VOLUME-7, ISSUE-1, JANUARY-2018 • PRINT ISSN No 2277 - 8160

Original Research Paper

Health Science

SCORE-TO-PERCENT COEFFICIENT: USE AND POTENTIAL IN HEALTH SCIENCES AND PSYCHOLOGY

Kastytis Šmigelskas

PhD, Department of Health Psychology, Medical Academy, Lithuanian University of Health Sciences

ABSTRACT Score transformations are usual and necessary part of different assessment tools in health research and psychology. Conditions to use transformations usually include the assumptions on normal distribution or taking into account the dispersion of the data. However, the distributions are not always normal, while dispersion is strongly dependent on the levels of sample's homogeneity-heterogeneity levels. In addition, the use of different scales for particular phenomena makes the comparability of the findings across different studies not always feasible. Therefore, in this article one new transformed indicator is proposed – the score-to-percent coefficient (STP). Its potential includes psychological as well as general health research, with potential to solve the issues of transformations mentioned above.

KEYWORDS : Methodology, Health sciences, Psychology, Score transformation, STP coefficient

Introduction

Health research is an area with a very wide range of phenomena under investigation. However, the concepts and constructs are not always feasible to evaluate objectively. The causes for this may be the complexity of phenomena, lack of agreement on secondary criteria and subjective nature of psychological or mental health constructs. Therefore, researchers have constructed many tools for measurement of particular phenomena under study. This has led to a broad variety of different scales being used in health research and clinical practice. In sum this creates a scientific and practical problem of comparability of the findings based on those measures across different contexts, since the scales have not only different calculation techniques but also different ranges. There are some standardized agreements, such as the intelligence quotient (Wechsler, 1958) set at the mean score of 100, but the majority of other psychological phenomena have much more variety and rules to subgroup. Only some phenomena have a clearly defined golden standard and application, such as Type D personality measured by DS14 scale (Denollet, 2005).

In contrast, an example of a phenomenon measured by scales that are not unified and therefore are with complicated comparison is the Type A personality assessed using the Jenkins Activity Survey (Jenkins, 1965), the Bortner Short Rating Scale (Bortner, 1969), the Framingham Type A Behavior Pattern Scale (Haynes et al., 1978) and others. Though they all measure the Type A behavior pattern, none of them is regarded being golden standard. This creates a situation in which the findings from studies that use different scales lead to restricted extrapolation and generalizability since it is not possible to properly compare not only the findings among the studies but even the studies samples mean indicators.

Similar issues arise even within one measurement tool that is designed to measure different dimensions of a certain phenomenon. For instance, in case of the Big Five Inventory (John & Srivastava, 1999) there are 44 items with various numbers of items per construct (extraversion has 8 items, conscientiousness – 9, openness – 10 etc.). This leads to a situation where one cannot compare the mean expression of separate sub-constructs within a sample, since the maximum scores of the sub-scales differ. Considering another example, the DS14 scale is a very convenient outlier both being a golden standard measure as well as having an equal number of items per dimension. However, it is a rare example of this kind.

Therefore, the scoring is one of the key steps and milestones in the assessment and analysis of phenomena that are approached using the standardized scales. In order to increase the comparability of scale-based findings, researchers use score transformations in biomedical as well as social research using standardized indicators instead of the original scores.

There have been some attempts to solve the issue of different scale ranges through standardization of the central position and dispersion indicators, mainly using the mean- and standard deviation-related measures. However, these transformations are very sensitive to the variation within a certain sample – therefore, the comparability of the findings across different studies is also limited due to different heterogeneity of the samples in studies and due to specificity of that samples.

Among descriptive score transformations, the main popular approaches include Z transformation, T transformation, standard score, stanine, percentile rank etc. The majority of such transformations are based on assumptions about normal distribution which may not always be correct, particularly in cases of specific samples. The distribution can be not only skewed, but have a large ceiling effect (Nasreddine, 2016) or outliers. There are also other transformations that are based on actual sample distribution, with the description of specific values in terms of where they lie in a sample or in the assumed general population. The example of this kind is percentile rank, informing about a single value's rank converted to percentile within the sample, again, based on assumptions of normal distribution. On the other hand, there are borderline situations where the assumption on normality solely based on P-value may be misleading, especially in case of under- or overpowered studies.

Thus, as mentioned above, the normality assumptions may not always be true and therefore should not be applied in such situations like skewed, clinical or pathology-related phenomena which are usually asymmetrical. On the other hand, rank-based transformations are very sensitive and dependent upon the sample's peculiarities, and especially – depending on the extent of homogeneity. Thus, if the study is conducted in a relatively homogeneous sample, even the minor differences in raw scores may end up as remarkable differences in rank-based scores, while comparing that same scale and its transformations in a rather heterogeneous sample that same raw score differences may end up in a smaller gap of percentile ranks.

Therefore, this article tries to review, discuss and propose an additional approach to score transformations, with the potential for use in health research and psychology. In this article it is referred to as the *score-to-percent (STP) coefficient*. One of the main potential advantages of the STP is its transferability across different samples, scales and phenomena.

Calculation of the STP Coefficient and its Use

The suggested coefficient has already been used previously in some studies. However, this method has had no consistent applications and name so far. The STP coefficient is an indicator showing where a certain score of the scale is within a theoretical (not just observed) continuum of that scale or subscale scores (see Formula).

$$STP = \frac{Actual\ raw\ scale\ score - Minimum\ possible\ scale\ score}{Maximum\ possible\ scale\ score - Minimum\ possible\ scale\ score}\ 100\ (\times 100)$$

This way of calculation has already been applied in health research but rather sporadically, with the majority of examples found in quality of life assessment. However, it was also used in other contexts, including various psychological studies (Table 1). Usually authors call it 'transformed score' even though it is not specific as a term – as mentioned above, there many different types of transformed scores. In addition, even though this way of calculation is seemingly optional or occasional, it is strongly recommended for the Short Form Health Survey (SF-36) questionnaire because the transformed scale scores can be compared with the norms (Ware et al., 1993).

Table 1. Examples of previous applications of STP coefficient.

Scale	Reference	Original formula
General Health Rating Index (GHRI)	Davies et al., 1988	Transformed score = ([actual raw score – lowest possible raw score]/ [highest possible raw score – lowest possible raw score]) × 100
Diabetes Quality of Life Measure (DQLM)	Jacobson, 2013; based on SF-36 principle of calculations [IRC 1991]	Transformed score = ([raw score – lowest possible scale score]/possible scale score range) × 100
Short Form Health Survey (SF-36)	Ware et al., 1993	Transformed scale = [(actual raw score - lowest possible raw score) / possible raw score range] × 100
Stroke Impact Scale (SIS)	primary version [Duncan et al., 1999]; current version [Mulder & Nijland, 2016]	Domain score = [(Mean item score – 1) / 4] × 100
World Health Organization Quality of Life Scale-Brief (WHOQOL-BREF)	Bonomi et al., 2000	Transformed scale = [actual raw score – lowest possible raw score / possible raw score range] × 100
Uterine Fibroid Symptom and Quality of Life (UFS-QoL)	Spies et al., 2002	Transformed score = ([actual raw score – lowest possible raw score]/possible raw score range) × 100
Three-Factor Eating Questionnaire (TFEQ)	de Lauzon et al., 2004	((raw score – lowest possible raw score)/possible raw score range) × 100
Restless Legs Syndrome Quality of Life Questionnaire (RLSQoL)	Abetz et al., 2005	([actual raw score – lowest possible raw score]/ [possible raw score range]) × 100
IKDC Subjective Knee Evaluation Form	Magee (2014), p. 800	IKDC score = [raw score – lowest possible score / range of scores] × 100
The Diabetes-39	Tulloch-Reid & Walker, 2009	Transformed scale = ([raw score – lowest possible score]/raw score range) × 100.
Overactive Bladder Questionnaire, Short Form (OABq)	Trabuco et al., 2016	Transformed score = [actual raw score – lowest possible raw score / possible raw score range] × 100

Strengths and Drawbacks

The main strength of STP coefficient is its standard nature ranging from 0 to 100, with the assumed average of 50. However, it is sensitive to symmetry, therefore it should be carefully used for phenomena that do not have theoretically or empirically established symmetrical distributions, for example, skewed depression scales. Nevertheless, accumulating evidence with STP estimates may propose some potential even for those asymmetrical scales, enabling for instance to compare phenomena. For example, if the feature has a usual STP of 20, it can be considered as highly expressed in samples where STP is 30 or 40. Very practical point is that the comparison of health-related phenomena and psychological indicators can be made both within and between the scales by analyzing subscales and inter-scale comparisons.

Another advantage of STP coefficient is its simplicity of calculation. Also, this indicator does not have the assumptions on normal distribution, rather describing the score within the scale's theoretical range without taking into account sample characteristics or assumptions. The transformed scores that are based on point estimates in the context of normal curve (such as percentile rank) are sensitive to the sample's actual distribution and level of homogeneity-heterogeneity. For example, when comparing the mean point estimates across different studies on the same scale but with a different variance, the same mean point estimate may transform to different scores simply due to different variability of the sample.

There is also a potential for larger applications if used for meta-analysis, since certain phenomena can have different instruments for assessment. However, since STP is not an effect size measure but a descriptive point estimate, coefficient would be more useful for combined description of phenomena rather than for inferential analyses.

Nevertheless, it should be noted that the STP coefficient should not be used to compare entirely different phenomena, e. g. stating that someone is more extravert based on the Big Five scale than he or she is Type A based on the Jenkins scale. It can also be supposed that the use of STP should be careful in case of relatively short scales due to their narrow range of possible values. However, the expression of similar phenomena or sub-constructs could be compared through the STP.

The STP is descriptive in its nature and does not interfere with well-known indicators of neither descriptive nor inferential statistics. This coefficient can be calculated for separate study subject as well as for subgroups or the whole sample under study. In addition, it enables to define the dominant sub-construct of the scale within certain individuals, such as in case of the Big Five.

Inter-scale comparisons based on the STP coefficient may reveal some consistent patterns with certain scales showing consistently lower coefficients than others. Theoretically it may suggest that the scales with the higher coefficients include items with more expression of the construct that is under study. However, this difference should not serve as an indicator for the appropriateness, reliability or validity of one scale versus another.

Another point of caution is the log-transformations of the score. For inferential statistics' purpose, it could be transformed; however, its practical sense in this case would be fully lost unless it is back transformed. However, in contrast to other transformations, the STP coefficient per se does not need back transformation, staying well interpretable with its transformative nature.

To conclude, it can be suggested that the STP coefficient is relatively simple transformed measure applicable for comparisons of phenomena and sub-constructs that are measured using different tools. Compared to the other transformed measures, it does not have assumptions on normal distribution and is not sensitive to heterogeneity of the samples. The best potential of the STP coefficient is possible (but not limited to) for quality of life in biomedicine and for personality assessment in psychology.

Acknowledgments

I am very thankful to my colleagues Gita Argustaitė-Zailskienė, Jolita Jonynienė, Tomas Vaičiūnas, and Justė Lukoševičiūtė for their useful

comments and suggestions.

References

1. Abetz, L., Vallow, S. M., Kirsch, J., Allen, R. P., Washburn, T., & Earley, C. J. (2005). Validation of the Restless Legs Syndrome Quality of Life questionnaire. Value Health, 8(2), 157-167.
2. Bonomi, A. E., Patrick, D. L., Bushnell, D. M., & Martin, M. (2000). Validation of the United States' version of the World Health Organization Quality of Life (QHOQOL) instrument. Journal of Clinical Epidemiology, 53(1), 13-17.
3. Bortner, R. W. (1969). A short rating scale as a potential measure of pattern A behavior. Journal Chronic Diseases, 22(2), 87-91.
4. Davies, A. R., Sherbourne, C. D., Peterson, J. R., et al. (1988). Scoring manual: adult health status and patients satisfaction measures used in RAND's Health Insurance Experiment. Santa Monica, CA: RAND Corporation.
5. Denollet, J. (2005). DS14: standard assessment of negative affectivity, social inhibition, and Type D personality. Psychosomatic Medicine, 67(1), 89-97.
6. Duncan, P. W., Wallace, D., Lai, S. M., Johnson, D., Embretson, S., & Laster, L. J. (1999). The stroke impact scale version 2.0. Evaluation of reliability, validity, and sensitivity to change. Stroke, 30(10), 2131-2140.
7. Haynes, S. G., Levine, S., Scotch, N., Feinleib, M., & Kannel, W. B. (1978). The relationship of psychosocial factors to coronary heart disease in the Framingham study: I. Methods and risk factors. American Journal of Epidemiology, 107, 362-383.
8. International Resource Center (IRC) for Health Care Assessment (1991). How to score the MOS 36-item Short-Form Health Survey (SF36). Boston, MA: New England Medical Center Hospitals, Inc.
9. Jacobson, A. M., & the Diabetes Control and Complications Trial Research Group (2013). The Diabetes Quality of Life Measure. In C. Bradley (Ed.), Handbook of Psychology and Diabetes: a guide to psychological measurements in diabetes research and management. Hove and New York: Psychology Press.
10. Jenkins, C. D., Friedman, M., & Rosenman, R. H. (1965). The Jenkins Activity Survey for Health Prediction. Chapel Hill, NC: University of North Carolina.
11. John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In Pervin & John (Eds.), Handbook of personality: Theory and research (Vol. 2, pp. 102–138). New York: Guilford Press.
12. de Lauzon, B., Romon, M., Deschamps, V., Lafay, L., Borys, J. M., Karlsson, J., Ducimetière, P., Charles, M. A., & Fleurbaix Laventie Ville Sante Study Group (2004). The Three-Factor Eating Questionnaire-R18 is able to distinguish among different eating patterns in a general population. Journal of Nutrition, 134(9), 2372-2380.
13. Magee, D. J. (2014). Orthopedic Physical Assessment, 6th edition. Saint Louis, Missouri: Saunders, an imprint of Elsevier Inc.
14. Mulder, M., & Nijland, R. (2016). Stroke Impact Scale. Journal of Physiotherapy, 62(2), 117.
15. Nasreddine, A. Y., Connell, P. L., Kalish, L. A., Nelson, S., Iversen, M. D., Anderson, A. F., & Kocher, M. S. (2016). The Pediatric International Knee Documentation Committee (Pedi-IKDC) Subjective Knee Evaluation Form: Normative Data. American Journal of Sports Medicine, 45(3), 527-534.
16. Spies, J. B., Coyne, K., Guaou Guaou, N., Boyle, D., Skyrnarz-Murphy, K., & Gonzalves, S. M. (2002). The UFS-QOL, a new disease-specific symptom and health-related quality of life questionnaire for leiomyomata. Obstetrics & Gynecology, 99(2), 290-300.
17. Trabuco, E. C., Klingele, C. J., Blandon, R. E., Occhino, J. A., Weaver, A. L., McGree, M. E., Lemens, M. A., & Gebhart, J. B. (2016). Burch Retropubic Urethropexy Compared With Midurethral Sling With Concurrent Sacrocolpopexy: A Randomized Controlled Trial. Obstetrics & Gynecology, 128(4), 828-835.
18. Tulloch-Reid, M. K., & Walker, S. P. (2009). Quality of life in Caribbean youth with diabetes. West Indian Medical Journal, 58(3), 250-256.
19. Ware, J. E., Snow, K. K., Kosinski, M., & Gandek, B. (1993). SF-36 Health Survey: Manual and Interpretation Guide. Boston, MA: The Health Institute, New England Medical Center.
20. Wechsler, D. (1958). The Measurement and Appraisal of Adult Intelligence (4th ed.). Baltimore, MD: Williams & Witkins.