Journal of Contextual Behavioral Science 6 (2017) 226–229

Contents lists available at ScienceDirect

Journal of Contextual Behavioral Science

journal homepage: www.elsevier.com/locate/jcbs

Brief empirical reports

# Adding a functional utility score to the evaluation of behavioral health screens in integrated care settings: What's all the FUS about?

Alexandros Maragakis[a,*], Jon Nolan[b], Sarah Lindeman[b]

[a] Eastern Michigan University, USA
[b] University of Central Arkansas, USA

A R T I C L E I N F O

*Keywords:*
Functional utility score
Integrated care
Screening

A B S T R A C T

As healthcare delivery is rapidly changing, the rise of integrated care systems is becoming more prominent. A vital pathway of care in these new integrated care systems is the screening and identification of behavioral health issues through the use of efficient and valid behavioral health screens. While research examining the utility of these screens tends to focus on traditional psychometric properties, such as sensitivity and specificity, these fail to measure how well these screens will generalize to medical settings outside the research context. This leads to the creation of behavioral health screens that have adequate psychometric properties but questionable functional utility in the medical setting. The current paper provides a justification and framework for examining the concept of a "functional utility score" (FUS), which assesses a screen's ability to be useful within its intended system of care. The use of such a metric can improve research on behavioral health screens, and ensure that new measures will be able to meet the needs of the integrated care setting.

## 1. The importance for effective behavioral screening in integrated care settings

Provisions in the Affordable Care Act (ACA) and recent calls for action by the World Health Organization indicate a shift in healthcare delivery services to a more holistic, biopsychosocial approach with a specific emphasis on preventative services (Rozensky, 2012; World Health Organization, 2016). Given these new foci, integrated care (IC), which embeds behavioral care providers (BCPs) in the primary care (PC) setting, has received much attention in recent years both in the United States (Rozensky, 2012) and internationally (Goodwin, 2016). It is hypothesized that this system of care delivery, if properly executed, can increase the quality of care received by a patient while simultaneously lowering the staggering cost of healthcare (Lionis & Petelos, 2015; Rozensky, 2012). Given the international shift towards IC systems, it is important the systemic processes required to create the most effective system of care be understood and refined in order to streamline its successful implementation.

A key process in a successful IC system is the use of behavioral health screens that can meet the fast-paced nature of the PC setting (Byrd & Alschuler, 2009; Kroenke & Spitzer, 2002; Parkerson, Broadhead, & Tse, 1990). Specifically, these screens require certain characteristics in order to have *functional utility*, a concept we define as a screen's ability to effectively operate within a system of care and achieve its goal of relaying meaningful information to the PC staff that a behavioral health concern is present. First, screens must be effective in the sense that they are sensitive and specific to detect a broad range of the behavioral health concerns (e.g., depression, anxiety, substance abuse, sleep hygiene issues, medication adherence, smoking, lack of physical activity/diet) (Byrd & Alschuler, 2009), yet be efficient in a way that they do not negatively impact PC work flow by slowing down the staff's ability to deliver care (Robinson & Reiter, 2015). They also must be designed so that other healthcare professionals can easily score and interpret the results (Robinson & Reiter, 2015). Finally, they must provide BCPs with useful information that will have treatment utility (Nelson-Gray, 2003). To address the unique needs of IC systems, many studies have been conducted around creating or testing psychometrically sound measures in the PC setting such as the PHQ-9 (Kroenke & Spitzer, 2002), GAD-7 (Spitzer, Kroenke, Williams, & Lowe, 2006), the Duke Health Profile (Parkerson et al., 1990), and The CAGE questionnaire (Dhalla & Kopec, 2007).

While examining the literature of behavioral health screens for the PC setting, many original screens, while psychometrically sound, were either too long or in some other way too burdensome for PC settings in general (Kroenke & Spitzer, 2002; Parkerson et al., 1990). This lack of functional utility resulted in screens being adapted in various ways (e.g., being shortened, removing complex scoring methods, adapting for automated use), and then reevaluated to assess their psychometric properties (Farzanfar et al., 2014; Kroenke & Spitzer, 2002; Parkerson et al., 1990). While we do not contend the importance of a psychome-

* Correspondence to: Eastern Michigan University, Department of Psychology, Ypsilanti, MI 48197, USA.

http://dx.doi.org/10.1016/j.jcbs.2017.04.008
Received 22 December 2016; Received in revised form 28 March 2017; Accepted 6 April 2017

trically sound screen, we believe that the current methods used in screen development do not adequately capture the larger PC context, resulting in having to alter and reevaluate screens multiple times. This inefficiency may be due to the notion that creating a psychometrically sound screen in one setting is sufficient evidence that it will generalize to others. However, this "one size fits all" practice is generally problematic within healthcare, and has resulted in small or unstainable improvements in practice (Chassin, 2013). To address this issue we believe the use of quality improvement (QI) methods, which promote the analysis of how a particular process operates within a specific system (Maragakis & O'Donohue, 2016), could greatly improve the utility of screens and create a more efficient development and validation process. Specifically, through the use of process maps (Mould, Bowers, & Ghattas, 2010) and project planning (Duffy, 2016), a new psychometric property, the *functional utility score (FUS)* may provide meaningful information for researchers, administrators, and clinicians about the context that a screen was used in and the processes required within a healthcare system to implement the screen effectively.

This paper will describe how to calculate a FUS, followed by how it could have been used with well-established behavioral health screens to identify pitfalls within system implementation and conduct more efficient screen development. We will also discuss how a FUS is consistent with trends in healthcare research and practices.

## 2. How is a functional utility score calculated?

A FUS contains two separate, yet equally important, aspects. The first involves screen designers to report the exact "steps" taken to administer the screen within the system it was tested in; this is typically done in the form of a process map (see Mould et al., 2010). This process should start with where the screen is delivered, and by whom, and end with how a detected behavioral health concern will be dealt with (e.g., appointment with BCP). Fig. 1 displays a hypothetical process of how a screen would be delivered and the training required to ensure it is properly executed. While it may be burdensome to describe the exact system and steps required for a screen, this information can be valuable for the researcher, because it may point to systematic barriers that were not accounted for when conceptualizing the screen or the design. Furthermore, providers and administrators would find a process map useful when determining if they have the infrastructure to ensure that particular screen could be successful within their system. Finally, the creation of a process map and completion of this step is required to engage in the second aspect of the FUS.

The second aspect involves calculating the "functional utility score". This score is obtained by multiplying the proportion of screens that were successfully used in step 1 by the proportion of screens that were successfully used in step 2, and so on, depending on how many steps were required to properly implement the screen as described in the process map. Using the example in Fig. 1, if 300 patients were eligible to receive the screen, and 250 actually completed it in the waiting room, then the score for step 1 ($FUS_1$) would be .83 (250/300). Of the 250 that completed, if the nurse or medical assistant properly scored 175, the score for step 2 ($FUS_2$) would be .7 (175/250). Finally, if 75 individuals indicated a behavioral health concern, and the nurse initiated a warm hand-off for 74, then the score ($FUS_3$) would be .99 (74/75). Therefore, the total score ($FUS_T$) would be .58 (.83*.7*.99).

The breakdown of a FUS by steps is important for multiple reasons. First, as can be seen in step 3, only the subset of individuals who have behavioral health concerns were included the analysis. Given this, simply dividing the total successes in step 3 (74) by the initial amount (300) would not provide the correct total FUS. Second, and more importantly, having a score for each step in the process provides useful information for the researcher and identifies problematic steps that impact the functional utility of the screen. Therefore, instead of discarding a screen due to a low FUS, researchers in this example can examine step 2 and identify barriers that reduced the nurses or medical assistants' ability to score the screen.

Fig. 1 provides a visual representation of how a FUS would be reported and calculated. This overall score, in conjunction with the process map, would allow providers and administrators to compare

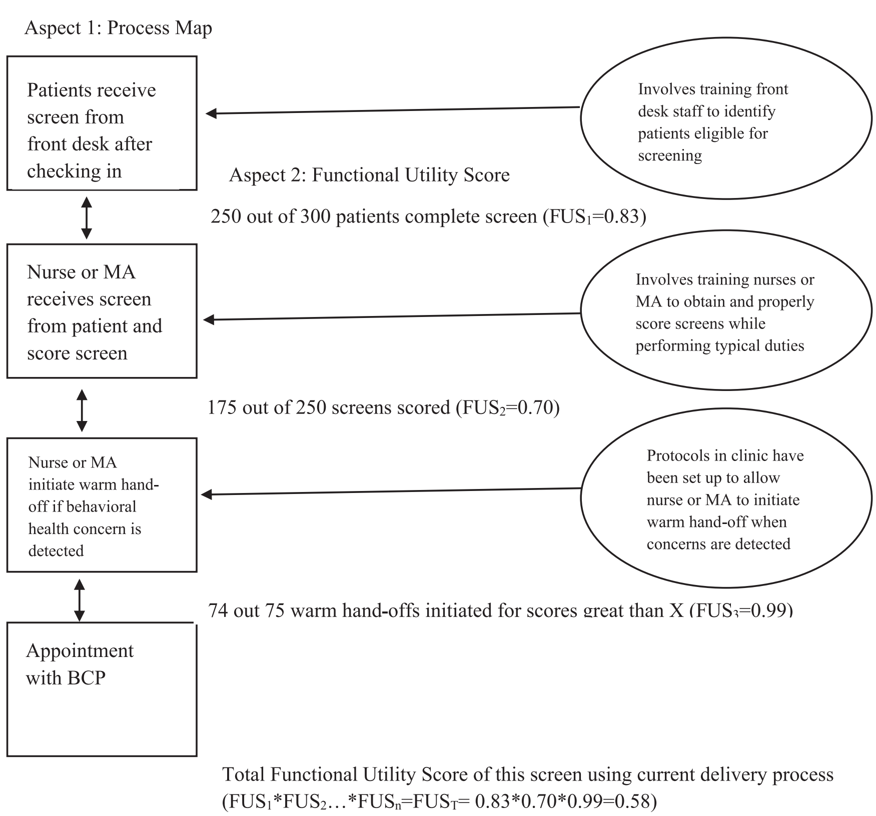


**Fig. 1.** Calculating a FUS.