Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Quantitative Characterization of Disease Severity in Diseases with Complex Symptom Profiles

George V. Kondraske, *Member IEEE* and R. Malcolm Stewart

***Abstract* – A number of clinical and research situations arise that require the integration of complex, multi-dimensional, performance-related information to determine a single quantity such as "disease severity" or "risk level" (for subsequent development of a disease). This process is generally carried out either by relying on a subjective gestalt approach or by using rating scales that combine scores for component measures using an additive process without justification. Concepts from General Systems Performance Theory have provided insights into this problem, teaching that a fundamentally multiplicative method of integrating components is often appropriate. In this paper, concepts and previous supportive experimental work are reviewed in the context of the quantitative characterization of disease severity. A Parkinson's Disease study (n = 114) is presented that mimics the "two-condition" situation encountered in the clinical trial of a new drug or other therapy. Traditional and performance theory-based composite scores are computed for each condition ("on" and "off" medications) and compared, with emphasis on the different quantitative "pictures" conveyed by each approach. It is concluded that performance theory based composites are not only more sensitive to therapeutic agents experimentally, but also have superior conceptual validity compared to traditional forms of single-number composites.**

I. Introduction

In clinical trials of drugs or other therapies targeting neurologic diseases such as Parkinson's Disease (PD), we have observed that it is not uncommon for disagreement to exist between clinicians' perception of efficacy and the quantitative results obtained from formal analyses. Clinicians may have a gestalt sense that the benefit is "substantial", when both individual and composite quantitative measures indicate relatively small changes (that may also be "statistically" significant). PD, for example, results in a diverse, complex array of motor and non-motor symptoms related to complex profiles of impairment. The fundamental challenge that arises in this and many other analogous situations is: "How does one systematically derive a single number that characterizes disease severity when multiple systems and attributes of performance are involved?"

Instruments, such as the Unified Parkinson's Disease Rating Scale (UPDRS) [1], are designed to capture this complexity and are often used in such clinical trials. Such scales are administered by a trained clinician making a subjective judgment to score each item of the scale. Each item generally reflects a different *symptom and/or impairment* (vs. *performance capacities*). In the UPDRS, sub-scores are computed by adding scores for items within categories. An overall score is computed by adding sub-scores. Similar methods are used in rating scales for other diseases or injuries (e.g., head injury). The use of addition to combine scores reflecting conceptually different quantities such as tremor, slowness of movement (bradykinesia), mental status, coordination, balance, and gait (for example) in such scales is by far the standard, but is never justified. It is apparently a "traditional thing to do", has been rather blindly applied, and has not been questioned. Recently, a task force [1] reviewed the UPDRS and recommended development of a new version that capitalizes on its strengths and rectifies its weaknesses. No mention was made of the method used to combine scores for individual items.

General Systems Performance Theory (GSPT) and application of it to human performance measurement [2] contain conceptual perspectives relevant to the stated challenge of forming composite scores such as those that reflect disease severity. GSPT provides a framework to address the complex, multidimensional and hierarchical nature of human performance. A key premise here is that disease impacts performance of affected subsystems. Briefly, GSPT requires that performance measures be defined using a resource construct (representing desirable quantities in contrast to impairments; e.g., speed vs. bradykinesia, steadiness vs. tremor, etc.). For a given system, these performance resources define the axes (or "dimensions of performance") of a multi-dimensional performance space in which a performance capacity envelope (PCE) is defined using measures of individual performance capacities (Fig. 1).

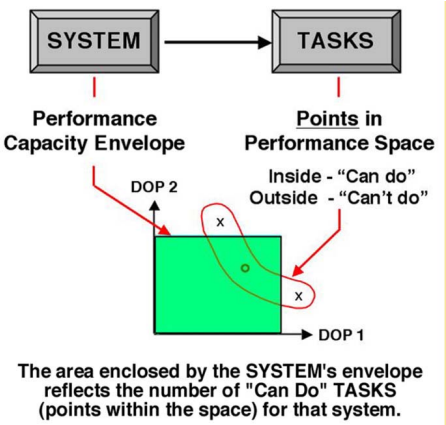

Fig. 1. Key concepts leading to formation of product-based measures of composite performance capacity are illustrated for a system with two dimensions of performance (DOPs). A larger product-based composite reflects a larger volume enclosed by the performance capacity envelope; more "points" (tasks) are thus enclosed.

Manuscript received April 24, 2006. This work was supported in part by the Horace C. Cabe Foundation and a Faculty Development Leave Award from the University of Texas at Arlington.

George V. Kondraske is with the Human Performance Institute, University of Texas at Arlington, Arlington, TX 76019-0180 USA (phone: 817-272-2335, fax: 817-272-2253, e-mail: gvk@hpi.uta.edu).

R. Malcolm Stewart is with the Human Performance Lab, Presbyterian Hospital of Dallas, Dallas, TX 75231 USA (e-mail: R.MalcolmStewartMD@texashealth.org).

Each such measure reflects the amount of availability of a given performance resource for use in tasks performed by the corresponding system. The logic of GSPT explains that the volume enclosed by the PCE represents the system's capacity to perform tasks that draw on the constituent performance resources. For example, if the dimensions of performance in Fig. 1 were speed and accuracy, the product of "available speed" and "available accuracy" would reflect the capacity to execute tasks that make demands on the system's speed AND accuracy resources. A key element of this logic is the recognition that the *enclosed* points represent specific tasks imposing demands that are "within the limits" of the system's capacity.

II. Relevant Previous Studies

The first effort to apply these concepts to experimental data addressed the problem of computing the functional capacity of the shoulder [3] and by inference, other neuromuscular systems. No effort was made to validate the result; there was no solid "gold standard" available for comparison. Results were explained conceptually using joint probability; i.e., what is the probability that the "shoulder system" would have enough strength AND range of motion AND .. to accomplish tasks it attempts? This type of application was subsequently expanded upon and further formalized for neuromuscular systems to consider performance capacity envelopes that are not "rectangular" but which have smooth curvilinear surfaces [4]. Methods for estimating the complex nature of the PCE were demonstrated with data for the knee extensor system.

In another study [5], three individual performance capacities (visual information processing speed, shoulder internal/external rotator movement speed, and shoulder abductor strength) were utilized, as well as the performance of a "more complex" higher-level task (speed of putting on a shirt) that made demands on the three individual performance resources (and others). Subjects ranged in age from 20-80 yrs. Regression equations, representing performance of each item as a function of age, were developed. Performance capacities for 20, 45 and 70 yrs were expressed as a fraction of 20 yr values. Various composite scores were formed by averaging and multiplying different combinations of the age-normalized individual capacities. It was found that the *slope of decline with age* for the product-based composite that included all three of the individual capacities exhibited the best agreement with the slope of decline of the speed of putting on a shirt.

GSPT concepts were also applied to the human speech production. One aspect of this effort focused specifically on the pitch control system [5]. Individual capacity measures reflected central processing associated with pitch control (speed of response to a stimulus) as well as a neuromuscular capacity (movement speed of the vocal folds). Four subjects (three healthy non-singers and one professional singer) were studied extensively. A composite formed as the product of the two individual measures was computed. While differences between the non-singers and singer were relatively small for component measures, the product-based composite indicated the average pitch control performance capacity across the non-singers was only about 18% of the singer. It was concluded that this composite began to reflect the "true difference" in vocal ability observed.

Motivated by the desire of physical therapists to "measure" motion quality in rehabilitation contexts, a study was conducted involving healthy subjects as well as those with various conditions affecting motion quality [7]. A set of individual measures associated with a generic upper extremity "motion producing system" (i.e., considering one arm as the "system of interest") was defined using GSPT constructs. These measures, each of which targeted different aspects of "quality" (e.g., speed, accuracy, smoothness, etc.), were acquired for a set of different "more complex" tasks (e.g., throwing, cleaning a surface, picking and placing, etc.). Execution of these tasks was also videotaped and subsequently rated by experienced professionals (who routinely must deal with the notion of "motion quality") on a single-dimensional visual analog scale of "overall motion quality". Individual performance measures and various composite scores (based on both additive and multiplicative combinations) were then correlated (across all subjects) with the subjective evaluation scores of the experts which served as an operational "gold standard". Product-based composites exhibited the best agreement (i.e., largest correlation coefficient) with this gold standard. Furthermore, the level of correlation increased as more individual measures were used to form the composite.

A well-established principle of human motion known as Fitts' Law was revisited [8], [9] from the perspective of GSPT and product-based composite performance capacities. Fitts' Law is generally characterized as a formal explanation of the speed-accuracy tradeoff that exists in human motion. This is interesting in that the traditional equation representing Fitts' law does not explicitly contain a speed or an accuracy variable. It does contain a "movement time" variable which can be considered to be speed related; i.e. the inverse of speed.

We have long used upper and lower extremity tests of coordination that are based on Fitts' Law. In these tests, a subject attempts to alternately move a limb segment between two targets "as fast and as accurately as possible". It was shown from experimental data drawn from our database (more than 1500 records) that the simple mathematical product of speed and accuracy (for a given task challenge) correlated perfectly with what Fitts' termed his "Index of Performance" (IP). The "IP" reflects a basic performance resource that can be used for speed or accuracy and thus results in a "speed-accuracy" tradeoff. Scaling the product of speed and accuracy by what Fitts' termed the Index of Task Difficulty (which is computed from the distance between targets and the width of the target) produced numerical values there were essentially in perfect agreement with Fitts' IP. The result is an equation that explicitly contains speed and accuracy variables to compute the equivalent of Fitts' IP, which we have termed "neuromotor channel capacity". Given the validity that is widely attributed to Fitts' Law, this result was viewed as a powerful endorsement of GSPT constructs pertaining to the formation of composite performance capacities.

More recently, GSPT was applied to explore the formation of composite scores in PD [10], [11] toward the ultimate goal of realizing a measure of disease severity. This paper represents the extension of these efforts.

III. Methodology

Subjects with Parkinsonism (n=114) were tested in "OFF" medication (so-called "practically defined off") and "ON" medication (~ 1 hour after medication) states. A battery of objective, computer-based performance capacity measures and the UPDRS was administered in both states. Six types of the objective measures representative of cognitive, motor, and balance domains were selected for analysis: 1) visual spatial short-term memory capacity, 2) visual information processing speed, 3) upper extremity movement speed (each arm), 4) index finger rapid alternating movement speed (each hand), 5) neuromotor channel capacity (four limbs), and 6) postural stability (each leg). Procedures and measures for these tests are described elsewhere; e.g., [12]. Each of these measures was developed in accordance with GSPT's resource construct; a larger numerical value always reflects "more performance resource availability" and therefore greater performance capacity.

The application of GSPT to traditional scales that are symptom or impairment oriented requires that all measures be transformed to represent "performance resources" that reflect *desirable quantities* and that have numerical values that are larger for "better performance". Generally, this is accomplished by simple inversion [9] or by reversing the numerical values associated with points on the scale. In the present study, two selected items from Subscale III of the UPDRS ("rapid alternating movements of the upper extremity" and "leg agility"; dominant and nondominant body sides) were used. These are rated on a five-point scale in which "0" represents "normal" and "4" represents "can barely perform the task", with three key-worded levels in between. These scores were transformed by subtracting each actual score from the numerical value "5", producing results ranging from 1 to 5 where "5" now represents "normal" (the "best performance" possible on this scale).

The "percent change" (from "OFF" to "ON" medication) was computed for each subject for each individual measure. Composite scores were computed by averaging various combinations of measures (i.e., a fundamentally additive method, consistent with the computation of subscale and overall scores on the UPDRS) and also by computing multiplicative composite performance capacities. This was done first for each condition (i.e., "ON" and "OFF"). Then "percent change" scores were computed for each composite measure. Finally, the average across all subjects was computed for each individual and composite "percent change" score. Better performance in the "ON" medication condition is reflected by positive percent change values.

IV. Results

For the objective, computer-based performance measures, "percent change" values ranged from 1.4-22.6% (greatest for lower extremity NMCC) for individual measures and averaging-based composites. For product-based composites, values range from 4.3-109% Larger percent change values were obtained for composites incorporating a greater number of individual performance measures. The percent change scores for representative individual and composite measures are illustrated in Fig. 2.

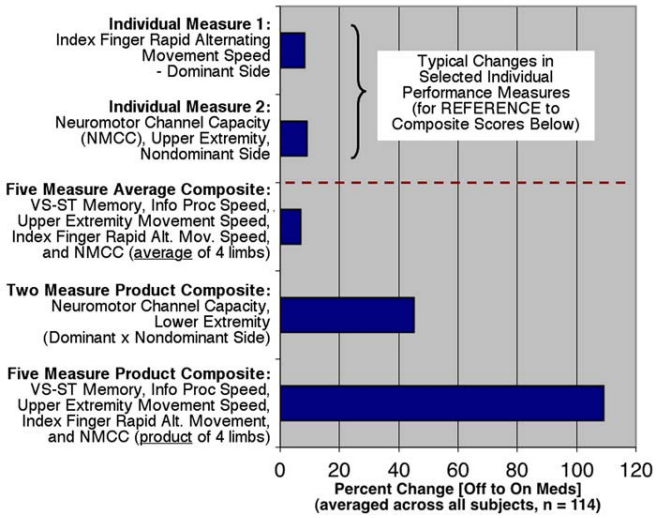

Fig. 2. In product composites, "percent change" values increase as additional performance resources, each of which exhibits a relatively small percent change value, increases.

A performance capacity envelope view is presented in Fig. 3 that includes, for reference, healthy subjects of a comparable age. Note that measures for two of the dimensions are, themselves, product-based composites.

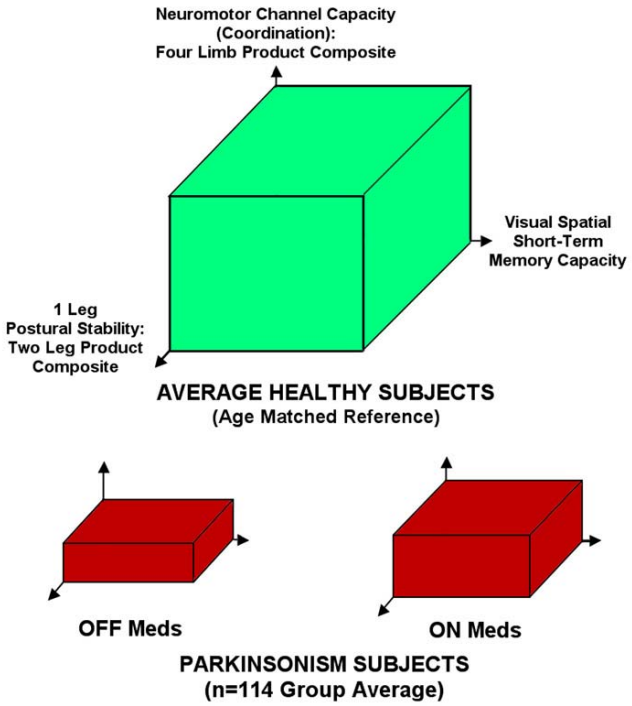

Fig 3. Change from OFF to ON meds along each dimension is relatively small; but change in volume, reflecting capacity to execute tasks that require the component performance resources) is 106%.

For UPDRS measures, the four individual measures indicated improved performance in the "ON" medication state. "Percent change" values ranged as follows: 1) individual measures (14.6-18.1%), 2) addition-based composites (15.2-17.2%), and 3) multiplication-based composites (30.8-63.7%). In the latter case, the magnitude of the percent change again increased with the number of individual measures included.

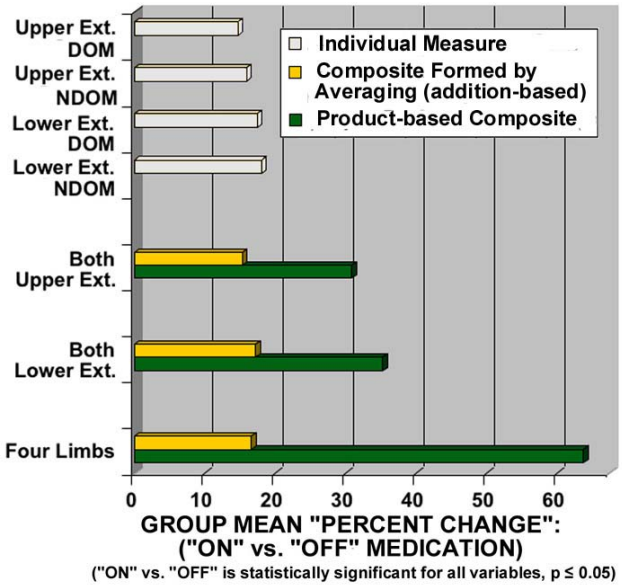

Fig. 4. Comparison of individual, traditional composites (additive), and GSPT-based composites using selected UPDRS items.

V. Discussion

Results support previous findings that product-based composites are more sensitive than addition based approaches. While no "operational gold standard" was formally included, it is argued that the amount of "change" (from "OFF" to "ON" medications) in ability execute daily tasks for these patients is more accurately reflected by numerical values of "percent change" in the range of 50-100%, rather than values of 10-15%.

Whereas the interpretation of product-based composites is clearly provided by GSPT (i.e., the capacity to execute tasks that make demands on the combination of performance resources included in the composite), the true interpretation of traditional addition-based composites is not clear. One obvious possible interpretation is that they reflect the average "availability" (or weighted average) across a selected pool of different performance resources. It is argued that this is not the information required to reflect disease severity as it does not characterize *the interaction of these performance resources* in the execution of more complex tasks. It is also not logical to add quantities that are fundamentally different (e.g., memory capacity and movement speed), even when normalizations are applied that give the appearance of "unitless" numerical values. Product-based composites inherently embrace the presence of units and preserve dimensionality when they are formed. This, aside from experimental findings, adds to their validity.

Our approach to characterizing disease severity is based on computing system performance capacity. Clearly, these are inverses of each other. The assumption is that the greater the *loss* in composite capacity (relative to a reference representing "healthy"), the greater is disease severity. This logic is consistent with subjective, *ad hoc* assessments of severity routinely made by clinicians and their patients.

VI. Conclusion

Quantitative characterization of disease severity is a complex and important problem. Formation of composite measures must be carefully studied using both *conceptual* and *experimental* tools. Collective results from this study and previous studies cited suggest that the traditional conceptual basis for computing composite scores should be revisited and scrutinized. Averaging (or simple addition commonly employed in rating scales) tends to diminish or dilute important differences exhibited by individual subjects across the array of performance variables and has conceptual problems as noted. GSPT-based composites appear to be more sensitive *experimentally,* and *conceptually* provide a more meaningful integration of information across multiple systems as well as their dimensions of performance.

References

[1] Movement Disorder Society Task Force on Rating Scales in Parkinson's Disease, "The unified Parkinson's Disease rating scale (UPDRS): status and recommendations," *Movement Disorders*, vol. 18, pp. 738-750, 2003.

[2] G. V. Kondraske, "The elemental resource model for human performance," in *The Biomedical Engineering Handbook 3rd Ed.: Biomedical Engineering Fundamentals.* J. Bronzino, Ed. Boca Raton: CRC Press, Taylor & Francis, 2006, pp. 75.1-75.19.

[3] G. V. Kondraske, "Computation of functional capacity: Strategy and example for shoulder," in *Proc. 9th Annual IEEE Eng. in Med. and Biol. Soc. Conf.*, 1987, pp. 477-478.

[4] P. J. Vasta and G. V., Kondraske, "An approach to estimating performance capacity envelopes: knee extensor system example," in *Proc. 19th Annual Conf. of the Eng. In Med. and Biol. Soc.*, Chicago, 1997, pp. 1713-1716.

[5] G. V. Kondraske, "Neuromuscular performance: Resource economics and product-based composite indices," in *Proc. 11th Annual Conf. of the Eng. In Med. and Biol. Soc.*, Seattle, 1989, pp. 1045-1046.

[6] M. Jafari, K. H. Wong, K. Behbehani, and G. V. Kondraske, "Performance characterization of human pitch control system: An acoustic approach," *Journal of the Acoustical Society of America*, vol. 85(3), pp. 1322-1328, 1989.

[7] C. A. Fischer and G. V. Kondraske, "A new approach to human motion quality measurement," in *Proc. 19th Annual Conf. of the Eng. In Med. and Biol. Soc.*, Chicago, 1997, pp. 1701-1704.

[8] G. V. Kondraske and P. J. Vasta, "Neuromotor channel capacity, coordination, and motion quality," in *CD-ROM Proc. World Cong. on Med. Phys. and Biomed. Eng.*, Chicago, 2000, (4 pgs).

[9] G. V. Kondraske, "Performance theory: implications for performance measurement, task analysis, and performance prediction," in *CD-ROM Proc. World Cong. on Med. Phys. and Biomed. Eng.*, 2000, (4 pgs).

[10] R. M. Stewart, G. V. Kondraske. and M. K. Sanghera, "Performance theory and formation of composite outcome measures: Implications for clinical trials," *Mov. Dis.*, vol. 19 (Suppl. 9), pp. S157-S158, 2004

[11] R. M. Stewart, G. V. Kondraske, and M. K. Sanghera, "Application of systems performance theory to the UPDRS: preliminary exploration," *Mov. Dis.*, vol. 20 (Suppl.10), pp. S82-S83, 2005.

[12] M. T. Gettman, G. V. Kondraske, O. Traxer, K. Ogan, C. Napper, D. B. Jones, M. S. Pearle, and J. Cadeddu, "Assessment of basic human performance resources predicts operative performance of laparoscopic surgery," *J. Amer. Col. Surg.*, vol. 197(3), pp. 489-496, 2003.