

Fundamental Measurement: The Need Fulfilment Quality of Life (N-QOL) Measure

Paul C Langley, PhD¹; Stephen P McKenna, PhD²

¹College of Pharmacy, University of Minnesota, Minnesota MN USA; Maimon Research, Tucson AZ USA

²Population Health, University of Manchester, Manchester UK; Galen Research, Manchester UK

ABSTRACT

The quality adjusted life year (QALY) has serious problems related to its failure to adhere to measurement theory. If a QALY is to be meaningful, the utility score that translates time spent to an equivalent time spent in so-called perfect health must have ratio properties (i.e., it must support multiplication). Multiattribute utility scores (e.g. those generated by the EQ-5D-5L) fail to meet this standard. The multiattribute instruments produce ordinal scores that lack a true zero and they generate negative values. The manifest deficiencies of multiattribute utility instruments render them unfit, not only as a measure of therapy response but also in generating QALY claims. After 30 years of belief in their use, utilities and QALYs are clearly analytical dead ends.

The purpose of this commentary is to demonstrate a coherent way forward in health technology assessment by focusing, not on clinical attributes as surrogates for quality of life, but on measures that are based on a conceptual model describing patient value in terms of need-fulfilment. Building on an extensive, yet often overlooked literature, need-based measures that fit Rasch Measurement Theory criteria are converted from ordinal scores to interval scores to evaluate response to therapy. These measures meet the requirements of single attribute fundamental measurement which is the standard in the physical sciences. It is proposed that a translation from a Rasch interval scale (defined by logits) can be transformed to a bounded ratio scale. Need based Quality of Life (N-QOL) scales bounded by 0 (where no needs are fulfilled) to 1 (where all needs are fulfilled) form such scales. The N-QOL supports the full range of arithmetic operations. Multiattribute utilities and mathematically invalid QALYs can be put to one side as unfortunate historical curiosities in favor of a disease or target population specific N-QOL scale. Such a scale has the required properties to evaluate disease specific response to therapy This can also support N-QOL adjusted life years with a need- fulfillment life year (NALY) metric with ratio properties.

Keywords: Need-fulfilment quality of life, N-QOL, need adjusted life years, NALY

INTRODUCTION

The quality adjusted life year (QALY) should no longer be considered the basis for health technology assessment claims, even when used to produce approximate information^{1 2}. The typical multiattribute utility instruments (e.g. EQ-5D-5L, HUI Mk3, SF-6D) fail to meet required standards of fundamental measurement³. In their development, no thought was given to their measurement properties, with the result that they all yield ordinal scores. As they fail to provide ratio or even interval measurement standards, they cannot be used to assess response to therapy. Consequently, they cannot support the creation of QALYs.

Apart from their ordinal properties, generic multiattribute instruments fail to provide a conceptual model for their content and so cannot have construct validity. Rather, they combine an assortment of health attributes, typically proposed by clinicians. The attributes cannot be added together to yield a meaningful total score. Further, there have been ongoing concerns about their failure to capture the patients' voice and their lack of sensitivity due to their generic nature. This often

affects disadvantaged and disability groups⁴. Other concerns relate to utility states worse than death, where negative scores are possible and the interpretation to be placed on the notion of perfect health.

There is a dichotomy between those who advocate health-related quality of life (HRQoL) instruments and those who propose patient-centric or need-based value measurement. The distinction between the terms 'score', 'scale' and 'measure' is important. HRQoL instruments are not measures but produce counts or scores. Scales have to be measures that fit Rasch Measurement Theory (RMT). A measure must have quantitative interval or ratio properties and must be consistent with the axioms of fundamental measurement. Inevitably, a move away from HRQoL instruments, with their symptom lists, to interval measurement is required scientifically and is perfectly possible. However, it would need a paradigm shift in health technology assessment which current practitioners would find difficult to accept. It requires a framework that converts counts of responses to true measurement. Concerns about clinically driven notions of quality of life (QoL) and the lack of fundamental measurement can be put to one side if the focus is on the needs of patients defined by a coherent latent trait or non-physical attribute. If need-based value is driven by whether the needs of patients and caregivers are fulfilled, specific measures can be developed that capture those needs within disease or target patient groups^{5 6}. This question has been addressed through the development, using RMT, of many need-based disease-specific measures over the past 25 years.

Corresponding author: Paul C Langley, PhD
Adjunct Professor, College of Pharmacy
University of Minnesota, Minneapolis, MN
Director, Maimon Research LLC; Tucson, AZ
Email: langley@maimonresearch.com
Website: www.maimonresearch.net

These measures are unidimensional and provide interval level measurement.

This commentary will make the case that it is possible to create ratio measures from ordinal raw scores via transformations that include Rasch analysis. The required analytical framework for a need fulfilment QoL (N-QOL) measure and the application of this to create time spent in needs adjusted health, the Need Fulfilment Life Year (NALY) measure will be introduced.

THE NEED-BASED MODEL OF QUALITY OF LIFE

Despite its crucial importance to patient-reported outcome (PRO) instrument development, RMT is rarely used by health outcome instrument developers. An exception is the body of measures developed to assess the extent to which individuals can fulfil their needs. The conceptual model underlying this approach is that patient value or QoL is the extent to which an individual's human needs are fulfilled. Rather than measuring a limited range of symptoms and functional limitations (as HRQoL scores), need-based QoL measures determine the impact of disease and its treatment on the overall lives of patients or caregivers. Need fulfilment is also affected by other variables such as formal and informal care services, environmental influences and financial resources⁷. This construct theory grew from qualitative work conducted in the development of the Quality of Life in Depression Scale (QLDS) a value measure specific to people with depression⁸. Analyses of patient interviews indicated that it was not the symptoms or inability to function that adversely affected the lives of depressed people but the impact of these symptoms and limitations on the ability of the interviewees to meet their needs.

In addition to having a coherent construct theory, all need-based value measures developed since the year 2000 have been developed and tested using RMT^{9 10 11 12 13 14}. The need-based model represents a genuine attempt to move beyond data-based instrument calibration, by ensuring that data collected with the scales fit the RMT framework. This is unlike HRQoL instruments (including multiattribute utility instruments), that seek a model to fit their data. The use of RMT represents a genuine attempt to match the quality of measurement in the physical sciences. Fit to RMT ensures that the measure is unidimensional and produces interval level data. Unidimensionality is a basic requirement of all measures but instruments that produce ordinal data do not meet this requirement.

THE REQUIREMENT FOR FUNDAMENTAL MEASUREMENT

In retrospect it was a mistake to create multiattribute utility instruments. The mistake lay in creating a utility score by matching the scoring algorithm to items from patient surveys, where the data have primacy, and the results are descriptive rather than forming measures.¹⁵ Statistical considerations and goodness of fit were given preference to the development of a coherent measure. Unfortunately, this approach is common in the social sciences with Classical Test Theory (CTT) as the

measurement model. This approach overlooks the question of intent. Is the objective to create a measure of response that meets the standards for fundamental measurement? Is the intent to create a measure with ratio properties, to support the creation of QALYs? If so, these intentions could never have been achieved using CTT. This methodology ensures that the question of fundamental measurement is ignored. It is inevitable that the outcome would be ordinal instruments, not measures capable of supporting the analysis of response to therapy. Meeting the standards for fundamental measurement must precede statistical analysis.

To achieve fundamental measurement, data must fit RMT. The Rasch model is confirmatory and predictive: *... a confirmatory model requires the data to fit that model ...by focusing on the size and structure of residuals where the principles of conjoint simultaneous measurement have been sufficiently realized in practice to justify the claim that the results can be used as a measurement scale with invariant, interval measurement properties*¹⁵. RMT corresponds to the requirements of measurement theory whereas CTT and Item Response Theory (IRT) are simply descriptive counts or ordinal scores¹⁶.

The Rasch model, developed in the 1960s, provides a framework to evaluate latent or non-physical attributes. It transforms ordinal level observations to interval-level measures. Responses to items captured in an RMT constructed questionnaire are interpreted in probabilistic terms, where the likelihood of a positive response reflects only item difficulty and the ability of the respondent. In need assessment measures, item difficulty captures the inherent difficulty of meeting a particular need while ability captures how likely it is that a patient will be able to meet that need. Therefore, a therapy intervention is judged in terms of its likelihood of meeting increasingly more difficult needs set against the likelihood that the respondent will be able to take advantage of that opportunity. The probability of a successful response increases with the ability of the respondent and decreases with the inherent difficulty of the need. Items are selected to meet RMT requirements with each item occupying a unique place in a 'hierarchy' of items. Where a pair of items occupy the same value (place on the measurement scale), one can be discarded as it does not provide additional information. The intent must be to create an instrument that yields interval properties; it is not something that can be assumed or contrived *ex post facto*. RMT takes as its starting point the long-recognized need to progress from counting observations to measurement; a point made by Thorndike over a century ago¹⁷.

Consider the widely used multiattribute utility instrument the EQ-5D-3L¹⁸. Its scoring algorithm was generated from the views of a community sample of respondents who were asked to score simple profiles of health states defined in terms of five symptoms or attributes (mobility, self-care, usual activity, pain/discomfort, anxiety/depression). Each attribute has three response levels (no problems, some problems, major

problems), yielding 243 possible health states. The algorithm takes as its base a score of unity which is labelled, 'perfect health' (defined by a 'no problem' response to the five symptoms). Ordinal utility scores are then created by decrements from the base depending on the response levels for the five items plus other constant values where rules govern their inclusion as part of the 'forced' fit to the data. The result is a range for utility scores from 1 to -0.59. Scores below zero (which is not a true zero) are labelled as states worse than death and are typically ignored to avoid the creation of negative QALYs. There is no concept of 'item difficulty' nor of the ability of a patient. Non-clinical factors that may impact the patient's QoL are considered irrelevant and are omitted.

When an interval measure of need-fulfilment is required, this must be built into the development of the measure. This is achieved by, as a first step, creating an ordinal score based on item response in the items in the measure. The steps in instrument development are well established, as illustrated in the development of the Alzheimer's Patient Partners Life Impact Questionnaire (APPLIQUE) ¹⁹:

- Define the target patient (or caregiver) population.
- Establish inclusion/exclusion criteria for a respondent sample.
- Undertake extensive unstructured qualitative interviews to explore key elements in need fulfilment for the target population (in this case caregivers).
- Generate a pool of potential questionnaire items (statements that can elicit a binary response, such as 'true' or 'not true').
- Construct a draft questionnaire for participant completion and review.
- Distribution of initial draft survey (30 items) to caregiver sample, followed by a second postal administration.
- Analyze survey data following RMT fit analysis to identify the final 25 item list for the unidimensional questionnaire ²⁰

In the case of the APPLIQUE, the only major concern was the distribution of responses between spousal and non-spousal caregivers. It became clear that there were differences in need fulfilment between the two groups. RMT assessment was undertaken for the spousal caregivers only, to confirm the relevance of the item hierarchy for the final questionnaire. Items that were only relevant to non-spousal caregivers were removed from the draft scale.

Representative items from the APPLIQUE interval instrument include:

- Organizing shopping is very difficult.
- There is no conversation between us.
- I have little freedom to do what I want to do.
- I feel I am losing my independence.
- It's like being with a stranger.

The possible score range for the APPLIQUE is 0 to 25. Overall, the psychometric properties were found to be excellent. Given a single attribute ordinal score based on the proportion of items that are responded to positively, the next steps would be to translate these to an interval score and then, for the N-QOL bounded ratio scale defined by a transformation algorithm.

INTERPRETATION OF THE N-QOL SCALE

The credibility of a ratio measure rests on its ability to defend or justify a true zero for the scale. In the physical sciences there is usually a strict criterion for defining a zero point; a true zero that does not allow negative values or observations below that point. For example, with temperature, Centigrade and Fahrenheit measures both have arbitrary 'zero' points meaning that they measure at the interval level. This contrasts with the Kelvin measure that has a technically justified measure of absolute zero (quantum quibbles aside). In the social sciences, the availability of 'natural' zeros for non-physical attributes is limited. Consequently, instruments must be developed that have a defensible zero point. The true zero equivalence in the social sciences rests on the application of fundamental measurement properties to non-physical attributes. RMT makes this possible by converting ordinal data to interval scales.

All PRO instruments produce ordinal data. This limits their usefulness, as it is not valid to add item scores together to produce total scores. As utility scales are ordinal, the distances between scores are unknown. This is a major problem for calculating QALYs. However, instruments that fit the Rasch model can be adapted to produce more useful scales by conforming to the axioms of fundamental measurement.

Consider, as an example, a need fulfilment questionnaire with a binary response option for each item. A score of 0 indicates that none of the needs of the respondent are met. Such an outcome can be contrasted with the 0 = death point on a utility scale but where the data fitting algorithm can create negative utilities. This 'states worse than death' situation is impossible with RMT developed instruments. The scale is also capped by the number of items in the instrument. A positive response to all items indicates all needs covered by the questionnaire are met. These scores or item responses are ordinal.

Rasch transformed scores produce values that are hierarchical. These values are called logits or natural log-odds units that produce equal interval, linear measures. Further transformations allow these logits to be expressed on a 0 to 1 interval scale. With the need-based measures, a score of one means that all needs are met. Defining the upper limit as all needs fulfilled avoids the notion of perfect health adopted by the multiattribute instruments. Reporting no problems on a five-item measure cannot be considered to indicate that the respondent has perfect health. The upper limit on a utility scale

is also an artifact of the fitting algorithm where utilities are defined as rule-based decrements from a fixed ceiling of unity.

The next issue is whether it is possible to convert an interval measure of a latent construct into a ratio scale with a meaningful 'zero' and a cap at unity. This is possible if there is a well-defined starting point that can be described as 'none' or 'zero'. While this is, of necessity, defined by the instrument, there may be frames of reference for which it is convenient to define such a position. If a Rasch based interval scale for a latent construct such as need fulfilment is considered, there is the option of translating the interval to a ratio scale. This point was made some 30 years ago by Wright and Linacre: *A ratio scale does have a clear origin. But that origin is usually of more theoretical interest than practical utility*²¹. It is a simple arithmetic operation to convert measures from interval scales to ratio scales and vice versa. This position was re-emphasized 25 years ago by Koch et al who pointed out that: *In general any interval scale can be transformed to a Stevens ratio scale by choosing an origin with some meaning. Further, any interval scale can be converted to a Rasch scale of ratios ('logarithmic scale') by choosing any origin and exponentiating. Consequently, the distinction between these scale types has no mathematical importance*²².

Logits transformed to a 0 to 1 scale can be further transformed to create the disease or target patient group specific N-QOL measure, a bounded ratio scale with well-defined endpoints. The N-QOL is in an important (yet trivial) sense, the analog of multiattribute utility ordinal scores. The difference is that multiattribute utilities are scores not potential measures and cannot be transformed to interval scales. While as a first step the N-QOL creates an ordinal scale based on the binary responses, the typical multiattribute utility scale is of lesser validity. It lacks unidimensionality and construct validity and it produces negative values. Also, as the EQ-5D-3L and similar scales are generic, they lack sensitivity to changes in health status in any specific disease population. In any event, it cannot be interpreted as a measure of therapy response, unless defined in non-parametric statistical terms.

Given modern methods of instrument development and validation it is reasonable to ask why multiattribute utilities and the mathematically invalid QALY have continued to be used? The answer is straightforward: there has not been an alternative to the QALY. A position that is bolstered by the lack of awareness of the axioms of fundamental measurement. This has the potential to change with the N-QOL, so that multiattribute utilities and the QALY can be retired at long last. If clinical attributes, including latent constructs, are important in assessing response to therapy each of these should be considered separately with their own scoring or measurement metric. They should not be bundled as an ersatz HRQoL instrument.

The N-QOL is the ideal patient-centric measure, not only for evaluating response to therapy but, if required, the basis for a 'true' time spent in a disease state equivalent measure. As the N-QOL is a ratio scale, multiplying it by time (also a ratio scale) yields a ratio disease-specific measure, the Need Fulfilment Life Year (NFLY). While the NFLY is similar to the QALY, the latter fails to meet the axioms of fundamental measurement.

APPLICATION OF THE N-QOL

The principal barrier to generating the N-QOL measure is the need to construct QoL instruments specific to different target populations. This is a time-consuming enterprise; but the combination of experience and RMT are available to undertake this activity successfully. Once the instrument has been developed and validated, it is then a question of translating the ordinal raw scores into a unidimensional bounded interval scale and then a bounded ratio scale that fits RMT. It can then be included as a primary or secondary endpoint in a pivotal clinical trial or phase 4 trial, or as a claim for need-fulfilment to be evaluated as part of an evidence base to support value contracting²³. What is certainly not recommended, given the importance that is attached to meeting the measurement standards of normal science, is to include these measures as part of modeled pseudoscientific simulation claims. The important point is that the N-QOL is a measure that is valid, credible, empirically evaluable and replicable. After some 30 years there is now a patient-centric PRO measure framework that meets the standards of normal science.

From a population health perspective there is now a measure that is fully consistent with fundamental measurement and that can provide a robust estimate of the extent to which needs are met in target populations. The N-QOL meets a long-standing concern with the required characteristics of a measure that can report on both patient and caregiver needs in patient groups. Manufacturers should be concerned with how their products help meet patients' unmet needs. The N-QOL provides a firm basis for such an assessment of therapy interventions.

It is now finally possible to reject QALYs and incremental cost-per-QALY claims. In contrast, patients and caregivers can be placed front and center in evaluating competing pharmaceutical product claims. There are over 30 disease-specific need-based measures that meet RMT standards for measurement that can be transformed into ratio scales. These include the Rheumatoid Arthritis Quality of Life (RAQoL) questionnaire²⁴, the Psoriasis Index Quality of Life (PSORIQoL)¹⁰; the Asthma Life Impact Scale (ALIS)²⁵; and the Crohn's Life Impact Questionnaire (CLIQ)¹³. Not only are these instruments easy to respond to and score, but the time required to complete them is below five minutes. This stands in marked contrast to some multiattribute utility instruments and the standard gamble, and time trade-off methodologies.

CONCLUSIONS

If health technology assessment is to retain credibility it must reject the creation of approximate information through the construction of imaginary lifetime simulations that defy the standards of normal science. Rather, it must shift focus to promote claims that can support hypothesis testing. The N-QOL can play an important part. Accepting the N-QOL means the acceptance of the axioms of fundamental measurement. Without measurement, the only option is to use models fitted to data to create ordinal scores. The requirements for developing the N-QOL are:

- The development of a QoL questionnaire that adopts a coherent measurement model of QoL, rather than a list of symptoms or impaired functioning.
- Demonstrating fit of the questionnaire to RMT.
- Converting logit values to create a bounded ratio scale

The proposed measures recognize that the likelihood that an individual can meet a need is dependent on only two factors; the difficulty of meeting the need and the ability of the respondent to meet that need. Any target population can be described by the distribution of needs fulfilled and the distribution of the ability of patients to meet that need. This is the meaning of patient-centric therapy response; not clinical responses that may have little relation to need fulfilment. After some 30 years it is time to reject instruments that support scores that fail to meet the axioms of fundamental

measurement. This includes not only multiattribute utility instruments but also most patient-reported outcome instruments. The benefits of adopting fundamental measurement to evaluate response to therapy have been well documented and are summarized here:

- Emulating the measurement standards of the physical sciences in single attribute claims.
- Accepting the application of RMT in measuring non-physical or latent attributes.
- Rejecting multiattribute clinical scores (not measures) as the basis for QoL claims.
- Putting patient and caregiver needs central to assessing their QoL and therapy benefits.

The N-QOL clearly meets these requirements for a measure of therapy response that is consistent both with fundamental measurement and the standards of instrument development that are recognized in the physical sciences. This is a major step forward. We can finally retire utility scores and the QALY. Whether we choose to do so is another question.

CONFLICTS OF INTEREST

PCL is an Advisory Board member and consultant to the Patient Access and Affordability Project, a program of Patients Rising. SPM develops QoL measures for use in clinical trials and routine clinical practice.

REFERENCES

- ¹ Neumann PJ, Willke R, Garrison LP. A Health Economics Approach to US Value Assessment Frameworks – Introduction: An ISPOR Special Task Force Report. *Value Health*. 2018;21:119-123
- ² Langley P. The Great I-QALY Disaster. *Inov Pharm*. 2020;11(3): No 7 <https://pubs.lib.umn.edu/index.php/innovations/article/view/3359/2517>
- ³ Langley PC and McKenna SP. Measurement, modeling and QALYs. *F1000Research*. 2020; 9: 1048 <https://doi.org/10.12688/f1000research.25039.1>
- ⁴ McKenna S, Wilburn J. Patient value: its nature, measurement, and role in real world evidence studies and outcomes-based reimbursement. *J Med Econ*. 2018;21(5): 474-80
- ⁵ McKenna S, Heaney A, Wilburn J et al. Measurement of patient-reported outcomes. 1: The search for the holy grail. *J Med Econ*. 2019;22(6): 516-22
- ⁶ McKenna SP, Heaney A, Wilburn J. Measurement of patient reported outcomes. 2: Are current measures failing us? *J Med Econ*. 2019;22(6):523-30
- ⁷ Doward LC, McKenna SP. Defining patient-reported outcomes. *Value Health*. 2004;7(Suppl 1):S4–S8
- ⁸ Hunt SM, McKenna SP. The QLDS: a scale for the measurement of quality of life in depression. *Health Policy*. 1992;22:307–19
- ⁹ McKenna SP, Doughty N, Meads DM, et al. The Cambridge Pulmonary Hypertension Outcome Review (CAMPHOR): a measure of health-related quality of life and quality of life for patients with pulmonary hypertension. *Qual Life Res*. 2006;15:103–15

- ¹⁰ McKenna SP, Cook SA, Whalley D, et al. Development of the PSORIQoL, a psoriasis-specific measure of quality of life designed for use in clinical practice and trials. *Br J Dermatol*. 2003;149: 323–31
- ¹¹ Doward LC, McKenna SP, Meads DM, et al. Translation and validation of non-English versions of the Ankylosing Spondylitis, Quality of Life (ASQOL) questionnaire. *Health Qual Life Outcomes*. 2007;5:7
- ¹² McKenna SP, Meads DM, Doward LC, et al. Development and validation of the living with chronic obstructive pulmonary disease questionnaire. *Qual Life Res*. 2011;20:1043–52
- ¹³ Wilburn J, McKenna SP, Twiss J, et al. Assessing quality of life in Crohn's disease: development and validation of the Crohn's Life Impact Questionnaire (CLIQ). *Qual Life Res*. 2015;24:2279–88
- ¹⁴ Wilburn J, McKenna SP, Heaney A, et al. Development and validation of the Parenteral Nutrition Impact Questionnaire (PNIQ), a patient-centric outcome measure for home parenteral nutrition. *Clin Nutr*. 2018;37:978–83.
- ¹⁵ Bond T, Fox C. Applying the Rasch Model; Fundamental Measurement in the Human Sciences 3rd ED. New York Routledge, 2015.
- ¹⁶ Panayides P, Robinson C, Tymms P. The assessment revolution that has passed England by: Rasch measurement. *Br Educational Res J*. 2010;36(4):611-26
- ¹⁷ Thorndike EL. An introduction to the theory of mental and social measurement. New York Teachers College. Columbia University, 1904
- ¹⁸ Drummond M, Sculpher M, Caxton K et al. Methods for the Economic Evaluation of Health Care Programmes. 4th Ed. New York: Oxford University Press, 2015
- ¹⁹ McKenna SP, Rouse M, Heaney A et al. International Development of the Alzheimer's Patient Partners Life Impact Questionnaire (APPLIQUE). *Am J Alzheimer's Disease Other Dementias*. 2020;35:1-11
- ²⁰ Hagell P, Rouse M, McKenna SP. Measuring the impact of caring for a spouse with Alzheimer's disease: Validation of the Alzheimer's Patient Partners Life Impact Questionnaire (APPLIQUE). *J Applied Measurement* 2018; 19(3): 271-282
- ²¹ Wright BD, Linacre JM. Observations are always ordinal: Measurements however must be interval, MESA Research Memorandum, No. 44. *MESA Psychometric Laboratory*. December 1989
- ²² Koch W, Schulz EM, Wright R et al. What is a ratio scale? *Rasch Measurement Transactions*. 1996; 9(4):457
- ²³ Langley P. Value Assessment, Real World Evidence and Fundamental Measurement: Version 3.0 of the Minnesota Formulary Submission Guidelines. *InovPharm*. 2020;11(4): No 12 <https://pubs.lib.umn.edu/index.php/innovations/article/view/3542/2613>
- ²⁴ Heaney A, Stepanous J, Rouse M et al. A review of the psychometric properties and use of the Rheumatoid Arthritis Quality of Life Questionnaire (RAQoL) in clinical research. *Current Rheumatology Rev*. 2017;13(3):197-205
- ²⁵ Meads DM, McKenna SP, Doward LC et al. Development and validation of the Asthma Life Impact Scale (ALIS). *Respir Med*. 2010;104(5):633-43