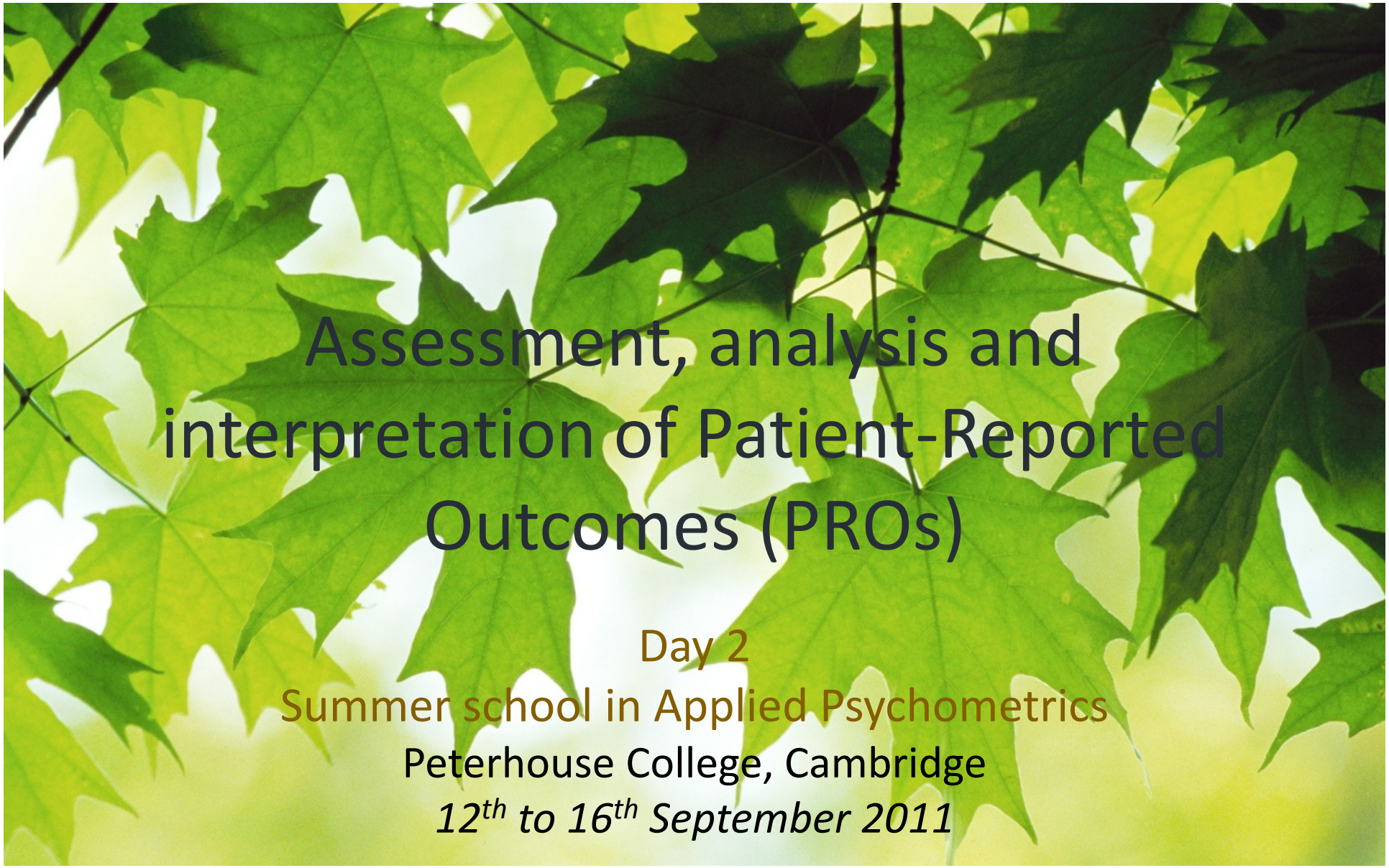

Assessment, analysis and interpretation of Patient-Reported Outcomes (PROs)

Day 2

Summer school in Applied Psychometrics

Peterhouse College, Cambridge

12th to 16th September 2011

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

This course is prepared by

Anna Brown, PhD ab936@medschl.cam.ac.uk
Jan Stochl, PhD js883@cam.ac.uk
Tim Croudace, PhD tjc39@cam.ac.uk
(University of Cambridge, department of Psychiatry)
Jan Boehnke, PhD boehnke@uni-trier.de
(University of Trier, Department of Clinical Psychology and Psychotherapy)

The course is funded by the ESRC RDI and hosted by
The Psychometrics Centre

Anna Brown

3. PRINCIPLES OF MEASUREMENT SCALES

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Measurement of an attribute

- The purpose of measurement is to quantify an attribute
- Measurement is the assignment of numbers to an attribute according to a rule of correspondence
- For example, the number of symptoms from a checklist would give a score to every patient according to a simple rule
  - This correspondence does not necessarily hold in the other direction
    - Patients with the same score might have different sets of symptoms
  - This rule might produce measurement of only a limited range of the attribute
    - Cannot measure below its *floor* or above its *ceiling*

Inferred measurement

- Psychometric tests are different from “proper” measurements we routinely use – such as temperature, weight, length etc.
- A questionnaire should be viewed as a series of small experiments (observations) outcomes of which are recorded
  - from which a measure is inferred (van der Linden & Hambleton);
  - These outcomes often have no metric of their own;
  - Observations in tests need to be mapped to numerical data

Experiments, observations, items…

- Questionnaires aim to gather information on "objectively scorable" items
  - It is decided before the test administration how responses to items should be scored
- Item is a stimulus to which a response is collected
  - Item stem
  - Response options
    - Many types (open-ended, multiple choice, binary response, graded response or Likert scale, ranking or forced-choice, etc.)

Scoring items – some initial questions

- Graded responses (*Likert** scales) typically assign consecutive integers to response categories
- Assumptions
  - Linearity
  - Equal intervals
  - All respondents interpret response categories in the same way
- Are these assumptions reasonable?

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Levels of measurement

- Ratio
  - Length (meters), or weight (kilos)
  - *Interval* between 15m and 16m is exactly the same as the interval between 1m and 2m
  - An object 2m long is "*twice as long*" as an object 1m long
- Interval
  - Temperature (Celsius)
  - Difference between 15° and 16° is exactly the same as between 1° and 2° with respect to the attribute
    - This might not be obvious from observations
- Ordinal
  - Hardness of minerals (Mohs scratch scale)
  - Ranges from the hardest (diamond) to the softest (talk)
  - The only meaning reflected in the scale is the *order* of hardness

Ordinal scales

- Let a be the measurement of attribute *A*, and b the measurement of attribute *B*
- Fundamental properties
  - Identity rules
  1. either a=b or a≠b
  2. If a=b then b=a
  3. if a=b and b=c then a=c
  - Order relations
  4. either a>b or a<=b
  5. If a>b and b>c then a>c
- Allowed operations
  - Any order-preserving (monotonic) transformations

Interval scales

- Let a be the measurement of attribute *A*, and b the measurement of attribute *B*
- Fundamental properties
  - All properties of ordinal scales plus
  - Additivity rules
  6. a+b = b+a
  7. If a=c and b=d then a+b=c+d
  8. (a+b)+c = a+(b+c)
- Allowed operations
  - Origin and unit of the scale are arbitrary
  - Linear transformations only

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Ratio scales

- Let a be the measurement of attribute *A*, and b the measurement of attribute *B*
- Fundamental properties
  - All properties of interval scales plus
  - Zero rules

  9. a+0 = a
  10. If a=c and b>0 then a+b >c

  - Zero is an absence of the attribute
- Allowed operations
  - Unit of the scale is arbitrary
  - Ratio transformations only

Choosing a metric

- Metric is a set of scale values for the observations
  - Includes choosing an origin and a unit of measurement
  - Decide which observation corresponds to number 0, and what difference between observations corresponds to number 1
- For our simple symptom-counting checklist, we can
  - Use the number of symptoms (*criterion-referenced measurement*),
  - or subtract the population mean, and divide by its SD (standardized, or *norm-referenced measurement*),
  - or take a natural logarithm of the odds (ratio of the number of criteria "met" to the number of criteria "failed"), etc.
  - ….and still satisfy the basic requirement of measurement
  - However, changing the scale by a transformation might alter some statistical hypotheses (e.g. linearity of a relationship)

Criterion-referencing

- Raw scores often have an absolute reference to behaviour

  I have had ("*very mild*") bodily pain during the past 4 weeks

  – Do we need to relate that report to others' reports?
  – If a patient meets all criteria for a diagnosis, this needs no comparison with other patients
  – Usefulness and virtue of raw scores are often neglected

Norm-referencing

- Choosing metric on the basis of distribution of scores obtained from a population of interest
  - Origin is the mean and unit the SD
  - Might make sense in large-scale public health programmes
- The same instrument can be referred to a criterion or to a norm
  - Depends on motivation: e.g. detection of psychopathology versus its general incidence in the country