Measuring counted fractions in healthcare

L R Pendrill
leslie.pendrill@ri.se
RI.SE Metrology, Göteborg, Sweden

Jeanette.melin@ri.se
RI.SE Metrology, Göteborg, Sweden

Abstract:

In establishing metrological quality assurance in healthcare, one major hurdle is the correct treatment of ordinal data typical of questionnaires, performance tests and other categorical data collected widely in care. Despite being known well over a century, there are still many examples of measurements in healthcare – for example, (i) on-line tables of percentage performance indicators (e.g. fraction of patients seeing a doctor within seven days) and (ii) correlation plots for Alzheimer sufferers of cognitive scores against biomarker concentration – where the 'counted fraction' distortion of scales is not compensated for. The Rasch form of generalised linear model not only handles counted fraction ordinality but also enables separation of object and instrument attributes (such as task difficulty and patient ability) essential for metrological restitution in measurement systems in healthcare. A perspective is given of a new kind of certified reference material employing causal Rasch models in terms of construct specification equations for metrological item banking in the social sciences. This is part of the response to a recent call for: 'a new international body to bring together metrology, psychometrics, philosophy, and clinical management to support the global comparability and equivalence of measurement results in patient centred outcome measurement to improve healthcare'.

Keywords: Metrology; Quality assurance; Rasch Measurement Theory; Psychometrics

1. Counted fractions in healthcare

Quality assurance of products and processes of all kinds rely on the availability of quality-assured measurement. To date, many observations in healthcare, as well as in other similar areas such as sustainability, material testing, etc., are considered to lie 'off the scale' of quantitative measurement. Alongside issues concerning the adoption and implementation of regular SI quantities and units in healthcare (Ferreira & Matos, 2015), developing and implementing quality-assured measurement of categorical and performance data therefore present considerable challenges.

'Counts' (non-negative integers), of for example the number of pills, play a key role in medicine – making sure that the patient takes the correct dose, for instance. Whether 'counts' (non-negative integers), of for example the number of pills, can be considered as 'dimensionless' quantities in the international system of measurement units (SI) for the purpose of quality assurance is still an active subject of debate in the international literature (Flater, 2017). Starting from the concept of 'dimension' of a quality, Kogan (Kogan, 2014) argues that "quantity for which all exponents of factors corresponding to base quantities in its quantity dimension are zero", is preferably called a quantity of dimension one rather than dimensionless. The unit of dimension one can for example be the counted entities, such as pills.

Yet more qualitative are 'counted fractions' (bounded by zero and one), but these are common and play essential roles as performance metrics for healthcare services, ability tests, customer satisfaction, and decision risks caused by uncertainty, and are the main subject of this paper.

2. Quality assurance of counted fractions

2.1 Quality-assurance standards

Current examples of areas where metrological requirements in general are set - for traceability and measurement uncertainty - but where metrological references are yet to be established in the healthcare sector include:

- EN 15224:2012 Healthcare services - Quality management systems;
- ISO 13485:2003 Medical devices - Quality management systems -- Requirements for regulatory purposes;
- ISO 112401 Health informatics - Identification of medicinal products — Data elements and structures for unique identification and exchange of units of measurement;
- FDA 2009 Patient-reported outcome measure for verifying new / improved drugs.

Common to most quality-assurance standards following the well-known ISO-9000 series are requirements on quality assured measurement, typically in paragraph 7.6 (Figure 1, quoted here in the version found in the Healthcare service standard EN 15224:2012).

Figure 1. Paragraph 7.6 of the standard EN 15224:2012 Healthcare services – Quality management systems.

Control of monitoring and measuring devices

The organisation shall determine the monitoring and measurement to be undertaken and the monitoring and measurement devices needed to provide evidence of conformity of product (healthcare service or other healthcare product) to determined requirements…. Where necessary to ensure valid results, measurement equipment shall:

a) Be calibrated or verified at specific intervals, or prior to use, against measurement standards traceable to international or national measurement standards…

Metrology – quality-assured measurement – is often defined in terms of two key concepts:

•Traceability, which provides measurement results which are comparable over space and time. This in turn is a pre-requisite for assuring that products and processes have comparable properties, necessary for interoperability and transparency;

•Declared measurement uncertainty, which in turn allows the risks of incorrect decision of conformity of products and processes to be assessed and minimised.

A principal challenge is that there are few, if any, recognized metrological standards to establish traceability when assuring human-based quality, for instance to assure that patients can expect the same level of healthcare wherever it is provided. Healthcare has been described as a "$1 trillion per year industry without a clear measure or definition of its main product" (Heinemann et al., 2006).

2.2 Counted fractions

The challenges of handling counted fractions have been known for many years. A classic quote is: "Beware of attempts to interpret correlations between ratios whose numerators and denominators contain common parts" (Pearson, 1897).

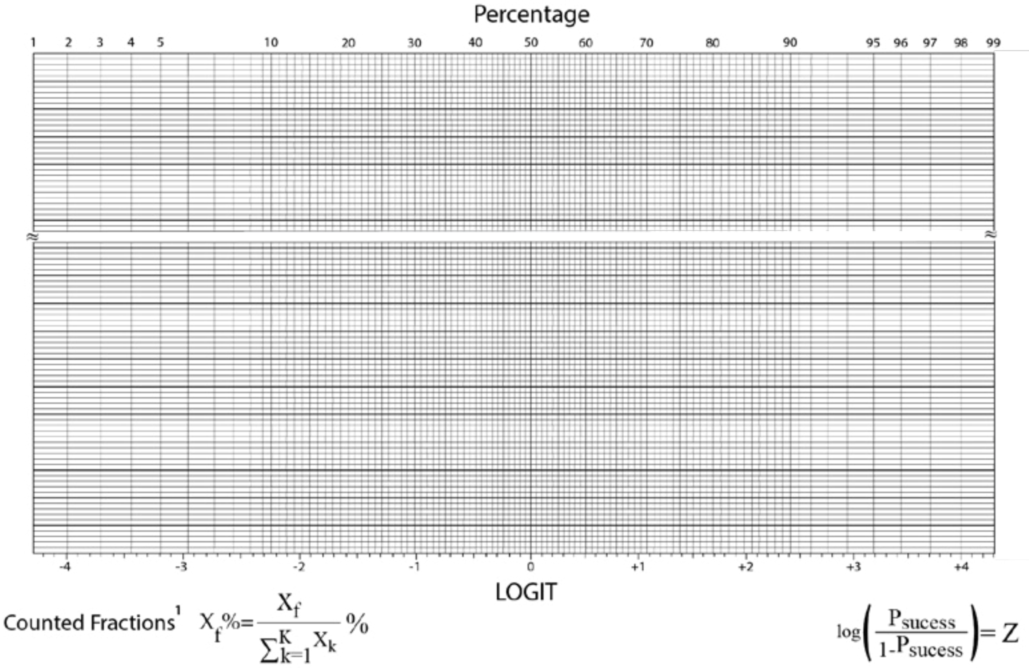
The kind of phenomenon characteristic of counted fractions being referred to has been described by Tukey (Jones, 1984):

> First, our experience-molded intuitions tell us clearly that it is not a mode where equal numerical changes correspond to equally important changes. A change of 5 % is not equally important across the scale. The difference, for almost all purposes except voting (where it is routine to use the qualifying expression "percentage points"), between 1 % and 6 % is very much more important than the difference between 48 % and 53 %. Once we break down our idea that "percentages are the proper mode," we come to feel quite clearly that we need to open out the scale for extreme percentages, as compared with percentages near 50 %.
>
> As a general consequence we should expect that scales which have a finite range are likely to give us trouble, unless our observations tend to be safely away from any ends which are present. Hence the fact that percentages go only from one end (at 0 %) to another (at 100 %) suggest that, whenever even moderately extreme percentages are likely to occur, we are likely to "stretch the tails", while, if really extreme percentages occur, we may have to stretch hard enough so that there are no ends (at any finite values) (Tukey quoted by Jones 1984).

Mathematically, in the counted fraction expression, $X_j\%=X_j/(\sum_i X_i)$ eq. (1), the presence of the amount X of component j appearing in both the numerator and denominator means – and increasingly where is either large or small compared with the other components – that any error in will be correlated with the other components, since of course there is the boundary condition $\sum_j X_j\%=100\%$. It is straightforward (Pendrill, 2018), using the method of Lagrange multipliers with this boundary condition as constraint, to derive the so-called 'link function' between ordinal, counted fraction raw data, such as percentages, p, and a more quantitative, linear scale: $l=\ln(p/(100-p)$.

In the days before computers in the first half of the 20th Century, such non-linearities at the scale extremities – both the high and low ends – although known to be linearisable through so-called logistic ruling, converting percentages p into logits, l, were arduous to calculate, and it was common to use monographs, such as exemplified in Figure 2.

Figure 2-Logistic ruling (reproduced and adapted from Tukey (Jones, 1984).



These days, Generalised Linear Models (GLM) which can handle ordinal scale properties can be readily implemented. Well-known examples are the Rasch psychometric programs which can be used to treat measurement system response, Psuccess, where mathematical distances between categories are not known exactly. Being based on the principle of least action and the Lagrange multipliers method, GLMs are not unique to human-based perception, but indeed apply to a wide range of ordinal and nominal data. Note however that counted fractions lead to an ordinal scale, but not all ordinal scales are related to counted fractions.

As will be seen in the case studies below, examples may be readily found in modern healthcare where data analysis is not done correctly. The consequences of incorrect decisions about care arising from improper treatment of counted fractions have in many cases not been analysed.

3. Restitution of counted fractions and metrological references

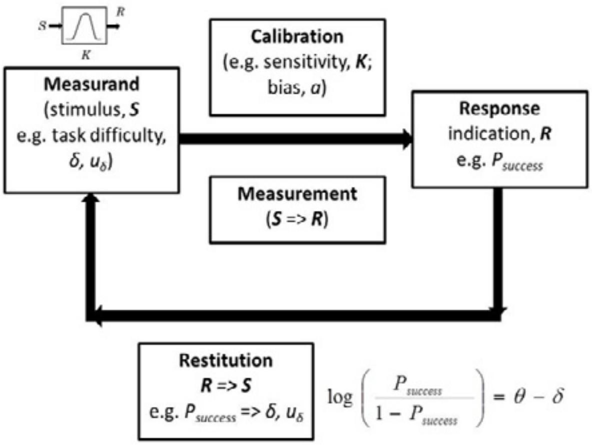
The counted fraction issue is an important part of the on-going discussion about how many concepts of traditional engineering and physical metrology can be extended to apply also to measurement in the social sciences (Cano et al., 2016, 2017).

When attempting a reconciliation between metrology in the physical sciences and measurement practices in the social sciences, a good start is to regard a human being (or other ‘probe’, when perceiving any ‘entity’ in the broadest sense) as a Measurement Instrument, as suggested in the pioneering multidisciplinary European Project ‘Measurement the Impossible Network’ (MINET; Pendrill et al., 2010). ‘Perceptive’ measurement system response is however not a simple instrument indication, but instead extends to the decision risks caused by uncertainty. Final restitution of the measurand (e.g., task difficulty) from this performance metric response in a form suitable for metrological quality assurance requires specific treatment.

Apart from handling counted-fraction ordinality, Rasch measurement theory (i.e., postulating the link function $z = l = \theta - \delta$) is unique amongst GLM in additionally enabling a separation of a ‘probe’ attribute θ and a ‘target’ attribute δ. This attribute separability is essential to underpinning measurement traceability and uncertainty, in the same way as the sensitivity of a weighing scale needs to be calibrated separately from the stimulus value of a mass (Pendrill, 2018).

In a conventional measurement system, where the input signal is measurement information from the measurement object, then the sought-after value of the quality characteristic of the object can be restituted from the observed output. The restitution process consists of inverting the observation equation of the measurement system to estimate the stimulus S in terms of the system output R, and other terms. An assumption is of course that system factors, such as the sensitivity K, remain unchanged from when calibration of the measurement system was performed. A simple example is a measurement system where the instrument sensitivity and there is an offset (bias), a, in the output, R. The formula for restitution of an unknown input, S, that is the stimulus value of the measurement object in this case is: , which might need to be evaluated at individual input levels if either sensitivity and/or bias vary with level.

Figure 3- Observation and restitution for performance metrics.



Source: Pendrill (2018)

The analogous separation of attributes of the measured object (“item”) from those of the instrument (“person”) measuring them, achieved with the Rasch model, brings invariant measurement theory to psychometrics.

A full picture of the measurement process when Man acts as a measurement instrument [Figure III], presents the process, step by step, from the observed indication (a performance metric, e.g. probability of success, Psuccess, of achieving a task, and restitution with Rasch Measurement Theory, through to the measurand (e.g. task difficulty) in a form suitable for metrological quality assurance:

In addition to the regular elements of any measurement system, such as sensing, signal conversion and processing, and data processing (Bentley, 2005), an important fifth category has to be added; namely, a decision-making element which models the performance metrics typical of measurements in the social sciences. Most measurements are not made solely for the sake of measurement, but because decisions are to be made about something – a product, a process, service or phenomenon - based on the measurements: Decision-making: algorithm producing an output on a categorical scale: the result of a decision, such as the binary, dichotomous response if the measured response Tm is above or below a specification limit TSL:

$$R = \begin{matrix} 0 \\ 1 \end{matrix} \; if \; \begin{matrix} T_m \leq T_{SL} \\ T_m > T_{SL} \end{matrix}$$

For such categorical response cases, including the important decision-making response measurement system ‘accuracy’ – as a performance metric - will be identified with decision-making ability: Accuracy (decision-making) = response categorisation – input (true) categorisation, where is a metric of measurement system performance in terms of the probability of making the ‘correct’ decision:

$$z = \theta - \delta = \log\left(\frac{P_{success}}{1-P_{success}}\right)$$

4. Case studies

Tukey (Jones, 1984) gives a number of examples where these effects of counted fractions can occur:

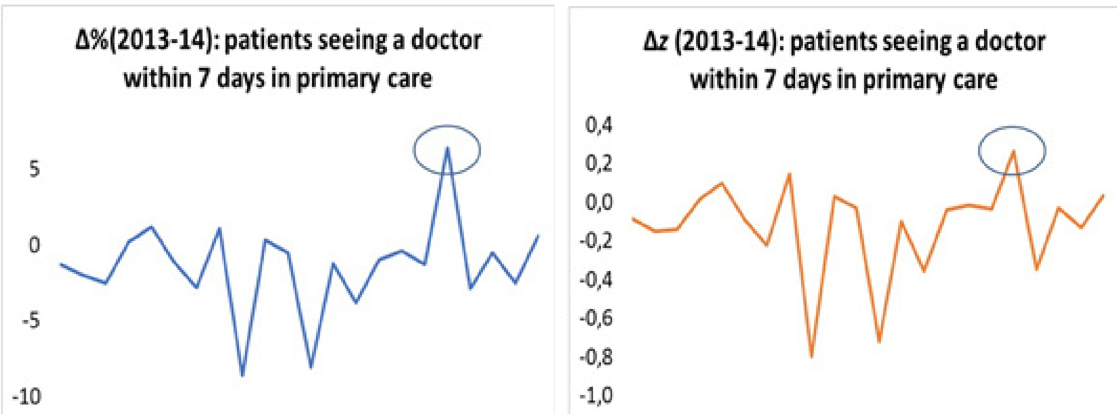
> «Counting sheep and goats, and reporting on the relative number of goats, still typifies much of behavioural science … As a consequence, the behavioural sciences have a very strong continuing interest in modes of expression of counted fractions, although they may appear to be unaware of this interest …. Experience with a rather wide variety of relative-number problems, varying from "how many were affected at this dose" to "how many of the pebbles are quartz", indicates that further analysis proceeds smoothly and thoroughly when other modes of expression are used instead of "percentages".»

Despite knowledge and available tools, counted fractions are still commonly ignored. As Aitchison (1982) notes in the context of the related field of compositional data analysis, so-called "describers" merely state: 'After all we are simply describing the data set in summary form, not analyzing it.' Another group are the "wishful thinkers": 'No problem exists or, at worst, it is some esoteric mathematical statistical curiosity which has not worried our predecessors and so should not worry us. Let us continue to calculate and interpret correlations of raw components. After all if we omit one of the parts, the constant-sum constraint no longer applies. Someday, somehow, what we are doing will be shown by someone to have been correct all the time.' (Aitchison, 1982).

4.1 Performance metrics in healthcare services

From only a few years ago, where annual pdfs were published on-line with detailed statistics of performance metrics for various healthcare services, these days websites are appearing where colourful tables warn in red or highlight in green negative and positive changes in regional healthcare services, almost in real time. But some of these have to be taken with a 'pinch of salt'.

•Figure 4a,b. Changes between 2013 and 2014 in the fractional number (y-axis) of patients seeing a doctor within 7 days for a range (x-axis) of regional primary care providers (a) in % terms according to eq. (1); (b) in logits, z, according to eq. (2) (SALAR, 2018).



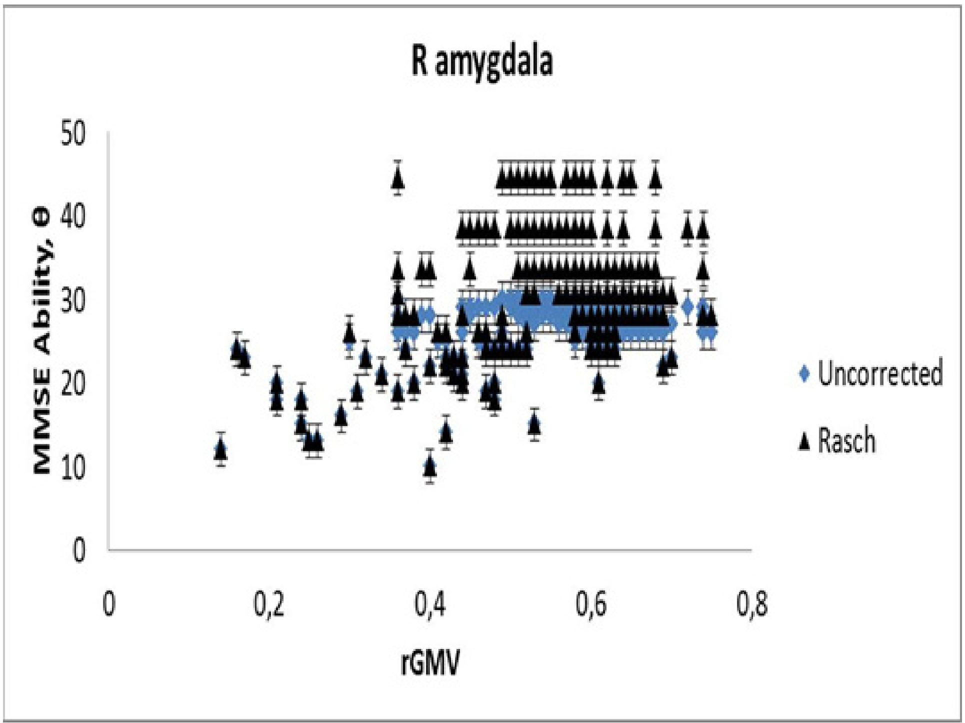
Source: SALAR (2018)

Figure IVa apparently shows a dramatic improvement by + 6 % for one particular regional primary care provider between the years 2013 and 2014 in the fractions of patients seeing a doctor within 7 days when the data is presented in percentages, according to eq. (1) (SALAR, 2018). But it is easy to reveal that this is just a counted fraction mirage, and is not matched by a corresponding change in logit metric z, according to eq. (2) (Figure IVb). The explanation is that one has to account for two separate factors; not only the change in percentage points but also the absolute level in percent for each quantity compared. In this example, the change for the regional provider in question spanned a range of absolute values 85 % - 91 % between the years 2013 and 2014, where each counted fraction scale has appreciable and different non-linearity according to eq. (1) at each absolute level. The corresponding logit change [Figure IVb] shows that this specific region's apparent improvement in care service provision performance is not significantly better than other regions, and indeed lies within intra-regional scatter.

4.2 Cognitive studies of Alzheimer's disease patient

As a second example where counted fractions have to be corrected for is regarding person-centred outcome measures, such as evaluations of abilities or leniency. Recent studies of possible correlations between neurodegeneration in patients suffering from Alzheimer's disease – specifically failing cognitive ability – and brain atrophy are part of the on-going EMPIR HLT04 NeuroMet project. A recent application of the Mini Mental State Examination (MMSE) test can be found of potential correlations between cognitive ability and brain atrophy studied by Dinomais et al., (2016) - their original ("uncorrected") data is plotted in Figure V. MMSE scale distortions arising from the counted fraction effect, surprisingly, do not seem to have been accounted in an otherwise extensive literature (Klein-Koerkamp et al., 2014) covering advanced correlation studies between cognitive and biomarkers of neurodegenerative diseases. But these can be corrected for in a metrological manner thanks to the Rasch invariant measure theory, which for MMSE has already been investigated by Hughes et al. (2003).

Figure 5- Correlation plots of cognitive ability (MMSE, dependent variable) versus regional grey matter volume (rGMV, independent variable).



original data (Dinomais et al., 2016); X Rasch corrected for distortion shown in Figure 4 (Hughes et al., 2003), and this work. Standard uncertainties are indicated for MMSE ability scores (Pendrill, 2018).

5. Metrological item banks

The need for better coherence between to date parallel initiatives in the metrology of healthcare from the traditional metrological organisations (predominantly in the physical and chemical sciences) and healthcare researchers has been highlighted recently. Cano et al. (2017) for instance: '….propose a new international body to bring together metrology, psychometrics, philosophy, and clinical management to support the global comparability and equivalence of measurement results in patient centred outcome measurement to improve healthcare'. In laboratory medicine, the standardization of measurements is of course a high priority, with the goal of comparability of results obtained using routine procedures (Jones & Jackson, 2016; Hallworth et al., 2015). 'A determination to accept patient outcome and patient experience as the primary measure of laboratory effectiveness' is a recommendation of a recent International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) Task Force on the Impact of Laboratory Medicine on Clinical Management and Outcomes. The IFCC in turn belongs to the Joint Committee for Traceability in Laboratory Medicine as a link to the metrology community.

The present work indicates that certified reference 'materials' (CRM) can apparently be synthesised in terms of construct specification equations, i.e., experimental intervention/manipulation on either person attribute (e.g., ability) or object attribute (e.g., difficulty) or both simultaneously, that yield successful prediction of the observed outcome (count correct). These causal Rasch models relate task difficulty and patient ability to explanatory variables such as the test sequence entropy and brain atrophy. This means more than merely running data through Rasch calibration software. The ability of the Rasch approach to yield separate and objective measures of task difficulty and person ability, for instance, is essential in establishing sets ("item banks") of metrological references based on proper use of the psychometric Rasch model (Choppin, 1968; Pesudovs, 2010). The bottom line is that 'counted fractions' do indeed belong to an extended quantity calculus, thus enabling quality-assurance of many essential measurements in healthcare.

Acknowledgments

This work has been performed as part of the 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.

Additional funding has been provided by VINNOVA, the Swedish Innovation Agency, in the project: Quality-assured measurement in Swedish healthcare, Reg. no. 2017-05011. The aim of that project is: "Care is increasingly delivered with the entire health condition of the person in focus, rather than just the disease. In this context, there is a growing need for quality-assured measurement of categorical measurements, but typical human responses, such as opinions on ordinal measurement scales, are more difficult to handle quantitatively. The purpose of the present project is therefore to start work on establishing a national metrology institute (NMI) for quality assurance of categorical measurements in healthcare. In the long run, the goal is to become a leading NMI in the area."

Thanks are also due, particularly, to:

•Members of the EMPIR NeuroMet consortium;

•Stefan Cano, Modus Outcomes, Boston (USA) & Stotford UK;

•William P. Fisher, Jr., Ph.D., Research Associate, BEAR Center, Graduate School of Education, University of California, Berkeley, CA (USA) & Principal, LivingCapitalMetrics Consulting.

References

Aitchison, J (1982), The statistical analysis of compositional data, *J. R. Statistic. Soc.* **44,** 139 -77.

Bentley, J P (2005), *Principles of Measurement Systems.* 4th ed. London: Pearson Education Limited.

Cano, S J, T Vosk, L R Pendrill, A J Stenner (2016), On Trial: the Compatibility of Measurement in the Physical and Social Sciences, *Journal of Physics: Conference Series* 772 012025, IMEKO2016 TC1-TC7-TC13, doi:10.1088/1742-6596/772/1/012025.

Cano, S J, L R Pendrill, S P Barbic and W P Fisher, Jr. (2017), Patient-Centred Outcome Metrology for Healthcare Decision-Making, *Joint IMEKO TC1-TC7-TC13 Symposium: Measurement Science Challenges in Natural and Social Sciences,* Rio de Janeiro, Brazil, July 31st - August 3rd 2017.

Choppin, B (1968), An item bank using sample-free calibration, *Nature* **219,** 870 -2.

Dinomais, M, S Celle, G T Duval, F Roche, S Henni, R Bartha, O Beauchet and C Anweiler (2016), Anatomic Correlation of the Mini-Mental State Examination: A Voxel-Based Morphometric Study in Older Adults, *PLOS ONE* | DOI:10.1371/journal.pone.016288.

EN 15224:2012, *Health care services – Quality management systems – Requirements based on EN ISO9001:2008,* European standard.

FDA (2009), *Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims,* U.S. Department of Health and Human Services, Food and Drug Administration, http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm.

Ferreira, M and A Matos (2015), Metrology in health: a pilot study, *Journal of Physics: Conference Series* **588,** 012018.

Flater, D (2017), "Redressing grievances with the treatment of dimensionless quantities in SI", *Measurement* **109,** 105–11.

Hallworth, M et al., (2015) Current evidence and future perspectives on the effective practice of patient-centered laboratory. *Med. Clin. Chem.* **61** 589–99.

Heinemann, A., W. Fisher, Jr., and R. Gershon (2006), Improving Health Care Quality With Outcomes Management, *J. Prosthet. Orthot.,* **18,** no. 1S, pp. 46-50.

Hughes, L F, K Perkins, B D Wright and H Westrick 2003 Using a Rasch scale to characterize the clinical features of patients with a clinical diagnosis of uncertain, probable, or possible Alzheimer disease at intake *J. Alzheimer's Dis.* **5** 367–73.

ISO 13485:(2003), *Medical devices - Quality management systems -- Requirements for regulatory purposes,* International standard.

Jones G, and C Jackson 2016 The Joint Committee for Traceability in Laboratory Medicine (JCTLM)—its history and operation, *Clin. Chim. Acta* **453** 86–94.

Klein-Koerkamp, Y, R A Heckemann, K Ramdeen, O Moreaud, S Keignart, et al., (2014), Amygdalar Atrophy in Early Alzheimer's Disease, *Current Alzheimer Research* **11,** pp.239-252, https://hal.archives-ouvertes.fr/hal-00959516.

Kogan, J. (2014), An Alternative Path to a New SI, Part 1: On Quantities with Dimension One, https://web.archive.org/web/20160912224601/http://metrologybytes.net/PapersUnpub/Kogan_2014.pdf.

Pendrill, L R, B Berglund et al., (2010) Measurement with Persons: A European Network, NCSLi Measure, (June 2010) 5, 42 – 55.

Pendrill, L R (2018), Assuring measurement quality in person-centred healthcare, *Measurement Science & Technology,* Volume 29, Number 3 , 034003 special issue Metrologie 2017, https://doi.org/10.1088/1361-6501/aa9cd2.

Pesudovs, K (2010),Item banking: a generational change in patient-reported outcome measurement, *Optom Vis Sci.,* 87(4) pp 285-93. doi: 10.1097/OPX.0b013e3181d408d7.

Tukey, J W [Chapter 8, Data analysis and behavioural science; quoted in "The collected works of John W Tukey, Volume III, *Philosophy and principles of data analysis: 1949 – 1964*", ed. L V Jones, Univ. North Carolina, Chapel Hill, Wadsworth Statistics and Probability Series. Wadsworth & Brooks/Cole, Belmont (1985).

SALAR (2018), Vård i siffror, (Swedish: Care in numbers), https://vardenisiffror.se/registry/registry?descriptionSource=V%C3%A4ntetider%20i%20v%C3%A5rden,%20Sveriges%20Kommuner%20och%20Landsting .

Authors Profiles:

Prof. L R Pendrill, Ph D Atomic Physics 1978, has until recently been R&D Director at Measurement Technology Division and the National Metrology Institute of SP at Borås. He plays a leading role, nationally and internationally, recently as Chairman (2009-12) for EURAMET (www.euramet.org), the European association of national metrology institutes in Europe. Pendrill is currently Chair of ISO/TC12 Quantities & Units. He develops new quality-assured measurement methods in which Man is focus by measuring PCOM (e.g. perceived comfort), as well as optimised measurement uncertainty, where the effects of measurement are weighed against the effects of product evaluations.

Dr. J Melin has the last years been working with measurements in health care. She is now working at the division for Metrology at RISE Research Institute of Sweden, where she is included in several project where person-centered metrology is used and further developed. During her PhD project she took part in the development of the PPRQ at the Gothenburg Centre for Person-Centred Care (GPCC). After her dissertation in December 2016 she has both continue the development of PPRQ and her skills in metrology.