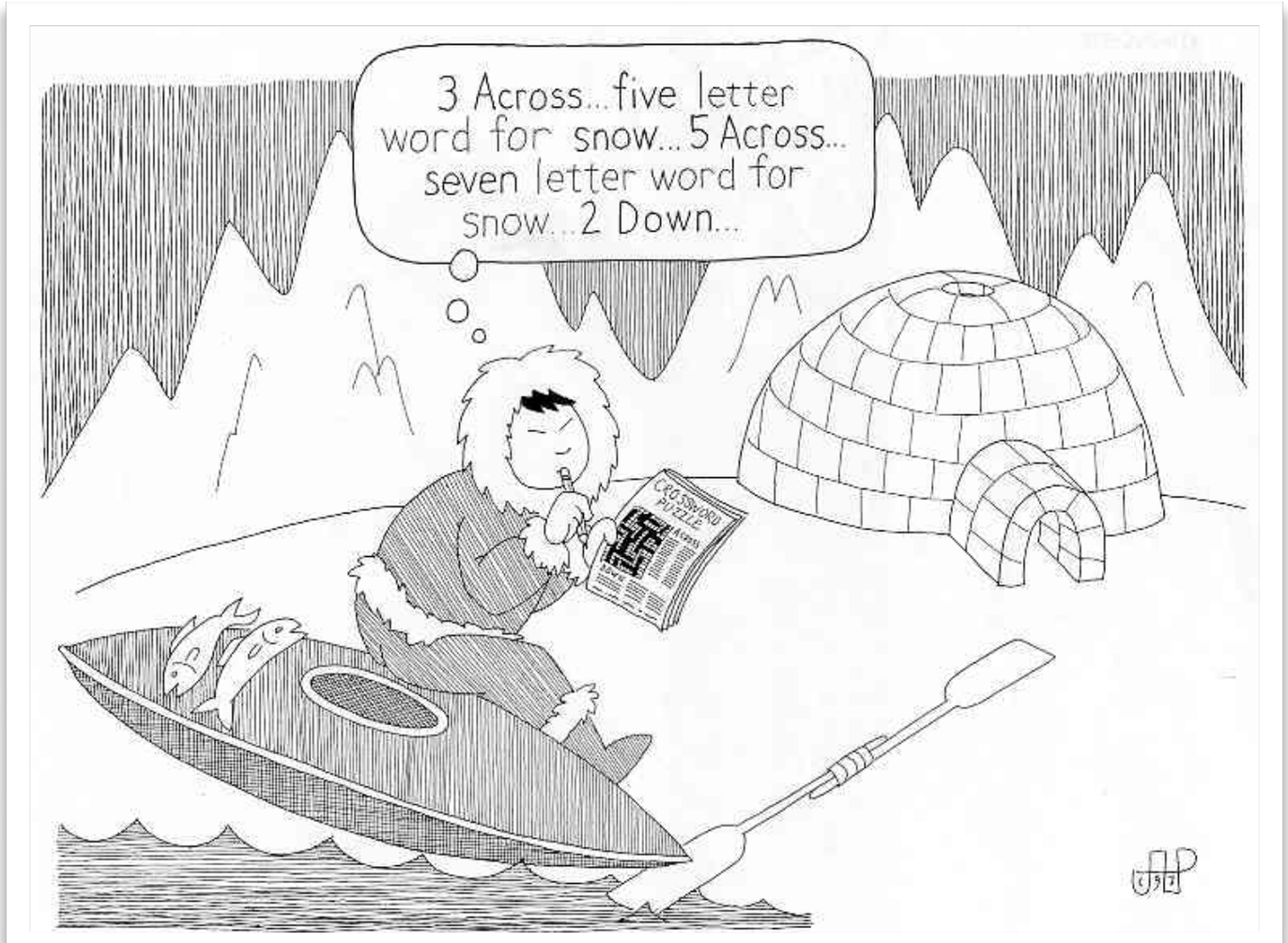


Whitney-Mahoney KJ, Guzman J, Feldman BM: The inter-rater reliability among health care professionals in the detection of joint swelling in children with juvenile arthritis. American College of Rheumatology. Annual Scientific Meeting, Boston, MA. Arthritis and Rheumatism 1999: 42 (9) S: pp S186

Bias (systematic error)



Validity (Truthfulness)

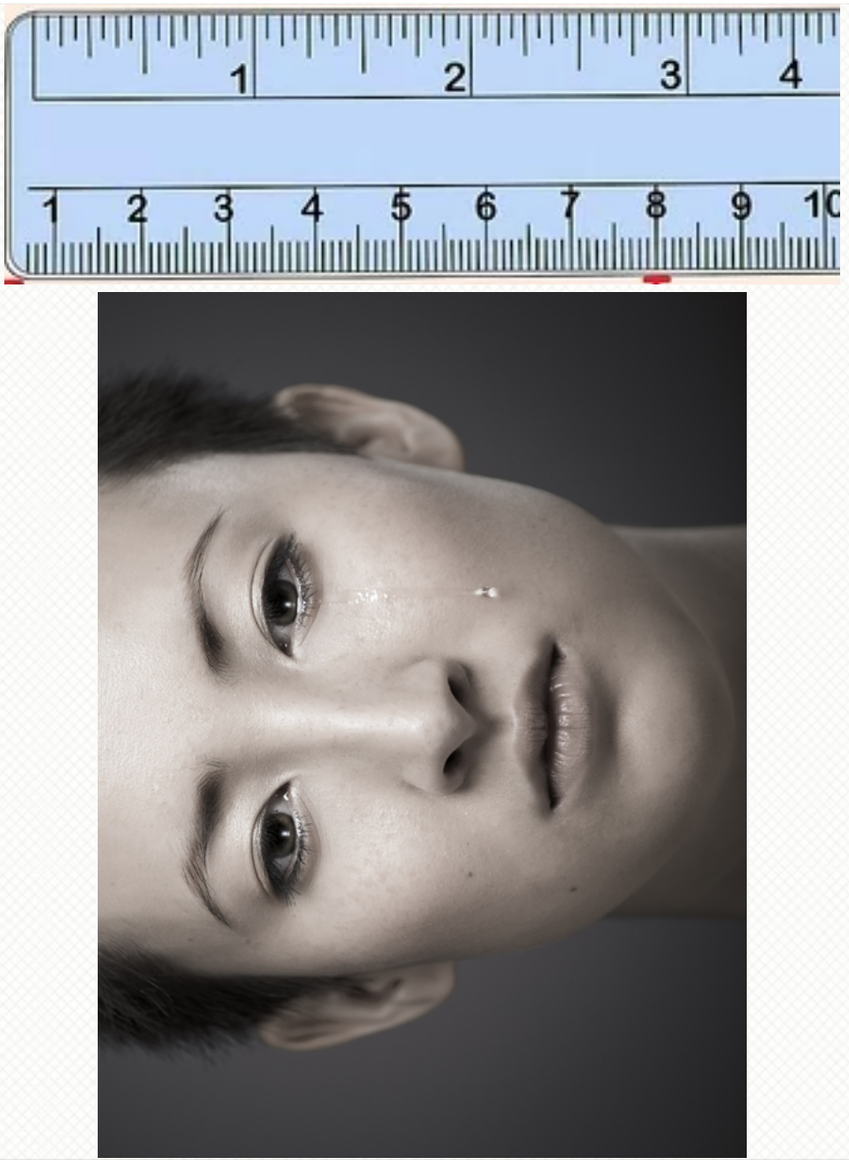


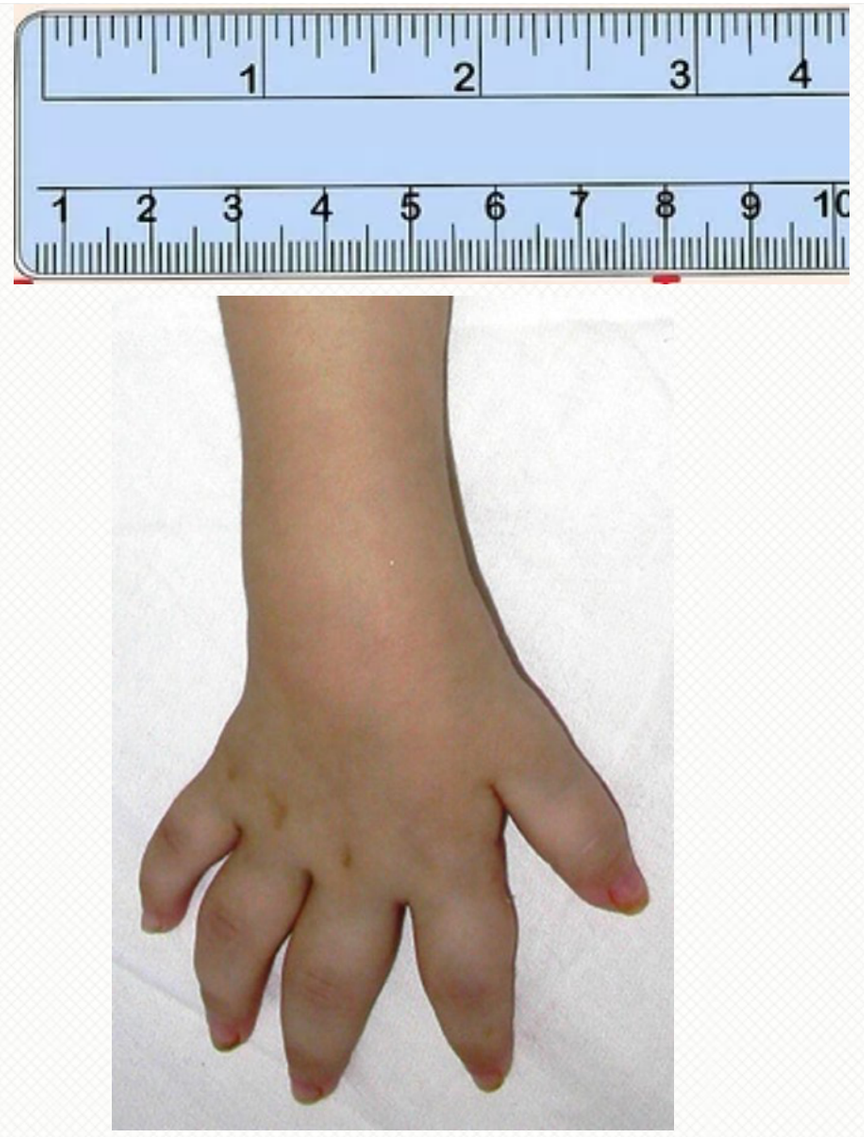
https://nunawhaa.wordpress.com/category/100-words-for-snow/page/4/

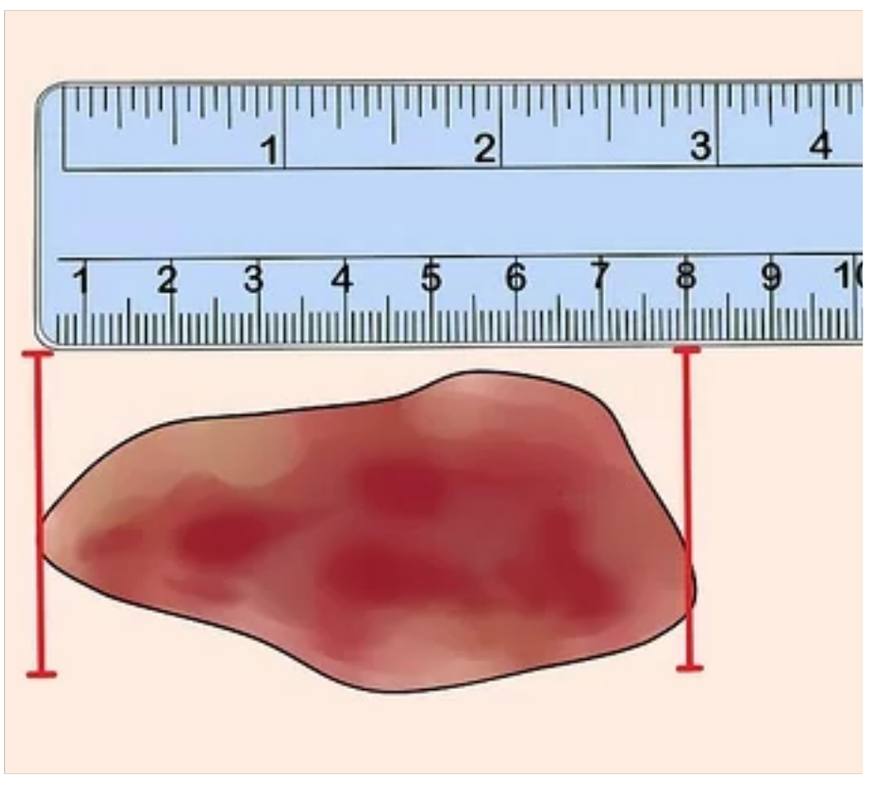
Validity!

Face
Content
Construct
Criterion
Responsiveness

2. Measurement tools must give us the truth.







Preference-Based Measurement of Health-Related Quality of Life (HRQL) in Children with Chronic Musculoskeletal Disorders (MSKDs)

H. I. Brunner, D. Maker, B. Grundland, N. L. Young, V. Blanchette, A-M. Stain, B. M. Feldman

Background. *Health-related quality of life can be measured by patients' health preferences (utilities or values). No method for measuring health state preferences has been standardized for children with arthritis or other musculoskeletal disorders (MSKDs). Such a method is needed for economic evaluations of current and new pediatric treatments.* ***Objectives****. 1) To assess the feasibility of utility measurements in children with MSKDs, 2) to test the validity of the Health Utility Index (HUI) for these children, 3) to assess whether rating scale values can be mathematically converted into meaningful standard gamble (SG) utilities, and 4) to study whether parents can act as proxies for their children with respect to health state preferences.* ***Methods.*** *Eighty parents of children with MSKDs were consecutively sampled. Their children, if 8 years of age or older (n = 55), were studied concurrently. Utilities of current health states were obtained by using the SG and the HUI in random order. In addition, health state preferences were assessed using categorical and analog rating scales. Traditional nonutility measures of health status (the Childhood Health Assessment Questionnaire [CHAQ] and the Activities Scale for Kids [ASK]) were also completed. Intraclass correlation coefficients (ICCs) were calculated to assess concordance between the different utility measures and also between the ratings of the parents and their children.* ***Results.*** *Children 8 years of age or older were able to express the strength of their health state preferences using the HUI and rating scales. Children older than 10 years of age were able to use the SG method. The health state utilities of the parents were higher than those of their children. The utilities varied widely depending on the elicitation method. The expected high agreement between the SG and the HUI was not found (ICC = 0.028 for parents, ICC = 0.016 for patients). Unlike the SG, the global utilities derived from the HUI agreed better with preferences derived from rating scales (ICC = 0.23–0.25) and correlated with traditional health status measures (with CHAQ, r = –0.56; with ASK, r = 0.46) both for parents and children. It was not possible to mathematically convert rating scale preferences into SG utilities. The SG utilities were unrelated to results from the rating scales, the CHAQ, and the ASK. Especially for parents, the SG utilities were very high, even when ratings of the other measures indicated poor health.* ***Conclusions.*** *Although it is possible to measure health utilities for children with MSKDs, the results are highly method dependent. The properties of the HUI in this population are more like those of the traditional health status measures rather than those of the SG. Preferences derived from rating scales, although easily performed, cannot readily be converted into SG utilities. Parents' ratings for their children are impaired by risk aversion.* ***Key words:*** *health-related quality of life; utility; standard gamble; Health Utilities Index; visual analog scale; child.* ***(Med Decis Making 2003;23:314–322)***

Received 14 November 2000 from the William Rowe Division of Rheumatology, Children's Hospital Medical Center, University of Cincinnati, Cincinnati, OH (HIB); and the Divisions of Rheumatology (DM, BMF), Hematology/Oncology (VB, AMS), and Paediatric Medicine (NLY), Hospital for Sick Children, Department of Pediatrics, Public Health Sciences and Health Policy, Management & Evaluation (NLY, BMF), University of Toronto, Toronto, Canada. Dr. Young is supported by the Hospital for Sick Children Foundation. Dr. Feldman holds the Canada Research Chair in Childhood Arthritis. Ms. Maker was supported by an Easter Seals summer studentship. We would like to express our appreciation to Hilary Llewellyn-Thomas for her helpful comments on an earlier version of this article. Revision accepted for publication 16 February 2003.

Chronic musculoskeletal disorders (MSKDs) in children can result from a variety of diseases, such as juvenile rheumatoid arthritis, juvenile dermatomyositis, hemophilia, sickle cell disease, and orthopedic conditions. The treatment regimens may be

Address correspondence and reprint requests to Brian M. Feldman, MD, MSc, FRCPC, Hospital for Sick Children, Division of Rheumatology, 555 University Avenue, Toronto, ON M5G 1X8, Canada; e-mail: brian.feldman@sickkids.on.ca.

DOI: 10.1177/0272989X03256008

British Journal of Cancer (2003) 88, 1185 – 1190

www.bjcancer.com

alidation of the Health Utilities Index and the Child stionnaire in children undergoing cancer py

Greenberg[2], JJ Doyle[2], NL Young[1,3], S Ingber[3], J Rubenstein[3], J Wong[3], T Samanta[3], d BM Feldman[1,3,4,5]

Policy Management and Evaluation, University of Toronto, Toronto, Ontario, Canada; [2]Division of Hematology/Oncology, Hospital , Ontario, Canada, M5G 1X8; [3]Department of Pediatrics, Population Health Sciences Program, Hospital for Sick Children, , M5G 1X8; [4]Division of Rheumatology, Hospital of' Sick Children, Toronto, Ontario, Canada, M5G 1X8; [5]Department of Public y of Toronto, Toronto, Ontario, Canada

tudy was to evaluate the construct validity of two questionnaire-based measures of health-related quality of life ndergoing cancer chemotherapy: the Health Utilities Index (HUI) and the Child Health Questionnaire (CHQ). n hospitalised for chemotherapy. To examine construct validity: (1) *a priori* expected relations between CHQ tributes were examined; (2) HUI and CHQ summary scores were compared to visual analogue scale (VAS) etion was rated using a 5-point categorical scale and completion time was recorded. A total of 36 subjects were m score was seen in 15 (47%) of HUI3 assessments. As predicted, CHQ body pain was moderately correlated 0.51), CHQ physical functioning was moderately correlated with HUI2 mobility ($r = 0.58$) and CHQ mental correlated with HUI2 emotion ($r = 0.53$). Only the CHQ psychosocial subscale (and not HUI) was correlated The CHQ and the HUI were both easy to use. The HUI questionnaires required less time to complete min) compared with CHQ (mean = 13.1, s.d. = 3.4 min, $P < 0.0001$). In conclusion, HUI and CHQ uct validity in children undergoing cancer chemotherapy. The Health Utilities Index is subject to a ceiling requires more time to complete.

er (2003) **88,** 1185 – 1190. doi:10.1038/sj.bjc.6600895 www.bjcancer.com
arch UK

erapy; pediatric; health-related quality of life; Health Utilities Index; Child Health Questionnaire

of life (HRQL) of children undergoing is becoming increasingly emphasised in ing differences in HRQL associated with egies may be particularly helpful to workers when these strategies are rvival.

generic and disease-specific instruments ssment of HRQL (Guyatt *et al*, 1993). The sures is that they provide a rating of comparisons across illnesses and often data (Spieth and Harris, 1996). However, responsiveness in specialised clinical n undergoing chemotherapy for cancer. nd a comparison of generic instruments in this specific group.

We evaluated two questionnaire-based measures of HRQL, the Health Utilities Index (HUI) and the Child Health Questionnaire (CHQ). The rationale for choosing these measures is that both the HUI and CHQ have been used to evaluate survivors of childhood cancer (Feeny *et al*, 1992, 1993; Billson and Walker, 1994; Kiltie and Gattamaneni, 1995; Glaser *et al*, 1997, 1999a, b; Barr *et al* 1999; Sawyer *et al*, 1999; Speechley *et al*, 1999; Felder-Puig *et al*, 2000; Sands *et al*, 2001) and they were both incorporated into a cross-Canada study of the long-term psychosocial and physical health of childhood cancer survivors (The Late Effects Study) (Gibbons *et al*, 1994). Additionally, the HUI has been included in every major Canadian population health survey since, 1990 (Furlong *et al*, 2001). These questionnaires therefore allow comparison of HRQL between children receiving cancer chemotherapy and long-term cancer survivors as well as enabling comparisons to population estimates of health.

Although HUI and CHQ both measure HRQL, they have important differences. They are based on different theoretical approaches with the HUI being a utility-based measure of overall HRQL, while the CHQ is a health-profile measure using summative categorical scaling to determine separate scores in two subscales. Also, different frameworks are used. The HUI uses a narrow 'within the skin' approach to the measurement of HRQL and does not include social interactions (which are considered to reflect phenomenon other than the strict health of the individual) (Feeny *et al*, 1996). Conversely, the CHQ is broader in scope and

*Correspondence: Dr L Sung, Division of Hematology/Oncology, Hospital for Sick Children, Toronto, Ontario, Canada, M5G1X8;
E-mail: Lillian.sung@sickkids.ca
Received 10 June 2002; revised 3 January 2003; accepted 22 January 2003

Clinical

Journal of Clinical Epidemiology

Journal of Clinical Epidemiology 56 (2003) 848–855

of a modified standard gamble elicited from parents of a hospital-based cohort of children

L. Greenberg[a], N.L. Young[b], M. McLimont[b], S. Ingber[b], J. Rubenstein[b], ong[b], T. Samanta[b], J.J. Doyle[a], A.M. Stain[a], B.M. Feldman[b,c]

[a]*Division of Hematology/Oncology, Hospital for Sick Children, Toronto, Ontario, M5G 1X8*
[b]*Population Health Sciences, Hospital for Sick Children, Toronto, Ontario, M5G 1X8*
[c]*Division of Rheumatology, Hospital for Sick Children, Toronto, Ontario, M5G 1X8*

Accepted 19 May 2003

the validity of a modified standard gamble (Mod SG) (nondeath baseline) by comparing these scores to SG off (TTO), visual analog scale (VAS), Health Utilities Index (HUI), and Child Health Questionnaire (CHQ). were parents of in-patients with cancer receiving chemotherapy and parents of children without cancer attending validity was determined by comparing *a priori* hypotheses to actual correlations between measures. Discriminant anticipating that in-patients with cancer would have lower HRQL than outpatients.
re included. Both Mod SG and SG were moderately correlated with TTO ($r = 0.50$ and $r = 0.49$; $P < .01$ for G were moderately correlated with TTO ($r = 0.47$ and $r = 0.05$, $P < 0.002$ for both).
G did not perform better than SG. Two nonoverlapping groups of HRQL measures were demonstrated. © 2003 erved.

ty of life; Standard gamble; Visual analog scale; Utility; Validity; Children

accepted method for measurement gamble (SG) [1]; the SG is consid-old standard for utility elicitation. the validity of the SG related to a among respondents or "gambling , several studies have found poor nd other measures of health-related in adult respondents [5–8]. This be particularly pronounced when dents for their children, an area of relatively little attention. A recent study of children with musculoskeletal disorders found that SG utilities were not related to other measures of HRQL, such as the Health Utilities Index (HUI) and categoric and analog rating scales [9]. The SG utilities were found to be very high, even when the other ratings indicated poor health. The authors postulated that the parents were unwilling to "gamble" with death as a possible outcome, and that this risk aversion affected the validity of the SG.

If the poor performance of the SG is related to risk aversion, and an unwillingness to "gamble" with death as a possible consequence, then one way to improve the SG might be to use a nondeath bottom anchor. In this modified SG (Mod SG) the respondent would choose between remaining in the current health state or taking a lottery between perfect health and an undesirable but nondeath state. The resultant utility would then need to be recalibrated to the traditional death baseline scale to make the results comparable to results obtained using the traditional SG. This method is often referred to as a chained utility elicitation, and is commonly used in the elicitation of preferences.

Our objective was to determine whether a Mod SG was a more valid measure of parent assessed HRQL compared to the SG. We hypothesized that the Mod SG, when compared to SG, should be more similar to other measures of HRQL, and should be better able to discriminate between in-patient

* Corresponding author. Tel.: 416-813-5977; fax: 416-813-5327.
E-mail address: Lillian.sung@sickkids.ca (L. Sung).

doi: 10.1016/S0895-4356(03)00160-4

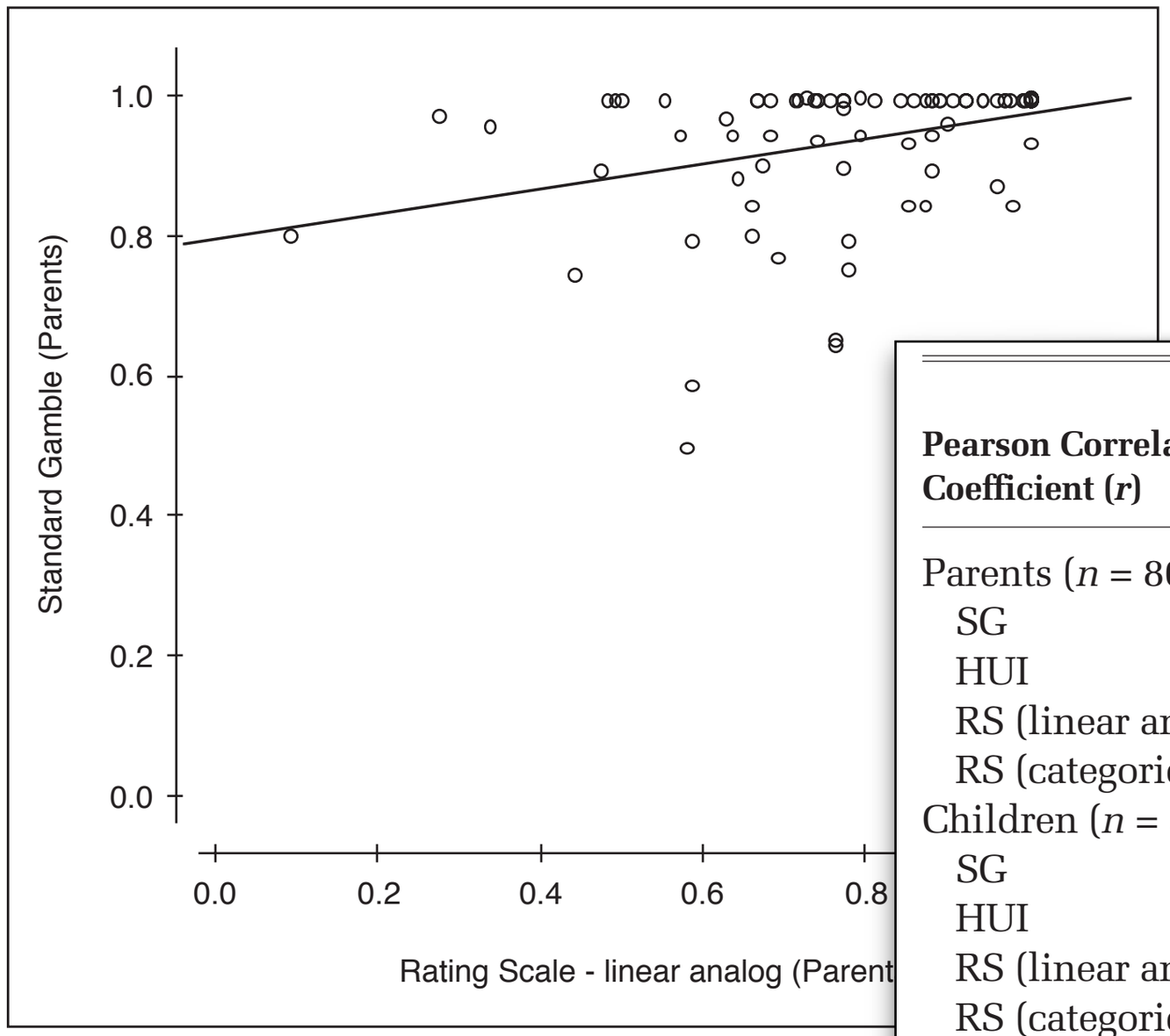


Figure 1 Comparison of standard gamble (SG) utility and the linear analog rating scale value as rated by the parents. The vertical axis shows the SG results on a scale of 0 to 1, where 0 represents the worst possible health state (death) and 1 represents perfect health. The horizontal axis shows the rating scale responses on a scale of 0 to 1, where 0 represents the worst possible health state (death) and 1 represents perfect health.

Pearson Correlation Coefficient (*r*)	SG	HUI	RS (linear analog)	CHAQ	ASK
Parents (*n* = 80)					
SG	1.0			–0.02	0.03
HUI	0.10	1.0		–0.56	0.46
RS (linear analog)	0.32	0.70	1.0	–0.32	0.33
RS (categorical)	0.32	0.64	0.83	–0.41	0.37
Children (*n* = 55)					
SG	1.0			–0.28	0.37
HUI	–0.06	1.0		–0.54	0.72
RS (linear analog)	0.04	0.50	1.0	–0.28	0.27
RS (categorical)	0.13	0.47	0.63	–0.35	0.30

Ontology

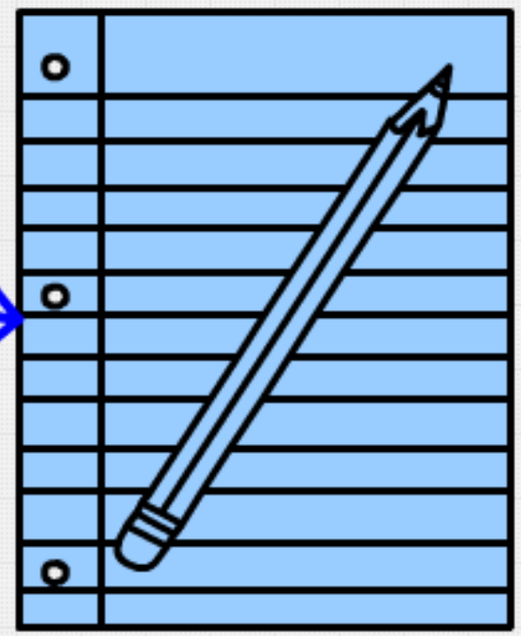
Classical Test

Realism

Anti-Realism

Latent Variable

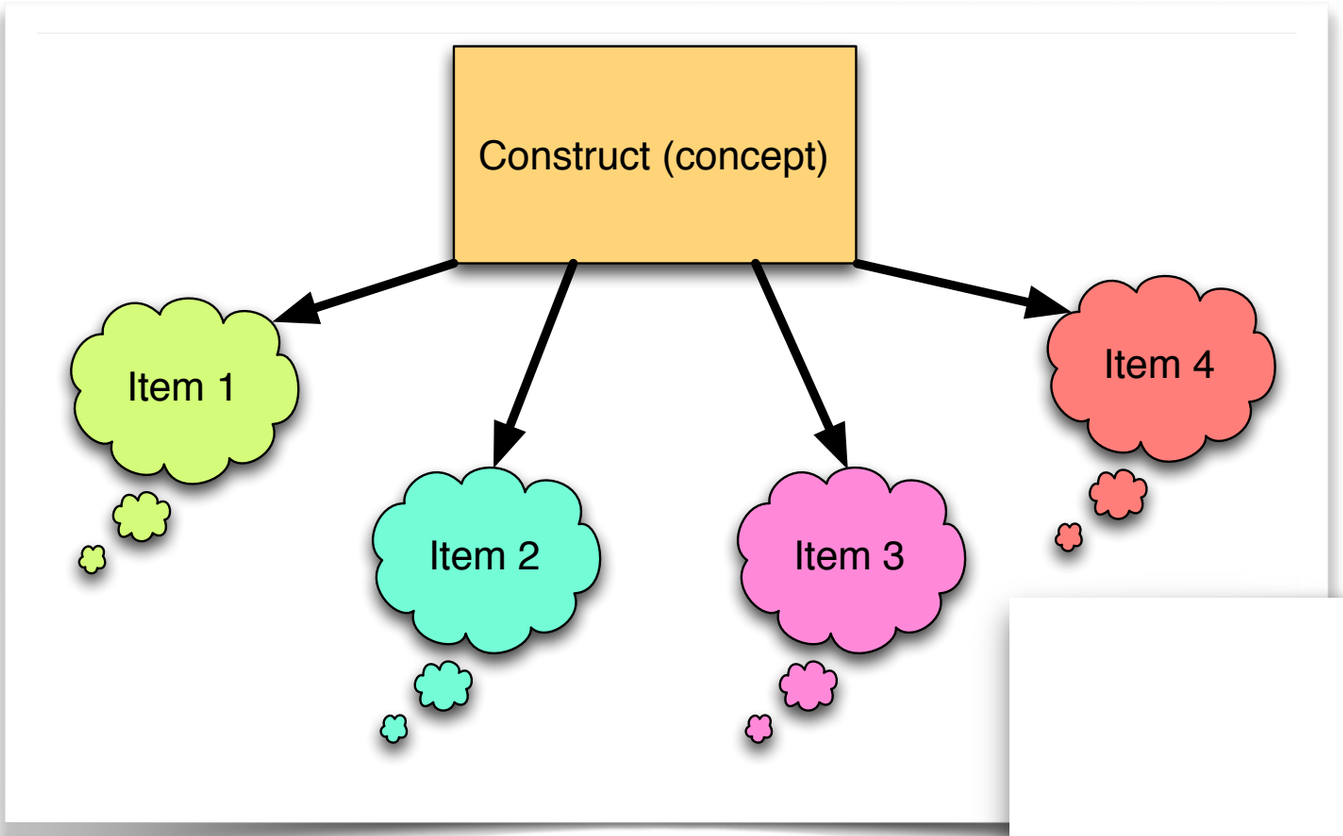
Constructivism

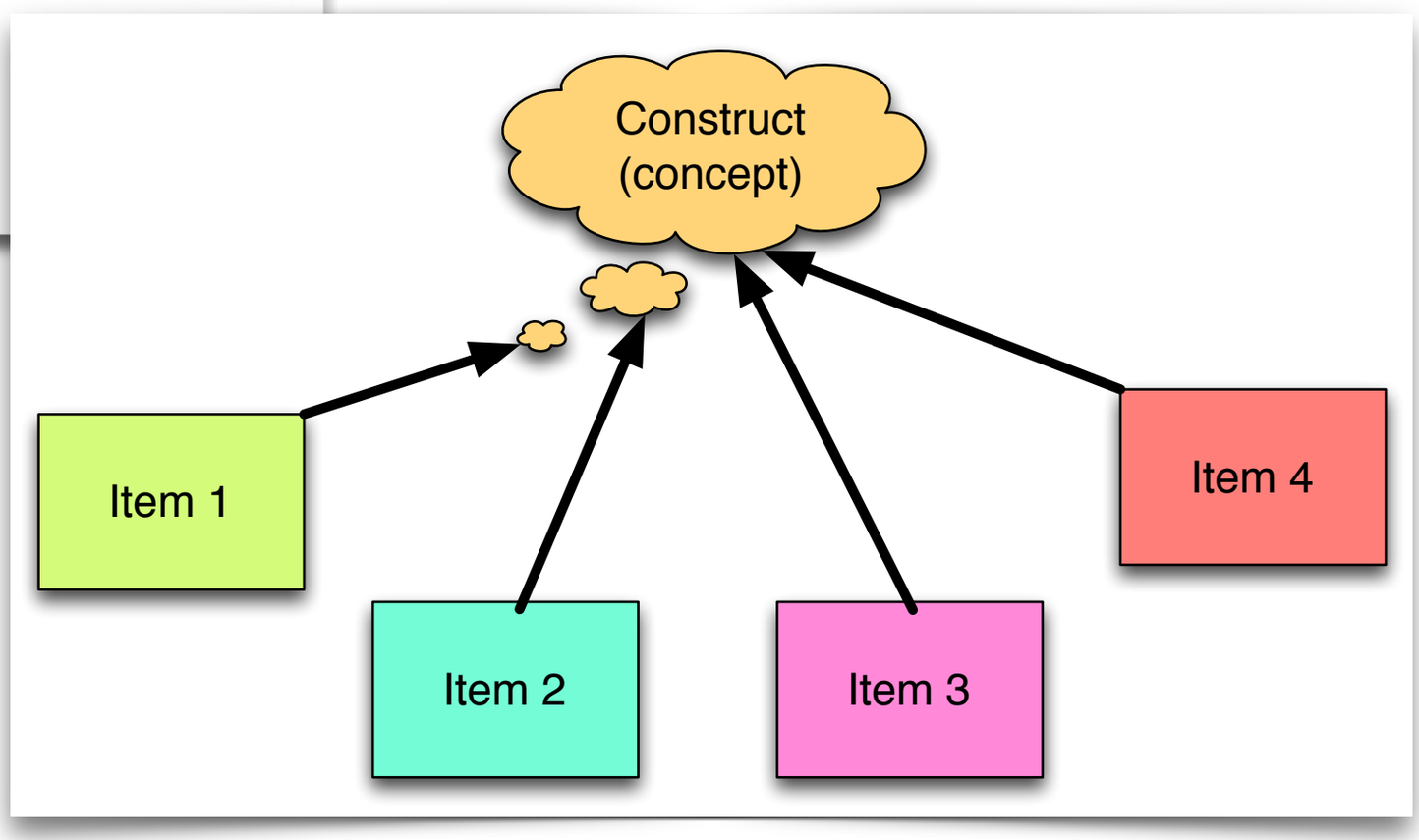


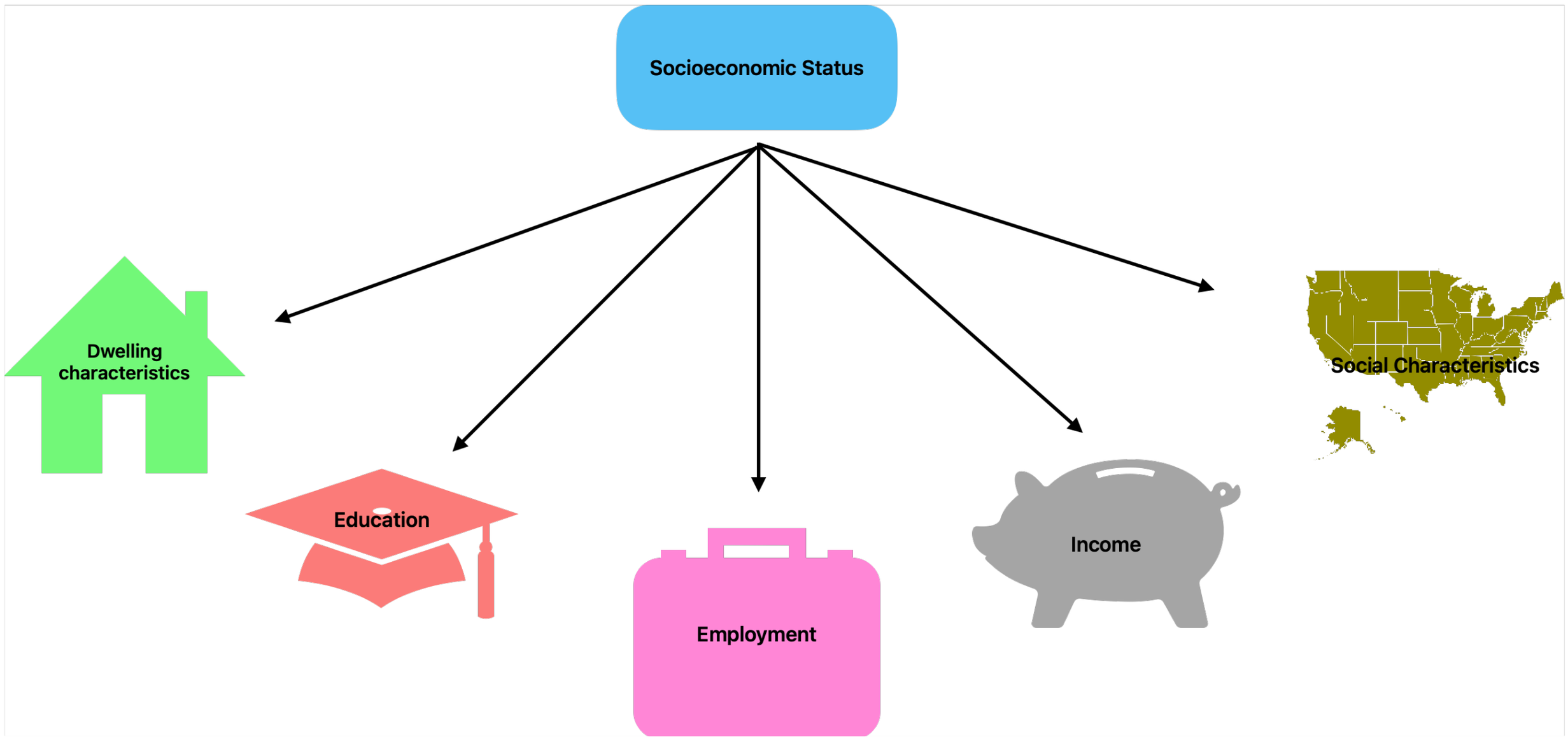
Instrumentalism

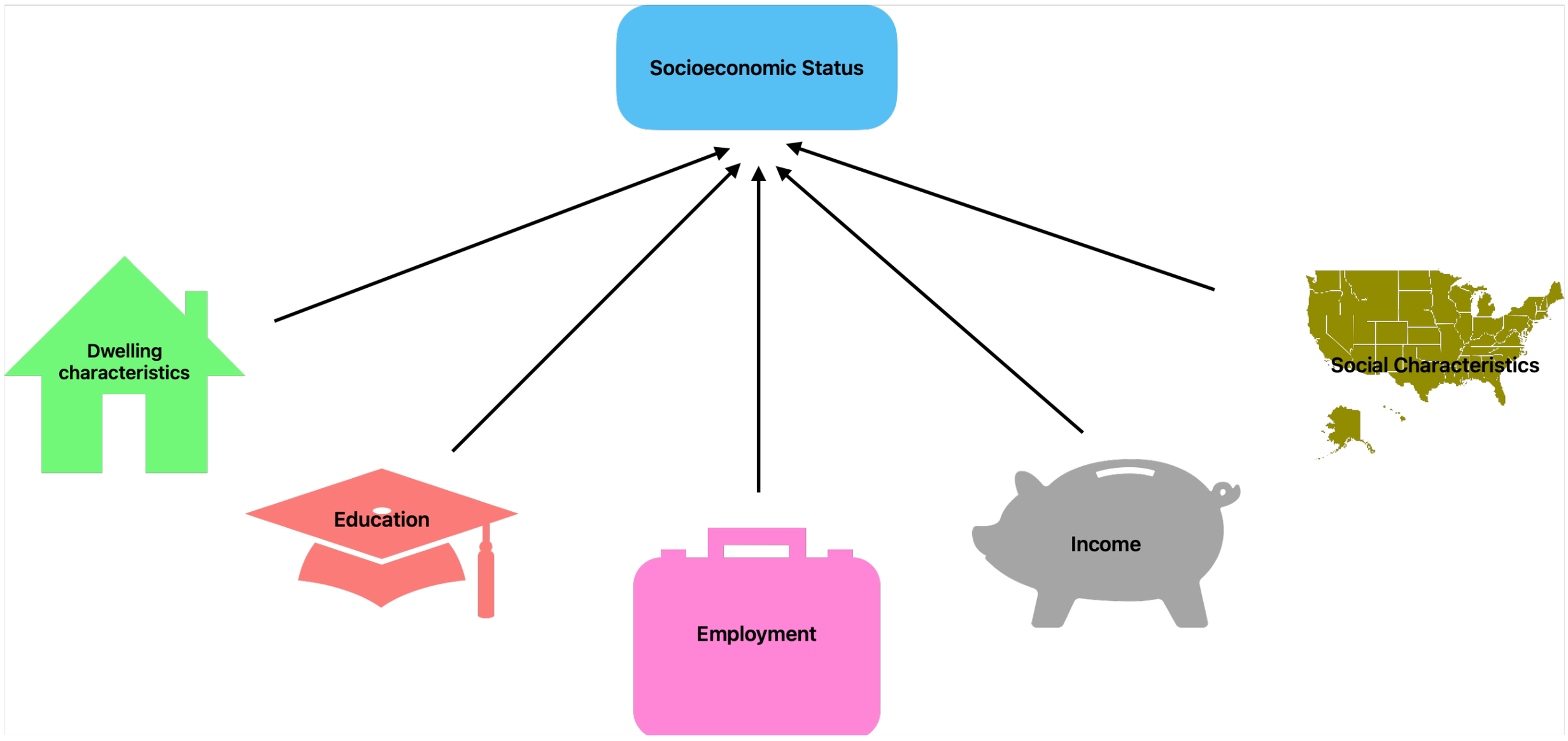

80
60
100
40
120
20
140
0
160
MPH
km/h
292
mph
km/h
MPH
km/h
OUTSIDE
TRIP
0.0

Reflective (psychometric) vs. Formative (clinimetric)






Socioeconomic Status
Dwelling characteristics
Education
Employment
Income
Social Characteristics


Socioeconomic Status
Dwelling characteristics
Social Characteristics
Education
Employment
Income