MAIMON RESEARCH LLC

RECONSTRUCTION OF HEALTH TECHNOLOGY ASSESSMEMNT



THE EQ-5D-3L: NUMBERS WITHOUT MEASUREMENT

Paul C Langley PhD Adjunct Professor, College of Pharmacy, University of Minnesota, Minneapolis, MN

LOGIT WORKING PAPER No 2042 NOVEMBER 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

Background: *The EQ-5D-3L valuation algorithm developed by Dolan (1997) occupies a central position in conventional health technology assessment (HTA). It generates numerical utility weights for 243 multidimensional health-state descriptions, subsequently employed in quality-adjusted life-year (QALY) calculations. This paper challenges the measurement legitimacy of those weights and the quantitative claims derived from them. Its central proposition is that statistical estimation cannot substitute for representational measurement.*

Methods: *Dolan's regression algorithm is examined against the requirements of representational measurement, including defined empirical attributes, representation and uniqueness conditions, scale properties, dimensional homogeneity, commensurability and admissible arithmetic. The analysis considers the transition from the original time trade-off (TTO) duration ratio to the regression dependent variable, the interpretation of dimension-level coefficients as additive utility decrements, and the claimed proportional meaning of the resulting tariff scores.*

Results: *Dolan's analysis employed 35,964 TTO-derived numerical observations from 2,997 respondents evaluating 42 selected health-state descriptions. These observations concerned hypothetical multidimensional classifications rather than independently measured empirical attributes. Positive TTO coefficients originated in lawful ratios of survival durations, while negative valuations arose through modified procedures. Neither established a common ratio measure of preference. The regression transformed these observations into decrements from unity and estimated coefficients for categorical dimension-level indicators. Although statistically additive, these coefficients were not demonstrated to represent commensurable, additive quantities of a single preference attribute. Assigning full health unity and death zero did not establish an absolute zero. Negative tariff values further contradicted their interpretation as non-negative preference proportions. Consequently, a predicted utility of 0.6 cannot be interpreted as 60% of a measured preference quantity.*

Conclusions: *The QALY fails two independent measurement requirements. First, Dolan's algorithm does not establish a lawful preference ratio capable of serving as a proportional adjustment factor. Second, even a hypothetical preference ratio would require an empirically demonstrated mapping into duration space before it could legitimately adjust survival time. Neither condition is satisfied. The foundational error was the decision to value multidimensional health-state descriptions rather than identify and measure specified attributes. From TTO through conditional interval utility to regression-generated tariffs, successive numerical procedures assumed measurement properties never established. Their global institutional acceptance has culminated in a reference-case framework built upon unsupported quantitative claims. The result is arithmetic without measurement: numerical storytelling rather than lawful quantitative science.*

INTRODUCTION

The EQ-5D-3L valuation algorithm occupies a central position in conventional health technology assessment (HTA). Developed from time trade-off (TTO) responses, it assigns numerical utility weights to 243 multidimensional health-state descriptions for use in quality-adjusted life-year (QALY) calculations and economic evaluation. The central question is not whether these numbers can be generated through regression, but what they measure and whether their subsequent arithmetic is scientifically admissible.

A preceding paper demonstrated that the original TTO coefficient, h = t/T, is a lawful ratio of survival durations, not an independently established measure of health-state preference [i] . Torrance's 1976 conditional interval argument did not demonstrate the required measurement properties, while the subsequent appeal to cardinal utility and comparative economic evaluation failed to resolve the distinction between interval and ratio measurement [ii] [iii]. These findings provide the starting point for the present analysis; the historical and axiomatic arguments will not be repeated.

This paper examines the decisive transition from the conditional justification of health-state utility to its statistical estimation with the EQ-5D-3L preference algorithm [iv].. The relationship between conventional HTA and representational measurement was tenuous from the outset. Torrance and colleagues began with preferences over multidimensional health-state descriptions rather than the identification and measurement of specified empirical attributes [v]. Nevertheless, Torrance's subsequent attempt to establish a conditional interval utility representation at least acknowledged that assigning numerical values to health-state preferences required theoretical justification. That justification failed to establish the measurement properties necessary for the subsequent arithmetic, but the question of measurement had not yet disappeared entirely.

Dolan's 1997 EQ-5D-3L regression algorithm marks a more fundamental departure [vi] . The numerical valuations derived from TTO are accepted as observations of a dependent variable without independently establishing what attribute they measure. Categorical health-state descriptions become explanatory variables, estimated dummy-variable coefficients become numerical decrements, and the resulting regression equation generates values for all 243 EQ-5D-3L health states. The process is one of statistical estimation and numerical assignment. No empirical measurement structure is established for the dependent variable, no representation or uniqueness conditions demonstrate the quantitative meaning of the coefficients, and no lawful measurement properties are conferred upon the predicted values. The remaining connection between health-state valuation and representational measurement is effectively severed.

The significance of this transition extends beyond the regression itself. The estimated coefficients are incorporated into an algorithm whose outputs are designated health-state utilities and subsequently employed as proportional adjustments to survival duration in QALY calculations. Yet neither regression fit, statistical significance nor predictive performance can establish that the resulting numbers measure a defined preference attribute or possess the ratio properties required for that use. The algorithm supplies numbers where measurement is required, treating the capacity to calculate as sufficient justification for the quantities calculated. What began with Torrance as an unsuccessful attempt to justify numerical preference representations becomes, with Dolan, a

system in which the measurement question is no longer addressed. This is the defining characteristic of the EQ-5D-3L valuation algorithm: numbers without measurement.

ATTRIBUTES, HEALTH STATES AND MEASUREMENT

Scientific measurement begins with the definition of an empirical attribute and demonstration of the properties required for its numerical representation. Manifest attributes, such as survival duration and event counts, may support linear ratio measurement. Latent attributes, including fatigue, pain severity and functional capability, require validated measurement structures. Rasch measurement provides one framework for constructing invariant linear measures of specified unidimensional latent attributes [vii] . These different measurement structures determine the arithmetic that can legitimately be performed.

The EQ-5D-3L does not begin with the independent measurement of these attributes. Instead, it classifies health states through five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension has three ordered response levels, producing 243 possible combinations. A description such as 11223 identifies a particular combination of reported problems; it does not establish the magnitude of a single empirical attribute.

The distinction is fundamental. A multidimensional classification can describe health without constituting a quantitative measure of health gain. Assigning a numerical utility weight to each description does not transform the classification into an invariant measurement scale. The scientific question is whether the resulting numbers possess independently demonstrated measurement properties, not whether the descriptive system can be represented numerically.

REPRESENTATIONAL MEASUREMENT: THE REQUIRED STANDARD

Representational measurement requires numerical assignments to preserve empirically established relationships among instances of a defined attribute. The axiomatic foundations establish the importance of representation and uniqueness conditions, admissible transformations and meaningful quantitative statements [viii] [ix] [x] [xi] . These requirements determine whether numbers support classification, ordering, differences or proportional comparisons.

Under Stevens' classification, nominal and ordinal scales do not establish equal quantitative intervals [xii] . Interval scales support meaningful differences but lack a non-arbitrary zero, while ratio scales support meaningful proportions. For interval measurement, admissible transformations take the form $\mathbf{x' = ax + b}$; for ratio measurement, $\mathbf{x' = ax}$, where a is positive. The distinction is decisive when a numerical coefficient is intended to represent a proportion of another measured quantity.

Dimensional homogeneity and commensurability impose further requirements. The TTO coefficient $\mathbf{h = t/T}$ is a lawful dimensionless ratio because both quantities measure duration. Its ratio properties belong to time space. They do not establish a corresponding ratio measure of preference or a lawful mapping from preference space to duration.

The EQ-5D-3L regression algorithm must therefore be assessed against a specific standard: whether its dependent variable and resulting numerical predictions represent a defined empirical attribute with demonstrated scale properties. Statistical estimation cannot establish those properties merely by producing coefficients or fitted values.

MEASUREMENT AS THE FOUNDATION OF QUANTITATIVE SCIENCE

It is important to emphasize that measurement is not an optional preliminary to quantitative analysis. It establishes the scientific conditions under which numerical statements can be formulated, tested, replicated and potentially falsified. A quantitative hypothesis cannot simply assert a numerical relationship between assigned scores. It must identify the empirical attributes concerned, establish their measurement properties and specify mathematical operations admissible for those attributes. Representational measurement provides these standards through empirical axioms, representation and uniqueness conditions, scale properties, dimensional homogeneity and commensurability. Stevens' distinction between nominal, ordinal, interval and ratio scales is fundamental: different measurement structures authorize different quantitative claims. The ability to perform an arithmetic operation does not establish that the operation has empirical meaning**.**

These requirements apply before statistical analysis begins. A hypothesis that treatment produces a measurable change must identify what attribute changes, how its possession or magnitude is measured, and what numerical comparison constitutes evidence of that change. The measurement structure determines whether differences, proportions, rates or other quantitative relationships can legitimately be hypothesized. Statistical estimation can then assess empirical relationships among the resulting observations. It cannot retrospectively establish the measurement properties of numbers supplied without the necessary empirical justification. **Hypotheses must be measurement-qualified before they can be statistically tested.** Otherwise, numerical precision, statistical significance and model fit may describe relationships among assigned numbers without establishing any corresponding quantitative relationship among empirical attributes.

This is why the abandonment of representational measurement is devastating for conventional HTA. Its principal quantitative constructs—health-state utilities, QALYs and incremental cost-effectiveness ratios—depend upon numerical operations whose empirical legitimacy must be demonstrated, not assumed. The central failure is not that the resulting estimates are imprecise or that the models require further refinement. It is that their foundational quantitative claims lack the measurement justification necessary to interpret them as lawful measures of therapeutic benefit. Without that justification, the reference-case framework cannot establish that its calculations represent the quantities they purport to evaluate. A discipline cannot claim quantitative scientific authority merely because it generates numbers.

Dolan's EQ-5D-3L regression algorithm illustrates the consequences. The dependent variable is accepted as health-state utility without establishing a measured preference attribute. Categorical health-state descriptions become dummy-variable predictors, estimated coefficients become decrements from unity, and the resulting algorithm generates numerical values for 243 descriptions. These values are subsequently treated as proportional adjustments to survival duration. Yet no regression coefficient, goodness-of-fit statistic or predicted utility establishes the measurement properties required for that interpretation. The model answers a statistical question

about the reproduction of assigned valuations while leaving the scientific question of what is measured unanswered.

The consequence is not merely a weakness in the Dolan algorithm but a foundational failure of quantitative HTA**.** Its hypotheses, estimates and economic comparisons proceed as though the requisite measurement structures had already been demonstrated. Once this assumption is accepted, the production of further coefficients, utilities, QALYs and cost-effectiveness ratios becomes an exercise in numerical storytelling. Arithmetic replaces empirical measurement as the organizing principle of evaluation. The result is a quantitative framework whose calculations may be mathematically reproducible but whose claimed empirical meanings remain unjustified. This is the defining failure examined in numbers without measurement.

THE DOLAN MODEL: THE DEPENDENT VARIABLE

Dolan's (1997) EQ-5D-3L regression analysis employed 35,964 individual TTO-derived numerical observations, comprising 12 valuations from each of 2,997 respondents with complete data. These valuations covered 42 selected EQ-5D-3L health-state descriptions, drawn from the 243 possible combinations of five dimensions, each with three ordered categorical levels. The regression was intended to estimate numerical weights for the complete descriptive system, including states not directly valued by respondents. The 42 descriptions were not measurements of 42 empirical attributes, nor were the 35,964 observations measurements of a common latent attribute. They were numerical judgments concerning hypothetical multidimensional health-state descriptions.

The observations were obtained through interviewer-administered TTO exercises rather than a questionnaire measuring respondents' possession of the five EQ-5D-3L attributes. Respondents were presented with health-state descriptions and guided through comparisons involving survival duration and full health. They were not required to have experienced the health states being evaluated. Their responses therefore represented judgments about hypothetical composite health circumstances, not measurements of independently specified manifest or latent attributes. Describing these judgments as assessments of health status or health-state preferences does not change their empirical character. The object of valuation remained a multidimensional description, not a defined single attribute possessing demonstrated measurement properties.

The dependent variable is consequently a dog's breakfast of incompatible numerical constructions. For states considered better than death, the original TTO coefficient, $h = t/T$, is a legitimate ratio of survival durations. It satisfies dimensional homogeneity because both quantities belong to time space, [T]/[T], and commensurability because both measure the same attribute. Its ratio properties are properties of duration measurement, not preference measurement. For states considered worse than death, modified TTO procedures introduce negative scores that cannot arise from the original non-negative duration ratio. Both sets of numbers are nevertheless treated as observations of a common quantity designated health-state utility, without demonstrating that they possess a common empirical measurement structure.

A further distinction is necessary concerning the actual regression specification. Dolan expressed the dependent variable as $1 - S$, where S represented the assigned TTO health-state valuation. The

regression therefore modelled numerical decrements from the full-health reference value rather than the original valuation directly. This arithmetic transformation did not alter the measurement status of the underlying observations. Subtracting an assigned valuation from unity cannot establish equal measurement units, a non-arbitrary zero or a common latent preference attribute. The transformation merely changes the numerical expression of the original scores. The regression coefficients subsequently estimated from these transformed observations cannot acquire measurement properties absent from the dependent variable itself.

The fundamental error is the unsupported transformation from a manifest ratio measure of duration to an alleged measure of preference for a multidimensional health-state description. Mobility, self-care, usual activities, pain/discomfort and anxiety/depression are distinct descriptive domains, not components of a demonstrated single quantitative attribute. Their categorical levels do not independently establish measured quantities of those domains, let alone a common latent preference continuum. There is no demonstrated empirical measurement structure through which a lawful duration ratio becomes a quantity of preference for this composite description. The negative valuations compound the problem by introducing numbers that cannot even retain the original interpretation as non-negative duration proportions.

The sample size does not resolve this foundational failure. The 35,964 observations provide extensive statistical information about respondents' numerical valuations of the selected descriptions, but increasing the number of observations cannot establish the measurement properties of what is being observed. Dolan employed a generalized least-squares random-effects approach to accommodate repeated valuations by respondents. Such procedures address statistical estimation and the dependence among observations; they do not establish representation or uniqueness conditions, identify a lawful preference scale or demonstrate commensurability between preference and duration.

The entire construction rests upon an unsupported measurement assumption: that TTO-derived numerical valuations, including differently constructed negative scores, can be treated as observations of a common preference quantity and transformed into regression decrements with legitimate quantitative meaning. No empirical measurement axioms establish that interpretation, and no lawful mapping from preference space to time space has been demonstrated. Dolan's regression does not resolve these deficiencies. It accepts the numerical observations, transforms them into decrements, estimates coefficients for categorical explanatory variables and proceeds to populate the 243 health-state descriptions.

The statistical model therefore begins after the decisive measurement question has already been bypassed. What enters the regression is not an established measure of preference; what emerges cannot become one merely through statistical estimation. The algorithm generates numerical predictions, but neither the number of observations nor the sophistication of the regression can transform unsupported valuations into lawful measures. This is the foundational problem of the EQ-5D-3L valuation algorithm: numbers without measurement.

THE DOLAN MODEL: STRUCTURE AND INTERPRETATION

Dolan's (1997) EQ-5D-3L regression model was designed to estimate numerical values for all 243 possible health-state descriptions from TTO valuations of a selected subset. The statistical task was straightforward: identify relationships between assigned health-state valuations and the categorical levels defining each description, estimate coefficients for those relationships, and use the resulting equation to predict values for descriptions not directly assessed. The critical issue is the interpretation of those coefficients. They are not independent measures of therapeutic attributes but estimated numerical decrements from a conventionally assigned full-health value of unity.

The regression structure

The EQ-5D-3L comprises five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each has three ordered levels, with Level 1 representing no reported problems, Level 2 moderate problems and Level 3 extreme problems. The full-health description, 11111, is assigned the reference value of 1. The remaining descriptions represent combinations of reported problems across the five dimensions.

Dolan's regression employed the transformed dependent variable $\mathbf{1 - S}$, where S was the TTO-derived numerical valuation assigned to a health-state description. The transformation converted the original valuation into a numerical decrement from unity. Thus, a state assigned S = 0.70 became a dependent-variable observation of 0.30, while a state assigned S = −0.20 became an observation of 1.20. These are arithmetic differences from the reference value, not independently established measurements of lost health, diminished preference or therapeutic disadvantage.

The regression can be represented schematically as:

$$\mathbf{1 - S = \beta_0 + \Sigma \beta_j X_j + \varepsilon}$$

Here, β_0 is the intercept, X_j represents a categorical indicator identifying a specified EQ-5D dimension and level, β_j is its estimated regression coefficient, and ε represents the unexplained statistical variation. In Dolan's modelling, additional terms were considered to capture patterns of severity across dimensions. The widely used N3 specification included an indicator identifying descriptions containing at least one dimension at Level 3.

The coefficients as decrements from unity

The categorical explanatory variables were constructed through dummy coding. For each dimension, Level 1 served as the reference category. Separate indicators identified Level 2 and Level 3, producing ten principal dummy variables across the five dimensions. Each indicator took the value 1 when the corresponding problem level appeared in a health-state description and 0 otherwise.

The estimated coefficient associated with each indicator represented an additional predicted decrement from the full-health reference, conditional on the other terms in the model. A coefficient of 0.08 for mobility Level 2, for example, would contribute 0.08 to the predicted decrement when that category appeared. It would not establish that mobility had declined by eight percent, that preference had fallen by eight percent or that eight percent of a measured quantity of health had

been lost. The coefficient is a statistical contribution to an assigned numerical decrement, not a measure of change in an empirical attribute**.**

The intercept represents the model's predicted decrement when all included indicators equal zero. In a specification with an intercept, this term contributes to the predicted value even before any dimension-specific decrement is added. The N3 coefficient adds a further predicted decrement when at least one dimension is classified at Level 3. Neither the intercept nor the severity adjustment possesses an independent interpretation as a measured quantity of health or preference. Their numerical values arise from the regression specification and the dependent-variable observations.

Populating the 243 health-state descriptions

Once the coefficients have been estimated, the algorithm can generate a predicted valuation for any EQ-5D-3L description. The relevant dummy variables are identified, their coefficients are added to the intercept and any applicable severity adjustment, and the resulting total is subtracted from unity: Predicted utility = 1 − predicted total decrement

For illustration, consider a hypothetical state with a predicted intercept of 0.05, a mobility decrement of 0.08, a pain/discomfort decrement of 0.12 and an additional severity decrement of 0.10. The total predicted decrement is 0.35, giving an assigned utility of 0.65. These illustrative coefficients demonstrate the arithmetic; they are not Dolan's estimated parameter values.

The procedure can be repeated for every possible combination of dimension levels. The algorithm therefore extends numerical valuations from the directly assessed descriptions to the complete set of 243 classifications. This is its statistical achievement. It provides a systematic method for assigning predicted values without conducting a separate TTO exercise for every description.

What do the coefficients actually measure?

The interpretation of the resulting numbers is the decisive issue. The regression coefficients are estimated relationships between categorical descriptions and a dependent variable consisting of TTO-derived numerical valuations. Their interpretation as decrements depends entirely upon the prior acceptance of those valuations as quantities of health-state utility. Yet the dependent variable has not been established as a measure of a defined latent preference attribute, and the categorical predictors are not independent quantitative measures of mobility, pain, anxiety or the other descriptive domains.

The arithmetic of subtracting coefficients from unity therefore does not demonstrate that the resulting differences represent measured losses of health or preference. The full-health value of 1 is a conventional anchor; the estimated decrements are model-dependent numerical adjustments; and the resulting utility values are predictions on the assigned valuation scale. Their decimal precision does not establish equal empirical units, a non-arbitrary zero or meaningful proportional comparisons.

There is a further problem of interpretation. The model treats the contributions of different descriptive dimensions as numerically additive, subject to its specified adjustment terms. A decrement associated with mobility can therefore be added to one associated with pain/discomfort or anxiety/depression. This addition is legitimate as an operation within the fitted statistical equation, but it does not demonstrate that the underlying domains are commensurable quantities of a single measured attribute. Statistical additivity is not evidence of empirical commensurability.

The distinction between a regression coefficient and a measured attribute is consequently fundamental. A coefficient quantifies a fitted relationship within a statistical model. It does not establish the measurement properties of the dependent variable, the quantitative status of the explanatory categories or the empirical meaning of their numerical combination.

From predicted decrements to utility arithmetic

The final step is the interpretation of the predicted valuation as a utility coefficient suitable for adjusting survival duration. A predicted value of 0.65 is treated as though one year in the corresponding health state represents 0.65 quality-adjusted life-years. Yet the regression has established no preference ratio of 0.65, no measured quantity of health equivalent to 65% of full health and no lawful mapping from the composite valuation to a proportional adjustment of duration.

The problem is not that the regression cannot perform the intended numerical calculations. It can estimate coefficients, combine them and predict values for the 243 descriptions. The failure is the assumption that these calculations establish quantities possessing measurement properties that were never demonstrated. Dolan's algorithm converts assigned valuations into predicted decrements and predicted decrements into utility weights. It does not convert them into lawful measures of therapeutic benefit.

The result is the defining feature of numbers without measurement: an internally coherent arithmetic procedure whose outputs are treated as empirical quantities without the representation, uniqueness and scale properties required to justify that interpretation. The coefficients are decrements from unity because the model defines and estimates them as such, not because independently measured quantities of health or preference have been demonstrated to diminish by those amounts. Their subsequent incorporation into QALY calculations extends the original measurement failure rather than resolving it.

Every coefficient in Dolan's EQ-5D-3L regression model carries an implicit quantitative claim. A coefficient assigned to mobility Level 2, pain/discomfort Level 3 or any other dimension-level combination is interpreted as the numerical decrement in overall health-state utility associated with that classification. The coefficients are not merely descriptive labels. They are numerical quantities intended to be added to other decrements and subtracted from unity to produce a health-state utility weight. Each coefficient therefore requires its own measurement justification as a quantitative contribution to a common empirical attribute.

The regression provides no such guarantee. Its coefficients are estimates of conditional relationships between categorical explanatory variables and TTO-derived numerical observations.

Their magnitudes depend upon the assigned dependent-variable scores, the selected explanatory variables, the specification of additional terms and the statistical estimation procedure. A coefficient of 0.10 indicates a predicted contribution of 0.10 to the modelled decrement. It does not demonstrate that the corresponding dimension and level reduce a measured quantity of preference by 0.10 units. That conclusion would require independent evidence that overall preference is a defined quantitative attribute, that its scale possesses the necessary measurement properties and that the dimension-level classification corresponds to a measurable change in that attribute.

The problem extends to the combination of coefficients. The algorithm assumes that numerical decrements associated with different descriptive dimensions can be combined within a common additive structure, subject to its specified adjustment terms. Yet mobility, self-care, usual activities, pain/discomfort and anxiety/depression are distinct domains represented by ordered categorical levels. No empirical measurement demonstration establishes that their regression coefficients are commensurable quantities of a single latent attribute. Their numerical additivity is imposed through model specification rather than derived from demonstrated empirical measurement relationships. The fact that coefficients can be added statistically does not establish that they represent additive quantities of health-state preference.

Nor can the TTO origin of the dependent variable supply the missing justification. For states better than death, the TTO coefficient begins as a lawful ratio of survival durations, $h = t/T$. Its ratio properties belong to time measurement. They do not transfer automatically to the respondent's preference for a multidimensional health-state description. Dolan's regression treats the resulting numerical valuations as though they already measure overall utility and then estimates the contributions of the individual dimension-level indicators. This requires an unsupported transformation at two stages: first, from duration equivalence to quantitative preference; second, from the fitted statistical coefficients to measured contributions of the descriptive categories.

The decisive failure is that the model estimates quantities whose claimed empirical meanings have never been established. Each coefficient is presented as a utility decrement, but no measurement structure demonstrates that the decrement exists as a quantity of the attribute it purports to represent. Statistical significance, model fit and predictive accuracy cannot repair this absence. Dolan's algorithm can estimate coefficients and generate numerical predictions for all 243 health-state descriptions. It cannot, by that process, establish that those coefficients measure the impact of particular dimensions and levels on overall utility. The quantitative claims are assumed in the interpretation of the model, not demonstrated by its estimation.

THE DECREMENT ALGORITHM: WHAT DOES THE RESULTING SCORE MEASURE?

Dolan's EQ-5D-3L regression relates TTO-derived numerical valuations to categorical indicators for five health-state dimensions. Level 1 serves as the reference category, while separate indicators represent Levels 2 and 3. Additional terms, including N3, accommodate specified patterns of severe problems. The purpose is to estimate coefficients that can be combined to generate a predicted valuation for each of the 243 possible health-state descriptions.

The decrement model can be represented schematically as:

$$1 - S = C + M + SC + UA + PD + AD + N3 + \varepsilon$$

The estimated coefficients generate the predicted tariff:

$$U = 1 - (C + M + SC + UA + PD + AD + N3)$$

Here, C represents the intercept, the five dimension terms represent the applicable estimated categorical coefficients, and N3 represents an additional severity adjustment when required. The resulting value is calculated by summing the applicable coefficients and subtracting that sum from the full-health reference of unity. The entire algorithm therefore depends upon treating the estimated coefficients as additive numerical decrements in a common quantity designated health-state utility.

THE REQUIREMENTS FOR ADDITIVE MEASUREMENT

The fundamental question is whether these coefficients possess the measurement properties necessary to justify their interpretation as additive losses of preference. Regression can impose an additive mathematical structure, but it cannot establish that the empirical quantities supposedly represented by its coefficients are themselves additive. That requires a measurement justification independent of the statistical specification.

First, every coefficient must refer to a change in the same defined empirical attribute. Mobility, self-care, usual activities, pain/discomfort and anxiety/depression are different descriptive domains. Their categorical levels do not establish measured quantities of those domains, much less quantities of a common latent preference attribute. A coefficient associated with mobility Level 2 and another associated with pain/discomfort Level 3 can be added mathematically because they are numbers. Their addition as measured losses of preference requires evidence that both express changes in the same quantitative attribute.

Second, the coefficients must possess a common unit. If a mobility coefficient is 0.10 and a pain/discomfort coefficient is 0.10, the interpretation of both as equal preference decrements requires that equal numerical differences correspond to equal empirical differences. The regression does not demonstrate this equivalence. It estimates numerical relationships with a dependent variable whose own measurement status remains unestablished. Equal coefficients are equal numbers, but they are not thereby equal measured losses of preference.

Third, the additive structure itself requires empirical justification. The model assumes that the relevant dimension-level coefficients, together with the intercept and any N3 adjustment, can be summed to represent the total decrement associated with a health-state description. Yet no demonstrated measurement structure establishes that the empirical contributions of these dimensions are additive in this manner. The N3 term is a statistical adjustment for patterns associated with severe classifications; it is not proof that the resulting equation reproduces a lawful empirical composition of preference quantities. Statistical additivity cannot substitute for demonstrated empirical additivity.

FROM ADDITIVE DECREMENTS TO A PROPORTION OF PREFERENCE

Even if the coefficients were shown to represent additive differences on a common interval scale, that would not establish that the final tariff represented a proportion of preference. An interval scale permits meaningful differences but does not possess a non-arbitrary zero. Its numerical values cannot be interpreted as ratios of the underlying attribute.

The intended proportional interpretation requires something more demanding. If P represented a measured magnitude of preference for full health and p the corresponding magnitude for an impaired health state, then:

k = p/P = 0.6

would mean that the impaired health state possesses 60% of the preference quantity represented by full health. This interpretation requires a specified preference attribute, lawful ratio measurement of both quantities, a common unit and a non-arbitrary zero. It also requires empirical justification for interpreting the regression-generated score as that ratio.

Dolan's algorithm establishes none of these conditions. A predicted utility of 0.6 results from subtracting estimated decrements totaling 0.4 from the assigned reference value of 1. It does not follow that 40% of a measured preference quantity has been lost. The coefficients are not demonstrated preference quantities, their sum is not an established total loss of preference, and the remaining value is not a demonstrated proportion of preference. Subtraction from unity is an arithmetic convention, not a measurement theorem.

THE FAILURE OF THE QUANTITATIVE CLAIM: THE MISSING ABSOLUTE ZERO

The decisive failure of Dolan's decrement algorithm is the absence of demonstrated ratio measurement. The algorithm assigns full health the value of unity and death the value of zero, then estimates coefficients representing numerical decrements from the full-health anchor. These assignments do not establish a ratio scale. A reference value of 1 is not evidence of a measured preference magnitude, and an assigned zero for death is not an empirically established absolute zero of preference.

Ratio measurement requires a defined empirical attribute, equal measurement units and a non-arbitrary zero representing the absence of that attribute. Only under these conditions can one magnitude be interpreted as a proportion of another. If a lawful ratio measure of preference existed, a reference magnitude P could be normalized to unity, and a second magnitude p could be expressed as the ratio p/P. However, this construction presupposes the prior demonstration of ratio measurement. It cannot be created by assigning the numerical values 0 and 1 to selected health-state descriptions.

The contradiction becomes particularly apparent when the algorithm generates negative utilities, including values approaching −0.6. If a tariff of 0.6 is interpreted as 60% of a measured preference

magnitude, what does a tariff of −0.6 represent? It cannot represent minus 60% of a non-negative preference quantity. Nor can the conventional death anchor of zero serve as an absolute zero of that quantity while the algorithm assigns values below it. Negative values may be meaningful on an interval scale, but their presence is incompatible with interpreting these tariffs as proportions of a non-negative preference magnitude measured from an absolute zero at death. The algorithm cannot simultaneously treat zero as the absolute absence of preference and assign negative proportions of that same quantity.

Dolan's regression supplies no demonstration of the required measurement structure. Its coefficients are estimated numerical contributions to a predicted decrement from unity, not established measurements of lost preference. Their addition does not demonstrate empirical commensurability, and subtraction from the assigned full-health anchor cannot establish an absolute zero. A predicted utility of 0.6 therefore means only that the algorithm has generated a numerical decrement of 0.4 from its reference value. It does not establish that 40% of a measured preference quantity has been lost. A negative utility of −0.6 similarly represents a model-generated decrement of 1.6 from unity, not a lawful negative proportion of preference.

The missing absolute zero, compounded by negative tariff values, defeats the proportional interpretation**.** Without an empirically established ratio scale, the tariff cannot justify treating its scores as proportions of preference. Even if such a scale were demonstrated, a separate empirical mapping would still be required to justify applying those proportions to survival duration. The algorithm establishes neither condition. Its outputs remain regression-generated numerical assignments rather than lawful proportional measures.

THE DISCONNECTION FROM TIME MEASUREMENT

The problem becomes clearer when the regression-predicted tariff is compared with the original TTO coefficient. For states considered better than death, an observed TTO value of 0.6 may originate from a comparison of six years with ten years. The ratio $h = t/T$ is a lawful dimensionless proportion of survival duration, satisfying dimensional homogeneity and commensurability within time space.

A regression-predicted tariff of 0.6 has a different origin. It is produced by combining estimated coefficients for categorical health-state descriptions and subtracting their sum from unity. It need not correspond to any directly observed six-to-ten-year duration comparison. More importantly, neither the observed TTO ratio nor the predicted tariff establishes a ratio of preference magnitudes. The lawful measurement properties of survival duration cannot simply be transferred to a composite judgment about health-state descriptions.

The algorithm therefore introduces successive unsupported quantitative interpretations. A duration trade-off is accepted as a preference valuation; the valuation is treated as a measured dependent variable; categorical regression coefficients are interpreted as additive losses of preference; their sum is treated as a total decrement; and the resulting tariff is designated a proportional utility weight suitable for adjusting survival duration.

COULD THE EQ-5D UTILITY SCORE BE ORDINAL?

The least demanding interpretation of the EQ-5D utility score is that it represents an ordinal ordering of health-state preferences. On this interpretation, a state assigned a score of 0.8 would be preferred to one assigned 0.6, provided that the predicted ordering corresponded to an empirically established preference relationship. The numerical difference of 0.2 would have no independent quantitative meaning. Nor could the scores be interpreted as proportions of preference.

Even this ordinal interpretation requires justification. The regression assigns numerical scores to multidimensional categorical health-state descriptions, but statistical prediction does not independently establish that the resulting ordering faithfully represents a specified empirical preference relation.

An ordinal interpretation would therefore substantially restrict the claims that could be made for the EQ-5D algorithm. It could potentially support the ranking of health states, but not quantitative differences in preference or proportional assessments of therapy impact. Those interpretations require additional measurement properties that cannot be inferred from the numerical output of the regression.

The critical question is not whether the EQ-5D algorithm can generate an ordering. It plainly can. The question is whether that ordering has been empirically validated as a representation of preference, and whether any additional measurement properties have been demonstrated.

THE ADDITIONAL FALLACY OF NORMALIZATION

The measurement problem is compounded when health-state utility scores are normalized or restricted to the interval from zero to unity. Such a transformation may create the appearance of a proportion, but numerical bounds do not establish ratio measurement. A score of 0.6 on a normalized 0–1 scale does not demonstrate possession of 60% of a measured preference attribute. Its interpretation depends upon the original measurement properties, the meaning of zero and the transformation employed. Where negative tariff values are truncated or rescaled into this interval, the procedure additionally changes the original numerical relationships and may eliminate the distinction between states assigned negative and non-negative valuations. Where the conventional death anchor remains zero, the original negative values contradict its interpretation as the absolute absence of a non-negative preference magnitude. Normalization cannot manufacture an absolute zero, establish equal empirical units or convert ordinal or interval representations into ratio measures. It merely reassigns numbers according to a specified rule. The resulting 0–1 scale may be convenient for calculation, but convenience provides no justification for interpreting its values as proportional quantities suitable for multiplication by survival duration.

THE IMPOSSIBLE QALY

The failure of Dolan's EQ-5D-3L decrement algorithm to establish ratio measurement has consequences extending far beyond the interpretation of regression coefficients. It destroys the measurement justification for the quality-adjusted life-year (QALY), the central quantitative construct of conventional health technology assessment. The QALY is conventionally expressed as:

$$Q = U \times T$$

where U is a health-state utility weight and T is survival duration. The equation assumes that U is a lawful proportional adjustment factor that can be multiplied by duration to produce quality-adjusted time. Yet this interpretation requires two independent measurement conditions, neither of which has been demonstrated. First, the utility weight must represent a legitimate proportion of a measured preference attribute. Second, that proportion must be shown to provide a lawful adjustment to duration in time space. Dolan's algorithm satisfies neither requirement.

The first impossibility: no proportion in preference space

The decrement algorithm generates utility scores by summing regression coefficients and subtracting the total from an assigned full-health reference of unity. These coefficients are statistical estimates associated with categorical health-state descriptions, not established measurements of changes in a common preference attribute. Their addition does not demonstrate empirical commensurability, and subtraction from unity cannot establish ratio measurement.

A legitimate preference proportion would require two measured magnitudes of the same attribute, p and P, such that:

$$k = p/P$$

The ratio would be meaningful only if preference were defined as an empirical attribute possessing equal units and a non-arbitrary zero. Neither the TTO valuation exercise nor Dolan's regression establishes such an attribute or its ratio-scale properties. Assigning full health the value 1 and death the value 0 cannot create an absolute zero of preference. Negative tariff values further contradict the interpretation of these scores as non-negative proportions measured from that zero. Subsequent normalization to a 0–1 range cannot repair the deficiency.

Consequently, a utility score of 0.6 does not demonstrate that an impaired health state possesses 60% of a measured preference quantity. There is no established preference proportion available to serve as the multiplier U in the QALY equation. The first requirement therefore fails before duration enters the calculation.

The second impossibility: no lawful mapping to time space

Even if, contrary to the available measurement evidence, a lawful preference ratio p/P were granted, the QALY would face a second independent obstacle. A proportion measured in preference space does not automatically become a proportional adjustment to survival duration.

Duration is a manifest ratio measure with a non-arbitrary zero and equal units. A lawful duration proportion can be expressed as:

$$r = T_1/T_2$$

Both quantities belong to time space, satisfying dimensional homogeneity, [T]/[T], and commensurability because they measure the same attribute. By contrast, p/P would belong to preference space. Although both ratios would be dimensionless, dimensional cancellation alone would not establish that they represent equivalent empirical relationships.

The QALY requires a demonstrated mapping from preference proportions to duration proportions:

$p/P \rightarrow T_1/T_2$

No such mapping has been established. The fact that two ratios are dimensionless does not authorize their substitution across different empirical attributes. Nor does a TTO indifference judgment establish the required equivalence. It records a respondent's choice between alternatives involving survival duration and multidimensional health-state descriptions; it does not demonstrate that a preference proportion is identical to a duration proportion.

CONCLUSION: THE CONSEQUENCES OF ABANDONING MEASUREMENT`

The failure of Dolan's EQ-5D-3L decrement algorithm is the predictable consequence of a foundational error: the decision to value multidimensional health-state descriptions rather than identify and measure defined empirical attributes. This choice placed conventional HTA outside the requirements of representational measurement from the outset. Instead of establishing attributes, measurement axioms, scale properties and admissible arithmetic, the discipline constructed numerical procedures around hypothetical descriptions of health states. The resulting quantities acquired the language of measurement without its scientific foundations.

The progression from Torrance to Dolan reveals a succession of forced assumptions. The TTO began with a legitimate manifest ratio of survival durations, which was reinterpreted as a measure of health-state preference. Torrance's subsequent attempt to justify conditional interval utility could not establish the required ratio properties. Dolan then accepted these numerical valuations as a dependent variable, estimated coefficients for categorical dimension-level descriptions, and combined them into an algorithm producing utility decrements from unity. At no stage was a lawful preference attribute demonstrated. Each successive procedure assumed the measurement legitimacy that its predecessor had failed to establish.

The resulting utility weights possess no demonstrated ratio-scale properties, no empirically established absolute zero and no justified interpretation as proportions of preference. Their use as multipliers of survival duration in QALY calculations introduces the additional, unresolved requirement of mapping between preference and time spaces.

Legacy HTA has therefore made a rod for its own back. By setting aside representational measurement and its focus on attributes, it created a framework dependent upon increasingly elaborate statistical and arithmetic constructions that cannot repair the original error. The unfortunate consequence has been the global acceptance and institutionalization of this measurement debacle, transmitted through HTA methods, textbooks, professional training and

reimbursement decision-making. Its culmination is the conventional cost-effectiveness reference case, which incorporates unsupported utility weights, impossible QALY arithmetic and model-generated claims of comparative value into an internationally accepted decision framework.

The impossible reference case will be the subject of the next paper. The necessary alternative is not another valuation algorithm but the reconstruction of quantitative HTA around defined attributes, lawful measurement and empirically testable hypotheses. Measurement must precede arithmetic.

MEASUREMENT MUST PRECEDE ARITHMETIC.

There is no scientifically defensible alternative to representational measurement as the foundation for quantitative health technology assessment. Reconstructing HTA requires abandoning the legacy dependence upon multidimensional health-state preferences and returning to independently defined attributes, lawful measurement scales, admissible arithmetic and empirically evaluable claims. This cannot be achieved through marginal revisions to utility algorithms or reference-case guidance. It requires a fundamental reconstruction of the knowledge base taught to faculty, researchers and students. Representational measurement is not one methodological option among many; it establishes the scientific requirements against which quantitative claims must be judged.

Maimon Research has developed a comprehensive nine-unit distance-learning curriculum comprising approximately 60,000 words of instructional material to support this reconstruction [xiii]The program begins with the foundations of representational measurement and the identification of empirical attributes, then develops the distinction between manifest and latent measurement, the properties of measurement scales, dimensional homogeneity, commensurability and admissible arithmetic. It addresses the construction and validation of linear ratio measures for manifest attributes and Rasch-based measurement for latent attributes, including requirements for unidimensionality, invariance and measurement fit. Subsequent units examine the formulation of therapy-impact claims, the measurement of possession and change, treatment compliance and persistence, resource utilization, prospective evaluation protocols, replication and falsification. The objective is not to replace one numerical convention with another, but to establish a coherent scientific framework in which every quantitative claim is supported by an explicitly defined attribute, a validated measure and an admissible mathematical interpretation. The program provides a practical foundation for reconstructing graduate HTA education, professional training and research assessment.

The curriculum is available now through Maimon Research, with structured instructional materials and continuing academic support. Participants can work through the nine units through distance learning, with opportunities for questions, clarification and discussion of measurement problems arising in their own research or teaching. Maimon Research provides continuing question-and-answer support, together with optional seminars and guided discussions to assist faculty, researchers and institutions in applying the framework. The program can support individual professional development, faculty preparation, graduate teaching and institutional curriculum reconstruction. Its significance is that the transition to lawful measurement does not depend upon waiting for established HTA organizations to revise their guidance or for another generation of methodological debate to conclude. The educational foundation and practical support already exist.

The responsibility now rests with universities, professional organizations and HTA institutions to replace the inherited curriculum of numerical preference modelling with a curriculum grounded in the measurement of empirical attributes. The reconstruction can begin immediately.

ACKNOWLEDGEMENT

I acknowledge that I have used OpenAI technologies, including the large language model, to assist in the development of this work. All final decisions, interpretations, and responsibilities for the content rest solely with me

ABOUT THE AUTHOR

Dr Paul Langley is a health economist. Originally from the UK, Dr Langley completed his graduate studies in Canada and has taught in the UK, Canada, Australia and the United States. After more than 20 years teaching at university in Australia, Dr Langley moved to take up academic positions in the United States. He now lives in Tucson, Arizona, to enjoy the Catalina Mountains, the Sonoran Desert, the absence of mosquitoes and the swimming pool.

REFERENCES

[i] Langley P. The TTO: A Failure of Representational Measurement. Logit Working Paper No 2041.

[ii] Torrance, G. W. (1976). Toward a utility theory foundation for health status index models. *Health Services Research*, 11(4), 349–369

[iii] Drummond M, Sculpher M, Claxton K et al. (2015) Methods for the Economic Evaluation of Health Care Programmes (4th Ed). New York: Oxford University Press

[iv] EuroQol Research Foundation. EQ-5D-3L: Frequently Asked Questions. Rotterdam, The Netherlands: EuroQol Research Foundation; n.d. https://euroqol.org/faq-eq-5d-3l/

[v] Torrance, G. W., Thomas, W. H., & Sackett, D. L. (1972). A utility maximization model for evaluation of health care programs. *Health Services Research, 7*(2), 118–133

[vi] Dolan P. (1997)Modeling valuations for EuroQol health states. *Medical Care*. 35(11):1095–110

[vii] Bond T, Zi Yan, Heene m.(2021) Applying the Rasch Model: Fundamental measurement in the Human Sciences (4th Ed). New York: Routledge

[viii] Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of Measurement, Volume I: Additive and Polynomial Representations*. New York: Academic Press

[ix] Suppes, P., Krantz, D. H., Luce, R. D., & Tversky, A. (1989). Foundations of Measurement, Volume II: Geometrical, Threshold, and Probabilistic Representations. New York: Academic Press

[x] Luce, R. D., Krantz, D. H., Suppes, P., & Tversky, A. (1990). Foundations of Measurement, Volume III: Representation, Axiomatization, and Invariance. San Diego: Academic Press

[xi] Luce, R. D., & Suppes, P. (2002). Representational measurement theory. In H. Pashler & J. Wixted (Eds.), *Stevens' Handbook of Experimental Psychology* (3rd ed., Vol. 4: *Methodology in Experimental Psychology*, pp. 1–41). New York: John Wiley & Sons.

[xii] Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103(2684), 677–680

[xiii] Maimon Research. *Reconstructing HTA Programs: Measurement before Arithmetic*. https://maimonresearch.com/hta-reconstruction-program-and-fees/