Towards patient-centred cognition metrics

J Melin¹, LR Pendrill¹, S J Cano², and the EMPIR NeuroMET 15HLT04 consortium³

¹ RI.SE, Eklundagatan 86, S-412 61, Göteborg, Sweden, jeanette.melin@ri.se, leslie.pendrill@ri.se
² Modus Outcomes, Spirella Building, Letchworth Garden City, SG6 4ET, UK, stefan.cano@modusoutcomes.com
³ NeuroMET coordinator Milena Quaglia, LGC, Milena.Quaglia@lgcgroup.com

E-mail: jeanette.melin@ri.se

Abstract. Metrological quality assurance is essential if reliable decisions about diagnosis, treatment and rehabilitation are to be made consistently throughout the healthcare system. However, many observations in healthcare are considered to lie ‘off the scale’ of quantitative measurement. We have previously argued that a fundamental reappraisal of metrology is needed if such ordinal or nominal properties are to be included in an extended quantity calculus on which the SI could be based. As just one example, metrological quality assurance in person-centred care now seems possible in a novel way for partners in care; physician ability (e.g., to diagnose cognitive impairment correctly) and patient ability (e.g., cognitive performance of patients with Alzheimer’s disease) using novel formulations of causal Rasch models articulated through construct specification equations (CSEs). Here, as part of our research in the EMPIR NeuroMET 15HLT04 project, we illustrate how construct specification equations and new metrological references for task difficulty for block tests, such as the Knox Cube Test, can be established in terms of entropy, reversals and average distance.

1. Introduction

Comparability through SI (System of International Units) traceability and uncertainty analysis is an, as yet, unmet requirement for regulatory approval of biomarkers, patient centred outcome measures (PCOMs), clinical thresholds and new therapeutic drugs. However, we have previously argued that: a) measurement, whether physical or social, should have the same definition (i.e., the ratio of two magnitudes of the same attribute, where the denominator is the unit, providing for invariant comparisons) and the same broad goal (i.e., quantification of meaningful variables) [1]; b) to ensure decision makers within health systems have access to the necessary objective evidence, we require the same kind of quality-assurance of PCOMs as is established in physics and engineering [2]; and c) Rasch Measurement Theory [3], and especially causal Rasch models [4], have a central role to enable references for traceability and means of evaluating measurement uncertainty to be established [5].

2. Patient centred cognition measurement

The European EMPIR NeuroMET 15HLT04 project [6] combines expertise from clinical, academic, and National Metrology Institutes to provide a better understanding of how to improve, combine and analyse measurements in Alzheimer’s disease (AD) diagnosis and treatment. As part of the project,
PCOMs of cognitive performance (e.g., recall, language and praxis) are examined. The ultimate goal is to improve the analysis of correlations between cognitive function and volumes of AD-related brain structures and neurometabolite concentrations. Rasch Measurement Theory is being used to develop a ‘NeuroMET Memory Metric’ based on results from administering an extensive battery of legacy cognitive performance PCOMs (e.g., Mini-Mental State Examination, Corsi Block Test, Digital Span Test). This is being achieved through the formulation of novel construct specification equations (CSEs) using multivariate principal component regression [7].

CSEs are defined as the highest level of construct theory in social measurement [8, 9]. At the lowest level 1, social measurement systems have no explicit construct theory and items are simply calibrated empirically from data analysis. At the highest level 5, measurement systems have a construct theory and CSE; a conjoint model that measures task difficulty and patient ability in a common unit. This is known as theory-referenced measurement. It affords three central benefits [7]: a) an observational outcome (i.e., a response); b) a causal mechanism that transmits variation in the intended attribute (e.g., cognitive test item difficulty), via an instrument to the observed response (e.g., the test person); and c) attribute measures (i.e., the measurand) expressed in some unit. Consequently, CSEs resemble ‘recipes for certified reference materials’ for traceability in chemistry [7]. They provide a comprehensive empirical understanding of: how a collection of items works together; what is being measured; and how validity is ensured.

3. A CSE for the Knox Cube Test – a worked example

A CSE breaks the relationship between intention (e.g., task difficulty) and attainment (e.g., person ability) by independently formalizing intent as an equation that can produce item calibrations (or ensemble means) that are in close correspondence with empirical estimates [4]. A construct-specification equation for task difficulty, \( \delta = f(x_1, \ldots, x_n) \), is developed by regressing object construct values \( \delta \) on selected characteristics, \( X_i \), (‘explanatory’ variables). The results from this multiple regression analysis yield two important sources of information: a) the \( R^2 \) index indicates the amount of variance in construct values \( \delta \) accounted for by the model; b) the analysis provides a regression equation that indicates which characteristics of the items are important in predicting construct values \( \delta \) [8].

The CSE approach provides a predictive tool for the design of, for instance, new cognitive tasks complementing existing tests and scales and demonstrating item equivalence. These sorts of cognitive performance PCOMs can often rely on the concept of entropy. Entropy can be described as a measure of order. The higher the order, the lower the entropy, and the other way around. Task difficulty should be expected to decrease with entropy, since more ordered tasks are more ‘informative’ and should be easier to perform. Information theory draws analogies between thermodynamic entropy in physics and the measures of information content in communication systems [10]. Information content can range from basic examples (e.g., the number of elementary symbols) to increasingly sophisticated information, through syntax, semantic and pragmatic aspects of meaningful information in many contexts.

Development of cognitive performance PCOMs includes considerations of the item factors that can explain the level of difficulty of remembering, say, a particular sequence (e.g., tapped blocks or a series of numbers or words). A basic idea is that task difficulty is proportional to the entropy, as follows. The Shannon entropy is proportional to \( \ln(P) \), where \( P \) is the probability of a message. The less expected a message is (i.e., smaller \( P \)), the greater the amount of information conveyed (‘surprisal’). Taking the logarithm facilitates addition and subtraction of different amounts of information. Consider in general a message in which a number, \( N_j \) (\( j = 1, \ldots, M \)) of symbols of \( M \) different types can be distributed in a number, \( G \), of categories (or cells) \( G = \sum_{j=1}^{M} N_j \). The probability of encountering the \( j^{th} \) symbol is \( p_j = \frac{N_j}{G} \), which can be summed to unity. The total number, \( P \), of messages that can be obtained by distributing the symbols at random over the \( G \) cells (with never more than one symbol per cell) is \( P = \frac{G!}{\prod_{j=1}^{M} N_j!} \) [11]. The information theoretical entropy, which is a measure of the amount of information in these messages, is then:

\[
\sum \text{e.g., the test person); and c) attribute expected a ry Metric’ based on results from administering an extensive battery of legacy cognitive performance PCOMs (e.g., Mini-Mental State Examination, Corsi Block Test, Digital Span Test). This is being achieved through the formulation of novel construct specification equations (CSEs) using multivariate principal component regression [7].

CSEs are defined as the highest level of construct theory in social measurement [8, 9]. At the lowest level 1, social measurement systems have no explicit construct theory and items are simply calibrated empirically from data analysis. At the highest level 5, measurement systems have a construct theory and CSE; a conjoint model that measures task difficulty and patient ability in a common unit. This is known as theory-referenced measurement. It affords three central benefits [7]: a) an observational outcome (i.e., a response); b) a causal mechanism that transmits variation in the intended attribute (e.g., cognitive test item difficulty), via an instrument to the observed response (e.g., the test person); and c) attribute measures (i.e., the measurand) expressed in some unit. Consequently, CSEs resemble ‘recipes for certified reference materials’ for traceability in chemistry [7]. They provide a comprehensive empirical understanding of: how a collection of items works together; what is being measured; and how validity is ensured.

3. A CSE for the Knox Cube Test – a worked example

A CSE breaks the relationship between intention (e.g., task difficulty) and attainment (e.g., person ability) by independently formalizing intent as an equation that can produce item calibrations (or ensemble means) that are in close correspondence with empirical estimates [4]. A construct-specification equation for task difficulty, \( \delta = f(x_1, \ldots, x_n) \), is developed by regressing object construct values \( \delta \) on selected characteristics, \( X_i \), (‘explanatory’ variables). The results from this multiple regression analysis yield two important sources of information: a) the \( R^2 \) index indicates the amount of variance in construct values \( \delta \) accounted for by the model; b) the analysis provides a regression equation that indicates which characteristics of the items are important in predicting construct values \( \delta \) [8].

The CSE approach provides a predictive tool for the design of, for instance, new cognitive tasks complementing existing tests and scales and demonstrating item equivalence. These sorts of cognitive performance PCOMs can often rely on the concept of entropy. Entropy can be described as a measure of order. The higher the order, the lower the entropy, and the other way around. Task difficulty should be expected to decrease with entropy, since more ordered tasks are more ‘informative’ and should be easier to perform. Information theory draws analogies between thermodynamic entropy in physics and the measures of information content in communication systems [10]. Information content can range from basic examples (e.g., the number of elementary symbols) to increasingly sophisticated information, through syntax, semantic and pragmatic aspects of meaningful information in many contexts.

Development of cognitive performance PCOMs includes considerations of the item factors that can explain the level of difficulty of remembering, say, a particular sequence (e.g., tapped blocks or a series of numbers or words). A basic idea is that task difficulty is proportional to the entropy, as follows. The Shannon entropy is proportional to \( \ln(P) \), where \( P \) is the probability of a message. The less expected a message is (i.e., smaller \( P \)), the greater the amount of information conveyed (‘surprisal’). Taking the logarithm facilitates addition and subtraction of different amounts of information. Consider in general a message in which a number, \( N_j \) (\( j = 1, \ldots, M \)) of symbols of \( M \) different types can be distributed in a number, \( G \), of categories (or cells) \( G = \sum_{j=1}^{M} N_j \). The probability of encountering the \( j^{th} \) symbol is \( p_j = \frac{N_j}{G} \), which can be summed to unity. The total number, \( P \), of messages that can be obtained by distributing the symbols at random over the \( G \) cells (with never more than one symbol per cell) is \( P = \frac{G!}{\prod_{j=1}^{M} N_j!} \) [11]. The information theoretical entropy, which is a measure of the amount of information in these messages, is then:
where $K$ is an arbitrary constant.

A CSE for task difficulty as a function of entropy can be developed for sequence memory tests, such as the Knox Cube (KCT), which involves patients replicating different tapping series on four black one-inch cubes placed in a row [12]. For this illustration, we used the WINSTEPS® Exam1.txt file, included 35 person ratings on 18 items. The set of items comprises different lengths of sequence, from 2 taps to 7 taps, and sequences with different orders. Thus, the KCT offers a suitable test in which expression (1) for entropy can be evaluated for a number of sequences, which not only have increasing numbers of taps but also some repeated taps of the same block (e.g., tapping sequence 1-4-2-3-4-1). The sequences with repeats can be expected to be easier to recall according to Brillouin’s expression than a sequence of the same length but without repeats [11]. Tentatively, as no suchlike tapping sequence is available in KCT, a tapping sequence of 1-4-1 would be expected to be easier than 2-4-1.

We found a high correlation (Pearson coefficient $r = 0.96$) in a linear regression of the measured task difficulty, $\delta$, against the CSE estimates $zR$ based on the three explanatory variables $X = \{\text{Entropy, Reversals, Average Distance}\}$ (Figure 1). Likewise, the measured task difficulty, $\delta$, is within the corridor of model uncertainties, i.e. $zR + UzR$ (Figure 2). The effects of dis-attenuation [13, 14] on the regression degree of correlation are quite small in the present case, reflecting that measurement uncertainty is small compared with other sources of scatter.

**Figure 1.** Linear regression of the measured task difficulty, $\delta$, against the CSE estimates $zR$ based on the three explanatory variables $X = \{\text{Entropy, Reversals, Average Distance}\}$ for the series of KCT sequences.
Predicted values of task difficulty, $\delta$, for series of KCT sequences of increasing difficulty from a CSE, $zR$, based predominantly on entropy [eq. 1] (with minor additional contributions from the number of reversals and average distance of each sequence) compared with corresponding measurement values (dots with uncertainty intervals). The corridor of model uncertainties is shown as $zR \pm UzR$, uncertainty coverage factor $k = 2$.

4. A CSE for person ability – the aspiration for person centred cognition metrics

As commented by de la Torre and Minchen [15] in the context of educational assessment, students’ scores are typically determined by identifying person locations along a single proficiency continuum using item response theory, do not in themselves naturally provide diagnostic information. To be able to provide diagnostic information to enhance classroom instruction and learning so-called cognitively diagnostic assessments are made where responses are modelled as a function of discrete latent variables.

Similarly, for patients with neurodegenerative diseases in health care, a CSE for person cognitive ability, $\theta = f[z_1, \ldots, z_m]$ can be formulated analogously to the task difficulty CSE. Instead of the explanatory variables, such as entropy, reversals and average distance, for task difficulty of KCT the CSE for person cognitive ability could be a function of diverse set of volumes of AD-related brain structures and neurometabolite concentrations.

A CSE for patient cognitive ability would be able to provide better metrological assurance of biomarker analyses compared to univariate studies. In the upcoming follow-up project, EMPIR NeuroMET2 18HLT09, this will further be examined to develop mathematical reference measurement procedures by combining the novel CSEs for patient ability and task difficulty which link metrological traceability in the cognitive and biomarker assessments thereby leading to improved measurement quality for better NDD early diagnosis and better monitoring of the effects of pharmaceutical intervention and disease progression.

5. Conclusion

The potential of synthesising novel certified reference ‘materials’ for metrological traceability for the separate probe and entity attributes based on ordinal or nominal responses becomes possible in terms of CSEs, i.e., experimental intervention/manipulation of either person attribute (e.g., ability) or entity attribute (e.g., difficulty) or both simultaneously, following logistic (Rasch) regression, and enabling prediction of the observed outcome (count correct). This work illustrates how CSEs and new metrological references for task difficulty for cognitive block tests, such as the Knox Cube Test, can be established in terms of, in particular, informational entropy as a dominant explanatory variable for cognitive task difficulty. Our analyses are in good agreement with earlier work such as the CSE proposed.
by earlier researchers, such as [16], using their LLTM model. We believe that the concept of entropy, as developed here, has extensive applications in a wide variety of contexts in explaining task difficulty and person ability when assuring the quality of categorical data on ordinal and nominal scales.

6. Acknowledgments
This work has been performed as part of the EMPIR 15HLT04 NeuroMET project which belongs to the European Metrology Programme for Innovation & Research (Horizon 2020), jointly funded by the EMRP participating countries within EURAMET (www.euramet.org) and the European Union.

References
[1] Cano S et al. 2016 J. Phys.: Conf. Ser. 772 012025
[2] Cano S et al. 2018 J. Phys.: Conf. Ser. 1044 012057
[3] Andrich D 2018 Advances in social measurement: A Rasch Measurement Theory Perceived Health and Adaptation in Chronic Disease ed F. Guillemin, et al. (Routledge: New York)
[4] Stenner A et al. 2013 Front. Psychol. 4 1-14
[5] Cano S et al. 2018 J. Phys.: Conf. Ser. 1065 072033
[6] EMPIR Project 15HLT04 NeuroMET 2016 Innovative Measurements for Improved Diagnosis and Management Neurodegenerative Diseases 2016–2019 https://www.lgcgroup.com/our-science/national-measurement-laboratory/european-metrology-programme-for-innovation-and-re/neuromet/#.XBPL2lz7SUk
[7] Pendrill L 2018 Meas. Sci Technol. 29 034003
[8] Stenner A and Smith M 1982 Percept. Motor Skill. 55 415-426
[9] Stenner A, Smith M and Burdick D 1983 J. Educ. Meas. 20 305-316
[10] Weaver W and Shannon C 1963 The Mathematical Theory of Communication (Univ. of Illinois Press)
[11] Brillouin L 1962 Physics Today 15
[12] Knox H 1914 JAMA LXII 741-747
[13] Spearman C 1904 Am. J. Psychol. 15 72-101
[14] Spearman C 1910 Br. J. Psychol. 3 271-295
[15] de la Torre J and Minchen N 2014 Psicologia Educativa 20 89-97
[16] Green K and Smith R 1987 J. Educ. Stat. 12 369-381