

Comparing Health Outcomes in Comparative Effectiveness Research: Do preferences matter?

Benjamin M. Craig, Ph.D

H. Lee Moffitt Cancer Center & Research Institute



Comparative Effectiveness Research

“Comparative effectiveness research is designed to inform health-care decisions by providing evidence on the effectiveness, benefits, and harms of different treatment options.”

- Agency for Healthcare Research and Quality

Where we are today: Clinical trials have amassed a wealth of evidence on health outcomes, yet how to summarize and compare outcomes is unclear.

The challenge: *How do we translate what we know about health outcomes into decision-relevant information?*



Tx A

Diarrhea

Fatigue

Depression

**Loss of
Appetite**

Tx B

Vomiting

Rash

Pain

Dizziness

“Which treatment has worse adverse events?” A or B



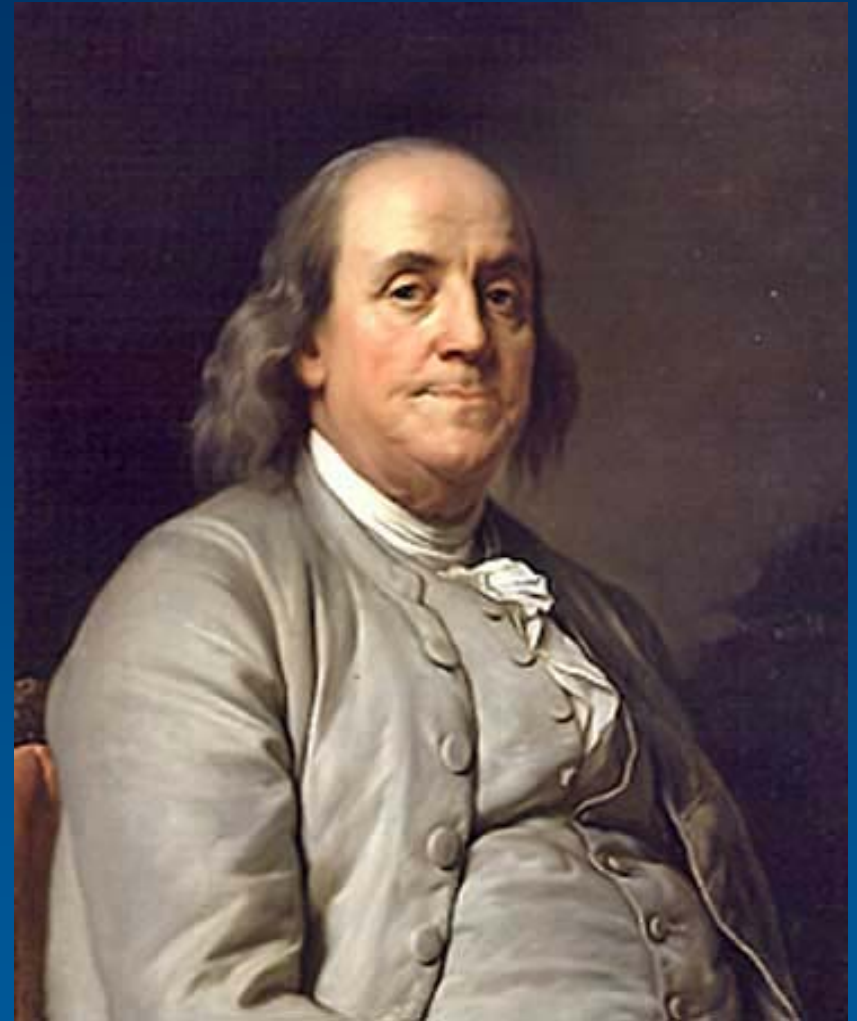
Comparing Health Outcomes In Comparative Effectiveness (CHOICE) Research

- Prudential Algebra
- Measurement of Health Preference
- US Valuation Survey
 - SF-12 and SF-6D Preliminary Results
- Applications
- Discussion



Benjamin Franklin
American Statesman &
Inventor (1706 - 1790)

- Writing to his friend, Joseph Priestley, Franklin described the process of "Moral or Prudential Algebra" to aid Joseph's decision on whether to accept a new position (Sept 19, 1772).
- Today, we know this process as a pro and con list



Pro

Con

~~Salary~~

~~Location~~

~~Colleagues~~

~~Moving~~

Library

~~Staff~~

New Lab

Family

Franklin often recommended this tool for handling personal dilemmas...



April 8, 1779

Dear Jonathan:

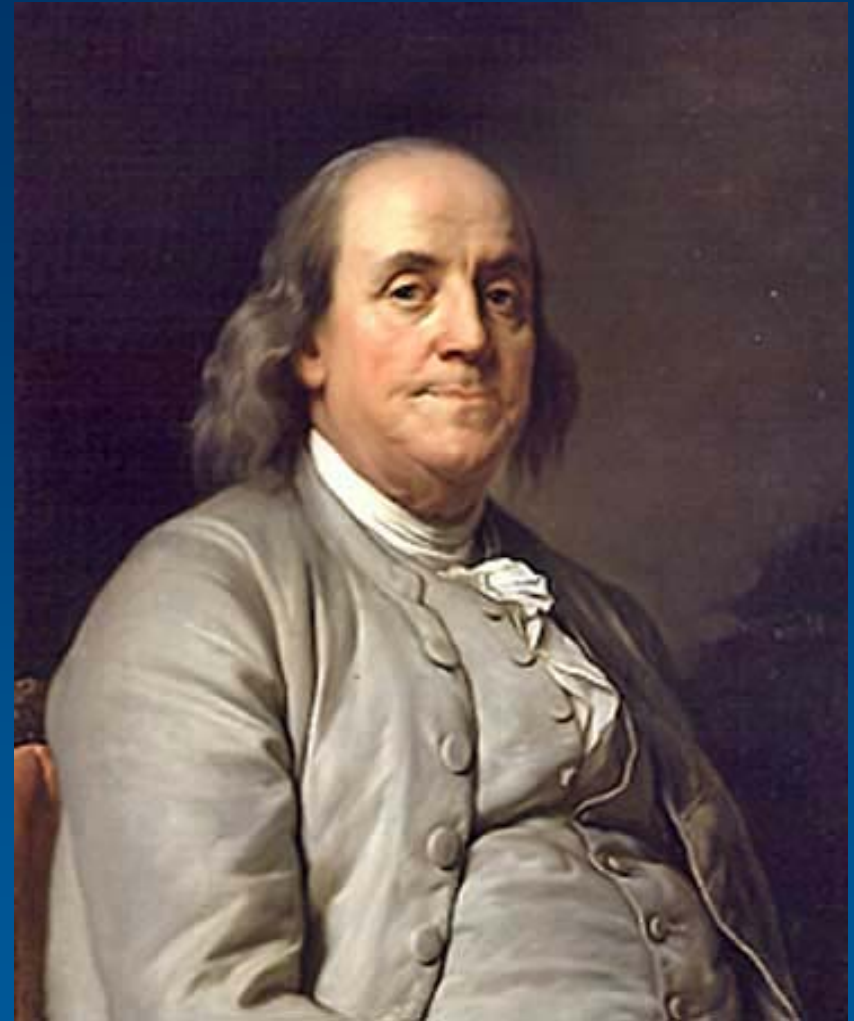
...Follow your own judgment.

If you doubt, set down all the reasons, pro and con, in opposite columns on a sheet of paper...

This kind of Moral Algebra I have often practiced in important and dubious concerns, and though it cannot be mathematically exact, I have found it extremely useful.

By the way, if you do not learn it, I apprehend you will never be married...

I am ever your affectionate uncle,
B. Franklin



Advice to his grand nephew
on whether to marry...



Jeremy Bentham (1748-1832)

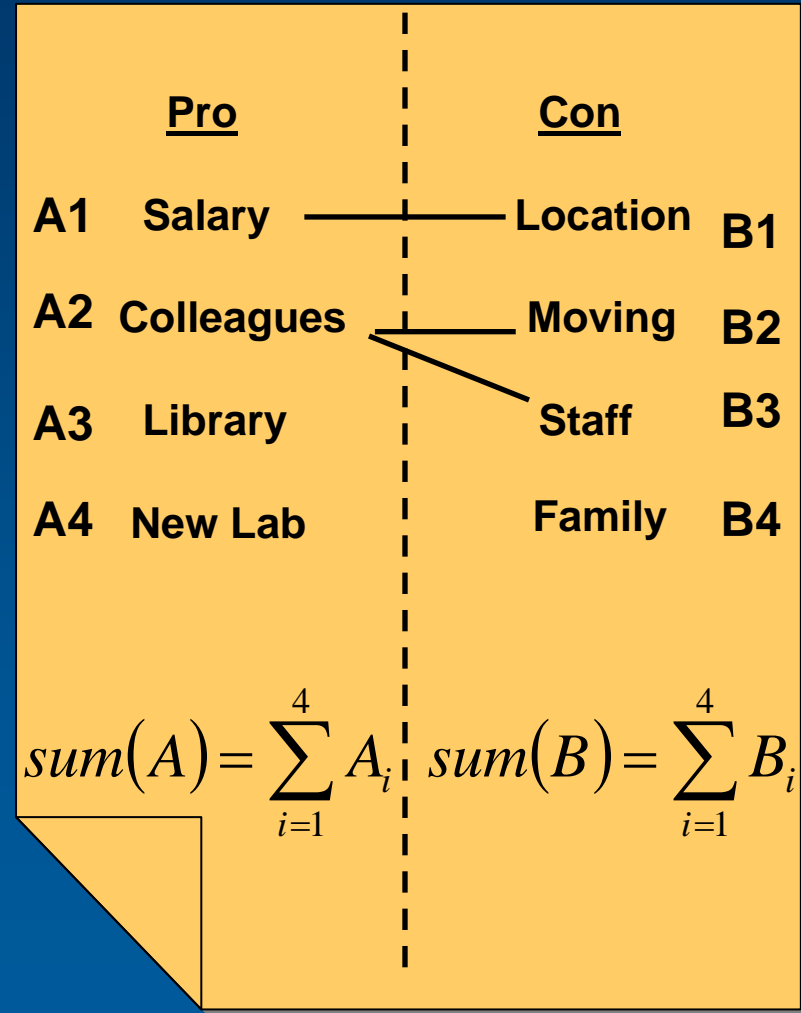
English jurist, philosopher, &
legal and social reformer

- He may be best known for his advocacy of utilitarianism and contributions to the development of welfarism
- "Benjamin Franklin was his source, direct or indirect, for this idea of classification by bipartition plus measurement of the relative weight of the two classes."
 - Viner, AER 1949



Weighting Items

- As Bentham suggested, "pro" and "con" items may be quantified
- The sums may inform the decision
 $H^0: \text{sum}(A) > \text{sum}(B)$
- Bentham took the phrase 'the greatest good for the greatest number' from Joseph Priestley's essay on government



What are the decision-relevant health outcomes?

- Quantity of Life
 - Number of Affected Persons
 - Longevity: duration alive
 - Mortality: risk of death
- Quality of Life
 - Patient-reported outcomes, such as a pain scale
 - Clinical outcomes, such as an adverse events or unconsciousness
- The value of these outcomes likely varies



Tools for the Valuation of Health Outcomes

- **Trade-off responses**

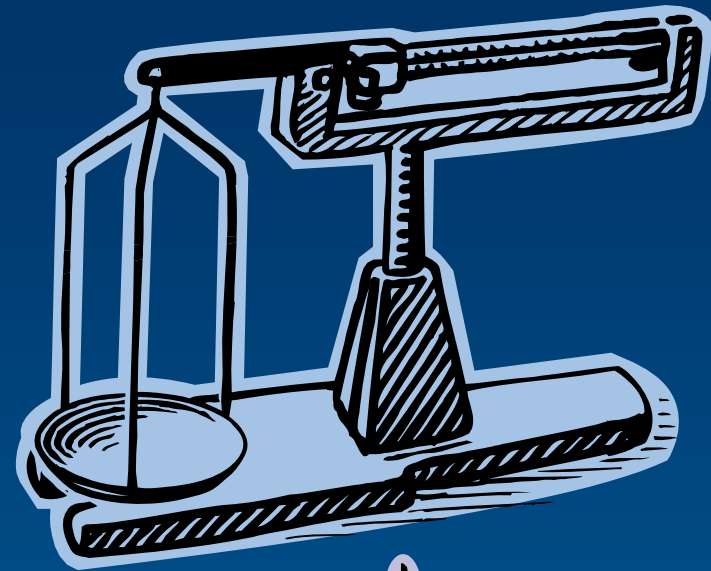
equate losses in quality of life to losses in quantity:

- Reduced longevity (time)
- Increased mortality (risk)

- **Discrete choices**

compare losses in quality to other losses in quality.

- depression vs. pain



Imagine a health scenario of 10 years with moderate depression followed by death

Trade-off response:

“How many years without depression is it worth?”



Discrete Choice:

“Is it worth more than 10 years with moderate pain?”



How much are you willing to suffer to avoid moderate depression?



Severe Pain

Moderate Pain

Mild Pain

% willing to suffer

Demand

Quantity



How to estimate weights: A & B

- Identify the decision relevant health outcomes
- Construct a series of meaningful choices
e.g. depression vs. pain
- Ask a target population to choose between them
i.e. valuation survey
- Estimate & sum the weights
a.k.a. part-worth utilities

	<u>Tx A</u>		<u>Tx B</u>	
A1	Diarrhea		Vomiting	B1
A2	Fatigue		Rash	B2
A3	Depression		Pain	B3
A4	Loss of Appetite		Dizziness	B4

$sum(A) = \sum_{i=1}^4 A_i$ $sum(B) = \sum_{i=1}^4 B_i$



For example...



SF-6D Survey: Version IV

Section 2: Paired Comparisons

A **paired comparison** is a choice between two alternative items. For example, which fruit do you prefer? Choose by clicking the circle under the preferred fruit.



Fruit



Fruit



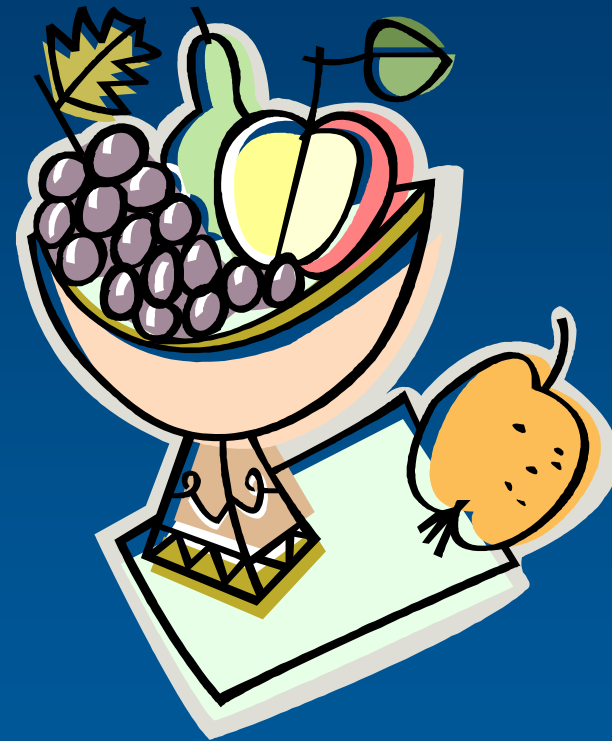
Next ▶

Completed:



Apple vs. Orange

- Results of 2,114 choices
 - 1113 preferred oranges
 - 1001 preferred apples
- Under a logit model, the part-worth utility of an orange is $\ln(1113/1001)$ or 0.106 utils
- Next we might compare other items from a fruit basket
- However, our study compares decrements in health, not fruit



For example...



SF-6D Survey: Version IV

Look below at the two changes in the original health scenario and read their descriptions carefully. Imagine that **you** must live in the health scenario for 10 years and then die. Which scenario do you prefer? Choose by clicking the circle under the scenario you prefer.

If you would like to read the description of the original scenario again, click [HERE](#).

Your health limits your social activities **most of the time**.

Your health limits your social activities **all of the time**.

You are limited in the kind of work or other regular daily activities as a result of your **physical health**.

You are **NOT** limited in the kind of work or other regular daily activities.

Scenario



Scenario



Health Valuation Study

- Descriptive systems measure health decrements using clinical and patient-reported outcome (PRO) items
 - e.g. Medical Outcomes Study 36-Item Short Form Survey (SF-36), version 1
 - Attributes are labeled from best (1) to worst (>1)
- The purpose of this study is to estimate the weights for two PRO descriptive systems: SF-6D and SF-12
- Weights need to be estimated for each decrement
 - e.g., level 1 (no pain) \rightarrow level 2 (some pain)
 - Weights are non-negative by definition (more pain is either insignificant or bad, when all else held equal)



Health Descriptive Systems

- SF-12 has 11 items (excluding EVGFP)
 - 11111111111 (best) to 33222256665(worst)
 - e.g., 1st item has 3 levels or 2 decrements: 1→2, 2→3
 - Across the 11 items, there are 31 possible decrements
- SF-6D has 6 items
 - 111111(best) to 645655(worst)
 - Across the 6 items, there are 25 possible decrements
- Using paired comparisons, each decrement is compared to multiple other decrements in order to collect decisional data
 - i.e., each fruit is compared to multiple other fruit



Internet Survey Design

- Over 100,000 email invitations were sent to panelists between May & June 2010
 - 3,287 subjects consented
 - 2,114 began the paired comparisons
 - 1,638 completed (SF-12v1 962; SF-6D 672)
- Pairs were randomly assigned and sequenced (i.e., order, top/bottom & left/right)
- Compensation
 - Respondents who qualify and complete the survey earn 900 points.
 - If don't qualify, entry for monthly cash sweepstakes



Survey Items

Components:

- 8 Demographic, SES, & Geographic items
- 36 MOS SF-36 items (MOS)
- 1st pair: Apple vs. Orange (practice)
- 24 pairs comparing health decrements
- 1 pair comparing EVGFP to risks of death (practice)
- 6 pairs comparing health decrements to risks of death
- 15 follow-up items of survey difficulty

Paradata were collected on every response

- e.g., response time/changes, browser, IP address
- Survey takes from 20 to 26 minutes (IQR)



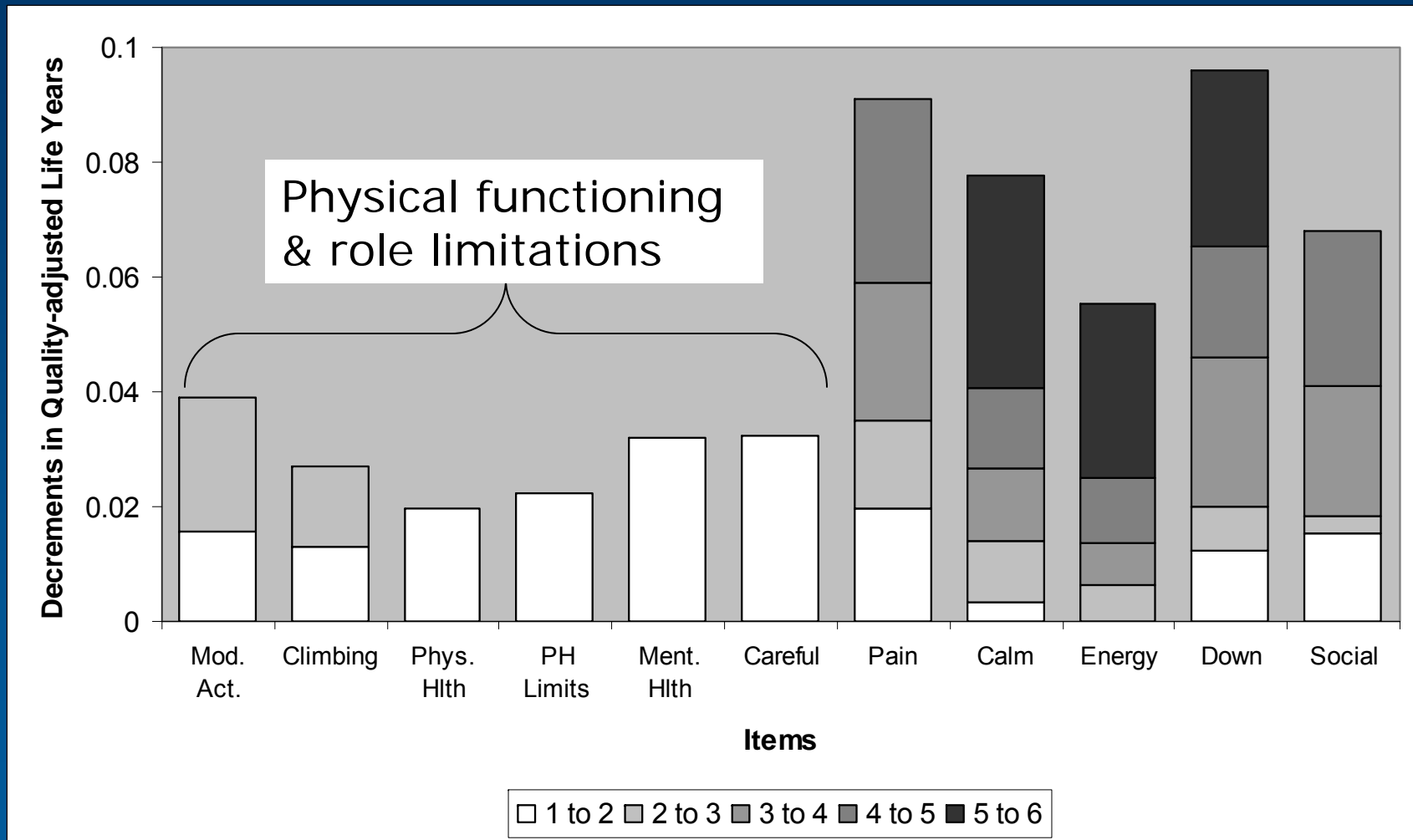
Paired Comparison Analysis

Preliminary

- Random Utility Model
 - Value= $\ln(\text{choice}/(1-\text{choice}))$ or log odds of choice
 - Value represents the relative difference in intrinsic utility as described by the paired comparison
- Values are typically translated into quality-adjusted life years (QALYs)
 - 1 QALY= a year of life in optimal health
- Rescaling of the estimated values into QALYs was incorporated into the model using an auxiliary model and responses on a series of pairs that included a risk of immediate death.



SF-12v1 Preliminary Results

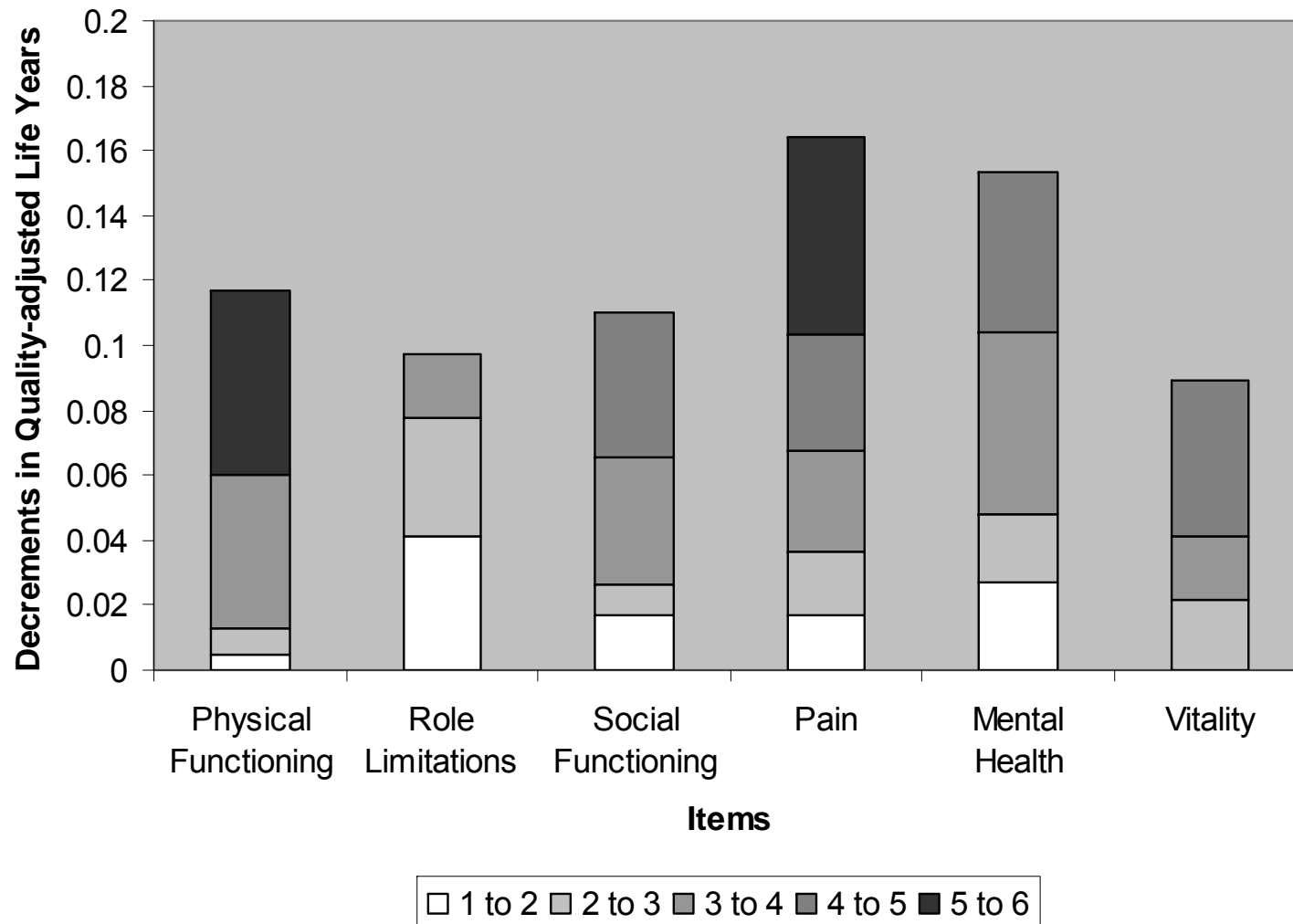


SF-12 Preliminary Results

- Largest and smallest decrements
 - Calm & peaceful a little(4) → none of the time(5) = 0.037
 - A lot of energy all(1) → most of the time(2) = 0
- Attribute Importance
 - Most important: Pain and Depression
 - Least important: Energy
- Decrement size tends to increase with severity



SF-6D Preliminary Results



SF-6D Preliminary Results

- Largest and smallest decrements
 - Pain interfered quite a bit(5) → extremely(6) = 0.061
 - A lot of energy all(1) → most of the time(2) = 0
 - Limited a lot in moderate activities(4) → limited a little in bathing or dressing(5) = 0
- Attribute Importance
 - Most important: Pain and Mental Health
 - Least important: Vitality
- The final decrements are often important.



Summary

- SF-12v1 and SF-6D responses can be translated into quality adjusted life years (QALYs)
 - 1 QALY= a year of life in optimal health
- QALY estimates vary by descriptive system due to differences in the definitions of best & worst health
 - SF-12v1 values ranges from 0.44 to 1 QALY
 - SF-6D values ranges from 0.27 to 1 QALY
- QALYs have become the gold standard measure in comparative effectiveness research according the USPSTF and UK NICE



Limitations

- National internet survey
 - Non-probability sample
 - Participation bias
 - Unweighted
- Insignificant decrements suggests that the sample sizes may be too small
- Sum of the parts may not equal the whole
- Statistical inference (bootstrap)
- Potential order effects

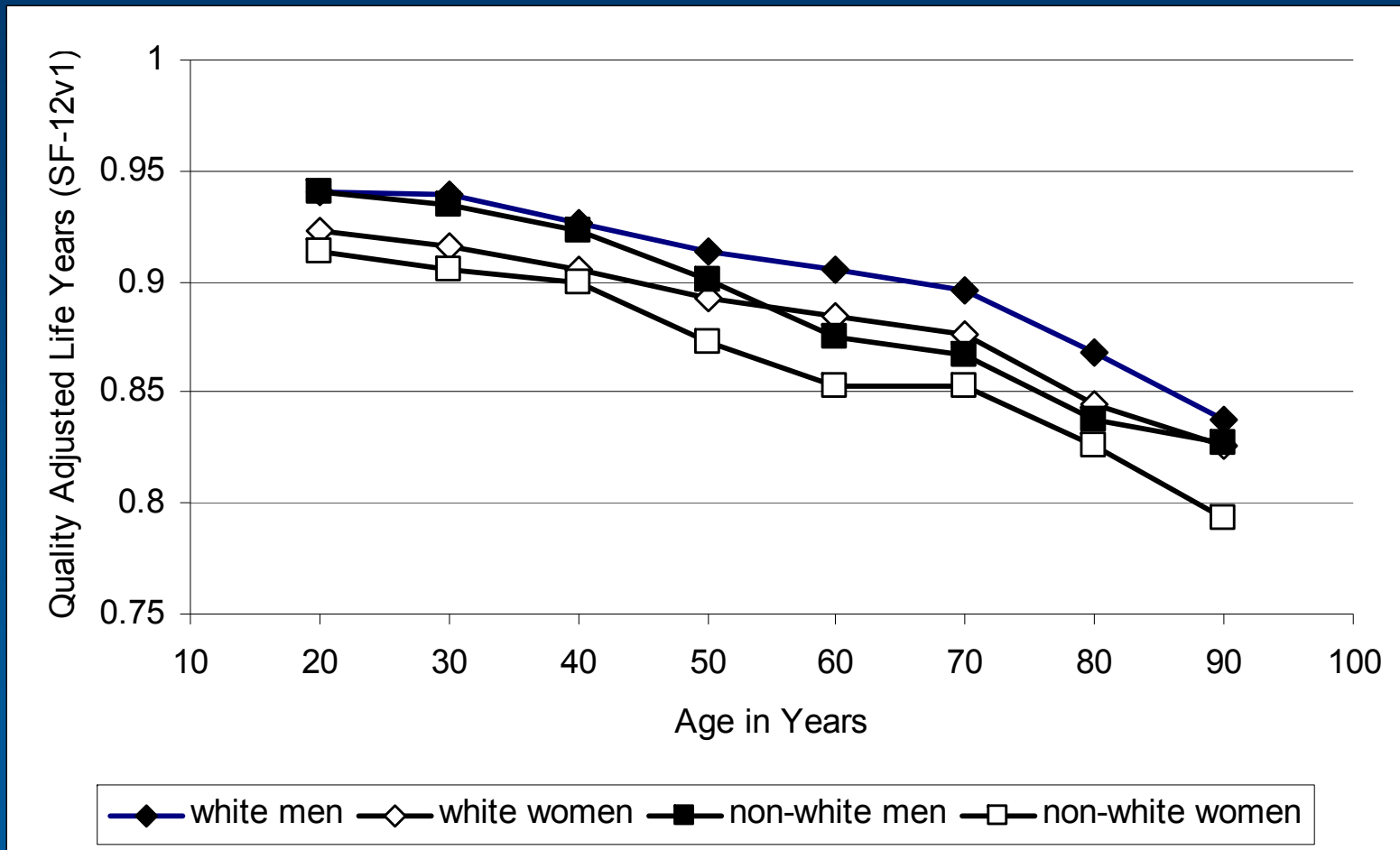




So what...How can I use this?

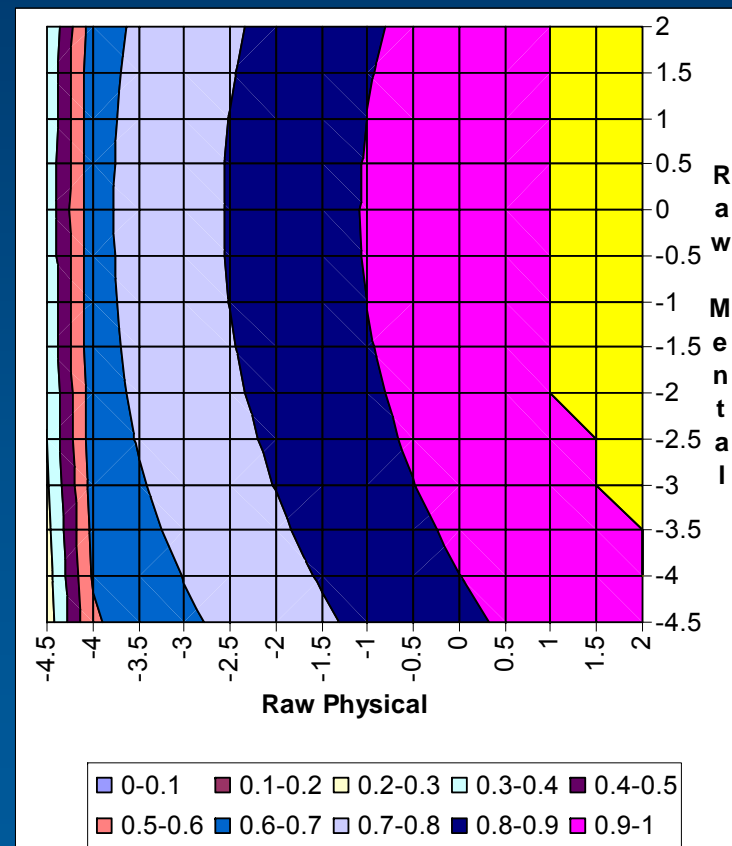


MEPS 2001-2003, SF-12v1



MHOS 1998-2003, SF-12v1

- 222,256 Medicare HMO beneficiaries completed 2 SF-36v1 responses
- **Yellow** represents optimal health (QALY=1)
- Psychometrics to QALYs:
 - Physical health largely determines QALYs
 - Mental health score has little effect until < -1
 - Mental health attenuates the value of physical health



Future Work

- Other descriptive systems:
 - e.g., EQ-5D, PROMIS, PRO-CTCAE
- Patient & culturally diverse populations
- Better understanding of method limitations...



Grand Nephew's Response

April 13, 1779

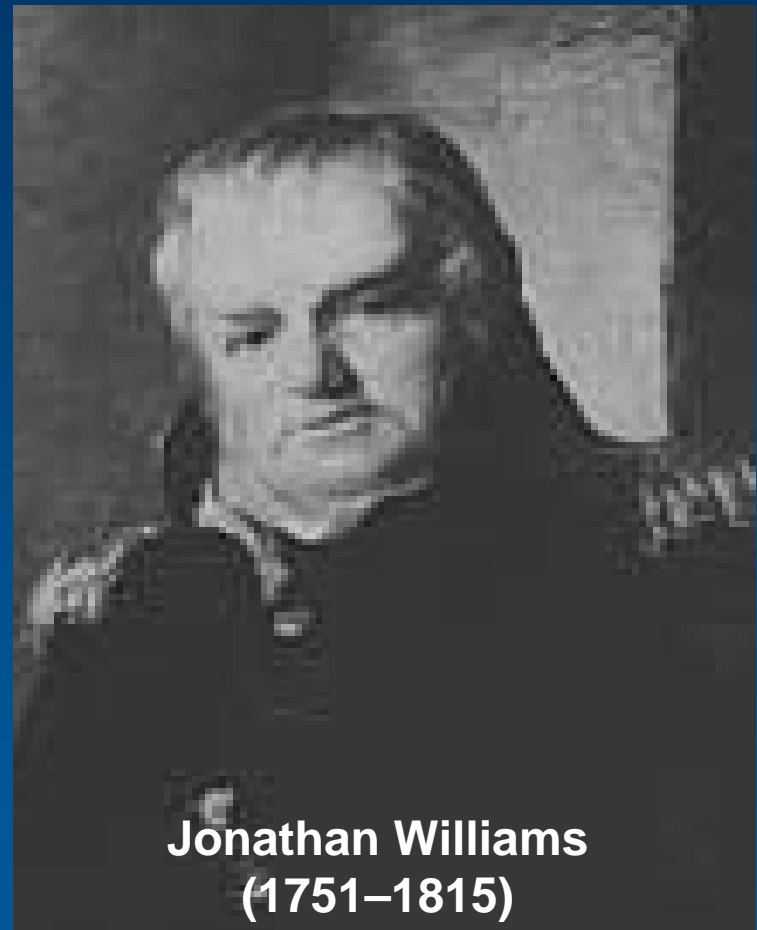
Dear & Honored Sir:

...I thank you for your advice as to the algebraic calculation & will follow it....

But in the matrimonial way, I a little differ from your opinion & instead of my never being married if I don't "use it" I am afraid I never shall be married if I do, for the negative column seems in this instance the weightiest.

Before a man is married, he must *fall* in love and this seems to be as involuntary an act as *falling* into a well...

Your dutiful & affectionate Kinsman,
Jona Williams

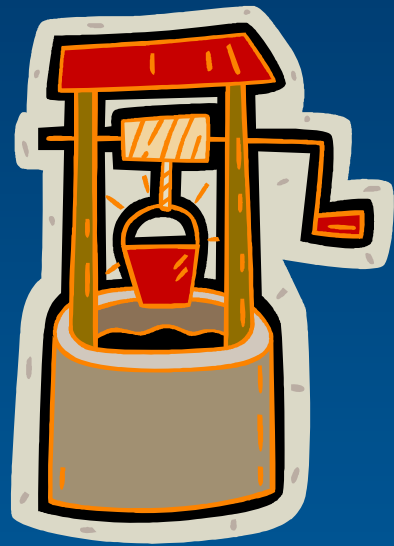


Jonathan Williams
(1751–1815)

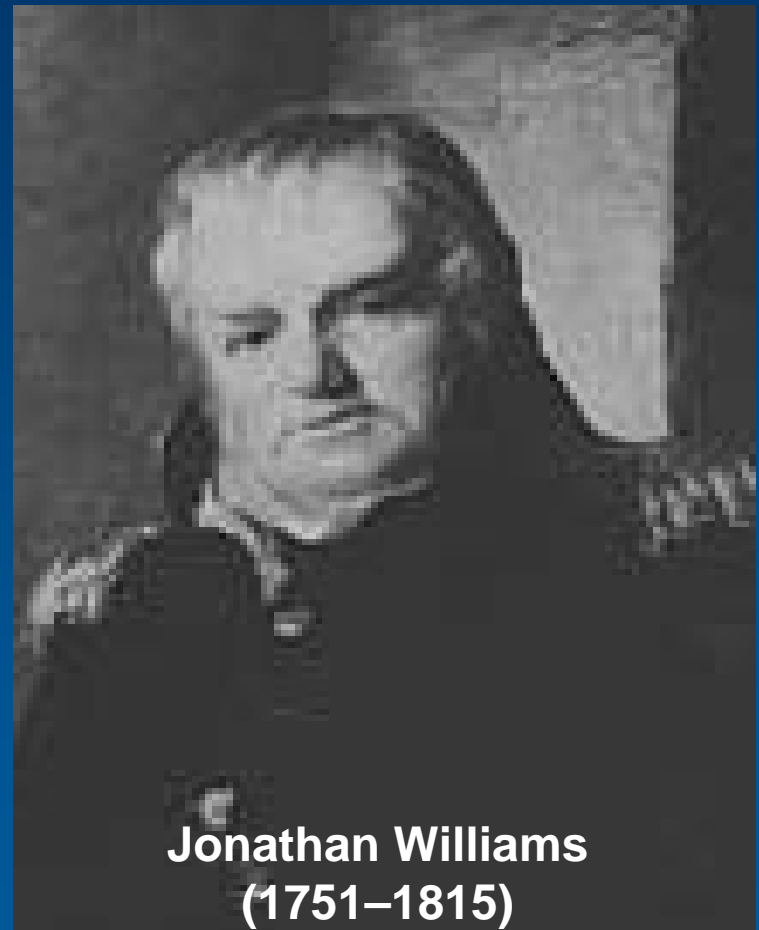


Grand Nephew's Response

In September 1779, Jonathan married Mariamne Alexander.



Dr. Franklin, then minister from the United States, was present at the ceremony in Paris, France.



Jonathan Williams
(1751–1815)



Thank you!

- Craig Lab
 - Andrea Collado
 - Liana Encinosa
 - Veronica Harville
 - Shannon Hendrix-Buxton
 - Riddhi Patel
 - Carol Templeton
- K25 Mentors
 - Paul Jacobsen (chair)
 - Kenneth Zuckerman
 - Lodovico Balducci
 - Joseph Lipscomb, Emory
- Collaborators
 - John Brazier, University of Sheffield
 - A. Simon Pickard, UI Chicago
 - Elly Stolk, Erasmus
 - Aki Tsuchiya, University of Sheffield



Questions?



Contact Information

Benjamin M. Craig, PhD
Assistant Member, Moffitt Cancer Center
Associate Professor, Economics, University of South
Florida

12902 Magnolia Drive, MRC-CANCONT
Tampa, Florida 33612-9416 USA
Benjamin.Craig@moffitt.org
Telephone: (813) 745-6710

