

## Evaluation of a Novel Digital Biomarker for Stress

### Introduction

Stress is a causative agent for multiple diseases and is also a predictor of worse outcomes for certain existing conditions [1,2]. It is also known to have a profound effect on longevity [3]. In the most recent Stress in America™ survey conducted by the American Psychological Association, forty-two percent of survey respondents said that they were not doing good enough or were not sure if they were doing enough to manage their stress [4]. One of the challenges in stress management is stress measurement. Currently prevalent biomarkers for stress focus on responses from essential stress-sensitive physiological systems like the hypothalamic-pituitary-adrenal axis (cortisol), the autonomic nervous systems (alpha amylase) and the immune system (pro-inflammatory cytokines). The invasive nature of blood collection for these markers, as well as their individual physiological rhythms, make it difficult to track them in a scalable manner outside the laboratory. This paper describes the evaluation of a novel voice technology that has the potential to utilize ubiquitous smartphones as non-invasive stress measurement tools.

### Technology

The Sharecare beta app, launched by Sharecare Inc. in the Google Play Store in Fall 2014, monitors stress indicators present in voice activity. The technology driving the app does not listen to, record, store or analyze the actual content of any call. After each call (Figure 1), the user is presented with a visual representation of stress intensity and a short phrase describing the stress type detected during their call. This is followed by a one or two sentence summary report of their stress as detected by Sharecare's proprietary voice stress analysis technology. In the latest version of the app (3.2.8) the user is presented with a five-star Likert scale to review the stress results presented by the app. The Sharecare beta app also has other health oriented functionality like sleep tracking and a self-reported health risk assessment called the RealAge® Test but the scope of this paper is limited to its voice stress analysis feature set.

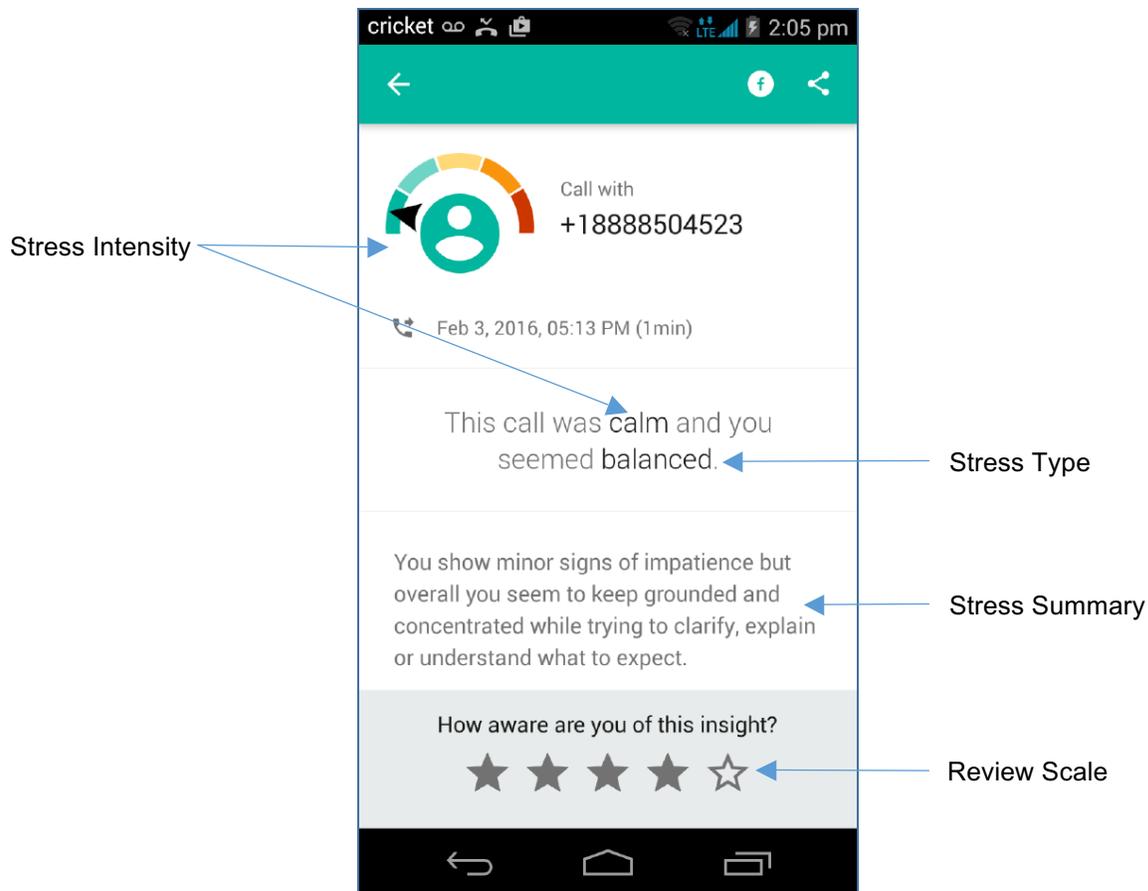


Figure 1

## Importance

Ubiquitous computing and low-cost biosensors are transforming health and wellness monitoring with the development of novel digital biomarkers; however, adoption by providers, payers and consumers requires scientific evidence based on outcomes [5]. Sharecare's voice analysis technology has the potential to become a validated biomarker for stress that is easier to deliver and is available at a fraction of the cost of existing gold standard measures of acute and chronic stress. As the user is not expected to wear a new sensor or alter their phone usage in any significant way to get feedback on their stress intensity and stress type, the technology has tremendous potential to get higher-than-average rates of short- and long-term adherence from end users. This will be especially impactful if high adherence rates can be demonstrated not just for stress monitoring but for stress reduction as well. Finally, the systematic validation of a novel digital biomarker will benefit the broader health-tech community that is interested in the scientific evaluation of other such newly available physiological data streams.

## Literature

Related work on the use of a physiological function to measure and ultimately reduce stress is prevalent in two domains, Biofeedback Therapy (BT) and Mindfulness-based Stress Reduction (MBSR). BT uses non-invasive sensors to measure, amplify and feed back information primarily from nervous system processes such as respiration, heart rate, muscle tension, skin temperature, blood flow and blood pressure, to the individual being monitored, thus promoting awareness of these processes in an individual to assist with their gaining voluntary control over body and mind. MBSR [8] is a well-established 8-week (2.5 hour sessions) mindfulness training that teaches individuals to observe situations and thoughts in a nonjudgmental, nonreactive, and accepting manner. Even though a consensus about an unequivocal operational definition of mindfulness was elusive in our review [6], one of most commonly employed definitions of mindfulness was one by Jon Kabat-Zinn who suggests that mindfulness could be described as a moment to moment awareness that is cultivated by purposefully paying attention to the present experience, with a non-judgmental attitude [7]. Recent meta-reviews of BT [9] and MBSR [10] trials have shown positive trends in reducing different types of stress, depression, anxiety and distress and in ameliorating the quality of life of healthy individuals. Trial designs from multiple randomized clinical trials focused on BT's and MBSR's stress reduction efficacy were used by our research team to arrive at an evaluation plan for Sharecare's voice analysis technology.

## Evaluation Plan

A multi-phase interventional randomized clinical trial is currently underway with the primary goal of measuring the effects of using the Sharecare beta app on psychological measures of stress, quality of life, and biological measurements. A secondary goal of the trial is to isolate the identified effects (if any) to the various voice-related feature sets of the app. The first phase of the trial, which has been approved by the Georgia Institute of Technology Institutional Review Board (Protocol H15247), is focused on the effects of using the app for a ten-week period on a validated stress perception questionnaire called the Perceived Stress Scale (PSS-14). Participants in the intervention group ( $n=1000$  healthy adults) are asked to use the app for a ten-week period or to make and receive a total of 350 phone calls, whichever happens first. The PSS-14 questionnaire is digitally delivered to them after 0, 140, 245 and 350 calls. A similar sized control group that does not use the app and is not part of any other structured stress reduction program, receives the same questionnaire at 0, 4, 7 and 10-weeks for comparison.

## Data Analysis

Sharecare Inc. provided the research team with a digital data log of 98,603 calls made by 1657 unique users from December 2014 through July 2015 to guide the development of a data analysis plan for the randomized control trial. Other than call characteristics like call duration and type (incoming or outgoing) the data collected for every call included proprietary scores for behavioral drivers, energy, stress intensity, stress type, mind stress and overall stress. Captured user data included age and gender which was self-reported during signup as well as system-generated app usage frequency. An exploration of the various proprietary scores in the provided dataset was performed which yielded the following interesting findings:

### Stress Intensity

Average Stress Intensity (as identified by the five categories Calm, Productive, Uneasy, Intense, and Very Intense) was analyzed for the first total 100 calls made by users for all calls, incoming calls and outgoing calls. For all calls (Figure 2), the trend line suggested a decline in average Stress Intensity with a low  $R^2$  value of 0.021. Similarly, for the first 100 outgoing (Figure 3) calls there was a downward trend but the  $R^2$  value of 0.044 was too weak to draw conclusions in stress reduction due to app usage.  $R^2$

value for incoming calls (Figure 4) was 0.001 and did not indicate a reduction in average Stress over time for calls that were received by the user.

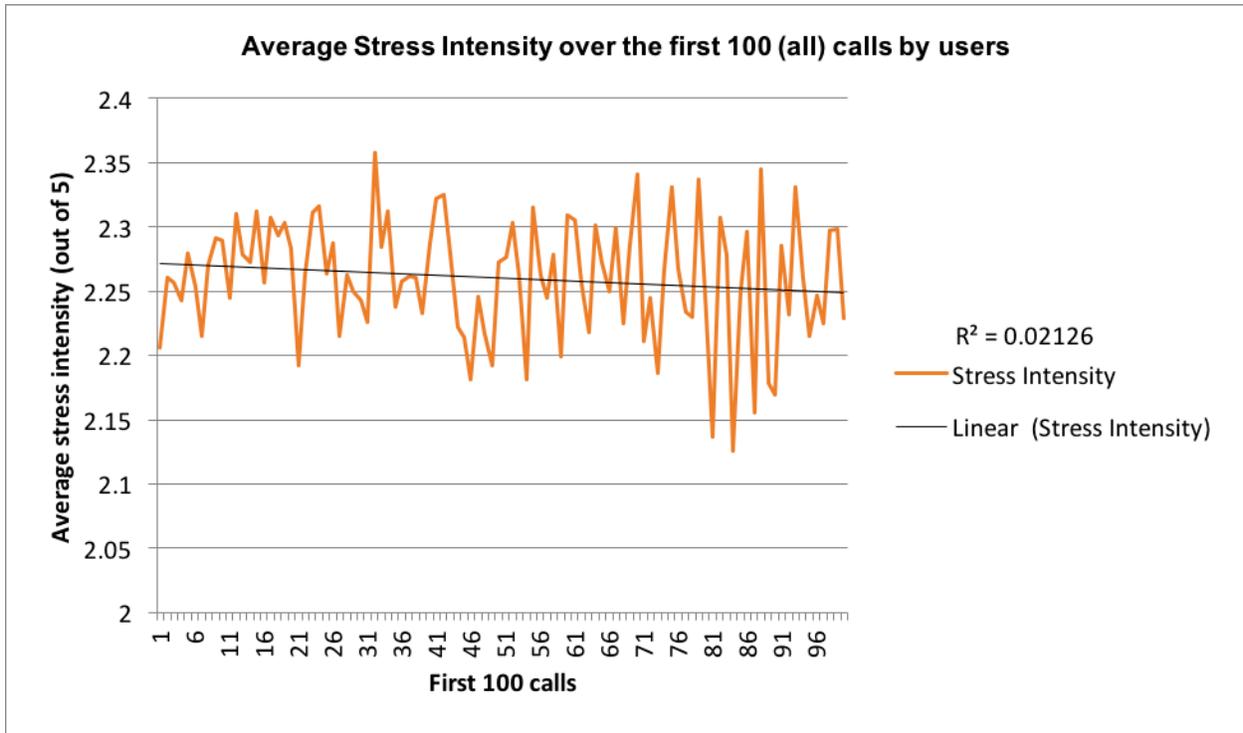


Figure 2

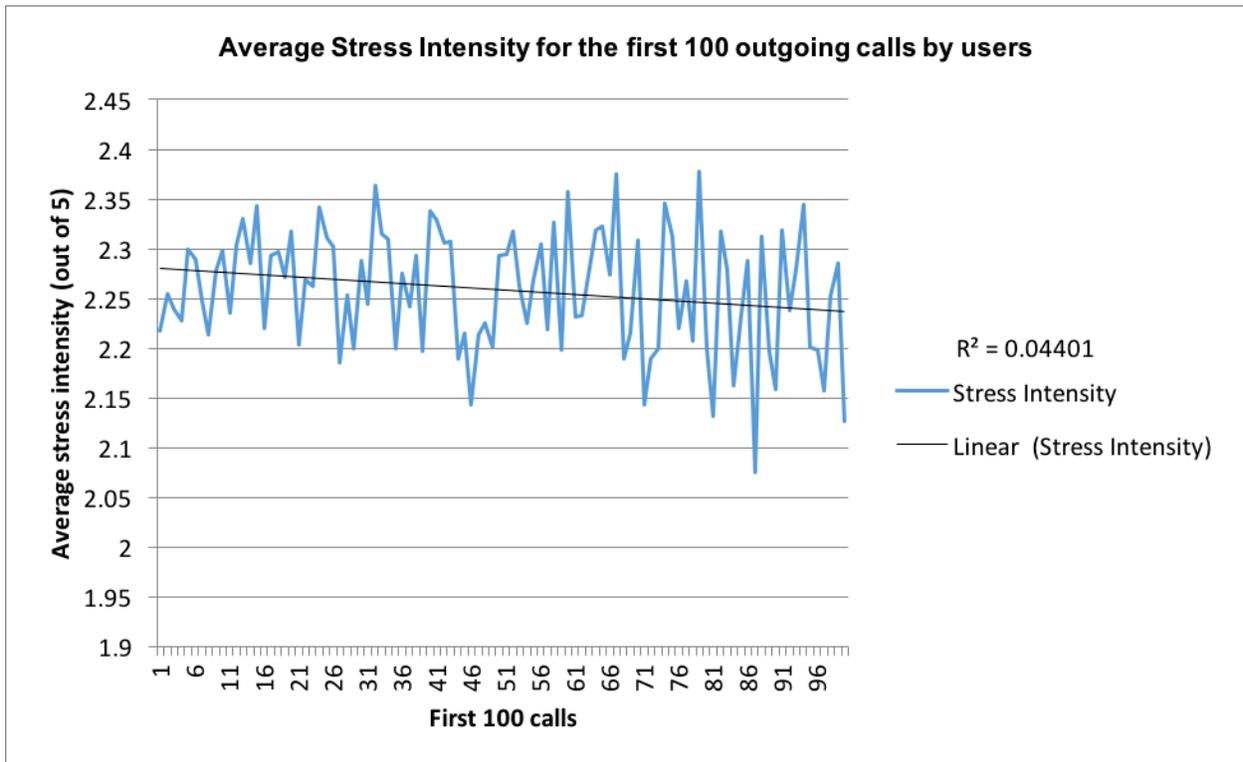


Figure 3

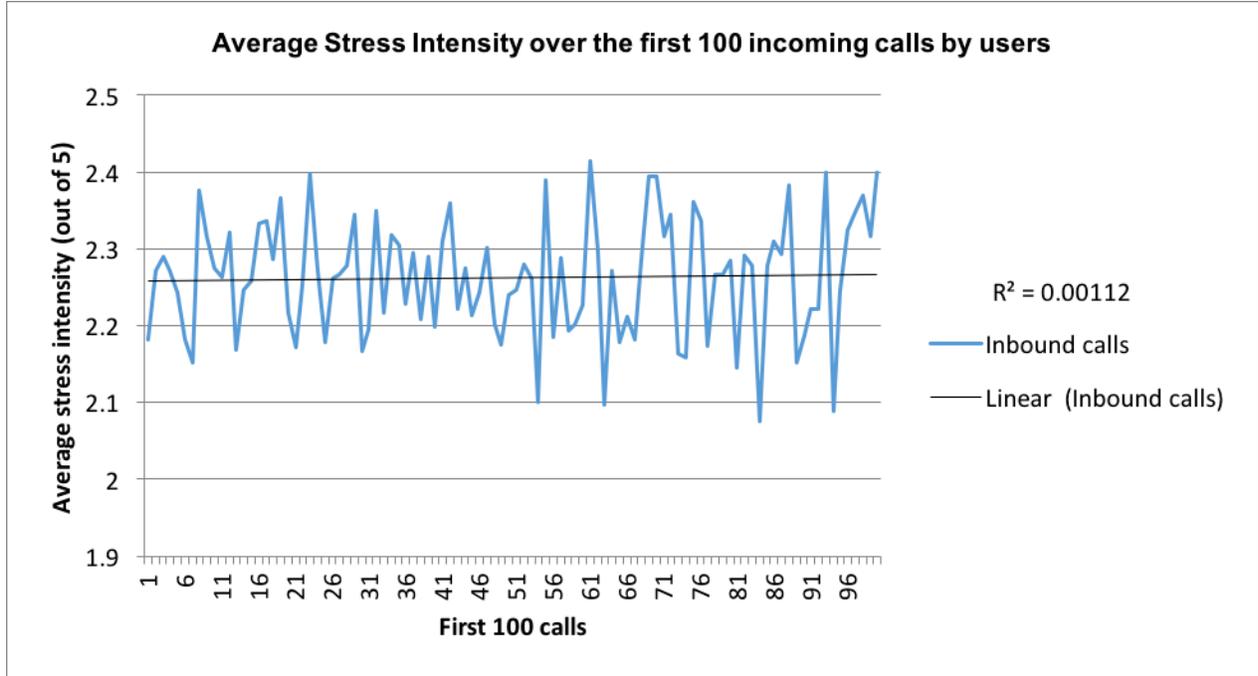


Figure 4

**Behavioral Drivers**

There were statistically significant differences found between incoming and outgoing calls for all three Behavioral Driver scores (Confidence, Emotionality, and Effort to Make Contact). Confidence scores were higher in outgoing calls (mean = 33.46, SD = 10.61) as compared to incoming calls (mean = 32.69, SD = 9.31),  $t(92, 346) = -11.559, p < 0.001$ . Emotionality scores were higher for incoming calls (mean = 33.17, SD = 7.68) as compared to outgoing (mean = 32.23, SD = 8.74),  $t(92,346) = 17.006, p < 0.001$ . Finally, scores on Effort to Make Contact were higher for outgoing calls (mean = 34.31, SD = 9.67) as compared to incoming calls (mean = 34.15, SD = 8.95),  $t(92,346) = -2.616, p < 0.001$  (Figure 5). It is important to note that our sample was skewed towards outgoing calls and equal variances were not assumed. Upon analyzing by gender, men showed a higher average Confidence score in their calls, as compared to women (men = 35.03, women = 31.93),  $t(91,222) = -45.405, p < 0.001$ . Women had higher scores in the Effort to Connect score, as compared to men (women = 35.48, men = 32.40),  $t(48,981) = p < 0.001$ . Note that, in our sample, there were uneven numbers of calls made by women as compared to men (60% of calls made by women). This is not necessarily because women were making more calls. Instead, due to missing qualifier data in the dataset indicating gender of the caller, this discrepancy was generated by lack of gender-identified data. Additionally, the variances were unequal in number of calls made by men and women for both of the behavioral drivers of Confidence and Effort to Make Contact. This could have exaggerated the effect and; therefore, further data collection is necessary to be confident in these conclusions.

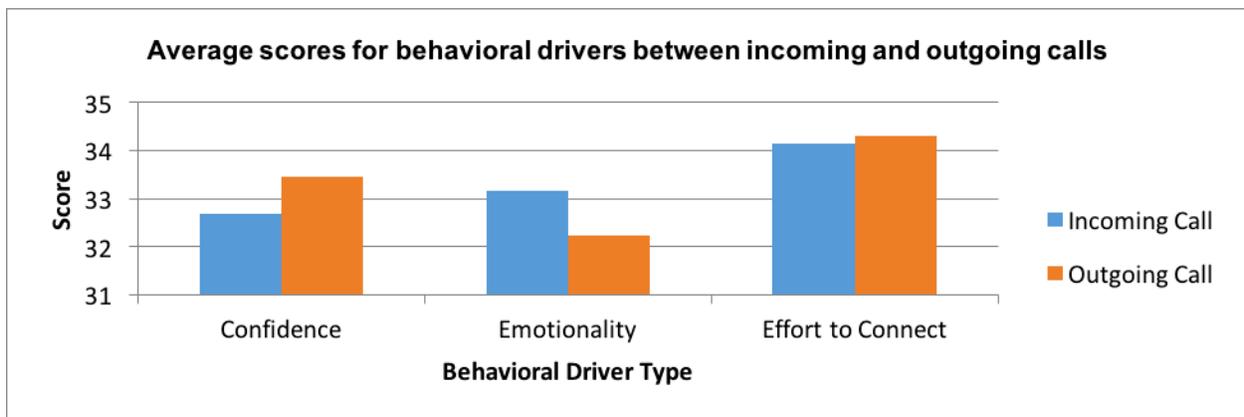


Figure 5

### Data Analysis Plan

For Phase 1 of the ongoing clinical trial we are using the PSS-14 questionnaire as our gold standard measure for psychological stress. Pre-post change analyses between groups is the validation test for the primary outcome which is stress reduction due to app usage. In Phases 2 and 3 self-reported stress symptoms paired with third party biosensor data and biological markers will be used for ground truth measures respectively. Correlational analyses for Phase 1 includes total change in scores for PSS14, Behavioral Drivers, Stress Intensity, Stress Type and Overall Stress. Stress reduction, if any, segmented by user and call characteristics like gender, app utilization, call type (incoming and outgoing), call duration, in-program time and utilization of star-based input for stress insights is also included in this phase.

### Conclusion and Future Work

Due to the multi-system nature of stress, there is no single, quantitative, gold standard scale for stress measurement that novel biomarkers for stress can be validated against with tests of agreement. For this reason, we have chosen an interventional design that is focused on validation through trend analyses for stress reduction in comparison to a control group.

App logs covering a span of 8 months during which 1657 app users made 98,603 calls, were provided to the research team by Sharecare Inc. to guide the data analysis strategy for the ongoing trial. Analysis of this dataset, which did not have a control group, showed a downward trend for the average of a proprietary Stress Intensity score generated by the app for users who made their first 100 incoming and outgoing calls ( $R^2=0.021$ ) and their first 100 outgoing calls ( $R^2=0.044$ ). The trend line for the same score for the first 100 incoming calls alone was almost flat ( $R^2=0.001$ ). Statistically significant changes were seen between incoming and outgoing calls in a multidimensional score generated by the app called the Behavioral Driver score, which is comprised of three sub-scores (Confidence, Emotionality, and Effort to Make Contact). The downward trend observed in the Stress Intensity score for the first 100 outgoing calls when considered in conjunction with the statistically significant differences between incoming and outgoing calls for the Behavioral Driver sub-scores shows a similarity to the stress reduction effects seen in MBSR literature. A broadly accepted definition of mindfulness in interventions that demonstrate clinical efficacy is the act of paying attention in a particular way: on purpose and in the present moment. One hypothesis that could explain the differences between the trends and scores for different call types is that, in the presence of the Sharecare beta app, outgoing calls produce a higher level of mindfulness than incoming calls. While this insight provides the research team with an interesting lens through which to view the data collected in the ongoing trial, the absence of a control group, unequal variances between incoming and outgoing calls and bias due to non-randomization in this dataset limit its conclusive value in and off itself.

In conclusion, we have described our evaluation path for a voice stress analysis technology developed by Sharecare Inc. in this paper. The product under evaluation is a smartphone app that unobtrusively generates a multitude of stress scores, as users make and receive phone calls, which show great promise to become digital biomarkers for stress. Since existing gold standard metrics for stress measurement either require self-reported assessments that leave room for excessive subjectivity or need laboratory procedures whose invasiveness limits their utility, the development and validation of such objective biomarkers presents an exciting opportunity to positively impact consumer health.

### Research Team

Jiten Chhabra, MD MS  
Research Scientist  
IMTC  
Georgia Institute of Technology

Laura Levy, MS  
Research Scientist  
IMTC  
Georgia Institute of Technology

Maribeth Gandy, PhD  
Director  
IMTC  
Georgia Institute of Technology

## References

1. Thorn, B.E., Pence, L.B., et al. (2007). "A randomized clinical trial of targeted cognitive behavioral treatment to reduce catastrophizing in chronic headache sufferers." *Journal of Pain* 8 , 938-949.
2. Krantz, D.S. & McCeney, M.K. (2002). "Effects of psychological and social factors on organic disease: A critical assessment of research on coronary heart disease." *Annual Review of Psychology*, 53 , 341-369.
3. Schneiderman, N., Ironson, G., & Siegel, S. D. (April 2005). Stress and health: Psychological, behavioral, and biological determinants. *Annual Review of Clinical Psychology*, 1, 607–628.
4. American Psychological Association. "Stress in America survey." (2015).
5. Rock Health. "The Emerging Influence of Digital Biomarkers on Healthcare Report." (2016)
6. J Kabat-Zinn. Full catastrophe living: using the wisdom of your body and mind to face stress, pain, and illness: fifteenth anniversary edition Delta Trade Paperback/Bantam Dell, New York, NY (2005)
7. The difficulty of defining mindfulness: current thought and critical issues *Mindfulness*, 4 (2013), pp. 255–268
8. J Kabat-Zinn. Wherever you go, there you are: mindfulness meditation in everyday life. Hyperion, New York (1994)
9. Neurofeedback and biofeedback for post-traumatic stress disorder, generalized anxiety disorder, and depression: a review of the clinical evidence and guidelines [Internet] Ottawa: Canadian Agency for Drugs and Technologies in Health;( 2012)
10. Khoury B, Sharma M, Rush SE, Fournier C. Mindfulness-based stress reduction for healthy individuals: A meta-analysis. *J Psychosom Res.* 2015;78(6):519–28. doi: 10.1016/j.jpsychores.2015.03.009.