



Interpretation for scales of measurement linking with abstract algebra

journal or publication title	Journal of clinical bioinformatics
number	4
page range	9
year	2014-06-10
URL	http://hdl.handle.net/10470/30828

doi: 10.1186/2043-9113-4-9

RESEARCH

Open Access

Interpretation for scales of measurement linking with abstract algebra

Jitsuki Sawamura^{1*}, Shigeru Morishita² and Jun Ishigooka¹

Abstract

The Stevens classification of levels of measurement involves four types of scale: "Nominal", "Ordinal", "Interval" and "Ratio". This classification has been used widely in medical fields and has accomplished an important role in composition and interpretation of scale. With this classification, levels of measurements appear organized and validated. However, a group theory-like systematization beckons as an alternative because of its logical consistency and unexceptional applicability in the natural sciences but which may offer great advantages in clinical medicine. According to this viewpoint, the Stevens classification is reformulated within an abstract algebra-like scheme; 'Abelian modulo additive group' for "Ordinal scale" accompanied with 'zero', 'Abelian additive group' for "Interval scale", and 'field' for "Ratio scale". Furthermore, a vector-like display arranges a mixture of schemes describing the assessment of patient states. With this vector-like notation, data-mining and data-set combination is possible on a higher abstract structure level based upon a hierarchical-cluster form. Using simple examples, we show that operations acting on the corresponding mixed schemes of this display allow for a sophisticated means of classifying, updating, monitoring, and prognosis, where better data mining/data usage and efficacy is expected.

Keywords: Scales on measurement, Stevens classification, Interpretation, Abstract algebra, Data-mining, Hierarchical cluster, Clinical medicine

Background

In 1946, S. S. Stevens devised his classification of "levels of measurement" [1], which subsequently has been used widely and has accomplished an important role in composition and interpretation of scales in medical fields. The systematics of levels of measurement seems to have been organized and validated by virtue of this classification. Nevertheless, we believe that an abstract algebra-like interpretation/systematization awaits introduction because of its logical consistency and unexceptional applicability in describing patterns and processes. We conjecture that it offers benefits in clinical medicine, especially, with respect to scales of measurement [2,3].

Thus, in the following, we re-interpret Stevens classification, and endeavour to give it meaning in some abstract algebra-like modelling. There, the most preferred construct is a vector-like structure of various sets of scores based on individual scales and operators that permit changes of score within the set. Additionally, classical datasets that are

classified in terms of the Stevens scales of measurement can be mined and combined on a higher abstract structure level based upon a hierarchical-cluster form. To explore this possibility, we provide simple examples to help readers understand this modelling tool.

S1. Application of group/field of abstract algebra to the various types of scales

Stevens classified the scales of measurement into four scale types [1]; I) "Nominal scale" that uses only labels or numbers (e.g., numbering of football players, blood type, nationality); II) "Ordinal scale" that introduces equality, rank-ordering (e.g., hardness of minerals, grading for efficacy of clinical treatment); III) "Interval scale" that is based on equally quantitative intervals (e.g., temperature as read in centigrade, duration, frequency); and IV) "Ratio scale" that assumes a 'zero' as an origin, equality, rank-order, equality of intervals, and equality of ratios (e.g., absolute temperature, speed of vehicles, and most physical values) that then admit manipulations using the four arithmetic operations.

* Correspondence: sawamura.jitsuki@twmu.ac.jp

¹Department of Psychiatry, Tokyo Women's Medical University, Tokyo, Japan
Full list of author information is available at the end of the article

For I), the “Nominal scale”, there seems to be little room where group theoretical operations apply because within that scale only a labelling scheme is permissible. Although some non-cyclic group might be definable, it seems that little meaning can be attached to operations for this sort of scale.

For II), the “Ordinal scale”, a ranking is realised by introducing a set with an N-graded scoring like ‘1, 2, 3,..., N - 1, N’ (N: positive integer) for a score deficient in (or with no absolute need for) a quantitative character, but not requiring a ‘0’ score according to the Stevens classification. Historically, the “graphic rating scale”, a grading from I to V, was proposed by Hayes and Patterson in 1921 [4], and Freyd in 1923 [5]. However, here, we envisage either operations that decrease the score by ‘1’ in an N-graded graphic scale necessitating a ‘0’, so that {0, 1, 2, 3,..., N - 2, N - 1} establishes the scoring scale, or simply adding the score ‘0’ as in {0, 1, 2, 3,..., N - 2, N - 1, N}. We focus on the former type. Then, for an arbitrary non-negative integer X, the operation giving the remainder of X after division by N, written $X \pmod N$, defines the cyclic group $Z_N = \{0, 1, 2, 3, \dots, N - 2, N - 1\}$, where modulo N addition is postulated. With this assumption, given two elements ‘ X_j ’ and ‘ X_k ’ ($X_j, X_k \in Z_N$) corresponding for example to the severity of a clinical symptom and/or finding, then composition (denoted by ‘*’) is taken to be modulo N addition; ‘ $X_j * X_{(j \rightarrow k)} = X_k$ ’ (with $X_{(j \rightarrow k)} \in Z_N$). Here ‘ $X_{(j \rightarrow k)}$ ’ is an operator that produces the change in score, ‘ $X_j \rightarrow X_k$ ’ (formally we have ‘ $X_{(j \rightarrow k)} = X_j^{-1} * X_k = X_k - X_j$ ’). Then, all scores ‘ X_j ’ and operators ‘ $X_{(j \rightarrow k)}$ ’ are composable within a single Abelian modulo additive group ‘ Z_N ’, where ‘ $X_j * X_k = X_k * X_j$ ’ holds, at least, in terms of operation ‘*’. Thus a patient’s state corresponding to a certain illness or disease can be changed through the application of a single operation determined by the two elements belonging to ‘ Z_N ’ [6,7] representing the previous and current state of the patient. A simple example is presented in Appendix A.

If a state of maximum severity is present, then the antithesis for any given disease Y is the ideal healthy state $E_Y = [0|0|0|0|0|\dots]$, the combination of all scores being ‘0’ and represented by the identity element for group $Y = \{Z_N^{*n}, *\}$. Here, Y is the n-fold Cartesian product of ‘ Z_N ’ (n: the number of components) that comprises all possible assessments related to each state of a given disease, for instance, ‘hypertension’, ‘hyperglycaemia’, ‘diabetes mellitus’, ‘acute pancreatitis’, ‘systemic lupus erythematosus’, and ‘cerebral artery stroke’. If in addition composition is given by modulo ‘N’ arithmetic, prime numbers (e.g., $N = 7$) are preferable [8,9] and considerable parts of components could be overlapping among individual diseases as was mentioned in our previous reports [6,7]. Note that, in practice, equal

increments within a grading scheme are not always postulated. Nevertheless, the scale represented by this Abelian modulo additive group ‘ Z_N^{*n} ’ will be called a “modular scale”. However, it may be an atypical case (partially weakened example) of a “Ratio scale” (type IV) without the strict requirement for equal calibration. Indeed, there are such scales because, like the ‘TNM classification (with a ‘T0’ entry) for malignant tumours’ [7,10], grades for scoring are determined for example according to histological characteristics, selection of treatment, and prognosis, having no strict linearity in scale, but which might be regarded as an “modular scale”. Based upon these results, for instance, the following are considered composable; Abelian modulo additive group $Y_1 = \{Z_7, *\}$ for ‘hypertension’, $Y_2 = \{Z_7, *\}$ for ‘hyperglycaemia’, $Y_3 = \{Z_7, *\}$ for ‘diabetes mellitus’, Y_4 for ‘acute pancreatitis’, Y_5 for ‘systemic lupus erythematosus’, Y_6 for ‘cerebral artery stroke’, $Y_{all} = \{Z_7 \times Z_7 \times Z_7 \times \dots, *\} = \{Z_7^{*n}, *\}$ (n: the number of components) for an entire body, and $Y_7 = \{Z_8 \times Z_4 \times Z_2 \times Z_2, *\}$ for the ‘TNM classification (with a ‘T0’ entry) for malignant tumours’ [7,10]. Additionally, these are treatable without exception within the abstract algebraic theory. For this case, an equal calibration for severity may have unbeneficial outcomes if used in clinical treatments. However, for ‘delirium’, ‘chronic liver dysfunction’, ‘acute pancreatitis’, and ‘diabetes mellitus’, for example, total scores based on equal calibration are desirable to assess disease severity.

For III), the “Interval scale”, differences in quantities are allowed. An example is ‘periods of time’ or ‘duration’, which, although can be measured with ratio scales, enables one period to be double another when compared. The same is true of ‘temperature’. If parameters ‘ X_j ’ and ‘ $X_l \in \mathbb{R}$ ’ (the continuous real number line) have ranges

$$-\infty < X < +\infty \tag{i}$$

we can consider an operator ‘ X_k ’ that causes changes from ‘ X_j ’ to ‘ X_l ’, and introduce a binary operation, denoted ‘ \circ ’, where ordinal addition and its inverse, subtraction, are assumed;

$$X_j \circ X_k = X_j + X_k = X_l \text{ (j, k, l; session numbers)} \tag{ii}$$

In this regard, as for ‘ X_j ’, it can also be expressed as a sum of an integer part and a decimal part,

$$X_j = 1m_j + c_j \tag{iii}$$

($m_j = [X_j]$, $c_j = X_j - [X_j]$, ‘ $0 \leq c_j < 1$ ’; ‘ $[X]$ ’ is the floor function meaning the highest integer below ‘X’). Similarly,

$$X_k = 1m_k + c_k \text{ (} m_k = [X_k], c_k = X_k - [X_k], '0 \leq c_k < 1') \tag{iv}$$

$$X_i = 1m_i + c_i \quad (m_i = [X_i], c_i = X_i - [X_i], '0 \leq c_i < 1') \quad (v)$$

'1' is a 'unit length' of the respective values. Thus, (iii) – (v) can be redefined using the unit length '1' as an interval scale,

$$\begin{aligned} X_j \circ X_k &= (1m_j + c_j) + (1m_k + c_k) \\ &= 1(m_j + m_k) + (c_j + c_k) = 1m_l + c_l \\ &= X_l \end{aligned} \quad (vi)$$

There exists an identity element ' X_0 ' ($=0$) that satisfies ' $X_j \circ X_0 = X_0 \circ X_j (=X_j + X_0 = X_j + 0) = X_j$ '. Additionally, the inverse element is ' $X_j^{-1} = -X_j$ ' satisfying ' $X_j^{-1} \circ X_j = X_j \circ X_j^{-1} = X_j + X_j^{-1} = X_j - X_j = X_0 (=0)$ '.

Naturally, commutativity and associativity are satisfied. Let U be the set that comprises all ' X_j 's, i.e., $U \equiv \{X_j \mid X_j \in \mathbb{R}\}$. Because ' $X_j, X_k, X_j \in$ set U , the closure law holds. Therefore, this operation defines a group $U = \{X_j, \circ\}$ [2,3]. "Body temperature readings", "clock time for the onset of sleep within a day" and "clock time for the onset of drip infusion within a day" are definable in this scale. Examples of the first two are provided in Appendix B. By making use of this procedure, the differences between quantitative values and operators are eliminated, and both can be regarded as elements belonging to a single group U . Moreover, a collection of additive Abelian groups $U_1 \equiv \{X_{1j} \mid X_{1j} \in \mathbb{R} \text{ (deg C)}\}$ based on an individual's clinical values can be described as, as for example $U_1 = \{X_{1j}, \circ\}$ for "body temperature readings", and $U_2 \equiv \{X_{2j} \mid X_{2j} \in \mathbb{R} \text{ (/24 hrs)}\}$ and $U_{2j} = \{X_{2j}, \circ\}$ for "clock time for the onset of sleep within a day", $U_3 \equiv \{X_{3j} \mid X_{3j} \in \mathbb{R} \text{ (/24 hrs)}\}$ and $U_{3j} = \{X_{3j}, \circ\}$ for "clock time for the onset of drip infusion within a day", ..., $U_N = \{X_{Nj}, \circ\}$, ..., (N : natural number). Those are considered readily treatable and recordable within an abstract algebraic context.

For IV), the "Ratio scale", the 'administration of medicine (with strict dosage regimes)' and 'International Statistical Classification and Health Related Problems' [11] were given as examples in our previous report [6,7]. Essentially, for this scale, because the four arithmetic operations are possible, 'rings' and 'fields' in abstract algebra are applicable so long as composition is given by modulo ' N ' arithmetic with ' N ' a prime. Although there could be scope where the four modulo arithmetic operations (denoted by '+' in ' $X_j + X_k = X_l$ ') are applicable in assessment scoring in clinical medicine, it might be preferable at this stage to confine the application of ratio scales to just modulo N addition '*' collectively for '+', similar in manner as established in Appendix A. For the example given in Appendix A, the difference in interpretation is the presence/absence of an equal calibration.

Whereas the scale of 'TNM classification for malignant tumours' [10] was regarded as an example of an "Ordinal scale", some of the scales defined as "Ratio scales" at initial glance should be regarded as "Ordinal scales" accompanied with '0'. It might be contentious whether clinical assessments performed using superficial scales based on the four arithmetic operations could have sufficient validity in clinical treatments or clinical research.

Nevertheless, other clinical scales range over a semi-open continuous interval like ' $0 \leq X < +\infty$ ' (X : real number), such as 'blood concentration of white blood cells: [WBC] (/mm³)', and 'administration of a certain drug like lithium carbonate: [Li⁺] (mEq/l), sodium: [Na⁺] (mEq/l), calcium: [Ca⁺⁺] (mg/dl), chloride: [Cl⁻] (mEq/l) and bicarbonate: [HCO₃⁻] (mEq/l)'. Also, there are clinical scales whose ranges are the open interval like ' $-\infty < X < +\infty$ ' (X : real number); 'Anion gap [AG] = [Na⁺] - ([Cl⁻] + [HCO₃⁻])' (reference range for blood tests: 12 ± 2 mEq/l) and 'Base excess [BE]' (reference range for blood tests: 0 ± 2 mmol/l). However, both can be treated using the notion of 'field' because those values are real numbers where all four arithmetic operations are included, with the exception of division by zero. Thus, the above clinical values could be definable over a 'field'. In this regard, we assume a rule that each unit like 'mEq/l' accompanies the value automatically with the results of operations regardless of types of operation among the four arithmetic operations (Note that there are cases when units vanish as when ratios are taken 'mEq/mEq (unitless)' or displayed in reciprocal form like '1/mEq'). Examples for '[WBC] (/mm³)', '[Na⁺] (mEq/l)' are presented in Appendix C.

In this case, we consider a set V and assume that '#' means one of 'addition, subtraction, multiplication, and division' collectively; thus, ' $X_j \# X_k = X_l (\in V)$ ', where ordinal arithmetic calculations are performed excluding of course division by zero.

For set V , addition is commutative: $X_j + X_k = X_k + X_j$, and associative: $(X_j + X_k) + X_l = X_j + (X_k + X_l)$. As for multiplication, set V meets the conditions of a 'monoid' [2,3]. Associativity: $(X_j \times X_k) \times X_l = X_j \times (X_k \times X_l)$, with Left and Right Distributivity: $X_j \times (X_k + X_l) = X_j \times X_k + X_j \times X_l$, $(X_j + X_k) \times X_l = X_j \times X_l + X_k \times X_l$. A nonzero Identity $X_0 (=1)$ for multiplication exists. The Inverse ' $X_j^{-1} = 1/X_j$ ' satisfies ' $X_j \times X_j^{-1} = X_j^{-1} \times X_j = X_0 (=1)$ '. For division, ' $X_j/X_k = X_j \times X_k^{-1} = 1$ ' is definable except for division by zero. Therefore, we can confirm that set V is a 'field'. It can be expressed as $V = \{X_j, \#\}$ or $V = \{X_j \mid X_j \in \mathbb{R}\}$.

Furthermore, different fields based on different sets of clinical values can be described as follows: field $V_1 \equiv \{X_{1j} \mid X_{1j} \in \mathbb{R} \text{ (/mm}^3\text{)}\}$ and $V_1 = \{X_{1j}, \#\}$ for "blood concentration of white blood cells: [WBC] (/mm³)", field $V_2 \equiv \{X_{2j} \mid X_{2j} \in \mathbb{R} \text{ (mEq/l)}\}$ and $V_2 = \{X_{2j}, \#\}$ for "administration

of a certain drug like lithium carbonate: $[Li^+]$ (mEq/l)", field $V_3 \equiv \{X_{3j} \mid X_{3j} \in R \text{ (mEq/l)}\}$ and $V_3 = \{X_{3j}, \#\}$ for "sodium: $[Na^+]$ (mEq/l)", field $V_4 \equiv \{X_{4j}, \#\}$ for calcium: $[Ca^{++}]$ (mg/dl), field V_5 for chloride: $[Cl^-]$ (mEq/l), field V_6 for 'Anion gap [AG] (mEq/l)', field V_7 for 'Base excess [BE] (mmol/l)',..., V_N, \dots , (N: natural number). For each, an independent abstract algebraic treatment is possible as for ordinal abstract algebra.

S2. A vector-like notation using group/field operations belonging to a single set

By making use of all types of scales of measurement, we propose a vector-like expression of a patient's state (denoted 'R_j', j = 1, 2, 3,...: number of sessions), where the mixed expression and its totality of operations that could be performed belong to a single set R. Because of the possible variety of operation rules, the genuine use of this set may be unwieldy at this stage.

Partially based upon our previous description [6,7], let us define 'R_j' to be a vector of five clinical values,

R_j = [severity for depression (within modulo 7 arithmetic) | clock time for the onset of sleep (/24 hrs) | blood concentration of white blood cell [WBC] (/mm³) | blood concentration of [Na⁺] (mEq/l) | a certain value (a certain operational unit)],

$$= [X_{(j)1} \pmod 7 \mid X_{(j)2} (/24 \text{ hrs}) \mid X_{(j)3} (/mm^3) \mid X_{(j)4} (\text{mEq/l}) \mid X_{(j)5} (\dots)] \quad (\text{vii})$$

Next, suppose the patient's state 'R_j' changes to 'R_{j+1}' effected by operator 'R_(j→j+1)'; we denote by '∆' the binary composition composed of the product of compositions for each component. Three possible states are:

$$R_1 = [(X_{(1)1} =) 2 \pmod 7 \mid (X_{(1)2} =) 21 (/24 \text{ hrs}) \mid (X_{(1)3} =) 5000 (/mm^3) \mid (X_{(1)4} =) 145 (\text{mEq/l}) \mid X_{(1)5} (\dots)]$$

$$R_2 = [(X_{(2)1} =) 5 \pmod 7 \mid (X_{(2)2} =) 19.5 (/24 \text{ hrs}) \mid (X_{(2)3} =) 18000 (/mm^3) \mid (X_{(2)4} =) 128 (\text{mEq/l}) \mid X_{(2)5} (\dots)],$$

$$R_3 = [(X_{(3)1} =) 3 \pmod 7 \mid (X_{(3)2} =) 22 (/24 \text{ hrs}) \mid (X_{(3)3} =) 7000 (/mm^3) \mid (X_{(3)4} =) 158 (\text{mEq/l}) \mid X_{(3)5} (\dots)].$$

For the 1st component, 'X₍₁₎₁, X₍₂₎₁' and 'X₍₃₎₁', modulo 7 arithmetic (addition) is used. For the 2nd components, 'X₍₁₎₂, X₍₂₎₂, X₍₃₎₂', operations of Abelian addition are used. For the 3rd component, 'X₍₁₎₃, X₍₂₎₃, X₍₃₎₃', 4th 'X₍₁₎₄, X₍₂₎₄, X₍₃₎₄', the four arithmetic operators (those operations denoted by '#') are required, and for the 5th, 'X₍₁₎₅, X₍₂₎₅, X₍₃₎₅', a certain operational unit is postulated. In the following examples, only addition/subtraction is presented; naturally, multiplication/division is also considered permissible.

With related operators $R_{(1 \rightarrow 2)} = [X_{(1 \rightarrow 2)1} \pmod 7 \mid X_{(1 \rightarrow 2)2} (/24 \text{ hrs}) \mid X_{(1 \rightarrow 2)3} (/mm^3) \mid X_{(1 \rightarrow 2)4} (\text{mEq/l}) \mid X_{(1 \rightarrow 2)5} (\dots)]$, and $R_{(2 \rightarrow 3)} = [X_{(2 \rightarrow 3)1} \pmod 7 \mid X_{(2 \rightarrow 3)2} (/24 \text{ hrs}) \mid X_{(2 \rightarrow 3)3} (/mm^3) \mid X_{(2 \rightarrow 3)4} (\text{mEq/l}) \mid X_{(2 \rightarrow 3)5} (\dots)]$

Then, using results in Appendix D, 'R_(1→2)' and 'R_(2→3)' from the three states given above are as follows:

$$R_{(1 \rightarrow 2)} = [3 \pmod 7 \mid -1.5 (/24 \text{ hrs}) \mid 13000 (/mm^3) \mid -17 (\text{mEq/l}) \mid X_{(1 \rightarrow 2)5} (\dots)]$$

$$R_{(2 \rightarrow 3)} = [5 \pmod 7 \mid 2.5 (/24 \text{ hrs}) \mid -11000 (/mm^3) \mid 30 (\text{mEq/l}) \mid X_{(2 \rightarrow 3)5} (\dots)]$$

Thus, we confirm the relation

$$R_1 \Delta R_{(1 \rightarrow 2)} \Delta R_{(2 \rightarrow 3)} = R_3 \quad (\text{viii})$$

Details are illustrated in Appendix E.

Note that, in general, there exists an identity 'E (=R₀) = [0 (mod 7) | 0 (/24 hrs) | 0 (/mm³) | 0 (mEq/l) | X₀ (...)]' such that 'R_j ∆ E = E ∆ R_j = R_j'. Additionally, there exists an inverse for any 'R_j', 'R_j⁻¹ = [X_{(j)1}⁻¹ (mod 7) | X_{(j)2}⁻¹ (/24 hrs) | X_{(j)3}⁻¹ (/mm³) | X_{(j)4}⁻¹ (mEq/l) | X_{(j)5}⁻¹ (...)] = [7 - X_{(j)1} (mod 7) | 24 - X_{(j)2} (/24 hrs) | - X_{(j)3} (/mm³) | - X_{(j)4} (mEq/l) | X_{(j)5}⁻¹ (...)]' that satisfies 'R_j⁻¹ ∆ R_j = R_j ∆ R_j⁻¹ = E'. However, commutativity, 'R_j ∆ R_k = R_k ∆ R_j' and associativity, '(R_j ∆ R_k) ∆ R_l = R_j ∆ (R_k ∆ R_l)' are not satisfied. Here, we assume that operators acting on 'R_j's should be performed from left to right, that is, from R₁ to R_m (m; number of session for assessment). They should not be applied between 'R_j's. For any assortment of 'R_j's with scales of measurement among types I)–IV), a single set R = {R_j | X_{(j)1} × X_{(j)2} × X_{(j)3} × X_{(j)4} × X_{(j)5}} ('×' means products among groups and fields) using a vector-like notation for the scoring of patient states can be structured where all possible assessments and/or clinical findings of the patient and treatment are included. The general form is the n-fold product; set R = {R_j | X_{(j)1} × X_{(j)2} × X_{(j)3} × X_{(j)4} × ... × X_{(j)(n-2)} × X_{(j)(n-1)} × X_{(j)n}} (n; the number of components).

As for the possible application to better data mining or data usage from the viewpoint of our reinterpretation, we provide a simple example that may help readers to follow an outline of the argument. Consider an example of 17 states "R₁, R₂, ..., R₁₇" (∈set R) each with four component ('n = 5') and arrows (only symbols) that indicate the possible changes among the 'R_j's, as displayed in Figure 1. The scheme covers the notation of our model, and also that of existing methods where (possible) results of data, 'R_j's, are not combined directly with each other in the sense of operations. Then, the arrows could be re-displayed according to our concepts as operators 'R_(j→k)' that can be regarded as elements 'R_j' belonging to a set R as in Figure 2. In ordinal data sets, the 'R_j's are merely a collection of values and the arrows in Figure 1

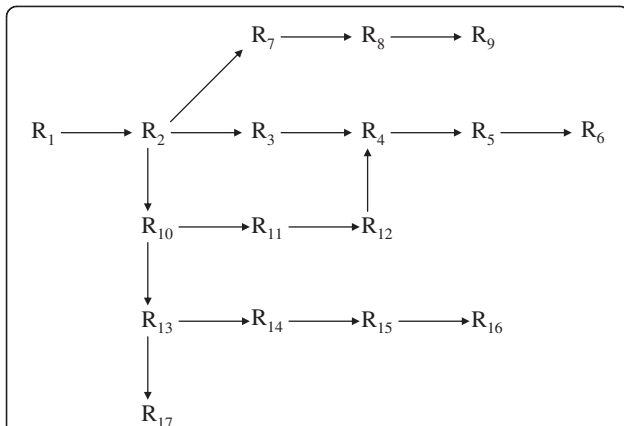


Figure 1 Example of a tree composed of a data set and ordinal arrows. The tree represents changes of states between 17 'R_j' data elements. In ordinal existing methods, data are only a collection of results that are not directly combined; manipulations of parts of the data are defined separately. The arrows merely indicate a change from one state 'R_j' to another 'R_k' and have no specific operational sense.

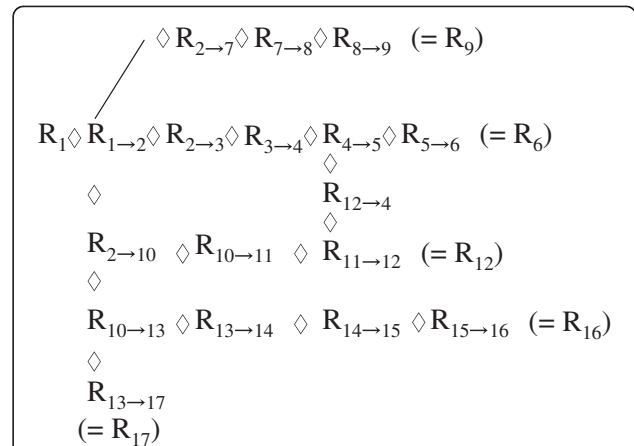


Figure 3 "Operational tree"—the compositional scheme using symbol '◊'. The tree of Figure 2 is re-illustrated using composition symbol '◊', where the operators are assumed to belong to the single set R. By admitting algebraic correspondences, this compositional scheme could potentially provide better data mining/data usage.

are only marks. However, in our interpretation, all 'R_j's and 'R_(j→k)'s are elements of a single set R subject to axioms of an abstract algebra as indicated using composition symbol '◊' in Figure 3. There, the changes 'from R_j to R_k' can be traced at each session. Displayed in this way, Figure 3 represents an "operational tree" that could offer potential for better data mining/data usage through a more generalized/concise treatment (e.g., withdrawing/recording correspond to schemes in Figure 3) that might be permissible. Practical improvements for efficacy, however, will need future investigations.

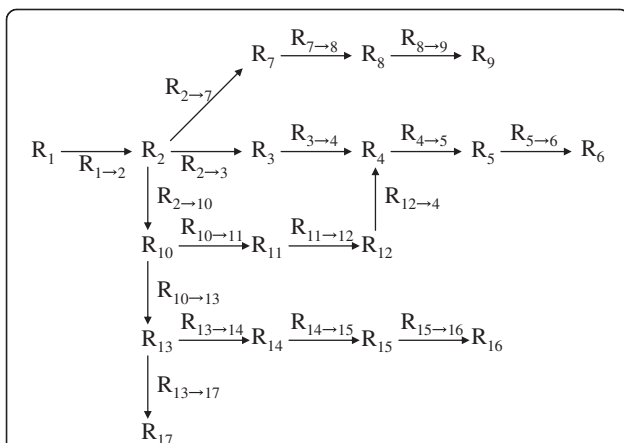


Figure 2 Scheme of the tree with arrows labelled by operators. Arrows are interpreted as operators 'R_(j→k)'s that could be regarded as ordinal elements 'R_j's belonging to a single set R. Each operator that changes 'R_j to R_k' can be traced and its degree for each session identified from initial and final states. The final states (R₆, R₉, R₁₆ and R₁₇) can be traced back to any initiating state 'R₁' by performing an appropriate sequence of 'R_(j→k)'s.

Here, consider the scenario of Figure 1 where from an initial value 'R₁' there are four outcomes 'R₆', 'R₉', 'R₁₆', and 'R₁₇' containing nodes at 'R₂', 'R₄', 'R₁₀', 'R₁₂' and 'R₁₃'. By making use of our previous examples 'R₁ - R₃', the next simplest examples with 'n (component number) = 5' can be confirmed easily:

$$R_{10} = [0 \pmod{7} \mid 17 \text{ (/24 hrs)} \mid 9000 \text{ (/mm}^3\text{)} \mid 130 \text{ (mEq/l)} \mid X_{(10)5}(\dots)]$$

$$R_{11} = [6 \pmod{7} \mid 20 \text{ (/24 hrs)} \mid 20000 \text{ (/mm}^3\text{)} \mid 149 \text{ (mEq/l)} \mid X_{(11)5}(\dots)]$$

$$R_{12} = [4 \pmod{7} \mid 23 \text{ (/24 hrs)} \mid 6000 \text{ (/mm}^3\text{)} \mid 140 \text{ (mEq/l)} \mid X_{(12)5}(\dots)]$$

$$R_{13} = [1 \pmod{7} \mid 18 \text{ (/24 hrs)} \mid 5000 \text{ (/mm}^3\text{)} \mid 135 \text{ (mEq/l)} \mid X_{(13)5}(\dots)]$$

$$R_{17} = [2 \pmod{7} \mid 23.5 \text{ (/24 hrs)} \mid 3000 \text{ (/mm}^3\text{)} \mid 150 \text{ (mEq/l)} \mid X_{(17)5}(\dots)]$$

Following these results, the next relations, according to the tree in Figure 3, can be obtained for instance:

$$R_1 \diamond R_{(1 \rightarrow 2)} \diamond R_{(2 \rightarrow 10)} \diamond R_{(10 \rightarrow 11)} \diamond R_{(11 \rightarrow 12)} = R_{12}$$

$$R_1 \diamond R_{(1 \rightarrow 2)} \diamond R_{(2 \rightarrow 10)} \diamond R_{(10 \rightarrow 13)} \diamond R_{(13 \rightarrow 17)} = R_{17}$$

The operator expressions are evaluated in Appendix F. Similarly, the next sequences are definable in principle,

$$R_1 \diamond R_{(1 \rightarrow 2)} \diamond R_{(2 \rightarrow 3)} \diamond R_{(3 \rightarrow 4)} \diamond R_{(4 \rightarrow 5)} \diamond R_{(5 \rightarrow 6)} = R_6$$

$$R_1 \diamond R_{(1 \rightarrow 7)} \diamond R_{(7 \rightarrow 8)} \diamond R_{(8 \rightarrow 9)} = R_9$$

obtain the complete set $R_j = \{X_{(j)1}, X_{(j)2}, X_{(j)3}, X_{(j)4}, X_{(j)5}\} = [X_{(j)1} \pmod{7} \mid X_{(j)2} (/24 \text{ hrs}) \mid X_{(j)3} (/mm^3) \mid X_{(j)4} (\text{mEq/l}) \mid X_{(j)5} (\dots)]$. A hierarchy has additional levels as necessary to reach single units at its base [12]. The top level is the entire dataset 'R_j' and that is always composable using base units. That is, arbitrary 'R_k' and 'R_l' can be combined into a single dataset as with 'R_k = [X_{(k)1} X_{(k)2} | ... | X_{(k)a}]' and 'R_l = [X_{(l)1} X_{(l)2} | ... | X_{(l)b}]', '{R_k, R_l} = R_j [X_{(j)1} X_{(j)2} | ... | X_{(j)a} | X_{(j)a+1} X_{(j)a+2} | ... | X_{(j)a+b}]' (a,b; positive integers). In this way, classical datasets that are classified in the Stevens scales of measurement could be mined and combined on a higher abstract structure level. To help better understand the concept, a sequence of schemes illustrating the principles of our model is presented in Figure 5.

Subject to future improvements, we envisage that this compact description is versatile to provide better data mining/data usage than from existing methods, although a final version is far from complete at this early stage.

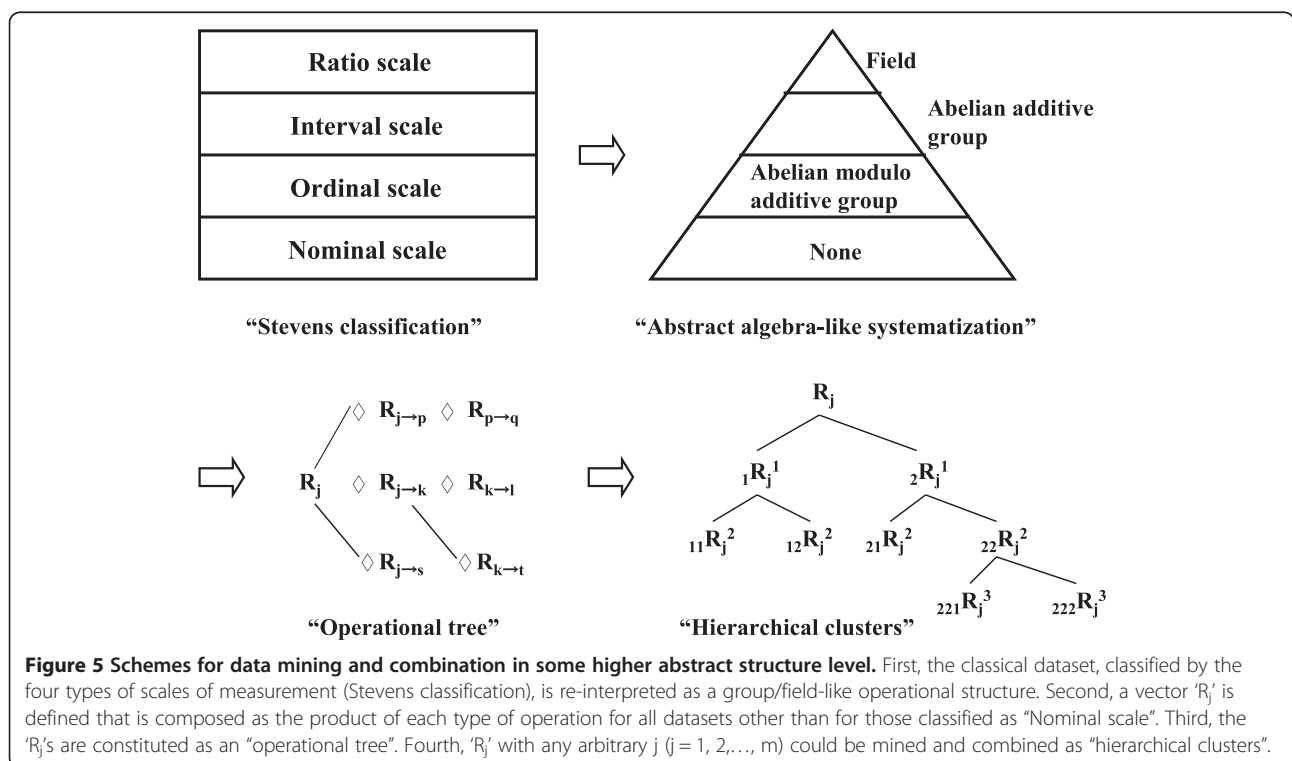
S3. Supplementary suggestions and limitations

If the four arithmetic operations are appropriate in handling the values from clinical assessments, representation by "Ratio scales" (in some cases, the "modular scale" with suitable modulo number previously mentioned) might be effective in describing the clinical treatments or studies. The "Numerical rating scale (NRS)" with range '0–10' [13,14] illustrates the point where the modulo 11 additive

group 'Z₁₁' arises as a natural modular scale. In contrast, similar approaches might be difficult for a "visual analogue scale" [15,16] where values could take any real number.

Whereas rating scales systemized as abstract algebra-like form may enable a more generalized/sophisticated understanding, establishing a link between fields of clinical medicine and abstract algebra, and mixed states and operators in vector-like notation as in (vii)–(xi), does not always assure more concise manipulations. A mixed treatment as exemplified in (vii)–(xi) might not always yield optimal results at present. In general, combining group and field-like structures within 'R_j' may cause some confusion in handling the 'R_j's although benefits accrue through operational compliance and convenience in dealing with the abstract algebra. For description and records, a vector-like definition 'R_j' may not always be advantageous in which only the four types I–IV) are used (particularly for 'I'); the "nominal scale", where systematization of operation seems to be impossible). Nevertheless, we infer that in the handling of operations in mixed-notation like 'R_j' the classification and synthesis of scales of measurement in some group/field-like form may be devised in a more rigorous methodology in future improvements.

That apart, similar, redundant, and obscure components may have been incorporated into the 'R_j's description without discretion. The 'R_j' in such instances loses validity and versatility in terms of a concise composition of scales. This is considered to result from the fact that a total state of a certain disease or a condition of a patient



is not always composable or describable via the combination of partial components. This implies that a larger number of components is not always desirable for assessment or rating scales.

Unfortunately, almost all current assessment scales in medicine are handled as if they were ratio scales although almost all are just ordinal scales. That might introduce considerable futility and/or waste of scientific resources. As previously indicated, some clinical scales (e.g., TNM classification) should be represented as an ordinal scale accompanied by '0' with no absolute need for a quantitative calibration (modular scale). Although a combination composed of entirely ratio scales seems to be difficult or impossible, we believe at least that appropriate operational structures (e.g., group, field) should always be selected that satisfied the conditions in instances like composition of scale, analysis, and interpretation of the results. These structures must be recognized clearly by users per each assessment to avoid misestimation, overconfidence, and complacency in scales.

Conclusions

The Stevens classification of scales of measurement can be re-interpreted and modelled as some abstract algebra-like systematization. Moreover, a vector-like notation using mixed types of operations and a hierarchical structure-like systematization are possible enabling a sophisticated means to classify, update, monitor, and forecast patient treatments. Better data mining/data usage and efficacy is expected and will be considered in future studies.

Appendix

Appendix A

Using 'N = 5' for the scale of a certain symptom or clinical finding with set $Z_5 \equiv \{0, 1, 2, 3, 4\}$, we suppose ' $X_1 = 1$ ' (\in set Z_5) for the initial state and ' $X_2 = 4$ ' (\in set Z_5) for the final state. Expressed as ' $X_1 * X_{(1 \rightarrow 2)} = X_2$ ', the change can be determined as ' $X_{(1 \rightarrow 2)} = X_2 - X_1 \pmod{5} = 4 - 1 \pmod{5} = 3 \pmod{5}$ ' ($\in Z_5$).

Appendix B

Suppose 'the body-temperature thermometer' (deg C; degree Celsius) changes from ' $T_1 = 36.7$ (deg C)' to ' $T_2 = 35.1$ (deg C)'. Because ' $T_1 \circ T_{(1 \rightarrow 2)} = T_2$ ', an operator part is calculated as ' $T_{(1 \rightarrow 2)} = T_2 - T_1 = 35.1 - 36.7$ (deg C) = -1.6 (deg C)'. For another example, when there are two clock times for the onset of sleep ' $t_1 = 21$ (/24 hrs)' and ' $t_2 = 19.5$ (/24 hrs)', the operator part is determined as ' $t_{(1 \rightarrow 2)}$ (/24 hrs) = $t_2 - t_1$ (/24 hrs) = $19.5 - 21$ (/24 hrs) = -1.5 (/24 hrs) = $24 - 1.5$ (/24 hrs) = 22.5 (/24 hrs)'.

Appendix C

Provided [WBC] changes in the following manner: '5000 (/mm³) (= W_1) \rightarrow 18000 (/mm³) (= W_2)'. Because ' $W_1 \# W$

$_{(1 \rightarrow 2)} = W_2$ ', the operator denoted by ' $W_{(1 \rightarrow 2)}$ ' for addition is derived from ' $W_{(1 \rightarrow 2)} = W_2 - W_1 = 18000 - 5000 = 13000$ (/mm³)'. Collectively, the operator is determined by division: ' $W_{(1 \rightarrow 2)} = W_2 / W_1 = 18000 / 5000$ (= 3.6) (/mm³)';

For another example, if '[Na]₁ = 145 (mEq/l)' changes into '[Na]₂ = 128 (mEq/l)', because '[Na]₁ # [Na]_(1 \rightarrow 2) = [Na]₂', the operator for addition is obtained from '[Na]_(1 \rightarrow 2) = [Na]₂ - [Na]₁ = 128 - 145 = -17 (mEq/l)'. Collectively, the operator for division is '[Na]_(1 \rightarrow 2) = [Na]₂ / [Na]₁ = 128 / 145 (mEq/l)'.

Appendix D

$R_{(1 \rightarrow 2)} = R_2 - R_1$
 = [5 (mod 7) | 19.5 (/24 hrs) | 18000 (/mm³) | 128 (mEq/l) | $X_{(2)5}$ (...)] - [2 (mod 7) | 21 (/24 hrs) | 5000 (/mm³) | 145 (mEq/l) | $X_{(1)5}$ (...)],
 = [5 - 2 (mod 7) | 19.5 - 21 (/24 hrs) | 18000 - 5000 (/mm³) | 128 - 145 (mEq/l) | $X_{(1 \rightarrow 2)5}$ (...)],
 = [3 (mod 7) | -1.5 (/24 hrs) | 13000 (/mm³) | -17 (mEq/l) | $X_{(1 \rightarrow 2)5}$ (...)].

$R_{(2 \rightarrow 3)} = R_3 - R_2$
 = [3 (mod 7) | 22 (/24 hrs) | 7000 (/mm³) | 158 (mEq/l) | $X_{(3)5}$ (...)] - [5 (mod 7) | 19.5 (/24 hrs) | 18000 (/mm³) | 128 (mEq/l) | $X_{(2)5}$ (...)],
 = [3 - 5 (mod 7) | 22 - 19.5 (/24 hrs) | 7000 - 18000 (/mm³) | 158 - 128 (mEq/l) | $X_{(2 \rightarrow 3)5}$ (...)],
 = [-2 (mod 7) | 2.5 (/24 hrs) | -11000 (/mm³) | 30 (mEq/l) | $X_{(2 \rightarrow 3)5}$ (...)],
 = [5 (mod 7) | 2.5 (/24 hrs) | -11000 (/mm³) | 30 (mEq/l) | $X_{(2 \rightarrow 3)5}$ (...)].

Appendix E

$R_1 \diamond R_{(1 \rightarrow 2)} \diamond R_{(2 \rightarrow 3)} = [2 \pmod{7} | 21 \pmod{24} \text{ hrs} | 5000 \pmod{mm^3} | 145 \text{ (mEq/l)} | X_{(1)5} \pmod{...}] \diamond [3 \pmod{7} | -1.5 \pmod{24} \text{ hrs} | 13000 \pmod{mm^3} | -17 \text{ (mEq/l)} | X_{(1 \rightarrow 2)5} \pmod{...}] \diamond [5 \pmod{7} | 2.5 \pmod{24} \text{ hrs} | -11000 \pmod{mm^3} | 30 \text{ (mEq/l)} | X_{(2 \rightarrow 3)5} \pmod{...}]$
 = [2 + 3 + 5 (mod 7) | 21 - 1.5 + 2.5 (/24 hrs) | 5000 + 13000 - 11000 (/mm³) | 145 - 17 + 30 (mEq/l) | $X_{(3)5}$ (...)],
 = [10 (mod 7) | 22 (/24 hrs) | 7000 (/mm³) | 158 (mEq/l) | $X_{(3)5}$ (...)],
 = [3 (mod 7) | 22 (/24 hrs) | 7000 (/mm³) | 158 (mEq/l) | $X_{(3)5}$ (...)].

Appendix F

For the 3rd and 4th components, only addition/subtraction is demonstrated collectively for ease in comprehension.

$R_{(2 \rightarrow 10)} = R_{10} - R_2 = [0 - 5 \pmod{7} | 17 - 19.5 \pmod{24} \text{ hrs} | 9000 - 18000 \pmod{mm^3} | 130 - 128 \text{ (mEq/l)} | X_{(2 \rightarrow 10)5} \pmod{...}] = [-5 \pmod{7} | -2.5 \pmod{24} \text{ hrs} | -9000 \pmod{mm^3} | 2 \text{ (mEq/l)} | X_{(2 \rightarrow 10)5} \pmod{...}]$,

$R_{(10 \rightarrow 11)} = R_{11} - R_{10} = [6 - 0 \pmod{7} | 20 - 17 \pmod{24} \text{ hrs} | 20000 - 9000 \pmod{mm^3} | 149 - 130 \text{ (mEq/l)} | X_{(10 \rightarrow 11)5}$

$$\begin{aligned}(\dots) &= [6 \pmod{7} \mid 3 \text{ (/24 hrs)} \mid 11000 \text{ (/mm}^3\text{)} \mid 19 \text{ (mEq/l)} \\ &\mid X_{(10 \rightarrow 11)5}(\dots)], \\ R_{(11 \rightarrow 12)} &= R_{12} - R_{11} = [4 - 6 \pmod{7} \mid 23 - 20 \text{ (/24 hrs)} \\ &\mid 6000 - 20000 \text{ (/mm}^3\text{)} \mid 140 - 149 \text{ (mEq/l)} \mid X_{(11 \rightarrow 12)5} \\ (\dots) &= [-2 \pmod{7} \mid 3 \text{ (/24 hrs)} \mid -14000 \text{ (/mm}^3\text{)} \mid -9 \\ &\text{(mEq/l)} \mid X_{(11 \rightarrow 12)5}(\dots)], \\ R_{(10 \rightarrow 13)} &= R_{13} - R_{10} = [1 - 0 \pmod{7} \mid 18 - 17 \text{ (/24 hrs)} \\ &\mid 5000 - 9000 \text{ (/mm}^3\text{)} \mid 135 - 130 \text{ (mEq/l)} \mid X_{(10 \rightarrow 13)5} \\ (\dots) &= [1 \pmod{7} \mid 1 \text{ (/24 hrs)} \mid -4000 \text{ (/mm}^3\text{)} \mid 5 \text{ (mEq/l)} \\ &\mid X_{(10 \rightarrow 13)5}(\dots)], \\ R_{(13 \rightarrow 17)} &= R_{17} - R_{13} = [2 - 1 \pmod{7} \mid 23.5 - 18 \text{ (/24 hrs)} \\ &\mid 3000 - 5000 \text{ (/mm}^3\text{)} \mid 150 - 135 \text{ (mEq/l)} \mid X_{(13 \rightarrow 17)5}(\dots)] = \\ &= [1 \pmod{7} \mid 5.5 \text{ (/24 hrs)} \mid -2000 \text{ (/mm}^3\text{)} \mid 15 \text{ (mEq/l)} \\ &\mid X_{(13 \rightarrow 17)5}(\dots)].\end{aligned}$$

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

JS conceived the main concept of this article and wrote the manuscript. SM revised the manuscript. JI gave advice on the potential validity from the viewpoint of clinical research and treatment. Additionally, all authors read and approved the final manuscript.

Acknowledgements

The authors wish to acknowledge Katsuji Nishimura, Kaoru Sakamoto, Takashi Oshimo, and Keiko Kojo for providing us with very useful advice.

Author details

¹Department of Psychiatry, Tokyo Women's Medical University, Tokyo, Japan.

²Depression Prevention Medical Center, Inariyama Takeda Hospital, Kyoto, Japan.

Received: 28 January 2014 Accepted: 2 June 2014

Published: 10 June 2014

References

1. Stevens SS: **On the theory of scales of measurement.** *Science* 1946, **103**(2684):677–680.
2. Judson TW: *Abstract Algebra: Theory and Applications.* Virginia: PWS Publishing Company; 1997.
3. Hungerford TW: *Abstract Algebra, An Introduction.* 2nd edition. Philadelphia: Saunders College Publishing; 1997.
4. Hayes MHS, Patterson DG: **Experimental development of the graphic rating method.** *Psychol Bull* 1921, **18**:98–99.
5. Freyd M: **The graphic rating scale.** *J Educ Psychol* 1923, **14**(2):83–102.
6. Sawamura J, Morishita S, Ishigooka J: **A group-theoretical notation for disease states: an example using the psychiatric rating scale.** *Theor Biol Med Model* 2012, **9**:28. July 9.
7. Sawamura J, Morishita S, Ishigooka J: **Further suggestions on the group-theoretical approach using clinical values.** *Theor Biol Med Model* 2012, **9**:54. Dec 19.
8. Tate J, Oort F: **Group schemes of prime order.** *Ann Scient Éc Norm Sup* 1970, **3**(1):1–21. 4e série, t.3.
9. Jullien GA: **Implementation of multiplication, modulo a prime number, with applications to number theoretic transforms.** *IEEE Transac Comput* 1980, **C-29**:899–905.
10. Sobin LH, Gospodarowicz MK, Wittekind C: *International Union Against Cancer (UICC), TNM classification of malignant tumours.* 7th edition. New York: Wiley-Liss; 2010.
11. WHO: *International Statistical Classification of Diseases and Related Health Problems. 10th Revision.* Geneva, Switzerland: World Health Organization; 1992.
12. Billard L, Diday E: *Symbolic data analysis, in 'Conceptual statistics and data mining.'* England: Wiley & Sons Ltd; 2006.
13. Turk DC, Rudy TE, Sorkin BA: **Neglected topics in chronic pain treatment outcome studies: determination of success.** *Pain* 1993, **53**(1):3–16.

14. Farrar JT, Young JP, LaMoreaux L Jr, Werth JL, Poole RM: **Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale.** *Pain* 2001, **94**(2):149–158.
15. Crichton N: **Information point: visual analogue scale (VAS).** *J Clin Nurs* 2001, **10**(5):697–706.
16. Langley GB, Sheppard H: **The visual analogue scale: Its use in pain measurement.** *Rheumatol Int* 1985, **5**(4):145–148.

doi:10.1186/2043-9113-4-9

Cite this article as: Sawamura et al: Interpretation for scales of measurement linking with abstract algebra. *Journal of Clinical Bioinformatics* 2014 **4**:9.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

