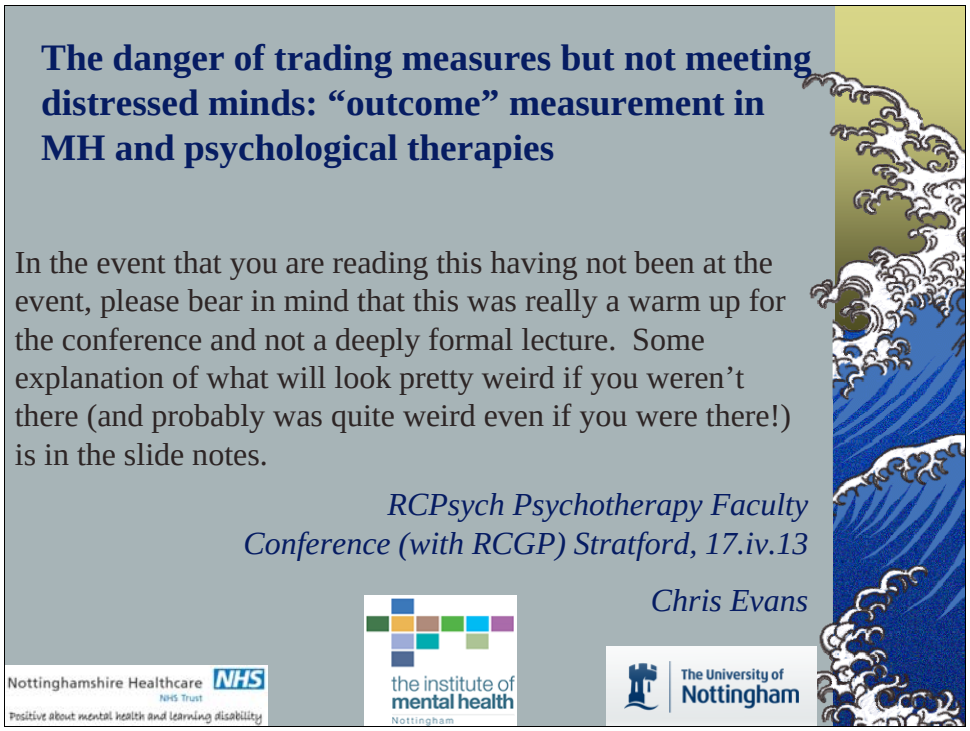
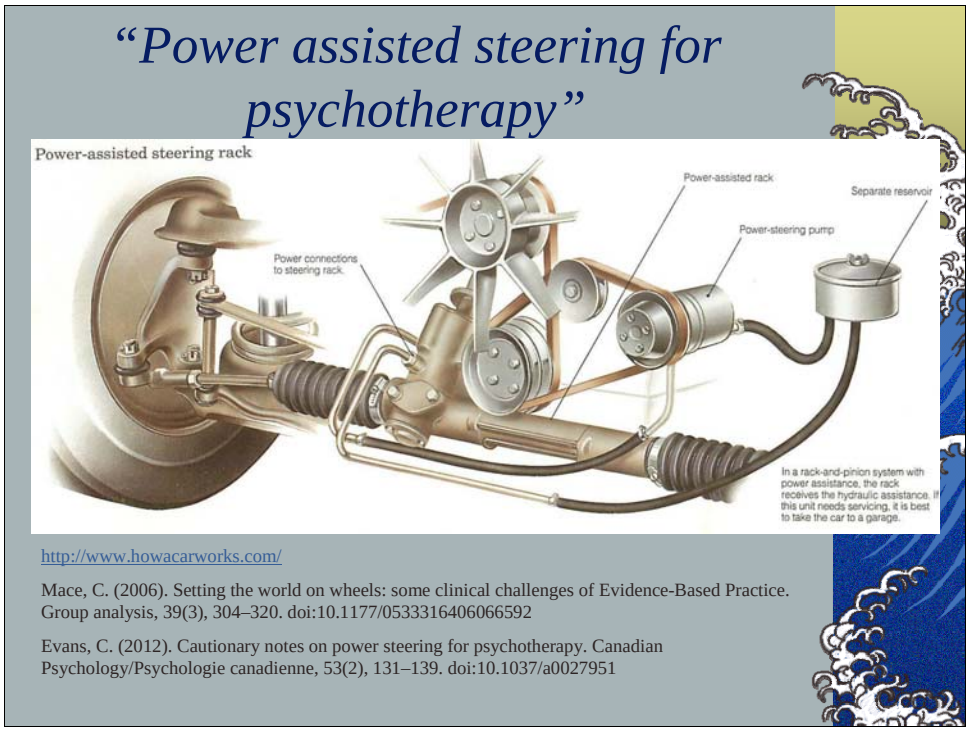


The reference copy should be available at http://www.psyctc.org/talks/2013/20130417_RCPsych_Faculty_talk.pdf and more information about other talks (perhaps) and licensing is at http://www.psyctc.org/talks/



This shows a power assisted steering system: the key thing is that hydraulic power is used to move the wheels more quickly when they don't reach the position the turning of the steering wheel wanted fast enough.

Why is this here? The main thrust of the lecture is that psychotherapy is not just a technology, not just a competence or a skill, at heart it is about meeting with someone in distress, it is a relationship of a certain duration with fairly clear boundaries, constraints and capacities but if it is a real meeting there is a potential for it to about "meeting", inferring the distressed state of mind of the other and it may involve some or more distress to be experienced and resonated with by the therapist. At one level this is about something almost technological: "mentalising" in modern jargon, but it generates anxiety on both sides and it can be hard at times to see the best, or even the least damaging, ways forward at times.

One modern vogue is for "sessional tracking" in which a self-report outcome measure is completed each session/week and corrective actions encouraged by protocol for the client and/or therapist if the score trajectory is "off track". I refer to this as "power steering for psychotherapy". Although I think it has a potential role in some therapies, I believe it is a development that has not been evaluated properly and is currently hugely overvalued and I believe it is largely, certainly when used outside theoretical modalities that might be congruent with it, merely an anxiety reducing device.

The rest of the talk tries to do four things:

1) Bring back good memories of Chris Mace and his work for the Faculty, for the Institute of Group Analysis and for the Society for Psychotherapy Research (the three settings in which he overlapped in the last fifteen years, probably in reverse order of the extent of our overlap) and more generally to remember him with great warmth as a generous and creative man deeply thoughtful about the development of psychological therapies.
2) Show, partly by expecting the audience to make use of this, that human interactions are fully of many parallel channels of information: verbal content, non-content verbal cues, cues from narrative form, facial and body language communications. Much some of these sources of information are of fairly simple face validity and don't draw on high levels of inference in the listener/receiver, much does rely on high levels of inference and empathy: everything from basic "theory of mind" to very high level examination of countertransference. Compared to these, the communication channel through a repeated "outcome measure" is very limited. That doesn't mean that it can't add anything within therapy, but I argue that we should be wary about distracting from the other communications by overvaluing such a source of information.
3) Encourage us to accept that "outcome measurement" is, ultimately, a post-mortem appraisal and never certain, however, also to accept that grossly simplified appraisals, including by self-report are vital then to describe the development of the one outcome measurement system I know really well (because I have been a co-author, co-leader and now co-trustee copyright holder of it): the CORE-OM and CORE system. I describe this in some detail not just to try to get a warm glow and share some of what's been achieved in the system, but more to show some of the complexities and encourage people both to respect simple self-report measures, but also to encourage people to see them as involving some psychological and other complexities that we should hold carefully in mind when using these measures.
4) Argue that any evaluation of within therapy "power steering" systems must be evaluated on a measure other than the one that is driving the system: doing otherwise is like evaluating a power assisted steering system not by finding where the wheels point but by seeing where the steering wheel is pointed: that doesn't really test the system at all.
5) Argue that, nevertheless, simple outcome measures remain useful and can be used sessionally as long as that is:
    1) Congruent with the theory of the therapy and aiming to "power steer" the therapy in which case the ultimate evaluative "outcome measure" should be a different one from that used in the main tracking
    2) or simply to get a trajectory description (in which case it probably won't be seen by therapist until after the end of the therapy),

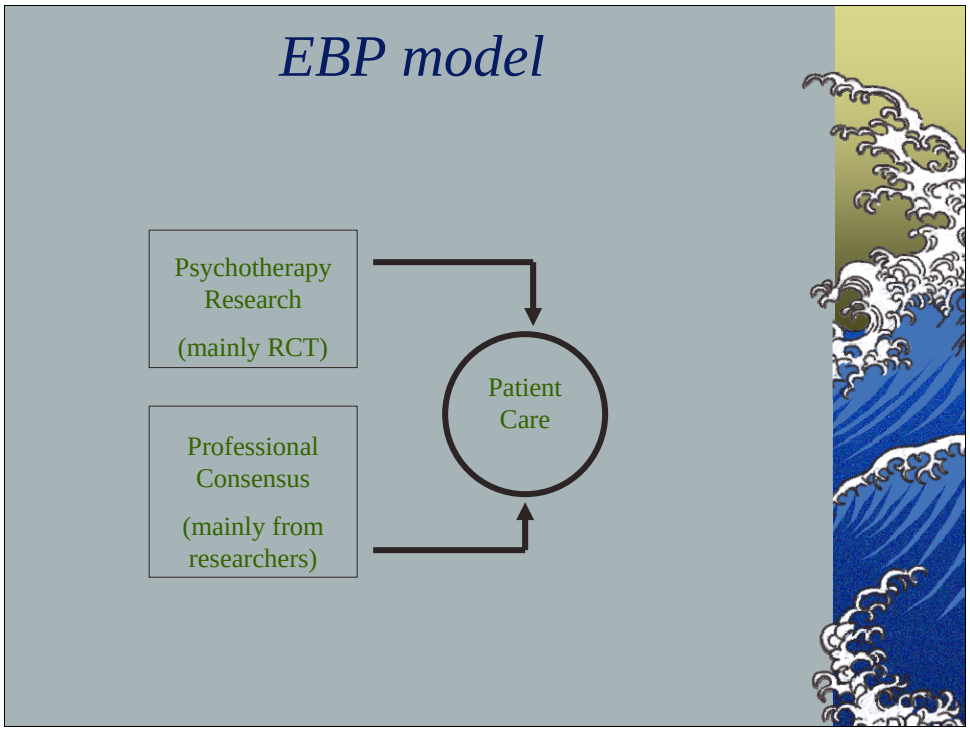
Power steering for cars is great for fast cars but there are times when we need to slow down for ducl
meeting I attended that Chris Mace attended. We walked together and Chris had been elected as the

CORE system history: plan

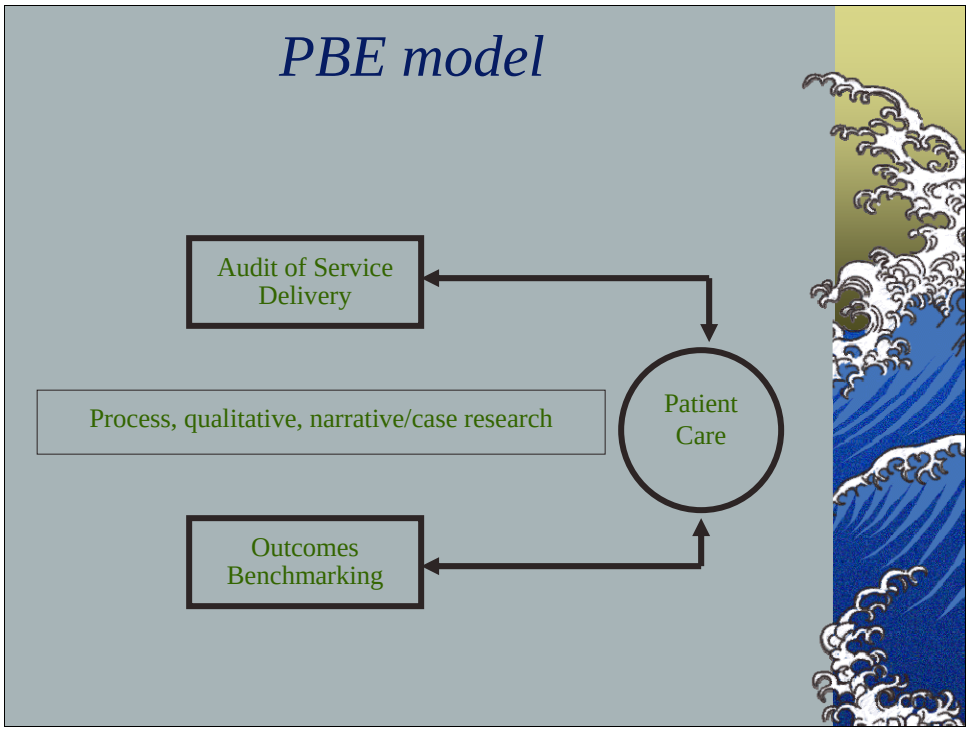
- CORE principles
  - Copyleft
  - "bottom up"
- DIY model:
  - D = Design
  - I = Implementation
  - Y = Yield
- "Il c(u)ore del CORE":
  - That the person completing the measure, who might have few or no problems or really very severe problems, might feel that someone scoring their questionnaire would gain some not irrelevant or unhelpful sense of their state over the last week.

This was back at the beginning of the CORE system about 18 to 19 years ago.

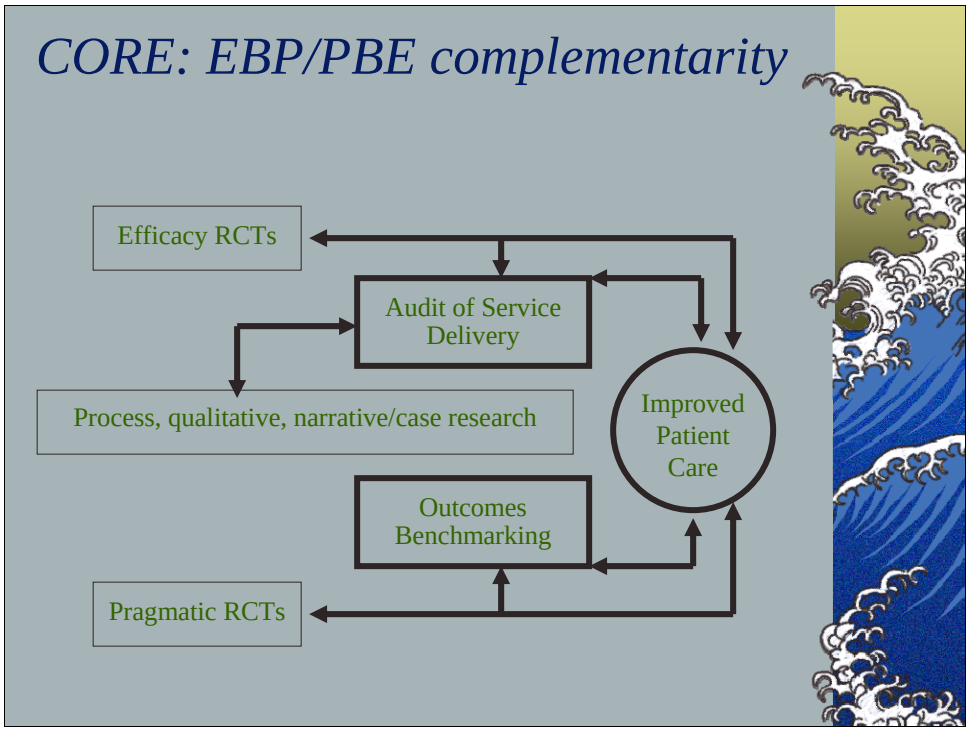
I added "Il c(u)ore del CORE": the heart of CORE (in medieval Italian, heart was spelt "core" where it is now "cuore"). I think that aim is worth remembering as it underlines the communicative nature of the enterprise and also that it is expansive (the range of people who might feel the measure might convey something useful) but not grandiose about precision.



19 years ago the EBP model was just cementing its huge political power into place.

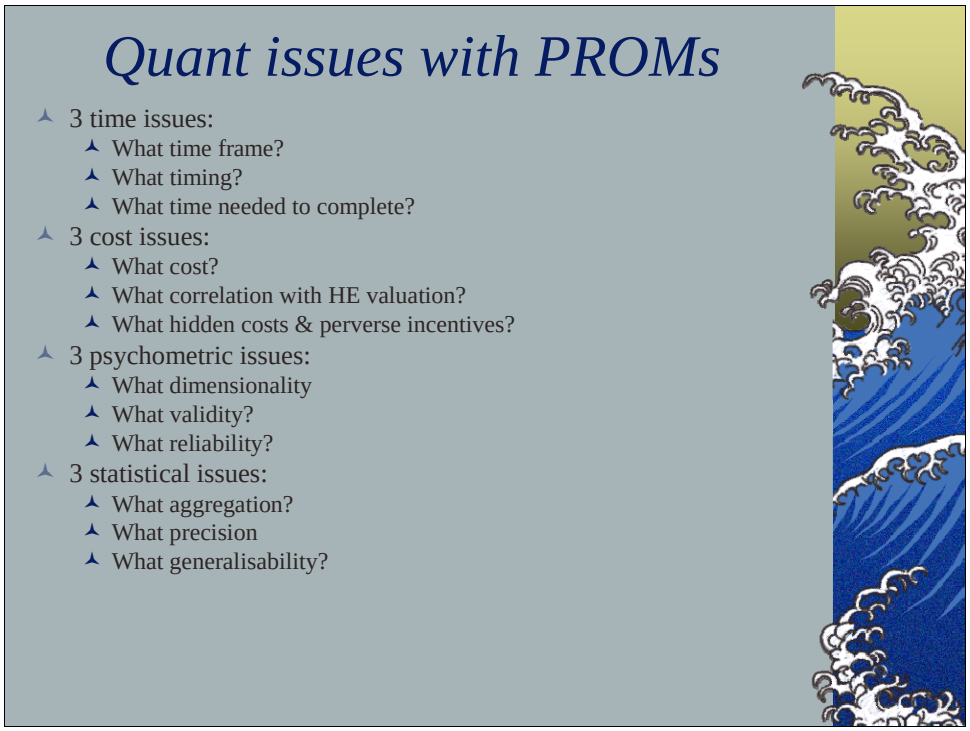


We believed that EBP had to be complemented with PBE (Practice Based Evidence), including qualitative research, and that no really healthy practice development and improvement of client/patient care would take place without PBE as this vital complement.

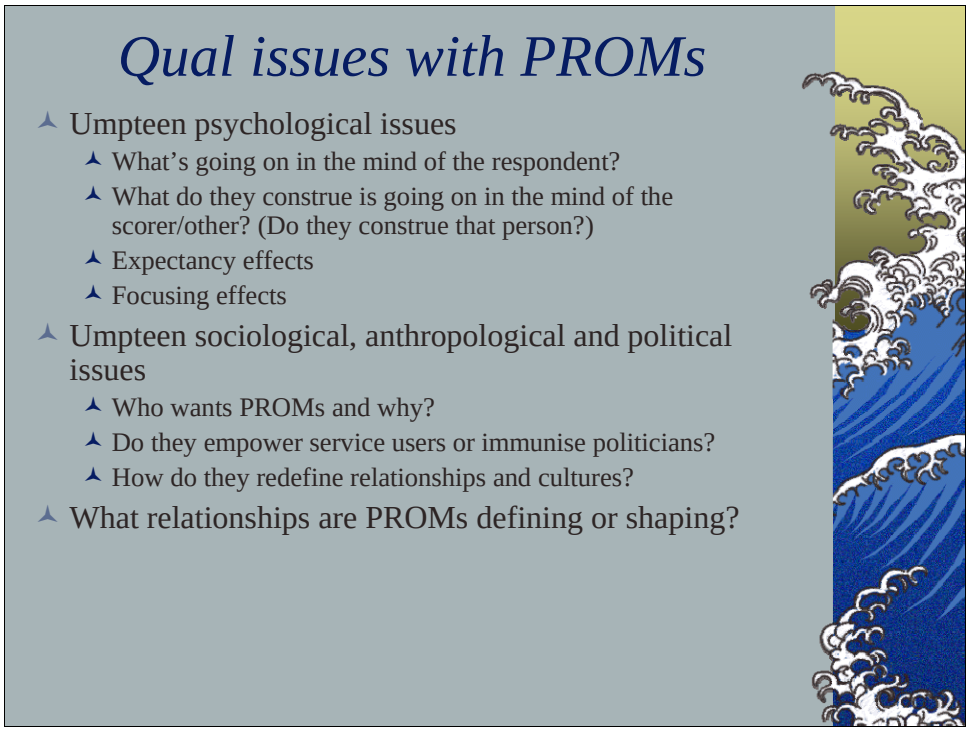


The CORE aspiration was that the CORE-OM would come to be useful in both forms of research and exploration of therapy That has happened and the CORE-OM *has* been used in both RCTs and in non-RCT research that is part of PBE not EBP.

This was further on in the walk, a fair way past the farm with the ducks, the hotel whence we'd come can just about be seen on the cliff in the background. Interestingly, it is house in which George III was confined during his illness and the regency. It's a beautiful place for a conference. You can see that a controlled trial of reduced footwear and of abbreviated trouser length was being designed.



Well, in a single lecture a wiser lecturer doesn't even mention all of this but I quite like my recent framework which argues that there are 4 sets of three quantitative issues with all PROMs, all self-report outcome measures. These aren't, unless the designers have done an absolutely dire job, fatal problems, but no measure, even the CORE-OM (!) can be be so good that it's without problems on any of these aspects.



I believe that the qualitative issues we need to consider about OMs are vastly more important than the quantitative ones. Sadly, the quantitative issues get vastly more exploration, sometimes quite obsessionally so in an "angels on the head of a pin" fundamentalist scholasticism; meanwhile these qualitative issues go largely or entirely neglected.

Note that qualitative issues can have strong ordinal relationships with other things: "vastly more important than" and that their counting systems are generally kept simple as here: one, some, umpteen!

Conversation with receptionist

- C – How are you?
- R – Fine … Thanks for asking.
- C – Would you say if you weren't?
- R – [..] Probably not?
- C – Hm. .. So what should I ask?
- R – [....] Well, what I say to the internal postie is "What are you today?" …. "from 1 to 10"
- C – [……] Hm. So … what are you today?
- R – [..] 7
- C – That sounds OK. …some people aren't happy with anything less than a 10
- R– Ah but I think I was down about a 3 before I was off sick.

Some things in interactive communication are both qualitative (particularly in form) and quantitative as with this, pretty accurately reproduced conversation with a receptionist who had been off sick some weeks earlier. The communication happened on the morning of the talk and is reproduced with permission.

Please note that it's the receptionist who makes the assessment quantitative, note also my attempt to give the rating some interpersonal comparison as well as within subject: "some people aren't" vs. "a 3 before I was off".

The speed warning on the other side of the farm.

Exercise

- Last therapy session you were in: how did you appraise how it was going:
  1. You realised you were very relaxed
  2. You realised you were bored … or other
  3. You were gauging the client's facial reactions
  4. You were gauging the client's body language
  5. You were using the client's expressed emotions (verbal content)
  6. You were guided by the discursive themes
  7. You were guided by the discursive form
  8. The session appraisal form
  9. Other …

The audience engaged with this generously on the day!

Exercise

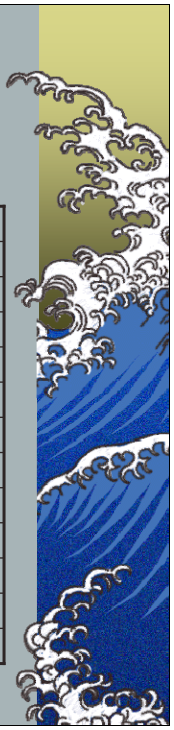
- As a “proper doctor” what information helped you appraise a patient’s state:
  1. Fixed, dilated pupils
  2. Non-response to pain
  3. Exercise tolerance ECG
  4. BMI
  5. ESR/viscoscity
  6. Hb level
  7. U&Es
  8. “I’m grand doc”
  9. Other …

Ditto

Exercise

Match the condition to a sensible investigation or sign to guide management

Asthma	Hb level
Depression	Respiratory rate
Ca Prostate	Plasma glucose level
Hypoglycaemia	Speech
Status epilepticus	Communication
Hyperglycaemia	PSA
Anaemia	Pulse
Rhematoid arthritis	Temp
Trigeminal neuralgia	Consciousness
Withdrawal	Fits
R sided CVA	CORE scores
Infection	Pain ratings
Atrial fibrillation	Mobility ratings



Skidded through this fast on the day.

Exercise

- Thinking of that client you considered in the first exercise, how did you gauge their state shortly before the session:
  1. The messages that were there for you from them/others before the session
  2. Snook a look into the waiting room
  3. The way the receptionist rolled his/her eyes and conveyed something telling you the client was there
  4. How they were dressed
  5. How they smelled
  6. How they came in
  7. What they said about the day/week before
  8. They were late/early
  9. The questionnaire data

Scaling and predicting

An anonymous caregiver said...

I am so glad that I came across this article several weeks ago, just in time to recognize that my 87-yo mother was in fact in the dying process. It cleared up so many things for me. She had almost every one of these signs. The hardest one for me was her change in personality and detachment. Mom passed away this evening 1/2 hr after I left her--I'd been with her for six hrs straight and I believe she was waiting for me to leave.

www.caring.com/articles/signs-of-death

This is another example of something that is both qualitative but also quantitative, personal and relational but referring to a protocol on the web that hopes to predict time to demise. (Oddly enough, one of my first publications ever was about the success health professionals working with people who were terminally ill had predicting time to death: Evans, C., & McCarthy, M. (1985). Prognostic uncertainty in terminal care: can the Karnovsky index help? The Lancet, i, 1204–1206. We found that we were (all) very bad at it. Just because it's difficult to measure something doesn't mean people wouldn't like to know more about it and the request for measurement doesn't always come, politically, from "above". This captures some of the importance and poignancy of how much some people want calibration and measurement.

The speed warning from the other side of the farm.

Outcome measures (PROMs)

- Trendy & plausible
- Necessary
- Very complex, many issues
  - Quantitative issues:
    - 3 time issues:
    - 3 cost issues:
    - 2 psychometric issues:
    - 3 statistical issues:
  - Qualitative issues
    - Umpteen psychological issues
    - Umpteen sociological, anthropological & political issues

Summarising

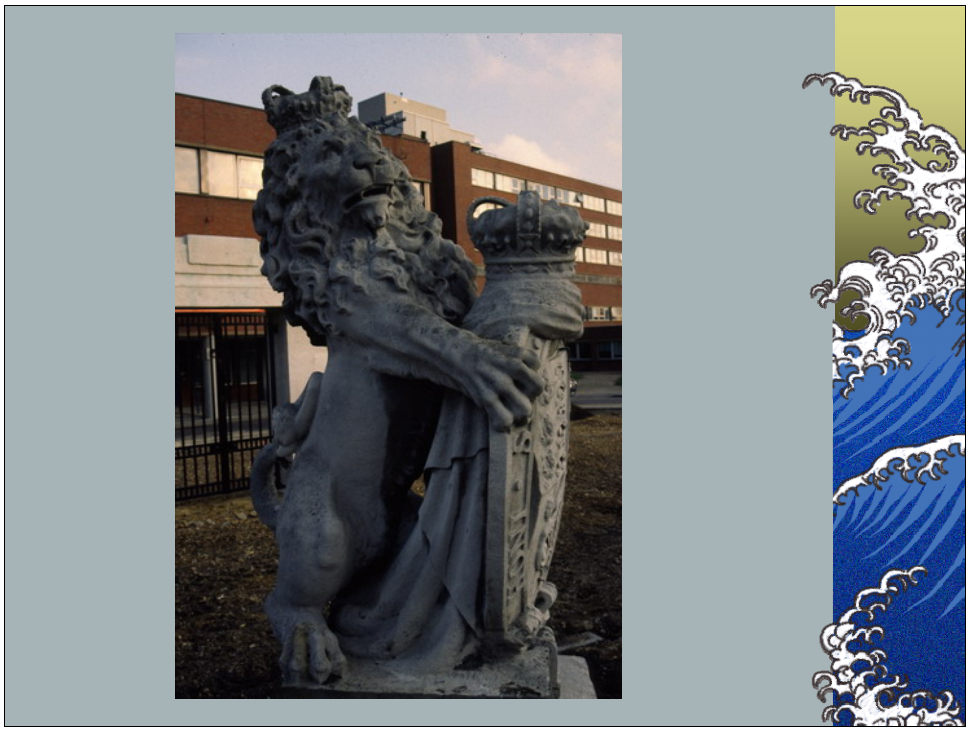
Quant issues with PROMs

- 3 time issues:
  - What time frame?
  - What timing?
  - What time needed to complete?
- 3 cost issues:
  - What cost?
  - What correlation with HE valuation?
  - What hidden costs & perverse incentives?
- 3 psychometric issues:
  - What dimensionality
  - What validity?
  - What reliability?
- 3 statistical issues:
  - What aggregation?
  - What precision
  - What generalisability?

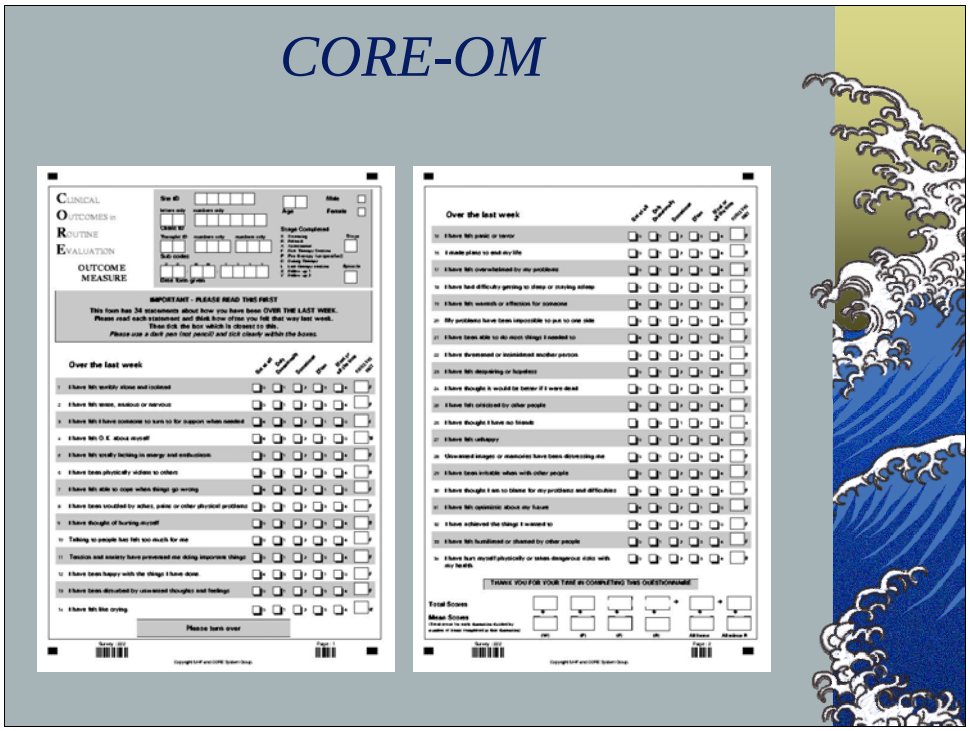
Qualitative issues with PROMs

- Umpteen psychological issues
  - What's going on in the mind of the respondent?
  - What do they construe is going on in the mind of the scorer/other? (Do they construe that person?)
  - Expectancy effects
  - Focusing effects
- Umpteen sociological, anthropological and political issues
  - Who wants PROMs and why?
  - Do they empower service users or immunise politicians?
  - How do they redefine relationships and cultures?

Reprise



A proud lion at St. George's Hospital Medical School: stomping ground for Chris Mace, myself and many, many other medical psychotherapists. I didn't realise until some time after I left St. George's just how lucky I'd been to have been in such a psychotherapy and mind friendly psychiatric training establishment and just how unusual (though not quite unique) it was.



A proud lion … oops, oh no, that was the last picture. A proud outcome, oops, no, transferred epithet: an outcome measure of which I'm somewhat proud!

"Outcome" challenges

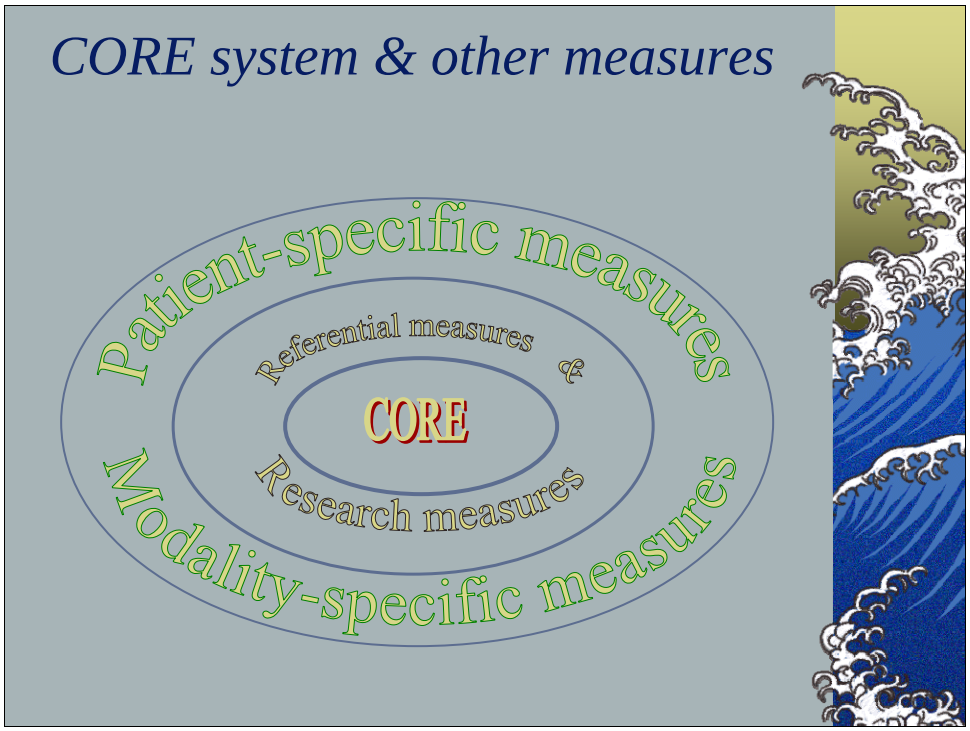
- Bandwidth-fidelity & length dilemma
- Dimensionality cf. domain coverage
- Theory-practice dilemma
- Readability
- Measure ownership
- Cost
- Lack of referential and normative data
- Generalisability to other cultures & service politics/pragmatics
- Lack of durability

These were challenges we recognised very early on in the process of designing the CORE system. Probably about 17/18 years back and Chris was one of the people to encourage us to persevere.

CORE project 1997 aims

- Improve patient care by:
  - facilitating communication between researchers and clinicians (i.e., familiarity);
  - providing comparability between studies;
  - facilitate audit of clinical services;
  - facilitate audit of individual therapists' work;
  - feeding into health economic and other research analyses.

Historical slide from around 2001: another proud lion?



Another very early slide, good to see that M$ borepoint has maintained backward compatibility enough that I can still edit such a slide and it'll still work. I've changed this. I've never been entirely happy with it and would like something a bit different still but it does capture part of the pun we had put into "CORE".

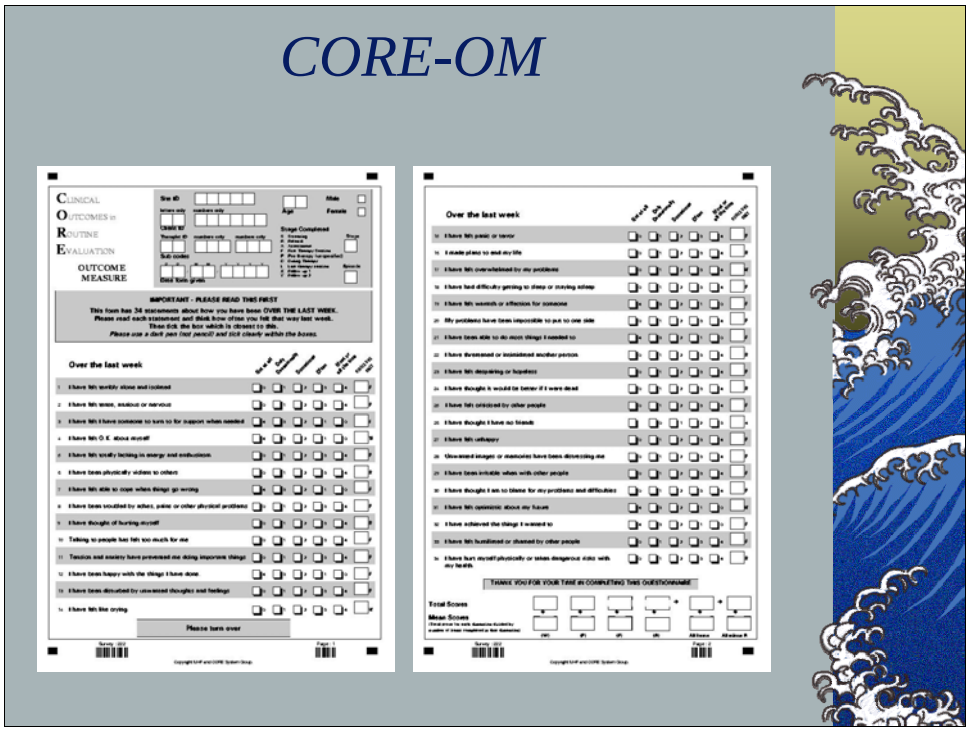
1998 CORE system measures

- "User" completed:
  - CORE-OM
  - CORE-GP, CORE-SF1 & SF2
  - Piloting: teen-CORE & translations
- Practitioner completed:
  - CORE-A comprising:
    - TAF (Therapist Assessment Form)
    - EOF (End of Episode Form)

This was another early description, c. 2001/2003 of where we were

1998 CORE system instruments

- "User" completed:
  - CORE-OM
  - CORE-GP, CORE-SF1 & SF2
  - Piloting: teen-CORE & translations
- Practitioner completed:
  - CORE-A comprising:
    - TAF (Therapist Assessment Form)
    - EOF (End of Episode Form)



Oh dear, maybe I'm flogging this, think what it's like getting it every week for say 16 weeks in a short term therapy.

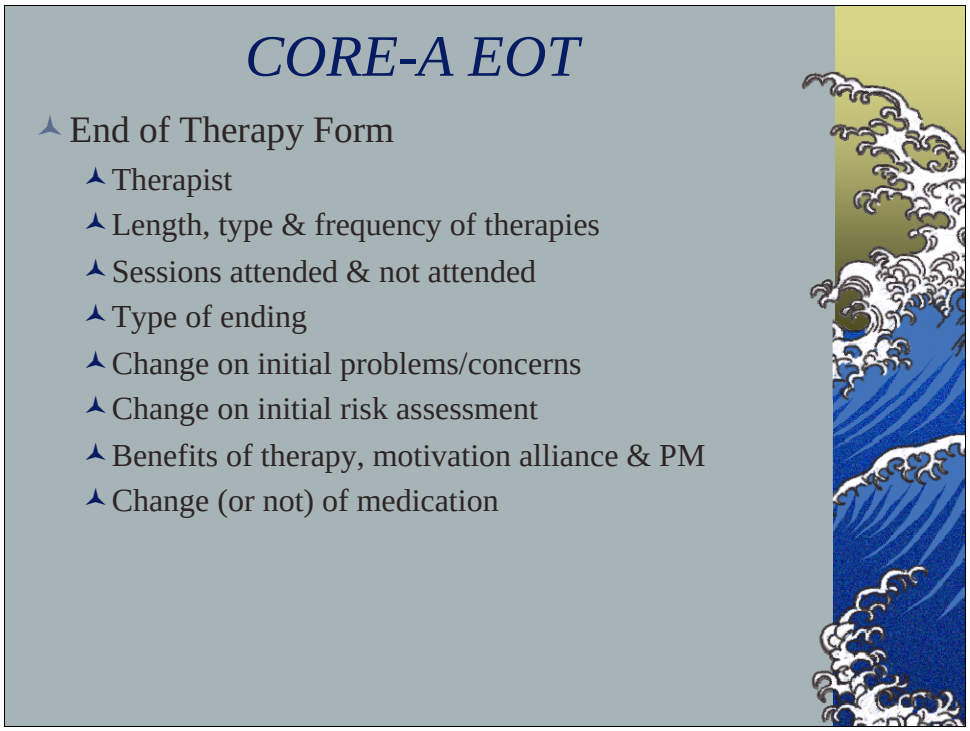
CORE-A

- Therapist completed
- As much to contextualise therapy as to be an “outcome measure”
- 2 sides of A4 for initial assessment
- 2 sides of A4 for end of therapy form
- Mostly tick box completion but room for some text for local use

This was the complement: the CORE, made up of an assessment and an end-of-therapy form was NOT a outcome rating scale (though the EoT form does have a tiny bit of that in it) it was designed to provide much contextual information with which analysis of aggregated CORE-OM data would be much more informative.

CORE-A TAF

- Therapy Assessment Form
  - Demographics
  - Referral history & pragmatics
  - Relationships/support
  - Clinical history (previous therapies & concurrent medication)
  - Identified problems/concerns
  - Risk
  - Result of assessment

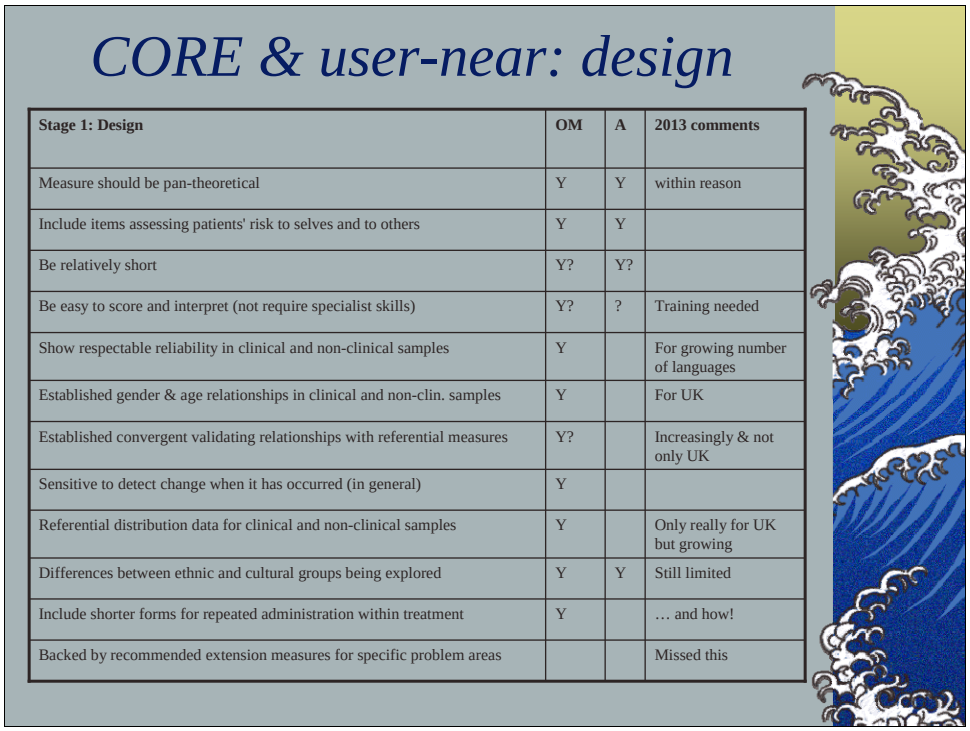


The main elements of the CORE-A EoT.

2013 CORE system measures

- "User" completed:
  - CORE-OM (34 items)
  - CORE-SF1 & SF2 (18 items)
  - CORE-GP (14 items)
  - CORE-10 (guess!)
  - CORE-5 (ditto!)
- YP-CORE (10 items)
- CORE-LD

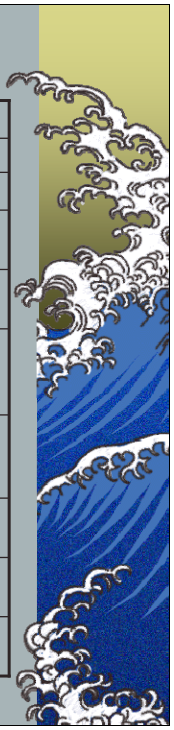
Where we've got to in 2013. Teen-CORE didn't work and was replaced with a 10 item "Young Person's" CORE and CORE-LD is designed for people with LD and is not just a translation of the CORE-OM but a new measure co-developed with PwLD that is influenced by the CORE-OM but taps other things that are crucial for PwLD but not so crucial for "neuronorms".



This was a review, done in 2003, of how we'd done against the DIY (Design Implementation & Yield) leonine aspirations for the system now updated ten years on for 2013. (By me alone, highly subjective, inter-rater reliability needs testing!)

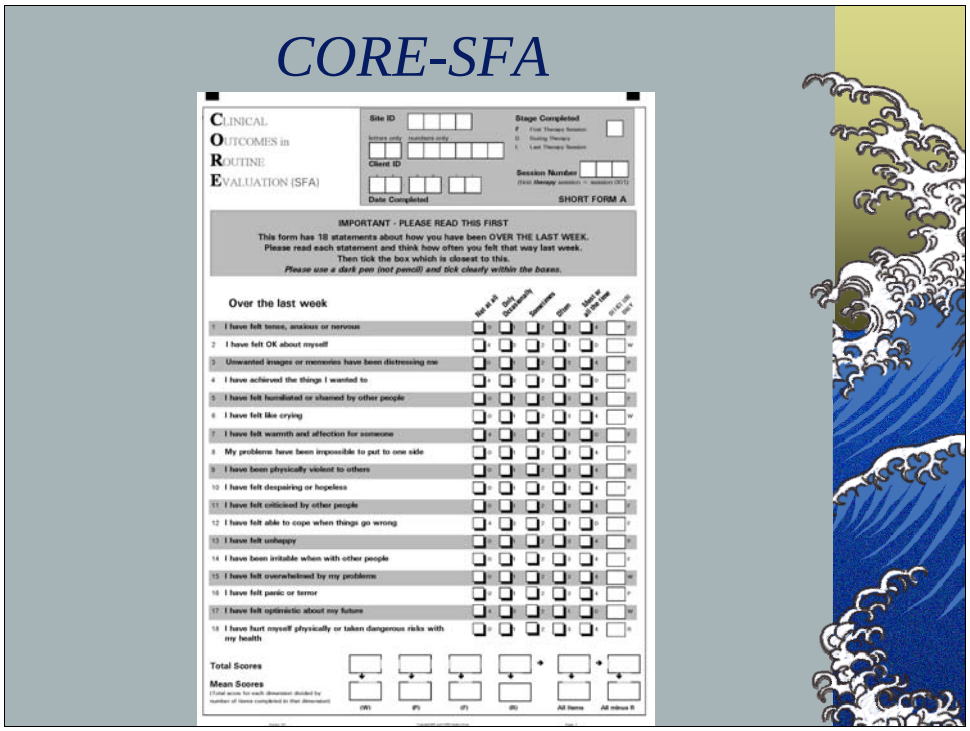
CORE: implementation

Stage 2: Implementation	OM	A	2013 comments
Measures should be “copyleft” i.e. zero cost	Y	Y	
Readable manuals provide clear guidance on use	Y?		More & training
Utilised by as wide a range of services as possible and data collated into an anonymised, referential database	Y	Y	Most UK
Monitoring of psychometric properties and “drift” in distributions established	Y		Y but no formal “drift” studies
Central database linked to batch processing with reporting of overall data and ID codes for and “best” case and “worst” case audits	Y	Y	Y but not batch processing: CORE-PC and CORE-Net
Contributing services encouraged to form peer review groups linking on geography, mode of therapy, specialist focus or other commonalities	Y?	Y?	Y but not really flourished?
Contributing services encouraged to link with research units to form 'practice research networks' (PRNs)	Y	Y	Ditto
Automated processing available for services too small for batch processing service	Y	Y	CORE-PC then CORE-Net
The batch and automated processing should become self-funding at lowest possible cost to practitioners	?	?	?

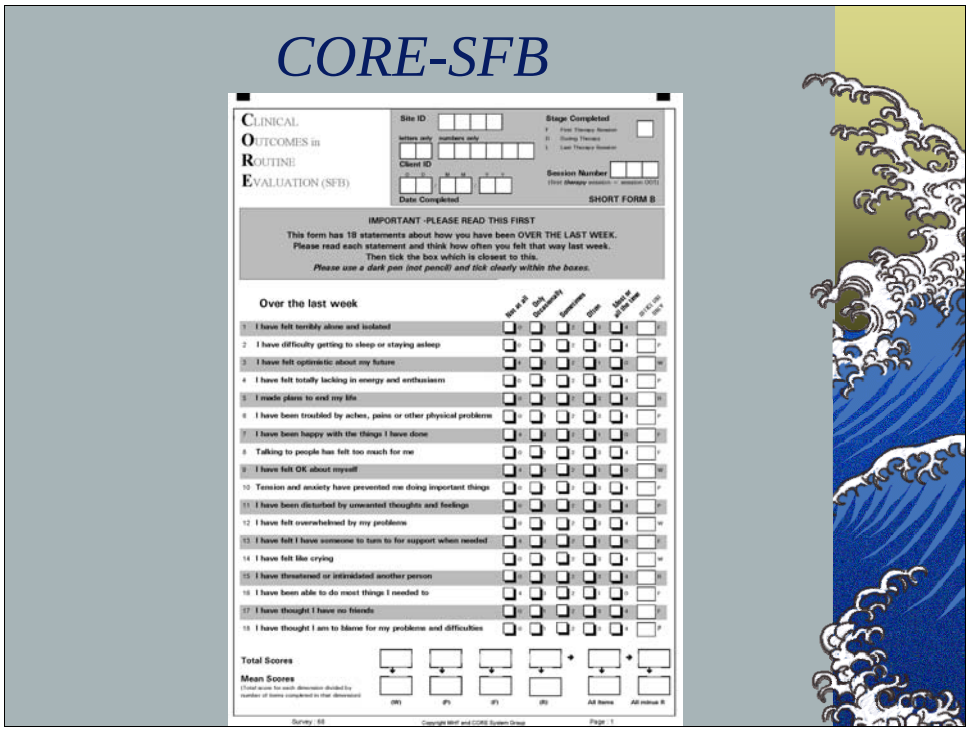


CORE & user-near: yield

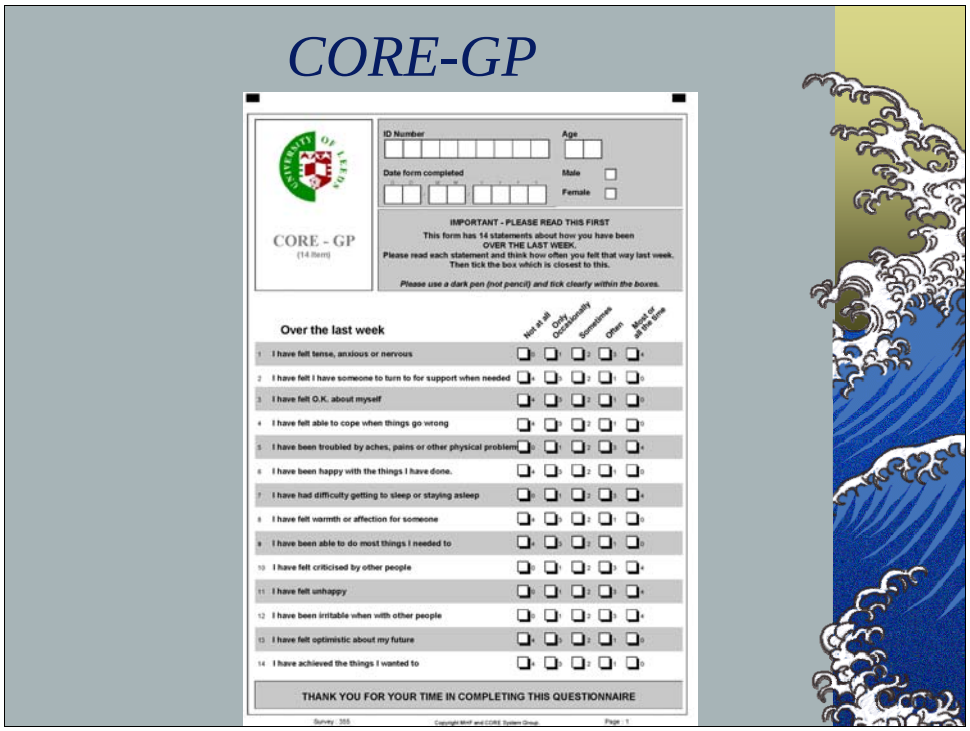
Stage 3: Yield	Overall
The measure should become the basis for outcome benchmarking	2003: "Still early to say how best to do it!" 2013: ditto
(Added 2013):	
Many publications	Very diverse and include qualitative deconstruction of what happens
Saved the NHS and others £100,000s	Cf. £2 per shot for BDI-II
Helped establish respectability of free and copyleft measures?	PHQ-9 & GAD-7 copyleft DoH says it doesn't pay for measures (not true)
Spawned other measures: CORE-GP, CORE-10, CORE-5, YP-CORE, CORE-LD & SCORE	Uptake developing
Translations led to real uptake in other countries	Norway, Netherlands, Denmark & Portugal
CORE-6D QALY scoring	
CORE-Net allowing total rethink of CORE-A	



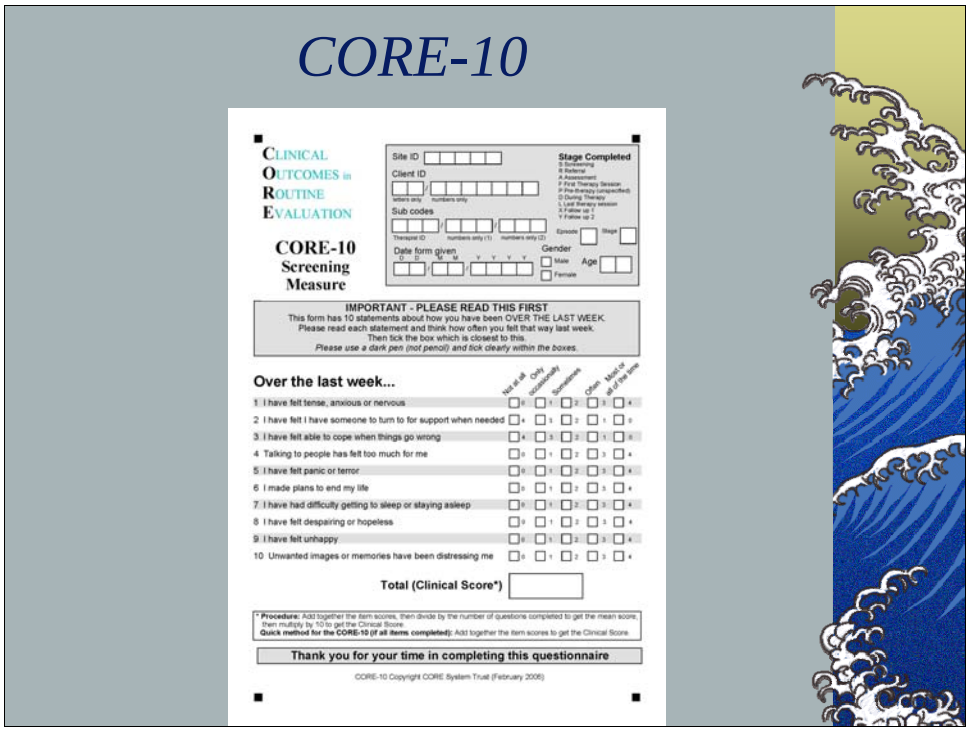
This was one of the two “parallel” short forms that we released with the CORE-OM. These were designed for sessional use and aimed to reduce the item load and, by not having the same items every time, reduce memory effects. They have been used but not widely.



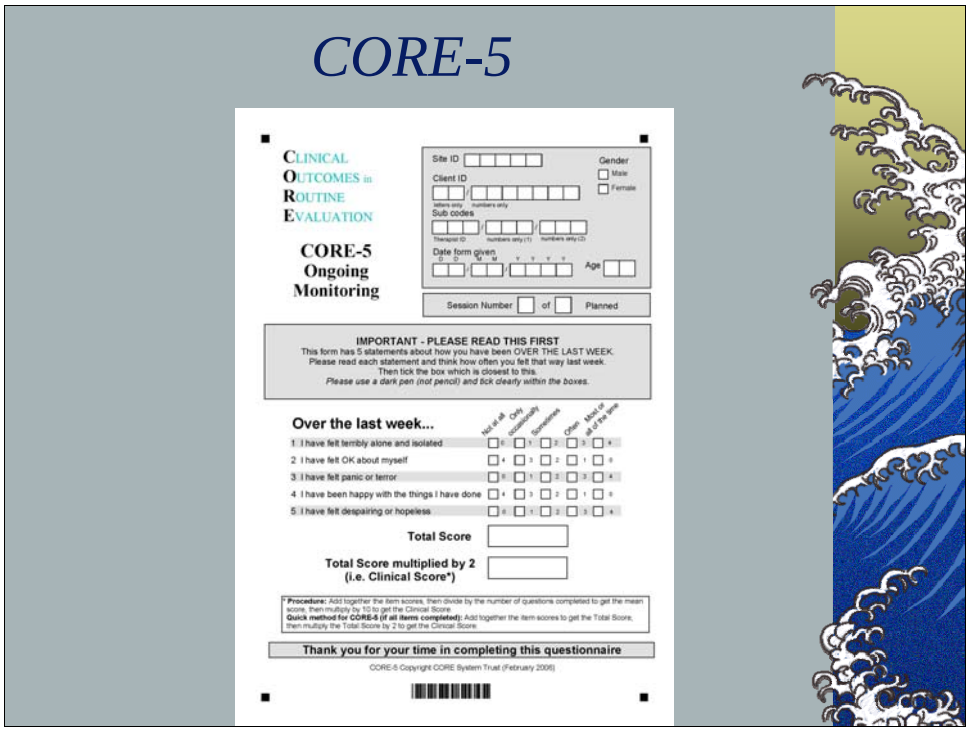
CORE-SFB, complement to the CORE-SFA



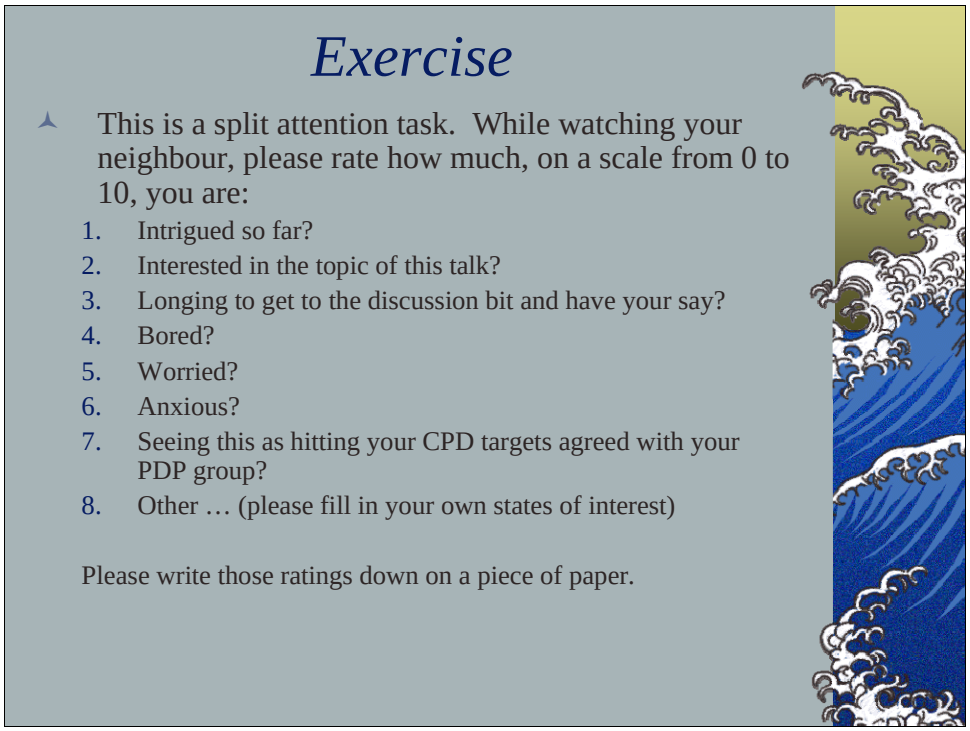
CORE-GP: for general population studies. As you can see from the Leeds University logo, this came into being when they wanted to survey their students' psychological states and didn't want to pay £2 per questionnaire but thought, rightly probably, that the full CORE-OM was a bit too clinical, at least politically, for their purpose.



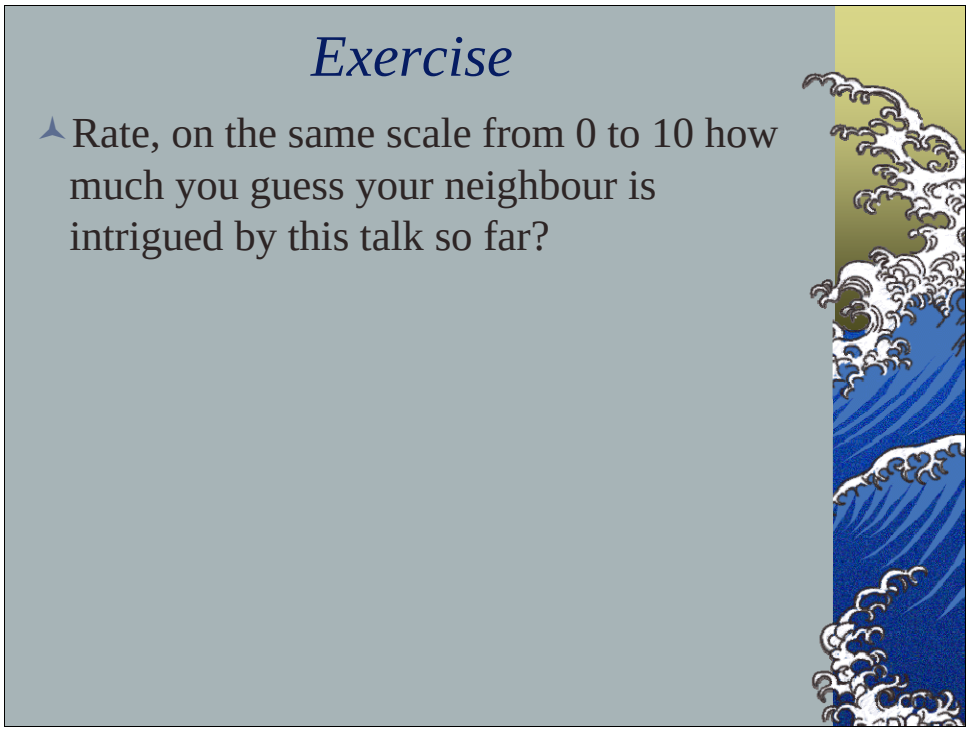
19 years ago 34 items was radically short, things have changed and that called into being the CORE-10 which has essentially replaced, a bit to my regret, the CORE-SFA/SFB.



… and even the CORE-5 though with only five items a great deal of psychometric strength really is lost.



On the night we dispensed with real slips of paper but people seemed to dive in with a real buzz. Not sure what they were saying and perhaps I don't want to know!

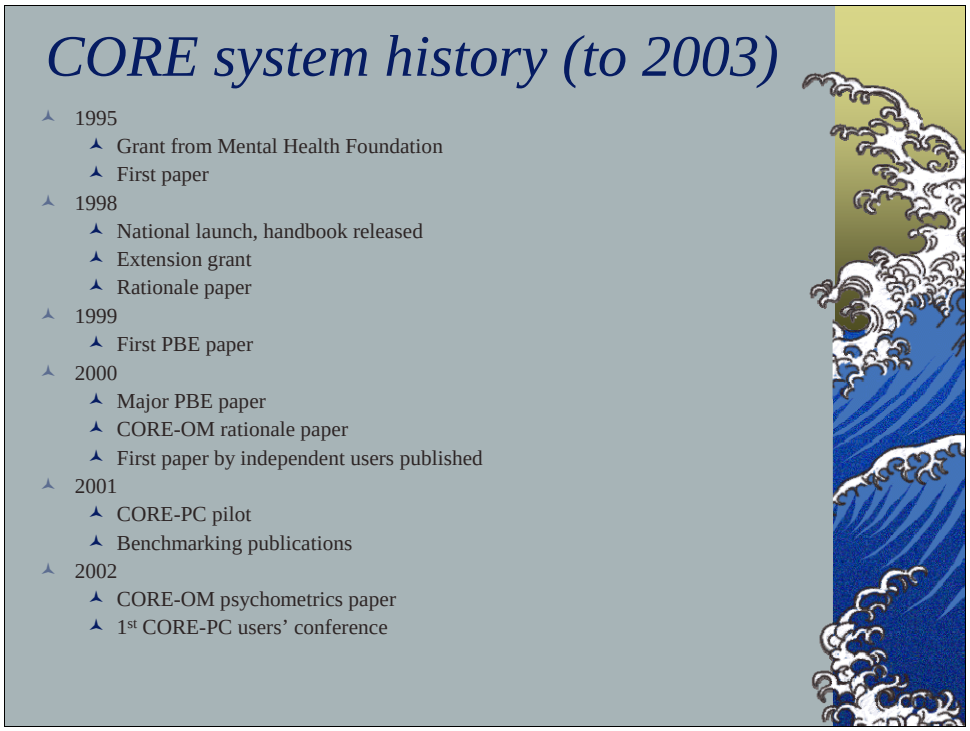


This was supposed to remind us of how much we do use inference. Clearly there are times when it is much better and simpler just to ask someone what they're thinking and feeling but there are also times when the skilled use of empathic inference can be as good or better: particularly it has that ability to get into that pane on Jokari's (?spelling?) window of the things about ourselves of which we are unaware but which are fairly visible to others.

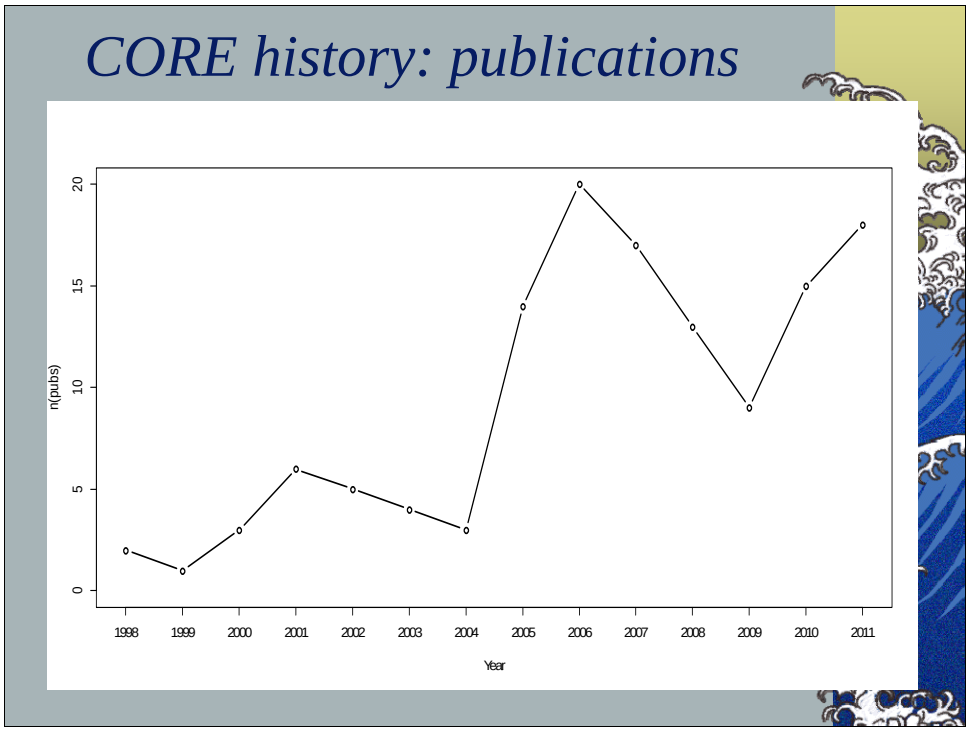
Exercise

- Compare the score you gave your neighbour with the score they had given themselves.
- Write down in brackets the score they gave you for that first reaction “Intrigued so far” alongside the one you had written down.
- Hold onto the paper for now.

More of St. George's but used to crunch us to a "fact question". I asked "What year was this picture taken?" John Hook, another medical psychotherapist from that cradle offered a year but I confess I've forgotten what he offered and I also confess I don't know the right answer. We joked about whether it could be inferred from, or a factual lower limit given by, knowing the date those sculptures arrived which I know was after I started there in 1984, or for those with eagle eyes, from the registration plate on the white minibus. We do work in a realm that has "facts" with "hard" numeric answers but which also needs softer and often entirely non-quantitative data.



Early history of the CORE system.

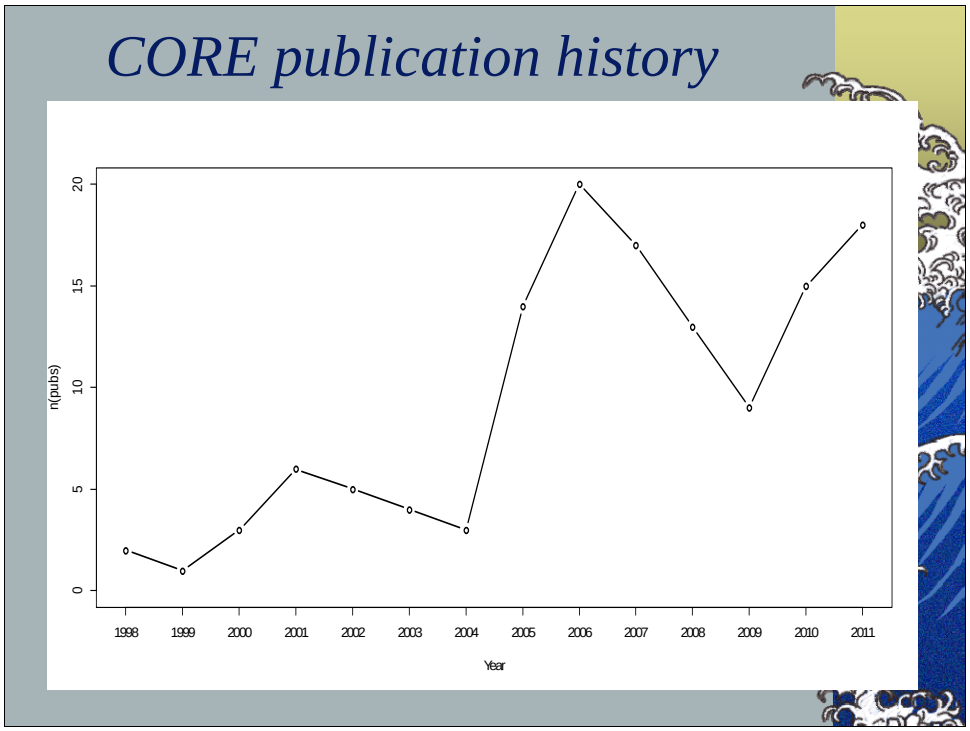


One count of the publications per year (from the list at www.coreims.co.uk) We had a quick discussion of how this looks fairly "hard" data …

Quant issues with PROMs

- 3 time issues:
  - What time frame?
  - What timing?
  - What time needed to complete?
- 3 cost issues:
  - What cost?
  - What correlation with HE valuation?
  - What hidden costs & perverse incentives?
- 3 psychometric issues:
  - What dimensionality
  - What validity?
  - What reliability?
- 3 statistical issues:
  - What aggregation?
  - What precision
  - What generalisability?

reprise



… and came to questions like: are these comparable publications (no) or are some peer-reviewed papers and other “lesser” publications (yes). Incidentally, it’s 133 publications in 63 different journals/books across a really wide diversity of topics, authors and with a slowly increasing non-UK, non-English component.l

Time issues with PROMs

- What time frame?
  - Only complete "outcome" evaluation is post-mortem …
  - … and it still changes with time (Jimmy Savile)
  - So compromise
  - "Over the last week"
- What timing?
  - Beginning & end *versus* within intervention?
  - If within, make sure you know why:
    - To avoid missing outcomes
    - To describe
    - To adapt intervention
    - Whether to use "parallel" or varying forms or not
  - Follow-up (oncology and proxy outcome measures/indicators)
- What time needed to complete?

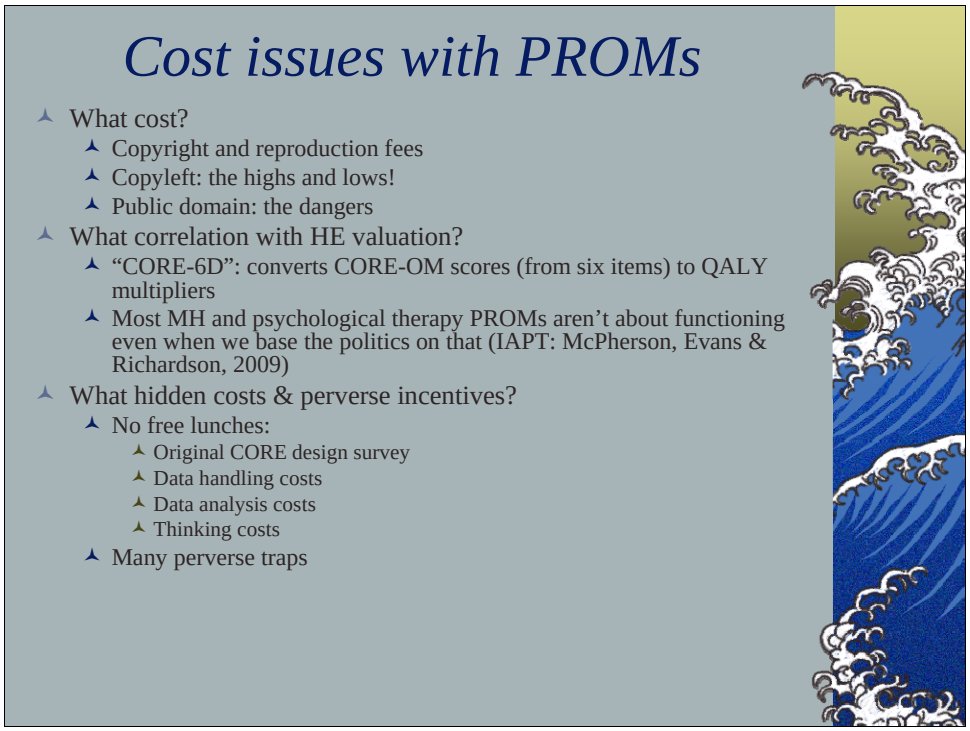
This just fleshes out some of the crucial time issues about OMs. Given that the we can never have the true, lifetime integrated, fully multidimensional "outcome" of lives of people who came for help all OM evaluation is about time compromises and choices.

Range finding vs. power steering

- C – How are you?
- R – Fine … Thanks for asking.
- C – Would you say if you weren't?
- R – [..] Probably not?
- C – Hm. .. So what should I ask?
- R – [....] Well, what I say to the internal postie is "What are you today?" …. "from 1 to 10"
- C – [……] Hm. So … what are you today?
- R – [..] 7
- C – That sounds OK. …some people aren't happy with anything less than a 10
- R– Ah but I think I was down about a 3 before I was off sick.

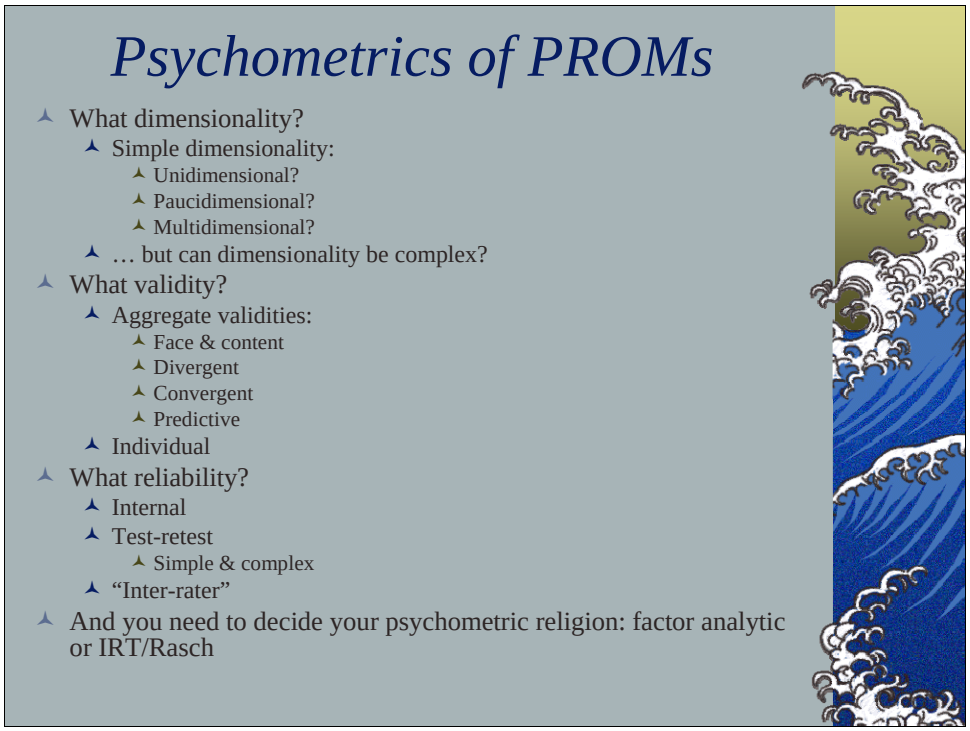
Evaluation of time frames in conversational encounters is negotiated and quite flexible. A time frame (5-8 weeks) is given by the "before I was off sick" with a change (3 to 7) just as I invite a lateral, cross-sectional way to see how the personal numerical frame might relate to how others might use 0-10 ("some people aren't happy with anything less than a 10"

(Please note for those who weren't there: I'm not suggesting that this sort of "scaling", which is used quite a lot in some areas of systemic therapy, behavioural and cognitive therapies, is any more or less good than self-report measures with their time framing.)



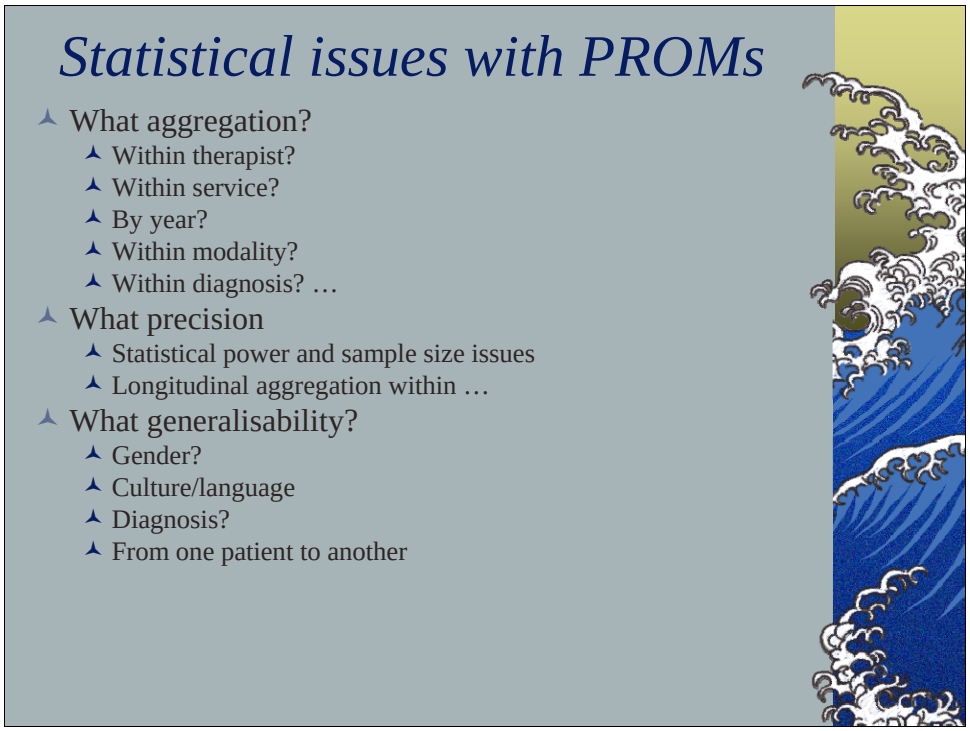
Cost issues: all investigation costs time. Gave the anecdote on the day about how few providers or purchasers (that was the jargon back then, not services and commissioners) were willing to answer a question about how much of the cost of a service should be given to routing change/outcome evaluation. I don't sense that the enthusiasm to set a figure to that was much greater in the room than it was those 18/19 years ago when we did that survey.

Tough, spiky choices all of these.



In psychiatry and psychotherapy we grotesquely ignore and don't train ourselves to understand psychomtrics. In so doing, we leave it to those who arguably overvalue it: largely psychologists and the very few people making an academic or commercial living in the niche in statistics of "professional psychometrician". Unfortunately, psychometrics is often polarised between two camps: those modelling things as if they were continuous scores and perhaps Gaussian in distribution (traditional factor analytic approaches in general); and those treating things as discontinous (the "item response theory" and Rasch approaches). Sadly, almost all of the very sophisticated mathematical models used by either camp are designed to do cross-sectional comparisons between people on measures. In psychotherapy we certainly do have an interest in such comparison and parameters that emerge from those psychometric analyses are not completely unhelpful or irrelevant to us, however, we are much more interested in the psychometric quality of the measurement of change *within* individuals and mathematical models and analytic methods to explore that are not well developed (not least because they're intrinsically harder to define and calculate).

All very tough and prickly, may be shoots on the trees beyond though.



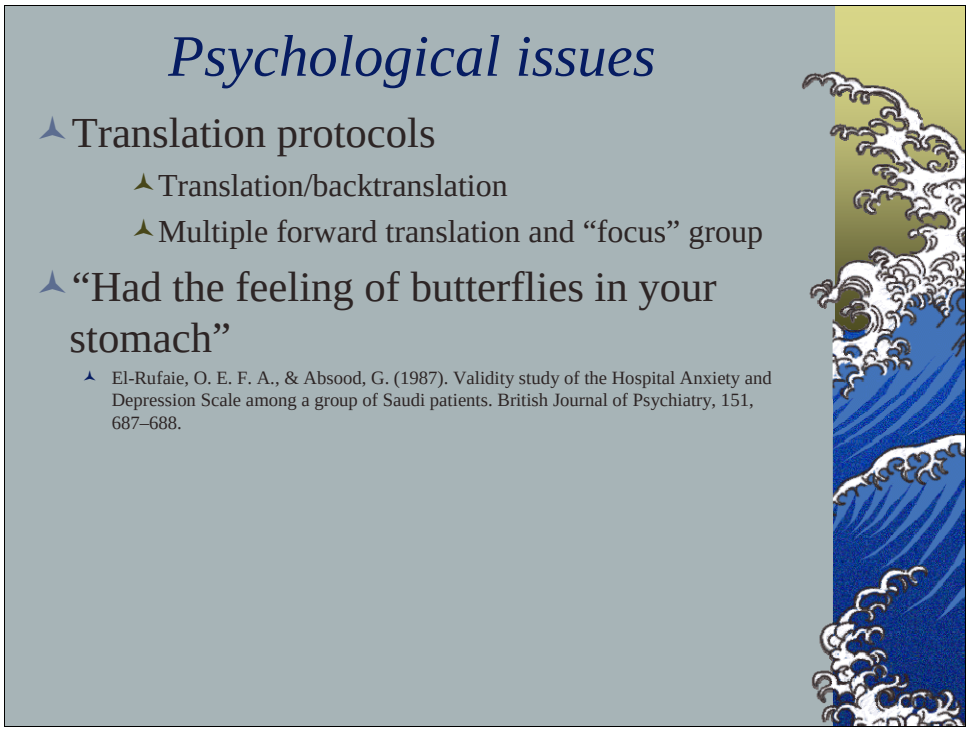
Statistics and psychometrics overlap but have rather different foci: psychometrics is largely about the value we can attach to scores: traditionally can we say they are reasonably reliable and valid; statistics is what we can say when we generalise from sample data to other data, the traditional statistical model, though not the only one, being of generalisation from sample to population and hence the creation of the whole toolkit of p values and inferential tests and of confidence intervals to define precision of estimation. Given that self-report OM data are of useful but very imperfect reliability and validity, statistical aggregation is potentially very useful to extract signal from within noise. However, there are many statistical issues about this that are ignored only a bit less often than psychometric issues but when used are just as often used incorrectly or misleadingly.

The walk back to the beginning goes fairly well: the hotel is much more visible now but that’s a tough old climb up from the beach looming at the end!

Qual issues with PROMs

- Umpteen psychological issues
  - What's going on in the mind of the respondent?
  - What do they construe is going on in the mind of the scorer/other? (Do they construe that person?)
  - Expectancy effects
  - Focusing effects
- Umpteen sociological, anthropological and political issues
  - Who wants PROMs and why?
  - Do they empower service users or immunise politicians?
  - How do they redefine relationships and cultures?
- What relationships are PROMs defining or shaping?

While I moaned the under-exploration of quantitative issues about OMs, the under-exploration of qualitative issues is scandalous. The key problem underlying this is the enormous pressure to reify and totemise questionnaire (and rating scale) scores.



One beautiful illustration of such reification is the attitude to translation of measures. The traditional gold standard way of doing this is independent forward translation and back-translation: you then compare the back-translation with the original and if they seem to match up well, you assume you have succeeded in getting a good enough translation and you proceed to conventional psychometric exploration of the translation. El-Rufaie and Absood translated the Hospital Anxiety and Depression Scales in this way and carried out an exploratory factor analysis of their translation into Arabic with "a group of Saudi patients". They found that 13 of the 14 items behaved broadly as they do in UK English samples but the one that mentioned "butterflies in your stomach" didn't. At that point they realised that the lovely English vernacular expression has no literal equivalent in Arabic and in Saudi culture.

We have used and developed a different method of getting multiple independent forward translations, one from a professional interpreter or translator, some from mental health professionals and, crucially, some from lay people. We then have a focus group and someone who knows the CORE-OM really well (me!) sits in and watches and listens and, though French is the only language which allowed me to claim to understand a lot of what happened, I can see when a group engage and a really good translation will emerge. I also learn a great deal about the complexities of language and culture dependence in the CORE-OM and in many other such questionnaires when the questions come:

"Why did you say that?"

"We don't have any one word that exactly matches …., we think we have to chose between … and … and we're wondering …."

"Is that grammatically correct in English?"

"Do we need so many words?"

"We just don't say 'feel' that much."

"What shall we do about gender?"

"How formal should we be? We have different linguistic forms for respectful 'you' and more familiar 'you'. We think we should go with …"

"We probably have different ways of saying this in different parts of the country"

"You don't have an item about connection with your [deceased] ancestors"

"We have huge differences between formal written language and spoken language and we're not sure …"

"You can't say 'please, please, please' like that in xxxx, people won't see this as professional, you have to say 'You must fill in all 34 items' then they will feel it is a respectable measure."

Questionnaires will never be all things and the same thing to all people even within quite tight linguistic and cultural groups.

This is one item from the French translation process: you can see the original English, 11 of the forward translations, some with comments, and what the focus group decided on which is different, and includes a rather nice French phrase "dans ma peau", "in my skin" that wasn't in any of the forward translations. That will go to back translation and "talk aloud" field testing with old, young, ill educated and some immigrants to check so isn't final yet.

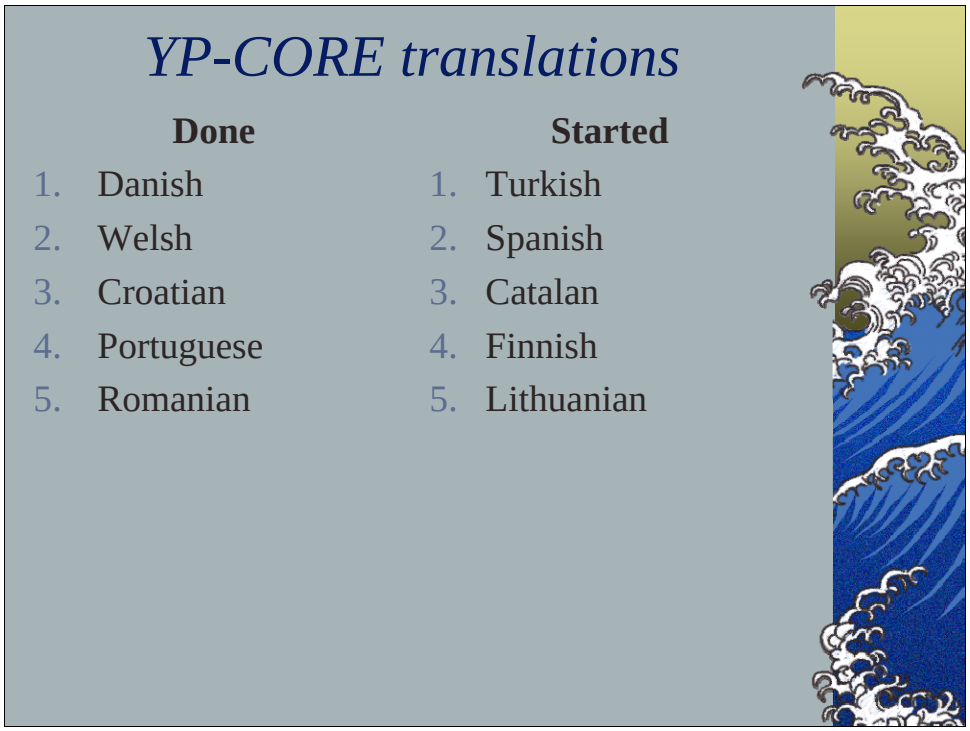
CORE-OM translations

Completed

1) Gujarati
2) Norwegian (Bokmal)
3) Italian
4) Slovak
5) Swedish
6) Icelandic
7) Albanian
8) Greek
9) Dutch
10) Danish
11) Portuguese
12) Croatian
14) Welsh
15) Serbian
16) German
17) Lithuanian
18) Polish
19) Turkish
20) Finnish
21) Spanish
22) Argentine Spanish
23) Xhosa
24) Romanian
25) **British Sign Language (BSL)**

Ongoing (from nearly finished to just starting)

1) French
3) Bulgarian
4) Hungarian
5) Russian
6) Brazilian (slightly different from the Portuguese)
7) Bosnian
8) Farsi
9) Catalan
10) Urdu
11) Punjabi
12) Kannada
13) Sami
14) Tamil
15) Czech
16) Japanese
17) Latvian
18) Somali
19) Arabic
20) Afrikaans
21) Korean
22) Hebrew
23) Maltese
24) Kurdish
25) Bahasa Malayu
26) Hindi



We haven’t translated the YP-CORE so extensively.

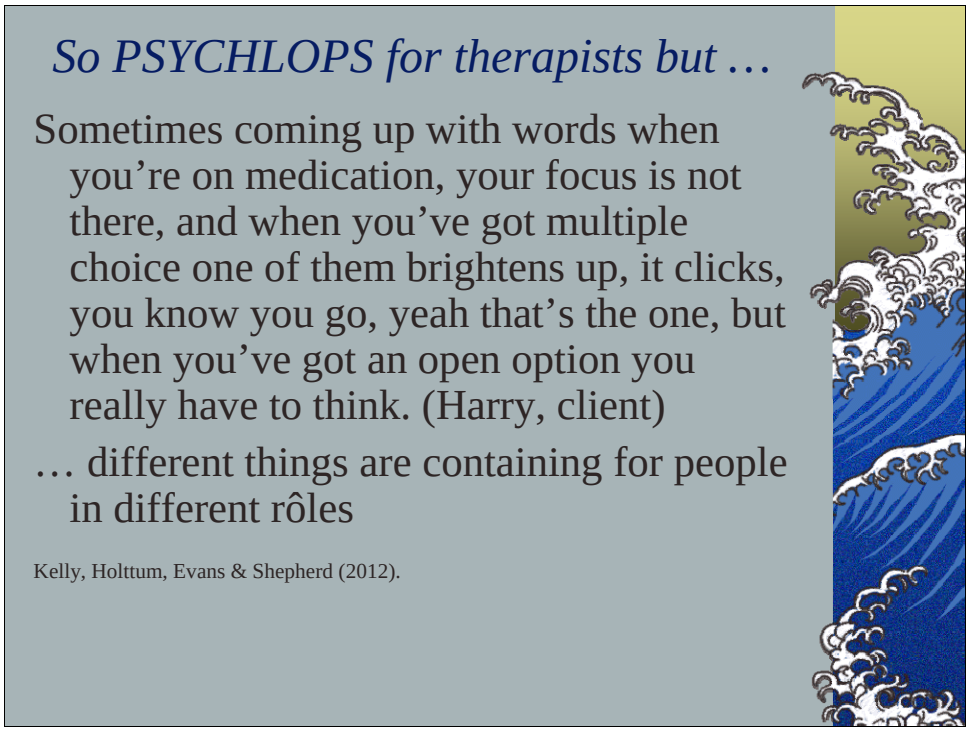
Sociological issues

I get the CORE filled in because we're expected to, and if I don't buy into it then I think there's something slightly unethical about doing it, 'cause if you're not using the information afterwards then why are you getting someone to fill it in?, and ultimately the CORE is about the service, it's about the Trust that I work in. . . for me it feels sometimes like the motive behind it is at a broader organisational level and it's not about that individual's life. (Beth, therapist)

These quotes are from Kelly, Holttum, Evans & Shepherd 2012. Small sample of CBT practitioners and of clients who have had psychotic problems from a large MH Trust in the UK.

What's containing for whom?

It's much easier to do the PSYCHLOPS than the CORE. I feel much more tentative about giving the CORE, partly because you just don't know how people are going to respond to it and it feels quite invasive in many ways, whereas the PSYCHLOPS feels much less invasive. (John, therapist)



The researcher, Vikki Kelly and I think her first supervisor and the therapists all felt pretty sure that PSYCHLOPS, a measure in which the clients define the problems they will rate would be more acceptable to clients and that it was much less controlling. The findings from the clients were almost the reverse. The sample is small, is a niche area, and any sense of pitting the one form of measurement against the other was restricting but the findings about the complexity of the experiences, thoughts and feelings of the practitioners and the clients are important and will replicate I'm sure and, when replicated and explored more, may help us stop reifying and totemising OMs and help us start using them more sensibly and thoughtfully.

Sometimes it’s hard to keep up on these working afternoons and evenings.

Exercise

- Still got that piece of paper with your score on top and the score your neighbour gave you?
- Compare the scores.
- We have to infer internal states: I like the analogy with watching the face not the abdomen when examining the abdomen.
- Discuss what cues you were using when each tried to rate how much the other was “Intrigued so far”.

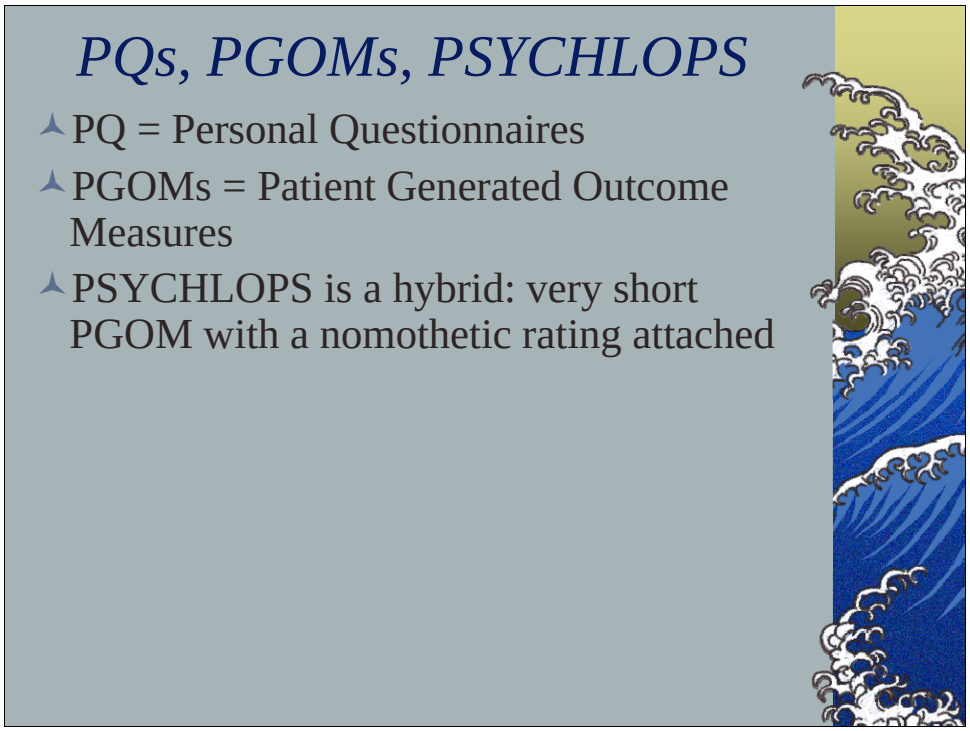
Sessional tracking/power steering

- Great if:
  - It's clearly tracking from outside to secure an ending score and a trajectory
  - It's used for steering and it's not **the** outcome measure and using it this way is congruent with theory:
    - CBT, BT, some or much systemic therapy
    - Some humanistic therapy
    - Some or much eclectic/integrative therapy
- Keeps an open stance to additional information

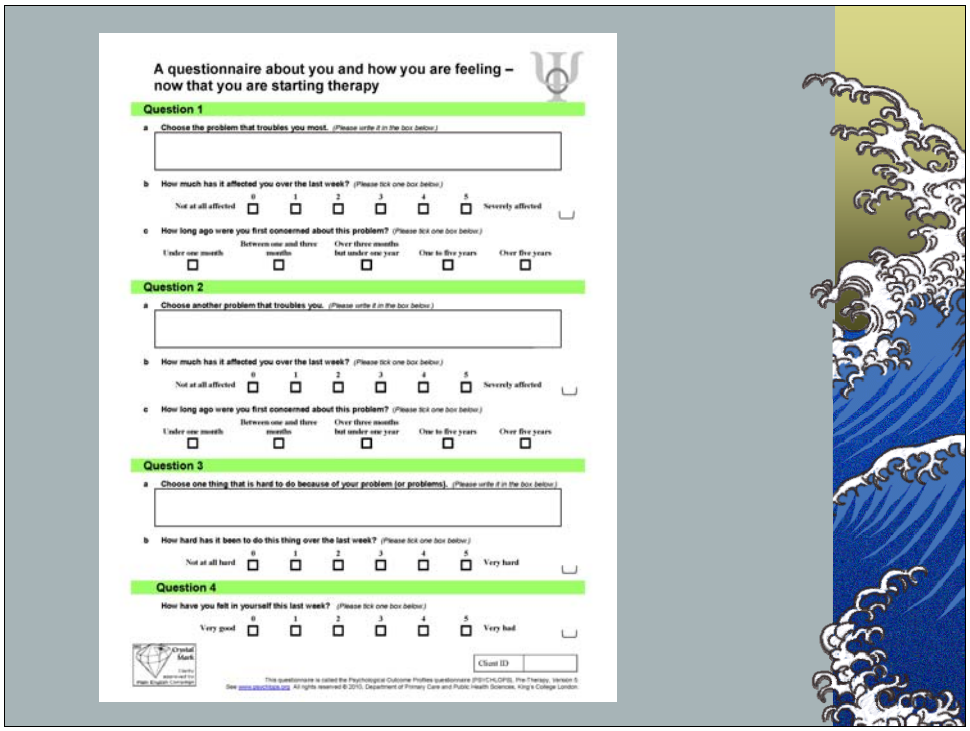
OK, let's get back to power steering in the light of some deconstruction (but not destruction I hope) of outcome and change measurement. My views on "sessional tracking"

Problems with power steering

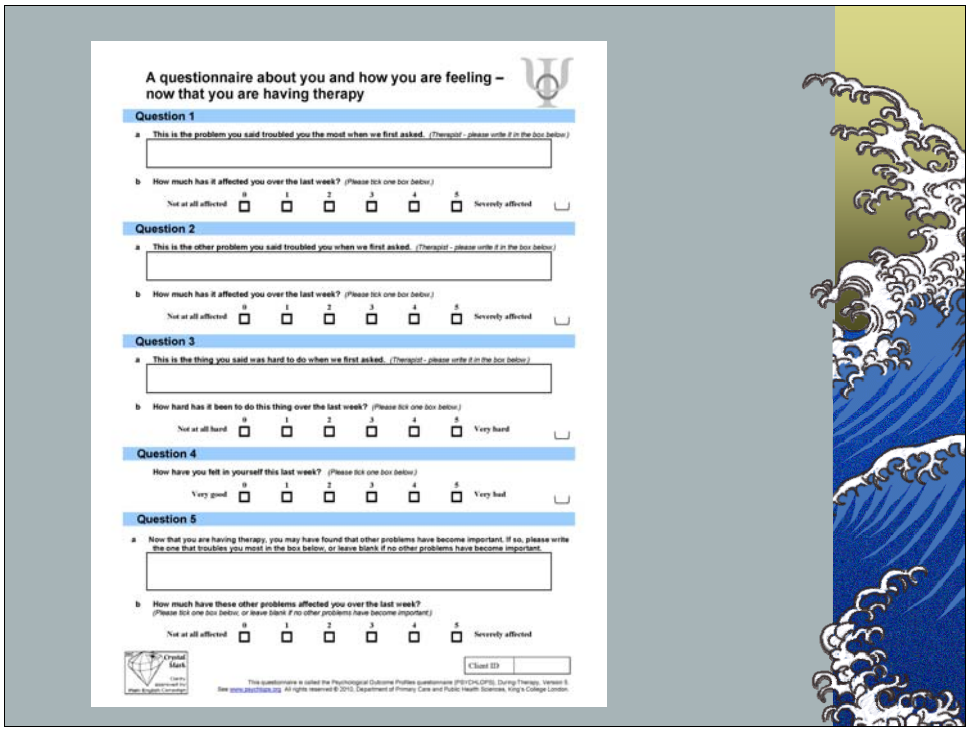
- Reduced channel of communication
- Nomothetic: i.e. same for all (not necessarily but little literature on use of idiographic or PQ steering)
- If a nomothetic measure like CORE-OM or OQ-45 the channel was defined by professionals (ultimately)
- Assumes that what's important is conscious (or reflected in the conscious completion of the measure)
- Offers anxiety containment when that may not be ideal
- All the research to date has used the communication measure as the outcome measure



One way to avoid some problems is to move to more personal measures. I'd already mentioned PSYCHLOPS which I see as an important complement to nomothetic measures like the CORE-OM, OQ-45 and all conventional questionnaires. A nomothetic measure imposes the same measuring frame on everyone, idiographic measures allow each individual to have his or her own frame while hoping to keep (and succeeding usually in keeping) some of the quantitative and psychometric utility of nomothetic measures.



First box is to write in a problem, 2nd box allows a 2nd problem, the third box is somewhat different: something it has been hard to do because of the problems: a neat way to get some sense of impact beyond the simple numerical ratings of impact. PSYCHLOPS is hybrid because the final rating is a standard nomothetic one: “How have you felt in yourself this last week?”



A questionnaire about you and how you are feeling – now that you are having therapy

Question 1

a This is the problem you said troubled you the most when we first asked. *(Therapist - please write it in the box below.)*

b How much has it affected you over the last week? *(Please tick one box below.)*

Not at all affected 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ Severely affected ☐

Question 2

a This is the other problem you said troubled you when we first asked. *(Therapist - please write it in the box below.)*

b How much has it affected you over the last week? *(Please tick one box below.)*

Not at all affected 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ Severely affected ☐

Question 3

a This is the thing you said was hard to do when we first asked. *(Therapist - please write it in the box below.)*

b How hard has it been to do this thing over the last week? *(Please tick one box below.)*

Not at all hard 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ Very hard ☐

Question 4

How have you felt in yourself this last week? *(Please tick one box below.)*

Very good 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ Very bad ☐

Question 5

a Now that you are having therapy, you may have found that other problems have become important. If so, please write the one that troubles you most in the box below, or leave blank if no other problems have become important.

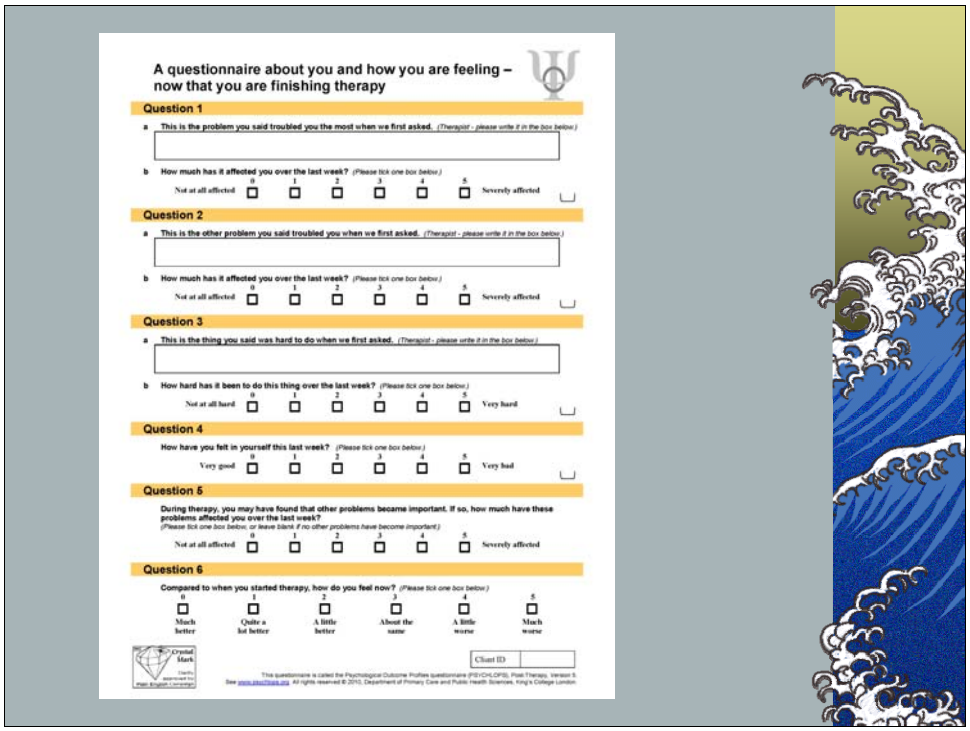
b How much have these other problems affected you over the last week? *(Please tick one box below, or leave blank if no other problems have become important.)*

Not at all affected 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ Severely affected

Crystal Mark

Client ID

Within therapy form which allows addition of an emergent problem



A questionnaire about you and how you are feeling – now that you are finishing therapy

Question 1

a This is the problem you said troubled you the most when we first asked. *(Therapist - please write it in the box below.)*

b How much has it affected you over the last week? *(Please tick one box below.)*

0 1 2 3 4 5

Not at all affected ☐ ☐ ☐ ☐ ☐ ☐ Severely affected

Question 2

a This is the other problem you said troubled you when we first asked. *(Therapist - please write it in the box below.)*

b How much has it affected you over the last week? *(Please tick one box below.)*

0 1 2 3 4 5

Not at all affected ☐ ☐ ☐ ☐ ☐ ☐ Severely affected

Question 3

a This is the thing you said was hard to do when we first asked. *(Therapist - please write it in the box below.)*

b How hard has it been to do this thing over the last week? *(Please tick one box below.)*

0 1 2 3 4 5

Not at all hard ☐ ☐ ☐ ☐ ☐ ☐ Very hard

Question 4

How have you felt in yourself this last week? *(Please tick one box below.)*

0 1 2 3 4 5

Very good ☐ ☐ ☐ ☐ ☐ ☐ Very bad

Question 5

During therapy, you may have found that other problems became important. If so, how much have these problems affected you over the last week?
(Please tick one box below, or leave blank if no other problems have become important)

0 1 2 3 4 5

Not at all affected ☐ ☐ ☐ ☐ ☐ ☐ Severely affected

Question 6

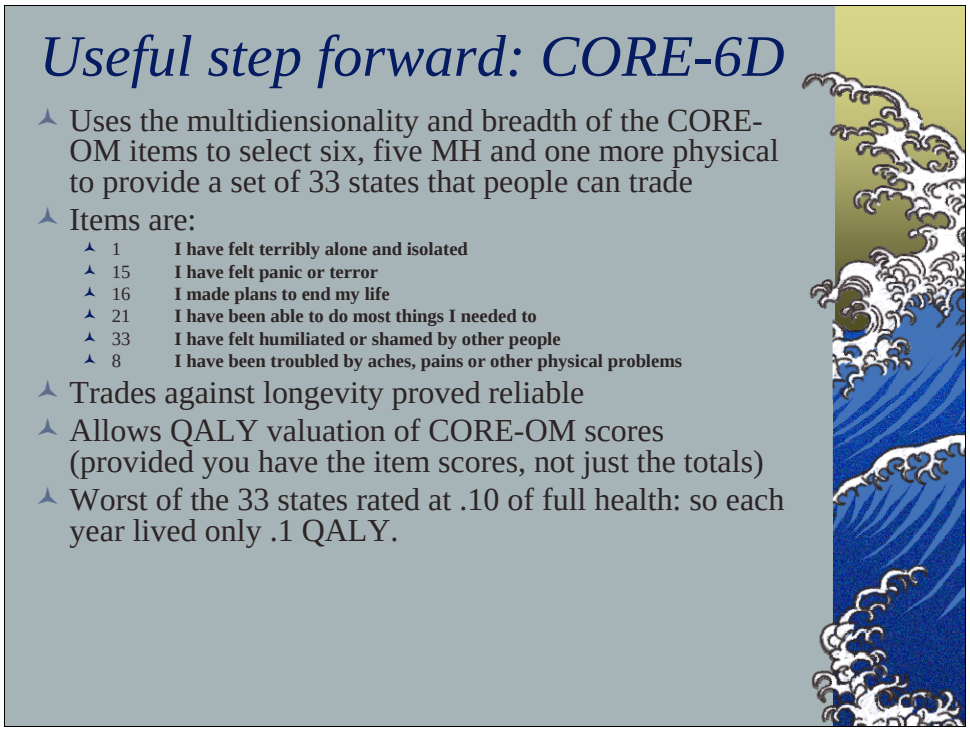
Compared to when you started therapy, how do you feel now? *(Please tick one box below.)*

0	1	2	3	4	5
☐	☐	☐	☐	☐	☐
Much better	Quite a lot better	A little better	About the same	A little worse	Much worse

Crystal Mark

Client ID

End of therapy form.



One huge issue in the current political climate is that most decisions about public health care are, or claim to be, made on health economic (HE), cost efficiency grounds. There are huge problems here but there has to be a basic honesty about this: there is almost infinite demand, there is huge need and to move from addressing demand to addressing need involves choices and rationing and there are finite and, though politicians equivocate, currently reduced funding and there is a clear actual underlying trend of increasing need based on the changing age structure of the population alone (not to mention obesity and other largely self-inflicted health problems).

Despite those harsh health economic realities, clinicians still, and particularly in the MH world, tend to make decisions and use measures that are not mapped to HE valuations, in fact, many decisions are made and argued for on functioning grounds (most of the stepped care and IAPT model for example) while only a tiny proportion of the change measures in the huge systematic review that drove the first IAPT/stepped care design (the NICE guidelines on depression for example) are measures of function. I'm biased but I think that McPherson, Evans & Richardson (2009) is an under-cited reference to that problem! (The idea was the late Phil Richardson's with Susan, she had done a lot of the hard work when Phil died, like Chris, and at the same age, sadly early, I got roped in and Susan and I finished the work.)

There are ways health economists use to define QALY valuations of different states of health and one, "time trade off" modelling, can be applied to questionnaire data. It's a huge amount of work (and I had no personal input to this so, excepting the fact that they happened to use the CORE-OM, I'm unbiased on this). The team from Sheffield, principally Ifigeneia Mavranezouli whose PhD I believe it was, did the hard work and have come up with a way of translating from CORE-OM data to QALY state valuations.

This showed that the worst state they rated was given only a tenth of the value of a well state, congruent with much other evidence that both sufferers and others rate the impact of psychological states of ill health as very severe in their impact on quality and valuation of life.

It will be interesting to see what applying this valuation method to existing and new CORE-OM will say about value gains from psychotherapy. Let's lobby that the collection of new data is sufficiently funded to allow at least a couple of years of follow-up for at least a good variety of settings and studies.

Concluding pleas(e) – technological

- Our current outcome measures are little better than asking oncology patients at the end of chemotherapy how long they think they'll live
- … or assuming that the temperature at the end of treatment of a severe meningitis will predict neurological outcome
- But that's better than nothing!
- There are a plethora of them (CORE, OQ, PHQ, GAD …)
- But CORE is the best (oops) …
- … partly because of the CORE-6D scoring

I hope the concluding pleas speak for themselves.

Concluding pleas(e) – technological contd.

- Now we need to separate tracking measures from "outcome" measures
- We need longer term outcomes and follow-up … predictive validity
- That won't come from RCT controlled separation of intervention arms
- We need to chip away at the "deathly hallows" of strong (DBRCT) causal attribution …
- … while recognising that NICE and the government will need the DBRCT throttle for medication costs (physical and MH)

Concluding pleas(e) – relational

- That's all well and good on the technological side but ..
- We also need to stop being embarrassed that our domain is about relationships with people in distress and about the arts of using the relationship, that it's not a technology
- We can usefully lead offering that understanding back into psychiatry (Bracken et al. 2012) and also into "general" medicine

References

Ashworth, M., Evans, C., & Clement, S. (2009). Measuring psychological outcomes after cognitive behaviour therapy in primary care: a comparison between a new patient-generated measure "PSYCHLOPS" (Psychological Outcome Profiles) and "HADS" (Hospital Anxiety and Depression Scale). Journal of Mental Health, 18, 169–177. doi:10.1080/09638230701879144

Ashworth, M., Robinson, S. I., Evans, C., Shepherd, M., Conolly, A., & Rowlands, G. (2007). What does an idiographic measure (PSYCHLOPS) tell us about the spectrum of psychological issues and scores on a nomothetic measure (CORE-OM)? Primary Care and Community Psychiatry, 12, 7–16.

Ashworth, M., Robinson, S. I., Godfrey, E., Parmentier, H., Shepherd, M., Christey, J., … Matthews, V. (2005). The experiences of therapists using a new client-centred psychometric instrument, PSYCHLOPS (Psychological Outcome Profiles). Counselling & Psychotherapy Research, 5, 37–42. doi:Journal-Article Empirical-Study Qualitative-Study Journal Peer-Reviewed-Journal

Ashworth, M., Robinson, S. I., Godfrey, E., Shepherd, M., Evans, C., Seed, P., … Tylee, A. (2005). Measuring mental health outcomes in primary care: the psychometric properties of a new patient-generated outcome measure, "Psychlops" ('Psychological Outcome Profiles'). Primary care mental health, 3, 261–270.

Ashworth, M., Shepherd, M., Christey, J., Matthews, V., Wright, K., Parmentier, H., … Godfrey, E. (2004). A client-centred psychometric instrument: the development of "PSYCHLOPS." Counselling & Psychotherapy Research, 4, 27–33.

Bracken, P., Thomas, P., Timimi, S., Asen, E., Behr, G., Beuster, C., … Yeomans, D. (2012). Psychiatry beyond the current paradigm. The British Journal of Psychiatry, 201(6), 430–434. doi:10.1192/bjp.bp.112.109447

El-Rufaie, O. E. F. A., & Absood, G. (1987). Validity study of the Hospital Anxiety and Depression Scale among a group of Saudi patients. British Journal of Psychiatry, 151, 687–688.

Evans, C., Ashworth, M., & Peters, M. (2010). Are problems prevalent and stable in non-clinical populations? Problems and test-retest stability of a patient-generated measure, "PSYCHLOPS" (Psychological Outcome Profiles), in a non-clinical student sample. British Journal of Guidance & Counselling, 38(4), 431–439. doi:10.1080/03069885.2010.503701

Evans, C. (2012). Cautionary notes on power steering for psychotherapy. Canadian Psychology/Psychologie canadienne, 53(2), 131–139. doi:10.1037/a0027951

Kelly, V., Holttum, S., Evans, C., & Shepherd, M. (2012). Social constructions of PSYCHLOPS (psychological outcome profiles) in the context of CBT for psychosis. Counselling and Psychotherapy Research, 1–10. doi:10.1080/14733145.2012.672439

Kelly, V., Holttum, S., Evans, C., & Shepherd, M. (2012). Social constructions of PSYCHLOPS (psychological outcome profiles) in the context of CBT for psychosis. Counselling and Psychotherapy Research, 1–10. doi:10.1080/14733145.2012.672439

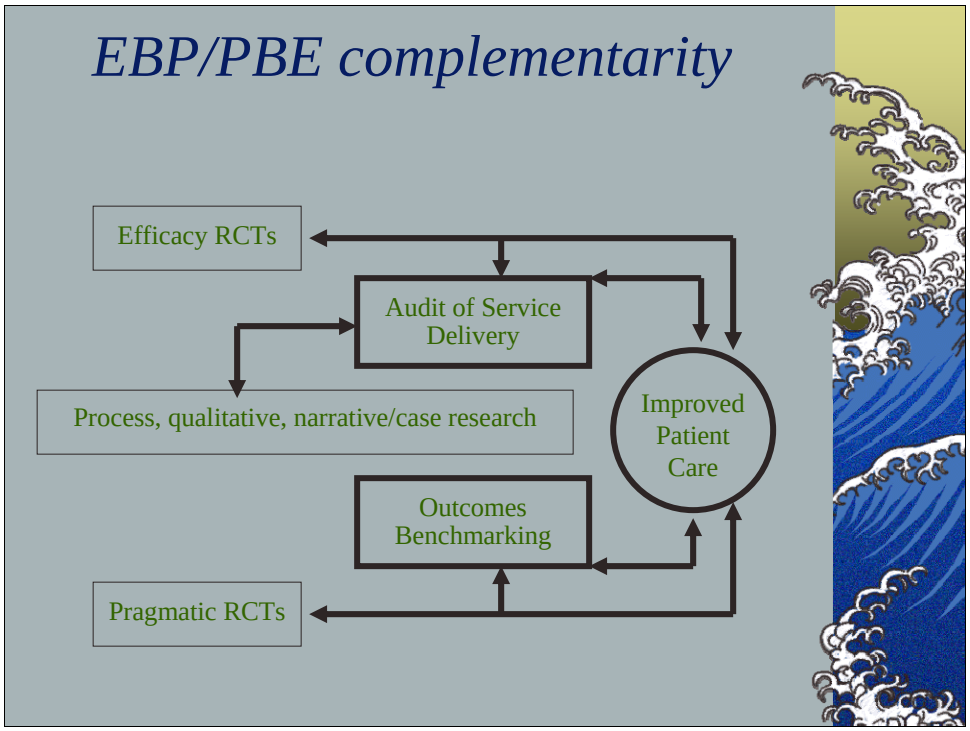
Mace, C. (2006). Setting the world on wheels: some clinical challenges of Evidence-Based Practice. Group analysis, 39(3), 304–320. doi:10.1177/0533316406066592

Mavranezouli, I., Brazier, J. E., Rowen, Donna, D., & Barkham, M. (2012). Estimating a Preference-Based Index from the Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM): valuation of CORE-6D. Medical Decision Making, 20, 321–333. doi:10.1177/0272989X12464431

Mavranezouli, I., Brazier, J. E., Young, T. A., & Barkham, M. (2010). Using Rasch analysis to form plausible health states amenable to valuation: the development of CORE-6D from a measure of common mental health problems (CORE-OM). Quality of Life Research, 20, 321–333. doi:10.1007/s11136-010-9768-4

McPherson, S., Evans, C., & Richardson, P. (2009). The NICE Depression Guidelines and the recovery model: is there an evidence base for IAPT? Journal of Mental Health, 18, 405–414. doi:doi: 10.3109/09638230902968258

Robinson, S. I., Ashworth, M., Shepherd, M., & Evans, C. (2006). In their own words: a narrative-based classification of clients' problems on an idiographic outcome measure for talking therapy in primary care. Primary care mental health, 4, 165–174.



A quick reminder that this isn't about overturning EBP: we need those more "laboratory model" studies though I personally think their allegiance to the RCT model needs to change, this is about complementary explorations and evaluations of what happens in psychotherapies and in the psychotherapy that can happen in any, even brief, GP consultation.

A picture of the warm discussions at the Society for Psychotherapy Research the evening after we finally clambered up the cliff. Chris as ever, evaded the camera. Photographic credits here to my daughter, now 19 and about a month old when we first started the CORE project.

I'd overloaded the presentation and there was on 15 minutes of discussion but it was very good and many of the themes recurred, some of them repeatedly and in very diverse sessions and in both talk about formal psychotherapy and about GP work, through the next two days.

If you want to pick any of this up or get copies of any papers I can give you, Email me on that address and harrass gently if I'm slow and harrass with cc's to chris.evans@nottshc.nhs.uk and c.evans@nottingham.ac.uk just in case my primary address is misbehaving.