Published in final edited form as:

Psychosomatics. 2010 November ; 51(6): 515–519. doi:10.1176/appi.psy.51.6.515.

Beyond the Global Assessment of Functioning: Learning From Virginia Apgar

Joel E. Dimsdale, M.D., Dilip V. Jeste, M.D., and Thomas L. Patterson, Ph.D.

Department of Psychiatry, University of California, San Diego

Abstract

Background—The Global Assessment of Functioning (GAF) scale is widely used in psychiatry, yet it has certain drawbacks.

Objective—The authors seek to generate further discussion and research around developing an improved successor to the GAF.

Method—The authors used the Apgar scale as a template for constructing a possible successor to the GAF. Consulting with 16 colleagues, they selected 5 domains that were felt to be central to functioning in psychiatric patients. Psychiatrists in diverse clinical settings then completed both a GAF and a Psychiatric Apgar scale on 40 patients.

Results—The two scales were found to agree significantly. Use of the Psychiatric Apgar, however, provides clearer guidance about assessing functioning.

Conclusion—The GAF was a brilliant addition to psychiatric practice. As we develop the next Diagnostic and Statistical Manual, it is pertinent to ask whether the GAF approach could be optimized even further by applying the lessons of Virginia Apgar.

The Global Assessment of Functioning (GAF) scale was introduced in the *Diagnostic and Statistical Manual of Mental Disorders, 3rd Edition* (DSM–III) as a method for rating patient functioning for all psychiatric disorders. The GAF is a direct descendent of the Global Assessment Scale,¹ with minor modifications. Implicit in the scale are assumptions about what constitutes mental health and mental illness and what constitutes well-being in society. Rather than specify all dimensions of functioning (existential, defensive/coping, sexual, occupational, etc.), the scale asks the rater for a “gestalt” number from 1 to 100 on which to rate patients’ overall functioning. The GAF is conceptually related to various quality-of-life measures, such as the World Health Organization Disability Assessment Schedule (WHODAS) and the 36-item Short Form (SF–36), but the GAF is unique by virtue of its specific focus on assessing global functioning and quality of life in psychiatric patients. It has been widely adopted as an important component of psychiatric assessment.

Like any scale, the GAF has its limitations. Given that the DSM is being revised, it is appropriate to examine whether the GAF may be further optimized for clinical utility. From our point of view, the GAF has two main problems: 1) the conceptual anchors are not well specified; and 2) the degree of specification (number of significant digits) exceeds the readily observable data. This article argues for a simplified GAF, with explicit domains and anchors, and it refers the reader to one of the most successful simple scales in contemporary medicine—the Apgar scale—as a model for modifying the GAF.

The GAF has limited psychometric properties in terms of test–retest reliability, interrater reliability, and validity.² These drawbacks are endemic to any unidimensional scale that tries to address a complex construct like “functioning.” It is a testimony to the GAF that it functions as well as it does while rating such extraordinarily diverse constructs. Researchers have commented on the fact that the GAF performs well in the hands of researchers but has

lower reliability when used in routine clinical settings.³ Clearer conceptual anchors would certainly help.

Part of the reason for the limited psychometric capabilities of the GAF is its avowedly gestaltist perspective. The rater must consider a host of factors simultaneously and condense these disparate ratings onto a 100-point scale. Such problems are endemic to assessing the construct of functioning,^{4,5,6} which merges patients' symptoms with their psychosocial performance. By parsing the overall score into a short explicit list of key measures of functioning, one may be better able to make rapid, reliable, and useful ratings of global functioning.

An alternate approach to the GAF was developed by the World Health Organization⁷ and then refined into the WHODAS, 2nd Version (WHODAS-II),⁸ which measures the impact of any disorder (not just psychiatric) on everyday functioning. The WHODAS-II is a 32-item tool that provides a profile of functioning across six domains of activity. Despite its considerable strengths, the WHODAS-II is difficult to use in everyday busy clinical practice situations. For a functional assessment to have clinical utility, it must be brief and easy to use. Thus, the "ease-of-use" criterion clearly favors a GAF-like approach over the WHODAS, despite the better specification of functioning in the WHODAS.

Even an apparently simple measure like the SF-36, which has been widely used in clinical research, is rarely utilized in non-research clinical settings.⁹ Although its widespread use and generic approach make it appealing, it is unlikely to gain widespread acceptance in clinical settings because it focuses on patients' self-report; it requires approximately 15 minutes to complete; and it relies on a complex scoring procedure. The shorter SF-12 is similarly limited by its complex scoring routine.

Another generic approach to measuring functioning uses utility-based measures, such as the Quality of Well-Being (QWB)¹⁰ and the EuroQol 5-Dimension Scale (EQ-5D).¹¹ These measures were designed to assess health outcomes from a wide variety of medical and psychiatric disorders on a common scale ranging from perfect health to death. Although they are useful for health-policy decisions, they are less useful clinically. Also, these scales require at least 10 minutes to administer.

Psychiatry could learn a great deal by examining the development of the Apgar scale to rate neonatal functioning. When Virginia Apgar developed the scale in 1953, she selected five core variables that were felt to reflect newborns' functioning.¹² She suggested that each of these variables be rated from 0 to 2, and also provided guidance on how these ratings were to be made. She then tested her scale on approximately 2,000 neonates, demonstrating that the scale accurately reflected risks to neonates. The simplicity of the scale, its face validity, simple anchors, and "metric feel" endeared it to obstetricians, anesthesiologists, and pediatricians throughout the world. It was rapidly incorporated into epidemiological and treatment studies of infants. Today, an Apgar score is used almost everywhere in the world to rate neonatal functioning.

Apgar did not report on the interrelationship among the various components of the scale and did not examine whether some variables were more important than others. Subsequent researchers have performed such studies, and some have favored a 4-item Apgar score as equivalent to or even superior to the original 5-item Apgar scale.¹³ However, the Apgar scale flourishes half-a-century later, more or less in its original form—a remarkably easy-to-use scale with clear implications for health.

Could psychiatry come up with an "Apgar-equivalent" to address the limitations in the GAF? It is useful to review the criteria Apgar developed and think about their applications

to psychiatry. Table 1 summarizes Apgar's approach. Following on her approach, are there four or five observable variables that most psychiatrists would think define functioning across the entire range of DSM? The purpose of this article is to generate further discussion and research around developing or refining a scale similar to the one proposed herein.

Method

We consulted with 16 psychiatrists and psychologists across the world to get their advice on key variables and anchors to use with a Psychiatric Apgar scale. We then asked eight clinicians in highly diverse practice settings to rate five of their patients with both the GAF and the Psychiatric Apgar; 40 patients with diverse diagnoses were assessed in settings such as the emergency room, consultation-psychiatry department, hospice, outpatient psychiatry department, inpatient psychiatry department, and drug-abuse units. The associations between GAF ratings and the Psychiatric Apgar ratings were examined with simple linear regression and with the generalized least-squares (GLS) model (which allows for correlated errors within each psychiatrist-rater).

Results

The experts' suggestions were to look for key domains that would follow directly from the mental-status exam and patient history. Furthermore, the consensus was that key variables should include: neurocognitive functioning, distress, psychotic features, everyday functioning, and social relationships. One advantage of this scope of variables is that virtually every patient would have these components assessed as a part of any psychiatric evaluation. One could add or subtract a domain for scoring on this putative new GAF, but, the longer the scale, the greater the response burden and the less likely that the scale would be routinely completed on every patient. Similarly, one could debate whether all domains should have the same "weights." For simplicity, we followed Apgar's model and merely added domains that seemed important and relatively independent. For example, she derives a score by adding ratings for heart rate and muscle tone together. By the same token, we add together ratings for distress and psychotic features, emphasizing that we are trying to retain the elegant simplicity of the GAF while delivering greater precision.

Although there can and must be interesting discussions concerning which specific dimensions should comprise the scale, the real challenge comes in defining anchors for the scale. Table 2 provides our suggested Psychiatric Apgar scale, together with anchors.

The first three dimensions (neurocognitive functioning, distress, and psychotic features) are extractable from the mental-status exam, and the last two dimensions (everyday functioning, and social relationships) would routinely be obtained from the psychiatric history.

We have combined affect, stress tolerance, and anxiety into a broader dimension, which we have called "distress." How resilient is the individual in the face of life's trials and tribulations? How much emotional suffering or distress is experienced?

Everyday functioning would apply to assessments of functioning at work, school, or home. Voluntary unemployment, part-time work, or retirement would not result in a lessened score. Job stress, protracted job dissatisfaction, as well as involuntary unemployment due to economic factors would be scored in the intermediate category. Similar criteria would be applied to functioning at school or as a homemaker.

Regarding "social relationships," the anchor is clearest for severe disruption of social relationships (score: 0), but the other scores would appear readily definable, given that anchor.

In emergency room and consultation-psychiatry settings, information on some of these dimensions could initially be difficult to obtain in a severely-agitated or confused patient who was unable to provide a history. On the other hand, if patients could provide no useful information, for instance on their everyday functioning, that functioning level would logically be rated as severely impaired at the time of the consultation. Later, for instance, after delirium had improved or when more data became available from the patient or other sources of information, that rating would be revisited. Thus, the Psychiatric Apgar score (like the GAF score) would be expected to change as the patient improves. Note that the GAF suffers from similar problems in a consultation setting if information is not available. The difference is that the Psychiatric Apgar clearly indicates how the various domains of functioning contribute to a total score.

How might this scale perform? How would scores be distributed? Is it truly (or should it be) a linear scale? These questions need empirical study. Note, however, that, to most people, a total score of 8-or-better would be considered to indicate reasonably high functioning, and a score below 5 would imply severe disability. The scores on the GAF and Psychiatric Apgar correlated well ($p < 0.0001$; Figure 1).

Discussion

The Apgar approach correlated highly with the GAF. Curiously, the range of scores was similar to the range of GAF scores, once such scores were divided by 10. What is different is that the Psychiatric Apgar explicitly tests a small number of specified domains of functioning that are keyed to clear anchors.

The Apgar-like approach requires assessment on each domain but does not assume that these domains are independent of each other. An individual with severe distress, for instance, would likely (but not necessarily) have impairment in vocational functioning, and so forth. Furthermore, such a rating could be based on self-report, caregiver report, and/or direct clinical observation. Its simplicity and practicality would be appealing to clinicians. Clinicians would, of course, need to familiarize themselves with its anchor-points; however, repeated use of the measure would lead to mastery of the procedure.

There is a natural trade-off between using a brief and simple scale versus a comprehensive scale. Nothing would preclude use of the Apgar approach with additional specialized scales that would be useful in certain practice or research settings. Iterative field-testing would be useful to examine reliability across different psychiatric disorders and practice settings, as well as different cultures and socioeconomic groups. To support the validity of the scale, prospective studies should establish whether the scale is sensitive to changes in patients' functioning levels in response to treatment. In a way, this "sensitivity-to change" aspect of the new scale is reminiscent of the Apgar-1 and Apgar-5 (assessing Apgar status immediately at birth and 5 minutes later). Thus, in a consultation psychiatry setting, one could obtain an Apgar rating during the time of the initial consultation as well as after initiating treatment.

A strength of this Apgar-like approach is also its limitation. The GAF mixes all domains of functioning into one index and leaves it up to the clinician to determine where to rate an individual. The Apgar approach, on the contrary, explicitly asks the clinician for a rating on a specific number of domains and provides more guidance than the GAF on how to rate each dimension. The Psychiatric Apgar, alas, is not as elegantly simple as Virginia Apgar's instrument. Our suggested rating of "distress" can never be as precise as her measurement of heart rate, for instance. On the other hand, her rating of respirations ("good, strong cry"

versus “weak cry”) involves the clinician’s subjective assessment, just as do our ratings of distress.

We hope this article will encourage clinicians and researchers to experiment with this approach: to define key domains of functioning that may be particularly important in special settings, such as consultation-psychiatry settings and, even more importantly, to articulate anchors for rating these domains. The anchors (as described in Table 2) “work,” that is, clinicians could use them, and they resulted in significant agreement with the GAF. They might be made to work better with further attention (see, for instance, the anchors for “everyday functioning,” which are clearest for work-related functioning, as opposed to functioning at home).

The GAF was a brilliant addition to psychiatric practice. As we develop the next DSM, it is pertinent to ask whether the GAF approach could be optimized even further by recalling the lessons of Virginia Apgar.

References

1. Finster M, Wood M. The Apgar score has survived the test of time. *Anesthesiology*. 2005; 102:855–857. [PubMed: 15791116]
2. Endicott J, Spitzer R, Fleiss J, et al. The Global Assessment Scale: a procedure for measuring overall severity of psychiatric disturbance. *Arch Gen Psych*. 1976; 33:766–771.
3. Frazee JC, Chicota CL, Templer DI, et al. The usefulness of the Axis V diagnosis: opinions of healthcare professionals. *J Nerv Ment Dis*. 2003; 191:692–694. [PubMed: 14555874]
4. Vatnaland T, Vatnaland J, Friis S, et al. Are GAF scores reliable in routine clinical use? *Acta Psychiatr Scand*. 2007; 115:326–330. [PubMed: 17355524]
5. McKibbin CL, Brekke JS, Sires D, et al. Direct assessment of functional abilities: relevance to persons with schizophrenia. *Schizophr Res*. 2004; 72:53–67. [PubMed: 15531407]
6. Moore DJ, Palmer BW, Patterson TL, et al. A review of performance-based measures of functional living skills. *J Psychiatr Res*. 2007; 41:97–118. [PubMed: 16360706]
7. National Center of Medical Rehabilitation Research. National Advisory Board on Medical Rehabilitation Research. Bethesda, MD: 1992.
8. World Health Organization. World Health Organization Psychiatric Disability Assessment Schedule (World Health Organization/DAS, With a Guide to Its Use). World Health Organization; Geneva, Switzerland: 1988.
9. World Health Organization. World Health Organization Disability Assessment Schedule (WHODAS-II). World Health Organization; Geneva, Switzerland: 2000.
10. McHorney CA, Ware JE, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36) II: psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care*. 1993; 31:247–263. [PubMed: 8450681]
11. Kaplan RM, Atkins CJ, Timms R. Validity of a quality of wellbeing scale as an outcome measure in chronic obstructive pulmonary disease. *J Chron Dis*. 1984; 37:85–95. [PubMed: 6420431]
12. Brooks R. Euroqol: the current state of play. *Health Policy*. 1996; 37:53–72. [PubMed: 10158943]
13. Apgar V. A proposal for a new method of evaluation of the newborn infant. *Curr Res Anesth Analg*. 1953; 32:260–267. [PubMed: 13083014]
14. Crawford, JS. Principles and Practice in Obstetric Anesthesia. 2. F.A. Davis; Philadelphia, PA: 1965.

Figure 1.
Correlation Between Apgar Score and GAF Score
GAF: Global Assessment of Functioning.

Table 1

Apgar Score: Signs and Definitions

Sign	Score		
	0	1	2
Heart rate	Absent	Slow (<100 beats/minute)	>100 beats/minute
Respirations	Absent	Weak cry; hypoventilation	Good, strong cry
Muscle tone	Limp	Some flexion	Active motion
Reflex irritability	No response	Grimace	Cry or active withdrawal
Color	Blue or pale	Body pink; extremities pale	Completely pink

Reprinted with permission from Finster and Wood.¹

Table 2**An Apgar-Like Scale of Functioning for Psychiatric Patients (The Psychiatric Apgar)**

Sign	Score		
	0	1	2
Neurocognition	Severe impairment ^a (e.g., MMSE score <24); cannot function independently	Some impairment (e.g., MMSE score 24–27), but able to function independently or with some assistance	Oriented ×3; no readily-apparent deficits (e.g., MMSE >27)
Distress	Severe impairment: suicidal, manic, incapacitated by symptoms, or cannot function independently	Some impairment due to prolonged distress from mood or anxiety, but able to function independently	Full range of affect; resilience in the face of stress; no current protracted severe anxiety or mood disruption
Psychotic features	Persistent hallucinations that patient cannot ignore; command hallucinations; delusions causing the patient to modify his or her behavior; thought-flow severely compromised	Some hallucinations that patient can disregard; some delusions that patient does not act upon; difficulties communicating thought-flow, but can make himself/herself understood	No evidence of hallucinations or delusions; good information flow
Everyday functioning (e.g., work, education, community, home)	Severe limitations (e.g., inability to work because of psychiatric illness)	Functional limitations such as working at a level below expected for training and background; difficulties with education	Satisfactory functioning commensurate with training and background
Social relationships	Social isolation or chaotic, disruptive social relationships; unable to maintain stable relationships	Has social relationships, but they are not satisfying	Satisfactory social relationships

^aImpairment could be assessed in any number of ways. As an example, the Mini-Mental State Exam (MMSE) could be used.