MAIMON WORKING PAPER No. 10 JUNE 2024

TIME TO ABANDON THE REFERENCE CASE: THE 2024 DUTCH GUIDELINE FOR ECONOMIC EVALUATIONS IN HEALTHCARE

Paul C Langley, Ph.D., Adjunct Professor, Graduate Faculty, College of Pharmacy, University of Minnesota, Minneapolis MN and School of Pharmacy, University of Wyoming, Laramie WY

Abstract

The latest revision to the Dutch Guidelines for economic evaluations in health care fails both the standards of normal science and fundamental measurement. This is not unexpected as the focus is on creating imaginary cost-effectiveness claims from assumption driven simulations in an attempt to provide a basis for the allocation of health care resources. It is surprising given the increasing criticisms directed to multiattribute reference case modelling that the expert panel advising the authorities have maintained the status quo. The panel appears to have had no understanding of the standards for fundamental or Rasch measurement. At the same time, the US has now effectively outlawed both multiattribute ordinal preference scores and the QALY with the prohibition of discrimination amendments to Section 504 of the Rehabilitation Act of 1973. Once the manifest deficiencies of the reference case are detailed, the Guideline ceases to have any relevance for health care decision making. The entire exercise not only fails to meet the standards of normal science in differentiating science from non-science but also fails to meet the standards for fundamental measurement in the focus on preference scores to create quality adjusted life years (QALYs) that are mathematically impossible because the scores are both composite and ordinal. The purpose of this brief note is to deconstruct the reference case as a composite construct that is essentially meaningless. This is not difficult as the failure of the reference case was obvious over 50 years ago. It fails because any value claim for therapy must be based on a measure that has unidimensional, linear, interval and invariant properties. There is no alternative; a requirement noted in the 17th century with the early development of the thermometer. Multiattribute instruments fail to grasp the necessity of rules to transform observations to measurement with unidimensional, interval and ratio properties. These guidelines need to be remastered to conform to the standards of normal science and fundamental measurement in focusing on patient centric, disease specific value claims and not impossible multiattribute preferences and the QALY.

INTRODUCTION

Although long overdue, the US has now effectively outlawed multiattribute preference or utility scores and quality adjusted life years (QALYs) [1] . The amendments to Section 504 of the Rehabilitation Act of 1973 state that: there should be no value claim entertained for medical treatment decisions by those that receive Federal financial assistance from the Department of Health and Human Services (HHS) if it is based on:

- biases or stereotypes about individuals with disabilities,
- judgements that an individual with a disability will be a burden to others, or

- beliefs that the life of an individual with a disability has less value than the life of a person without a disability [2] .

In practice, given the financial reach of DHHS, this means any value claim that discriminates on access to care is disallowed. Multiattribute generic preference or utility scores, driven by community valuations of health states, must be abandoned in favor of patient centric or disease specific instruments. Value claims involving QALYs, including cost per QALY thresholds and reference case assumption driven simulation creating non-evaluable cost-effectiveness claims are by definition disallowed.

The irony, however, is that it has taken 40 years for reference case models and QALYs to be declared surplus to requirements in health technology assessment [3] [4]. If the many advocates of reference case models, particularly those in single payer health systems, had been aware of (i) the standards of normal science; (ii) the standards for fundamental measurement and (iii) Hume's problem of induction, the development of multiattribute instruments such as the EQ-5D-3L/5L and the assumption driven modelled simulations should have been abandoned when first proposed in the 1980s. At this late stage, outlawing multiattribute scores and the QALY on the grounds of their discriminatory role is precisely what these instruments were intended to achieve This is to be welcomed: a universal metric that rationed or denied care given community preferences for the value of health states has no place in health system decision making.

The reference case model is the mainstay of those who seek, unwisely, a single metric that will, with the application of cost-effectiveness tools, drive the efficient allocation of health care resources. This Holy Grail of resource allocation is a will o'the wisp; it is impossible to conceive of a single generic metric that will achieve this 'central planning' objective. To extend the metaphor: the search is truly a Pythonesque endeavor. The purpose of this brief note is to make those supporting the Guidelines aware of their irrelevance. At the same time this will ensure that the Guidelines are not put forward as the exemplar for health care resource allocation in other counties of the European Union.

STANDARDS OF NORMAL SCIENCE

Judged by the standards of normal science, with the focus on the evolution of objective knowledge, reference cases are an analytical dead end. They fail the demarcation criteria that distinguish science from non-science [5] . This is not inadvertent but by design; perhaps aptly described as *nonsense on stilts*. There was no intention that hypothesis testing had any place in reference case modelling because the value claim, in cost-effectiveness terms, was not potentially falsifiable. The assumption driven simulation for a hypothetical population could stretch decades into the future with the label cost-effectiveness in mathematical terms just nonsense. Instead, the case rested on approximate information; although it was never clear as to what the approximate information reference point actually was so that any therapy could be driven by any number of assumptions, justified by being taken from the literature, with each creating a different cost-effectiveness claim.

Presumably one reference case model will be a closer approximation than another to a Platonic model ideal, although we would never know. Obviously, this opened the door to sponsor supported models which, presumably by happenstance, supported the sponsors product [6]. The door was open

to false and frivolous claims with reference models explaining things rather too well. A situation characteristic of non-science endeavors such as intelligent design, Marxism and astrology.

Nevertheless, this belief system has endured for some 40 years with the latest relativist incarnation the CHEERS 2022 guidance for the creation of reference case imaginary claims, presented in a form that maximized their chance of acceptance in journals that should know better [7] [8] . The good research practices of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) support this analytical dead end of assumption driven imaginary value claim simulations [9] . The result is that we have thousands of peer reviewed and published Markov simulation models that make no contribution whatsoever to our understanding of therapeutic response or as inputs to ongoing disease area and therapeutic class reviews. They fail to provide any foundation for value claims despite what must be seen as recent rearguard actions [10].

STANDARDS OF FUNDAMETAL MEASUREMENT

As the basis for the rejection of reference case modelling, we don't have to go as far as to note its rejection of demarcation and its home in non-science. The focus on generic multiattribute instruments such as the EQ-5D-3L, algorithmic utility and preference scores and the QALY are sufficient to ensure membership of the non-science fraternity. For those familiar with the contributions of Luce and Tukey, and Rasch to fundamental measurement, the challenge for patient reported outcomes is to transform observations to measurement [11]. This was resolved by the 1960s in the application of Rasch rules as the necessary and sufficient condition for transforming observations to measurement where the value claim was unidimensional or a single attribute, with linear, interval and invariant (and empirically evaluable) properties [12]. Central to the concept of measurement is to measure one attribute at a time. A measure is invariant for a single attribute *where the readings will remain invariant across all suitable contexts; and for any one context, all suitably calibrated devices will yield invariant readings*. Composite measures must be avoided. In failing to recognize the standards of measurement theory they produce invalid and misleading ordinal scores.

Judged by the standards of fundamental measurement, multiattribute utility or preference scores are a waste of time. The score that is produced has ordinal characteristics: the score can be ranked but differences between the scores are meaningless. There is no meaningful unit of difference. The reason for this is obvious: there was no intent when these instruments were developed to focus on a single attribute, a measure to be created with linear, interval and invariant properties. Instead, the development followed traditional procedures in fitting a model to describe the data where these have primacy. The Rasch model is the reverse: items or data inputs are selected by the application of Rasch rules to fit the Rasch model. If the size and structure of residuals meet Rasch standards then we are justified in claiming to have a measurement scale with linear, interval and invariant properties.

The Rasch model is patient centric. For a latent construct such as quality of life, the focus may be on measuring an attribute or manifestation of quality of life such as need fulfillment. This will be a characterization specific to a disease state or target patient population. The value claim will be for that attribute. Finally, the Rasch model is probabilistic: the success in responding to a need

fulfillment item will be a function of the difference between the difficulty of the item and the ability of the patient. In other words, a true measure of response to therapy for a target population with a distribution of abilities and a distribution of items with levels of difficulty. If thought had been given to the creation of a metric to guide health system resource allocation, it should have been obvious that a composite health state instrument, valued by community preferences, was impossible. Attempts to allocate health system resources with a single metric such as the QALY should have been seen as a wasted effort.

Instruments such as the EQ-5D-3L have an unfortunate eugenic legacy; they reflect community preference, the value of pre-determined health state descriptions, as inputs to algorithms that decide who wins or loses in the allocation of, presumably scarce, health care resources. An unfortunate, and fatal, aspect of the scores generated by multiattribute instruments is the presence of negative preference scores. While the scores are capped at unity there is no lower bound as the individual scores are decrements from unity and can overshoot zero. Euphemistically described as 'states worse than death' the negative score has been the focus of concern by those, unfortunately, who have no idea of the standards for fundamental measurement [13]. In an UK valuation of the EQ-5D-3L with a total of 243 health states (5 health dimensions and 3 reporting levels), 34.6% were states worse than death with the range of preference scores for health states from perfect health 1 to -0.594. To describe a health state as worse than death implies that, with a negative score, disallowing health care for those patients will increase, as measured in QALYs, the overall health of the population. This is what the Section 504 amendments are focused on: prohibiting discrimination by application of instruments that can be deliberately discriminatory. All multiattribute instruments fall into this category. The resolution is obvious: abandon generic instruments in favor of Rasch standard instruments for each disease state or target patient group where the concept of a state worse than death is absent; there are no negative measures.

Debating different valuations for instruments, such as the perennial comparisons of the EQ-5D-3L and EQ-5D-5L, agonizing over the meaning of states worse than death, attempting to tweak utility scores to avoid the more egregious discrimination and the weird attempts to crosswalk between instruments with ordinal scores, are all the result of an obsession with a universal resource allocation metric. An obsession that is set to continue with the EQ-Health and Wellbeing Instrument (EQ-HWB). The answer is that, again, there is no universal metric; this search for a Holy Grail is inconsistent with the standards for fundamental measurement. This was known 50 years ago.

PROBLEM OF INDUCTION

David Hume, the Scottish enlightenment philosopher, in 1748 raised the question of what has become known as the problem of induction; drawing inferences from data and the expectation of uniformity [14]. The problem, as Russell pointed out in his seminal 1912 essay, is that we cannot prove or disprove the principle of induction from experience [15]. Arguments in probabilistic terms fail because likelihood is always relative to data or known cases. We are, therefore, in the position that to anticipate uniformity, the verification of an expectation of a fresh instance, we have to assume the principle of induction. Unfortunately, unless we make this assumption, we cannot expect that future futures will resemble past futures.

Given the role of assumptions in the construction of a reference case, the belief that a particular set of assumptions are 'realistic for the future' is pushing the boundaries of credibility. Even if we argue in probabilistic terms, the number of past instances on which an assumption is justified is typically limited. In many cases, where the model derives from pivotal trial the instances used to justify an assumption on the future is in low single digits. Add to this the likelihood that an application of the trial protocol in a new patient group has, at best, a 50:50 chance of replication.

In logic, we cannot argue that the fact that all past futures have resembled past pasts that all future futures will resemble future pasts. We cannot establish the validity of inductive procedures. The fact that the claim is typically made that a reference case model has been validated goes against the modern Popperian scientific method: the logical asymmetry between verification and falsification[16] . Irrespective of any number of confirming statements universal statements are impossible because one non-verification observation implies possible falsification.

The reference case denies the possibility of falsification not only because the cost-per-QALY claim is non-evaluable but because a judicious choice of assumption can create any desired outcome. As adoptees of the NICE reference case, the Dutch guidelines deny or obstruct the evolution of objective knowledge; the denial of progress in understanding therapeutic impact.

IS THE DUTCH GUIDELINE FIT FOR PURPOSE

As a key element in the Dutch commitment to health technology assessment, the 2024 Guideline is not fit for purpose. As a guide to economic evaluations for therapy and care interventions from a societal perspective the Guideline is not only a disappointment but from the standards of normal science and fundamental measurement, irrelevant. The focus of the Guideline on the reference case is an error of the first magnitude. It denies the intent of the new philosophy of the scientific revolution of the 17th century to discover new facts, the evolution of objective knowledge, and the associated need for accurate measurement.

The reference case in HTA is a degenerate decision; rather than the idea of progress in scientific endeavor, and many go to lengths to insist that HTA is a science, the focus of belief is on a construct, the quality adjusted life year (QALY) that epitomizes a pre-scientific mind set. Rather than a commitment to science and the criteria for demarcation, HTA is committed to non-science. Approximate information takes center stage; the reference claims for non-evaluable cost-effectiveness are the decision criteria. It is unclear if this is what the authors of the Dutch guideline intended or health system decision makers.

The reference case is a cover-up. Rather than a commitment of time and resources to meet the evidence gaps that are inevitable when a new product is launched, with provisional formulary acceptance and pricing, the approximate information creed dictated that if evidence was unavailable then it could be invented. A relativist position that evidence is created within a social community, not discovered; where a research program depends not on is generation of new knowledge but its mobilization of and support from a community where truth is consensus [17].

The reference case is the perfect vehicle: an assumption driven model simulation for a hypothetical population that captures, typically for the lifetime of the imagined patients, the costs and outcomes

for an unknown future. Once the reference model has been created, often in support of a sponsor's product, there is no incentive to go any further; it is, by design, an analytical dead end.

The reference case model focuses on estimated future costs as the basis for incremental cost-per-QALY claims. While costs are clearly interval measures with linear and invariant properties, they are typically constructed as composite measures with an aggregate cost reported for assumed future resource utilization. They are not unidimensional and fail the standards for fundamental measurement. Value claims should be for specific elements of resource use (e.g., drug utilization, physician visits, hospitalization) there should be a value claim that is empirically evaluable lined, in the case the Dutch guidelines to DRG procedure and ATC drug codes.

The commitment to the reference case is reinforced by the uncritical acceptance of multiattribute instruments where the community dictates the weights that should be applied to health state bundles of health dimensions and symptom levels. Despite a prior 30 or more years of establishing the standards of fundamental measurement these were ignored. They were antitheical to the search for a Holy Grail community weighted composite metric to drive resource allocation irrespective of its failure to meet fundamental measurement properties.

The reference case model had to embody composite multiattribute utility and preference scores that failed the standards for fundamental measurement. To accept those standards, if any were aware of them, would be to accept that measurement must be for single attributes with linear, interval and invariant properties. To entertain this would be to abandon multiattribute instruments. There is no option. If observations are to be transformed to measurement, then, for the attribute of interest, we have to apply the necessary and sufficient Rasch rules [11].

Alongside reference case modelling, the Guidelines recognize what they describe as empirical economic evaluations. This is in an important sense a misnomer; it refers only to cost-effectiveness evaluations alongside clinical trials. Two points are worth noting: first, any protocol supporting an economic evaluation alongside a clinical trial must recognize the importance of empirically evaluable single attribute linear, interval and invariant claims that meet Rasch measurement standards with particular reference to PRO claims and, second, that the results of the evaluation must not be taken at face value but as input to a formulary submission protocol for a value claim in an attempt to replicate the claim.

In support of the empirical evaluation framework the Guideline references the ISPOR Good Research Practices Task Force Report for Cost-Effectiveness Analysis Alongside Clinical Trials [18]. This should not be taken too seriously because the ISPOR authors clearly have no notion of fundamental measurement in recommending weighting clinical end points by utilities (an ordinal scale) to create the impossible QALY. Effectiveness expressed in these terms should be rejected as completely absurd.

Indeed, ISPOR is not alone in a lack of understanding of fundamental measurement. The 2023 EUnetHTA Individual Practice Guidelines Document Guideline appears completely unaware of Rasch rules for measurement to capture PRO outcomes [19].

Unless the standards of fundamental measurement are applied, claims for relative effectiveness of competing therapies lose any meaning. Certainly, claims can be made for changes in a score but as these fail Rasch standards for measurement, the comparative effectiveness claim, even of supported by traditional statistical techniques for effect size, are meaningless as the score lacks the required measurement properties and, of course, cannot support parametric statistical techniques. Once these concerns are detailed, the Dutch Guidelines should be withdrawn. While many may regard them, in NICE terms, as the state of the HTA art, they are an analytical dead end.

A NEW START IN HTA

The Section 504 discrimination amendments to the Rehabilitation Act of 1973 have created a vacuum for disease and target patient group specific outcomes instruments. Once the standards for fundamental measurement are applied then there are few instruments that meet the required Rasch standards.

A recent proposal for a new start in HTA for comparative effectiveness value claims provides three premises for all value claims [4]. These premises are patient centric, making the case that claims must be for disease states or target patient populations**.** The premises are:

- All value claims must refer to single attributes for defined patient populations that meet the demarcation standards for normal science: they must be credible, evaluable and replicable
- All value claims, notably for patient or caregiver reported outcomes. must be consistent with the limitations imposed by the standards of fundamental measurement: they must be unidimensional with linear, interval and invariance properties
- All value claims must be supported by an agreed protocol detailing how they are to be assessed in a meaningful timeframe for clinical, PRO, drug and resource utilization submissions,

The first premise eliminates assumption driven modelling with imaginary cost-effectiveness claims; the second premise eliminates the QALY and successor attempts to create multiattribute instruments; and the third premise requires all claims to be empirically evaluated and monitored to eliminate false claims and support reproduction of claims in different target patient populations. These standards for a new approach to HTA are detailed in a recently released Certificate Program from the University of Wyoming *A New Start in Health Technology Assessment* (attached).

CONCLUSIONS

The Dutch Guideline has no practical relevance. If value claims are to be considered for a new therapy or care interventions, then then these claims should be presented individually to cover clinical, PRO, drug utilization and resource use outcomes. Each should be accompanied by a protocol for evaluation. This gives the required analysis platform consistent with standards for normal science and fundamental measurement. Accepting this means a commitment to a research strategy to meet evidence gaps and set the stage for ongoing disease area and therapeutic class reviews.

Whether this new start in HTA will be adopted is a moot point; there are too many opinion leaders in HTA who have too much to lose; the baggage of 30 years of wasted effort. To admit that the thousands of peer reviewed and published papers that have embraced the QALY and promoted reference case models are a waste of time and resources, does not put HTA in a good place; latest PubMed count (5 June) is 26,218 . Debates, for example, in the UK over access to cancer drugs when denied by NICE are to be seen as totally misplaced; cost-per-QALY thresholds are meaningless.

If the objective is to deflate time spent in a disease state by a score that gives the equivalent to perfect health, then the requirements and techniques in measurement terms are straightforward. We start with a latent concept, call it quality of life as judged from a patient or caregiver perspective in a disease state. Then, we select a manifest of that latent construct that we want to measure. Call it need fulfillment (this goes back 35 years). Then we apply Rasch modelling to produce, if possible, a need fulfillment measure which is unidimensional, linear, interval, invariant and fixed to a range from 0 to 1 [20]. Rasch rules can meet these standards; they have been applied to create disease specific instruments for over 30 years. The EQ-5D-3L/5L certainly do not meet these standards; they are not a metric as defined by fundamental measurement. Multiattribute scores are an analytical dead end. They lack interval properties and cannot support statistical analysis. They show a complete lack of understanding of fundamental measurement; as evidenced by the will o'the wisp community weighted false QALY which is finally being abandoned.

In the history of science there are many examples of beliefs being overthrown as part of what Popper describes as the process of conjecture and refutation; the reference case stands out as an analytical framework that creates non-falsifiable claims. The irony, and this applies in all applications by health systems, is that if those promoting and accepting the reference case had been aware of the standards of normal science and the standards for fundamental measurement, multiattribute ordinal composite scores and the reference case should not have been considered in the first place.

UNIVERSITY OF WYOMING CERTIFICATE PROGRAM

A NEW START IN HEALTH TECHNOLOGY ASSESSMENT

For those who are interested in following up the arguments presented here for Rasch standard patient centric value claims, the recently released on-line University of Wyoming Certificate Program: A New Start in Health Technology Assessment is recommended.

The Certificate Program is in three parts:

- Part I: Required evidentiary standards for product and therapy assessment
- Part II: The failure of approximate modelled information for therapy decisions
- Part III: Formulary submission value claims and protocols for a new start in product evaluation in health system management

The Certificate Program package includes extensive notes (overall for the 14 modules 85,000 words), audiovisual presentations and a short true-false and multiple-choice assessment for each module. The cost of the Certificate Program is $875 USD with 20.5 hours of ACPE credit. For those who do not need ACPE accreditation, the University of Wyoming will provide a Certificate of Completion. Following interest already expressed, for those introducing the proposed new start standards for technology assessment there will be a program of one- and two-day workshops and on-line seminars to support course development and alternative program structures to meet local needs. There will also be a series of working papers to explore specific aspects of the new start program.

The link to register in the Certificate Program is:
https://www.uwyo.edu/pharmacy/resources/certificate-program-a-new-start-in-healthtechnology-assessment.html

The Certificate Program was developed by Dr Paul C Langley, a health economist, who presently resides in Tucson, Arizona. If more information is required on course content and the need for a new start, he can be contacted at langleylapaloma@gmail.com

REFERENCES

[1] Langley P. The United States outlaws multiattribute preference scores and the QALY. *Maimon Working Paper.* No 7 May 2024
https://maimonresearch.com/wp-content/uploads/2024/05/MAIMON-WORKING-PAPER-No.-7-MAY-2024-V21.pdf

[2] US Government. Health and Human Services Department. Nondiscrimination on the Basis of Disability in Programs or Activities Receiving Federal Financial Assistance: Final Rule. Federal Register May 9, 2020. 89 FR 40066 https://www.federalregister.gov/public-inspection/2024-09237/nondiscrimination-on-the-basis-of-disability-in-programs-or-activities-receiving-federal-financial

[3] Langley PC and McKenna SP. Measurement, modeling and QALYs [version 1; peer review: 2 approved] *F1000Research* 2020, 9:1048 https://doi.org/10.12688/f1000research.25039.1

[4] Langley P. Nothing to Cheer About: Endorsing Imaginary Economic Evaluations and Value Claims with CHEERS 22 [version 1; peer review: 2 approved]. *F1000Research* 2022, **11**:248 (https://doi.org/10.12688/f1000research.109389.1)

[5] Pigliucci M. Nonsense on Stilts: How to tell science from bunk. Chicago: University of Chicago Press, 2010

[6] Langley P. Facilitating bias in cost-effectiveness analysis: CHEERS 2022 and the creation of assumption-driven imaginary value claims in health technology assessment [version 1; peer review: 3 approved]. *F1000Research* 2022, **11**:993 (https://doi.org/10.12688/f1000research.123709.1)

[7] Husereau D, Drummond M, Augustovski F, *et al.*: Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) Statement: Updated reporting guidance for health economic evaluations. *ValueHealth.* 2022; **25**(1): 3–9

[8] Husereau D, Drummond MN, Augustovski F: Consolidated Health Economic Evaluation Reporting Standards (CHEERS) 2022 Explanation and Elaboration: A Report of the ISPOR CHEERS II Good Practices Task Force. *ValueHealth.* 2022; **25**(1): 10–31

[9] Langley P. Are the practice guidelines of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) fit for purpose? *Maimon Working Paper* No. 9 June 2024 https://maimonresearch.com/wp-content/uploads/2024/06/MAIMON-WORKING-PAPER-No-9-2024-V5-1.pdf

[10] Wilkie R, Pizzi L, Rand L et al. The Value of the Quality Adjusted Life Years. *ValueHealth*. 2024;27(6):702-705

[11] Bond T, Yan Z, Heene M. Applying the Rasch Model: Fundamental Measurement in the Human Sciences (4th Ed.) New York: Routledge, 2021

[12] Wright B, Linacre J. Observations are always ordinal; measurements, however, must be interval. *Arch Phys Med Rehabil.* 1989; 70(12):857-60

[13] Schneider P. The QALY is ableist: on the unethical implications of health states worse than dead. *Qual Life Res*.2022;31:1545-52

[14] Hume D. An Enquiry Concerning Human Understanding. 1748

[15] Russell B. The Problems of Philosophy. 1912

[16] Popper K. Objective Knowledge: An Evolutionary Approach (Rev. Ed.). Oxford: Clarendon Press, 1979

[17] Wootton D. The Invention of Science: A New History of the Scientific Revolution. New York: Harper Collins, 2015

[18] Ramsey S, Wilkie R, Glick H et al. Cost-effectiveness analysis alongside clinical trials: An ISPOR Goor research Practices Task Force Report. *ValueHealth*. 2015;187:161-72

[19] EUnetHTA: Guideline Endpoints. Individual Practice Guidelines Documents. D4.4 Outcomes (Endpoints). V1.0 2023 https://www.eunethta.eu/wp-content/uploads/2023/01/EUnetHTA-21-D4.4-practical-guideline-on-Endpoints-v1.0.pdf

[20] Langley P. McKenna S. Fundamental Measurement: The Need Fulfilment Quality of Life (N-QOL) Measure. *InovPharm*.2021;12(2):No. 6
https://pubs.lib.umn.edu/index.php/innovations/article/view/3798/2697