HERO GUIDANCE ON CRITICAL REVIEW OF RESEARCH EVIDENCE • APPLICATION TO 2019 JAMA STUDY BY SONG AND BAICKER

EXECUTIVE SUMMARY

In 2014, HERO asked recognized experts to address the question, "Do workplace health promotion (wellness) programs work?"[1] Their collaborative paper provides an overview of the substantial evidence base indicating that well-designed, comprehensive workplace health and well-being (HWB) initiatives integrated into supportive organizations yield high participation rates, sustained behavior change, improved health and workplace performance, and positive economic outcomes. Despite a large and growing evidence base demonstrating positive outcomes of best-practice HWB initiatives and providing guidance on their design, implementation and evaluation,[1] media coverage of occasional studies with negative findings can create confusion among employers about what really works in wellness. This commentary is intended to help HERO members better assess such findings by providing tips on how to critically examine research on program effectiveness. To illustrate how these tips can be used in practice, we apply them here to the 2019 randomized clinicial (RCT) trial by Drs. Song and Baicker of a wellness intervention implemented for employees of BJ's Wholesale Club.[2]

Critical review of published studies should pose questions about the study and underlying HWB intervention rather than relying solely on media stories. When details of the study and intervention are not readily available, related media coverage should be approached with these questions in mind:

	How is this single study similar to or different from previously published studies on HWB program effectiveness?		Did the study use an appropriate evaluation approach?
	Does the tested intervention represent a best practice approach?		Are there indications the researchers had preexisting beliefs about HWB initiatives that may have influenced their research methods or interpretation of results?
	Was enough time allowed for the intervention to meaningfully impact the outcomes studied?		Were there unexpected or important findings not reported in media coverage of the study?

The study by Drs. Song and Baicker represents a rigorous evaluation,[3] but one applied to the first 18 months of a very basic wellness program. For this reason, the results are not generalizable to best-practice approaches combining comprehensive, evidence-based HWB initiatives with broad organizational, cultural, and leadership support. The study found that participation in at least one wellness module lasting 4 to 8 weeks yielded statistically significant improvements in physical activity and weight management behaviors, but did not impact downstream clinical, financial, and business outcomes within the 18-month study timeframe. Although longer follow-up would be required to determine if the initial health behavior changes amongst a relatively small number of participants would ultimately yield population-level clinical and financial outcomes, initiatives associated with such positive organizational outcomes are typically much more comprehensive and include social, environmental, and cultural supports[1] that appear to have been lacking in the organization on which the study was based.[4]

DETAILED REVIEW OF STUDY

The following analysis applies HERO's guidance on critical review of published research studies to the 2019 randomized clinical trial (RCT) by Drs. Song and Baicker of a wellness intervention implemented for employees of BJ's Wholesale Club.[2]

Be skeptical of claims from a single study that appears to refute a large body of previously conducted peer-reviewed research.

Media headlines often imply that a single study proves something is or is not true, so our first tip is to be very skeptical of claims based on one study that appear to refute a large body of well-designed research. In the four decades since the workplace health promotion field emerged, hundreds of studies published in peer-reviewed scientific journals have identified organizational policies, interventions, and cultural supports that yield positive health and business outcomes. Systematic literature reviews have critically evaluated this research, finding strong evidence that well-designed initiatives are effective. One of the most recent reviews—a 2010 meta-analysis by Baicker, Cutler, and Song—concluded that investing in workplace disease prevention and wellness programs yields significant impacts on medical costs and absenteeism.[5] In this context, any new study must be weighed against this large body of previous research and subjected to scientific scrutiny. For example, if a single clinical trial on a promising new drug failed to support its effectiveness, researchers would be cautious about rushing to judgment. Perhaps the drug was not administered in the proper dose; perhaps patients were not compliant with taking it as directed. We must be similarly cautious about jumping to conclusions when a single wellness study fails to report positive results. A single study addresses the merits of one intervention, which may or may not have incorporated evidence-based, best-practice approaches. It is also important to note that lack of a statistically significant finding in one study does not support conclusions often reached by media outlets that wellness programs in general do not produce financial cost savings. It just means the particular intervention tested in that specific study was not robust enough to yield such an outcome within the study's timeframe.

When more than one study appears to reach a similar conclusion, it is important to assess common themes amongst them and contrast these themes against the existing larger body of evidence on wellness initiative effectiveness. The Song and Baicker study represents the second recently published RCT study[6] used to assess a relatively basic wellness program over a similarly short timeframe against an arsenal of potential health, financial, and business outcomes. The remainder of this critical review focuses primarily on the Song and Baicker study while also commenting on the similarities and differences between it and the previously published RCT of the University of Illinois Workplace Wellness Program.[6] A more detailed analysis of the University of Illinois study is available to HERO members on the HERO website.

Determine whether the tested intervention represents a best practice approach.

Best-practice, evidence-based approaches to workplace HWB initiatives are characterized by a comprehensive set of strategies offering all employees—and ideally their family members—a menu of programs, services, and resources that engage them in their own HWB and increase their awareness of the influence of their behaviors on others. These strategies must rely on evidence-based behavior change principles. Best-practice approaches also include organizational policies, workplace environment, comprehensive communications, leadership practices, and social norms that make healthy choices easier.[7-10] No single element or wellness program can effectively meet the variety of HWB needs of all employees in an organization. Multiple strategies are needed and, ideally, each is supported by an established body of evidence demonstrating its effectiveness.

Best-practice initiatives are also supported by strategic planning and ongoing evaluation to identify opportunities to improve upon the programs or their implementation. Some studies that fuel media claims that wellness does not work are based on programs that are poorly designed, poorly implemented, not evidence-based, or are incorporated into unsupportive workplace environments. Every peer-reviewed study includes a section describing the intervention design. Critically assess whether the intervention

aligns with what is known about best-practice approaches. If you are not familiar with best practices, numerous scorecards[11-14] and well-conducted systematic literature reviews[15-18] identify elements of best-practice initiatives. These best practices elements cannot be assembled into a "one-size-fits-all" approach to wellness program design. The basics of program planning start with an assessment of the problem one is solving specific to an organization, and this should include the involvement of stakeholders the program is intended to serve, particularly those who are not likely to be drawn to wellness programs. One study showed, for example, significant variation in how best practices produced differential outcomes based on the demographic differences in the populations studied.[19]

The Song and Baicker study describes the wellness intervention as completion of at least 1 out of 8 modules offered sequentially on a variety of health topics, with each module lasting 4 to 8 weeks. These modules were delivered by registered dietitians assigned to the intervention worksites. Of those employed at any time during the study period, only 35% participated in at least one module and fewer still (21%) participated in three or more modules, with an overall average of 1.3 modules completed. The most highly attended modules were those focused on physical activity, weight loss, and stress management. Tailored and personalized approaches have been demonstrated to improve behavior change outcomes[20-24] but no information was provided to indicate such approaches were used. Health assessment and biometric screening elements were offered to all members of the study population as a way to evaluate health behavior and clinical health outcomes for both participants and non-participants so cannot be considered as part of the intervention. Based on supplemental study information available, the registered dietitians who delivered the wellness modules also led various activities such as group fitness activities and cooking demonstrations, with the flexibility to tailor such offerings around the themes emphasized in the modules. While employees randomized to be offered the educational modules were eligible to receive a $25 BJ's gift card for completing each module, there is no other mention in the paper of other critical best-practice components such as health-supporting policies, leadership support, and comprehensive communication strategies. Moreover, not enough information is provided about the wellness modules to determine if they align with guidelines for evidence-based behavior change initiatives. There is scientific consensus that behavior change programs informed by theory are optimal, [25,26] but the information provided about the wellness modules does not specify the behavior change theory upon which they were based.

Ask whether enough time was allowed for the intervention to produce desired outcomes.

One of the most common mistakes we have observed in evaluations of worksite HWB initiatives is an insufficient evaluation period following the launch of interventions vis-à-vis the outcomes being studied. Some studies include comprehensive interventions but measure results before allowing sufficient time for study participants to complete programs and put newly learned behaviors into practice. Every comprehensive initiative should assess leading and lagging indicators of program success.[27-29] For example, the Song and Baicker study provides insights on who was attracted to participate in certain aspects of first-year program, which represents a leading indicator of program performance. Best-practice process evaluation uses such information to improve the program to ensure lagging measures of performance are likely to be met as the program matures over an extended timeframe. Instead, the Song and Baicker study conducted process, impact and outcome evaluation simultaneously. Furthermore the highest participation rates in the wellness modules were not observed until months 8 and 9 of the study. As a result, most participants were exposed to the intervention with less than a year of follow up. Most new programs suffer deficiencies that need to be identified and corrected before focusing on outcomes several years into the program, and behavior change after exposure to skill-building programs often takes time. Past evidence indicates that some of the outcome measures the researchers examined would not be expected to change in the short 18-month evaluation period.[29,30]

Do not assume all research uses the most appropriate evaluation approach.

Random assignment of study individuals into treatment and control groups has long been considered the ideal scientific approach for evaluating interventions when feasible because it provides the strongest evidence of cause and effect. This RCT, this study design is commonly used to assess the effectiveness

of pharmaceutical drugs and some other medical treatments. However, it is typically impractical to use in workplace health promotion evaluation because most employers are reluctant and may even find it unethical or possibly illegal to withhold valuable programs and information from specific groups of employees.[28] The Song and Baicker study was rigorously evaluated[3] and improved upon the methods used in the University of Illinois Workplace Wellness Study by randomizing worksites to the intervention rather than randomizing individuals. This design is suitable for testing critical social, environmental, and cultural influences on study outcomes. However, none of these critical influencers appear to have been measured in the study, so it is not possible to assess the degree to which they were present in any of the intervention or control worksites. This omission leaves us to question the value of worksite-level randomization in this study.

Be watchful for "confirmation bias."

Confirmation bias is the tendency of researchers to draw inferences from their study that align with their preexisting beliefs but that are not well supported by their data. One of the ways to identify confirmation bias is by looking for study findings within the research study to support each of the assertions or conclusions made by the authors. The Song and Baicker study indicates that, despite limited exposure to wellness content and unusually conservative decisions about statistical controls, significant population-level impacts were observed for physical activity and weight management behaviors. Additional promising trends were identified for smoking, alcohol use, and the overall health behavior index, even though these trends did not achieve statistical significance when subjected to statistical adjustment. Moreover, non-significant reported medical spending trends were directionally supportive of programmatic effect. The non-significant pattern of increased physician visits and decreased hospital utilization was also consistent with desirable changes in the use of care expected after introduction of awareness raising, educational wellness program elements. It is also unfortunate that the researchers did not include the short timeframe for medical impact as a study limitation, since previous research has demonstrated that health care cost impact lags behind health behavior change.[31]

In the University of Illinois study, even though the authors indicated the intervention design was not intended to assess selection bias, they nevertheless focused a substantial amount of discussion in their paper on selection bias and the potential for wellness programs to shift inordinate benefits to healthy employees. Specifically, they paid special attention to their finding that program participants were lower cost health care users and suggested that employers may view wellness as a cost shifting opportunity or a way to attract and retain their healthy employees and, by implication, discourage unhealthy employees from working there. Since these researchers were economists, this may be an understandable example of confirmation bias, an example we viewed as pernicious. One noteworthy finding from the Song and Baicker study is that those self-selecting to participate in the wellness modules had similar health care costs as those who voluntarily chose not to participate. This does not support the cost shifting conjecture that journalists found of interest in the University of Illinois workplace wellness study. Media coverage to date of the Song and Baicker study has not mentioned this non-cost shifting finding and, as our guidance has emphasized, one study does not make for a definitive conclusion. Even if the authors had paid special attention to this finding, it is doubtful journalists would lead with a non-controversial headline such as: "Wellness programs found to be fairly distributed and of interest to high cost and low cost employees alike."

Identify unexpected findings to inform your future approach.

Sometimes research yields lessons or observations not represented in the original research questions. For example, a study's primary goal might be to examine the influence of a HWB initiative on health care costs but leads to discoveries about who is drawn to participate or which program elements most affect program impact. These unexpected discoveries are sometimes important enough to generate media headlines of their own but are overlooked because they are only mentioned briefly when discussing findings. Even if a study's primary research question is not clearly answered, such observations could make the study valuable for future program design or implementation strategies.

CONCLUSION

Research on the effectiveness of workplace HWB initiatives continues to evolve and every new study – whether its results are supportive of HWB or not – must be subjected to critical scrutiny to determine whether its design and findings support media claims.

The Song and Baicker study represents findings from a very limited intervention, and study findings must be interpreted with that caveat in mind. The study found that participation in at least 1 wellness module lasting 4 to 8 weeks yielded statistically significant improvements in physical activity and weight management behaviors but was insufficient to impact downstream clinical, financial, and business outcomes within the first 18 months of the program. Appropriately, the study is ongoing, and we will be following it closely to better understand how the program may evolve toward best practice in a way that is likely to ultimately impact the outcomes being tracked.

Cited References

1. Goetzel RZ, Henke RM, Tabrizi M, et al. Do workplace health promotion (wellness) programs work? Journal of Occupational and Environmental Medicine. 2014; 56(9): 927-934.
2. Song Z and Baicker K. Effect of a workplace wellness program on employee health and economic outomes: A randomized clinical trial. JAMA. 2019; 231(15): 1491-1501.
3. Abraham JM. Employer wellness programs – A work in progress. JAMA. 2019; 231(15): 1462-1463.
4. Heart of Human Capital Blog. Wellness program headlines miss he point: The importance of context in enabling program effectiveness. April 25, 2019. Available at: https://www.heartofhumancapital.com/blog/2019/4/25/wellness-program-headlines-miss-the-point-the-importance-of-context-in-enabling-program-effectiveness
5. Baicker K, Cutler D, Song Z. Workplace wellness programs generate savings. Health Affairs. 2010;29(2): 1-8.
6. Jones D, Molitor D, Reif J. What do workplace wellness programs do? Evidence from the Illinois Workplace Wellness Study. National Bureau of Economic Research. Available at: www.nber.org/papers/w24229
7. Kent K, Goetzel RZ, Roemer EC, Prasad A, Freundlich N. Promoting healthy workplaces by building cultures of health and applying strategic communications. Journal of Occupational and Environmental Medicine. 2016;58(2): 114-122.
8. Kent KB, Goetzel RZ, Roemer EC, McCleary K, Henke RM, Head MA, Fabius R. Developing two culture of health measurement tools: examining employers' efforts to influence population health inside and outside company walls. Journal of Occupational and Environmental Medicine. 2018;60(12): 1087-1097.
9. Henke RM, Head MA, Kent KB, Goetzel RZ, Roemer EC, McCleary K. Improvements in an organization's culture of health reduces workers' health risk profile and health care utilization. Journal of Occupational and Environmental Medicine. 2019;61(2): 96-101.
10. Goetzel RZ, Fabius R, Roemer EC, Kent KB, Berko J, Head MA, Henke RM. The stock performance of American companies investing in a culture of health. American Journal of Health Promotion. 2019;33(3): 439-447.
11. Safeer R, Bowen W, Maung Z, Lucik M. Using the CDC Worksite Health Scorecard to assess employer health promotion efforts: A case study at Johns Hopkins Medicine. Journal of Occupational and Environmental Medicine. 2018;60(2):e98-e015.
12. Weaver GM, Mendenhall BN, Hunnicutt D, Picarella R, Leffelman B, Perko M, Bibeau DL. Performance against WELCOA's worksite health promotion benchmarks across years among selected US organizations. American Journal of Health Promotion. 2016. DOI:10.1177/0890117116679305
13. Goetzel RZ, Henke RM, Benevent R, et al. The predictive validity of the HERO Scorecard in determining future health care costs and risk trends. Journal of Occupational and Environmental Medicine. 2014;56(2):136-144.
14. Grossmeier J, Fabius R, Flynn JP, Noeldner SP, Fabius D, Goetzel RZ, Anderson DR. Linking workplace health promotion best practices and organizational financial performance: Tracking market performance of companies with highest scores on the HERO Scorecard. Journal of Occupational and Environmental Medicine. 2016;58(1):16-23.
15. Sorensen G, Sparer E, Williams JAR, et al. Measuring best practices for workplace safety, health and wellbeing: The Workplace Integrated Safety and Health Assessment. Journal of Occupational and Environmental Medicine. 2018. DOI:10.1097/JOM0000000000001286
16. Goetzel RZ and Ozminkowski RJ. The health and cost benefits of work site health-promotion programs. Annual Review of Public Health. 2008;29: 303-323.
17. Goetzel RZ, Roemer EC, Liss-Levinson RC, and Samoly DK. Workplace health promotion: policy recommendations that encourage employers to support health improvement programs for their workers. Partnership for Prevention, December 2008.
18. The Community Guide. Evidence-based findings on worksite health promotion interventions. Available at: www.thecommunityguide.org/content/evidence-based-findings-worksite-health-promotion-interventions
19. Terry P, Grossmeier J, Mangen D, Gingerich S. Analyzing best practices in employee health management: how age, sex, and program components relate to employee engagement and health outcomes. Journal of Occupational and Environental Medicine. 2013; 55:378–392.
20. Broekhuizen K, Kroeze W, van Poppel MN, Oenema A, Brug J. A systematic review of randomized controlled trials on the effectiveness of computer-tailored physical activity and dietary behavior promotion programs: an update. Annals of Behavioral Medicine. 2012;44(2): 259-286.
21. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, Oenema A. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. Journal of Medical Internet Research. 2011;13(1): e2.
22. Brug J, Oenema A, Campbell M. Past, present, and future of computer-tailored nutrition education. American Journal of Clinical Nutrition. 2003;77(4 Suppl): 1028S-1034S.
23. Kreuter MW, Wray RJ. Tailored and targeted health communication: strategies for enhancing information relevance. American Journal of Health Behavior. 2003;27(Suppl 3): S227-S232.
24. Soler RE, Leeks KD, Razi S, et al. A systematic review of selected interventions for worksite health promotion. American Journal of Preventive Medicine. 2010;38: S237-S262.
25. Michie S, Johnston M, Abraham C, Lawton R, Parker D, Walker A. Making psychological theory useful for implementing evidence based practice: a consensus approach. Quality and Safety in Health Care. 2005;14(1): 26-33.
26. Johnson SS, Paiva AL, Cummins CO, Johnson JL, Dyment SJ, Wright JA, Prochaska JM, Sherman K. Transtheoretical model-based multiple behavior intervention for weight management: effectiveness on a population basis. Preventive Medicine. 2008;46(3): 238-246.
27. Grossmeier J, Terry PE, Cipriotti A, Burtaine JE. Best practices in evaluating worksite health promotion programs. American Journal of Health Promotion. 2010;24(3), TAHP 1-9,iii.
28. Goetzel RZ, Roemer EC, Kent K, et al. Health promotion in the workplace program evaluation. In: O'Donnell M (ed.). Health Promotion in the Workplace, 4th Ed. 2014. Troy, MI: American Journal of Health Promotion: 199-263.
29. HERO, PHA. Program Measurement and Evaluation Guide. 2015. Available at: https://hero-health.org/resources/committee-publications/
30. Grossmeier J, Terry PE, Anderson DR, Wright S. Financial impact of population health management programs: Reevaluating the literature. Population Health Management. 2012; 15(3):129-134.
31. Nyce S, Grossmeier J, Anderson DR, Terry PE, Kelley B. Association between changes in health risk status and changes in future health care costs. JOEM. 2012;54(11): 1364-1373.