



“Needs More Research”—Implications of the Proteus Effect for Researchers and Evidence Adopters

As a researcher, I try to avoid starting the concluding paragraph of the manuscript reporting my study's findings with, “We need more research to better understand....” Every researcher wants to perform the definitive study that provides the definitive answer, and funding agencies prefer not to support researchers to repeat studies again and again. Yet Alahdab and colleagues¹ show us in this issue of *Mayo Clinic Proceedings* that treatment trials published early in the chain of evidence can present exaggerated effects. This phenomenon is termed the *Proteus effect*.^{2,3} Because of the *Proteus effect*, it means that knowing the answer to a question involves a body of evidence, not a definitive trial. Early trials should be viewed with caution. Trials that yield inconclusive and negative findings are important, and even studies that seem to definitively answer a question will need to be replicated, both in a similar manner and with different populations and variations in the intervention.

To objectively measure the Proteus effect, Alahdab and colleagues¹ identified randomized controlled trials that evaluated a drug or device in patients with a chronic condition and which were included in meta-analyses published in the top 10 general medical journals over the past 8 years. They determined the prevalence of having the largest effect size in the first 2 published trials and assessed for associations of an exaggerated early effect with several a priori explanatory variables. A priori explanatory variables included factors known to influence study results beyond being an early study such as sample size, total number of events, study length, duration of follow-up, presence of publication bias, risk of bias, whether trials were stopped early, funding source, number of study sites, and study settings.

What they found from 930 trials published in 70 meta-analyses was that 37% of early

trials had an effect size that was on average 2.67 times larger than the overall pooled effect size. This finding was beyond what would be expected due to random chance ($P < .001$). Although it occurred more often for medications than for procedures and for endocrine than for other chronic conditions, no other explanatory variables related to study design were associated with the presence of the Proteus effect. The authors discussed ways that earlier trials could have greater effects than later trials, including shifting over time from efficacy to effectiveness evaluations, expanding inclusion criteria, reducing intervention intensity, and reducing intervention fidelity.

The findings from this study raise important questions about whether the Proteus effect occurs in other scenarios. Will it occur for prevention and acute care? Will it occur for counseling interventions? Will it occur for diagnostic, prognostic, and observational studies? Will it be similar for continuous outcomes? Is there a way to predict when it will occur or which early studies are at greatest risk for exhibiting the Proteus effect? Assuming these findings apply to other conditions, interventions, studies, and outcomes, the results reported by Alahdab and colleagues¹ have interesting implications.

For researchers, when designing and fielding studies for new interventions, extra caution is needed to limit the risk of bias and ensure valid results. Adequate randomization, effective allocation concealment, and not stopping studies early for benefit can all limit bias. Designing a study for a new intervention to ensure the greatest possibility of success may involve including highest-risk patients with the greatest potential for benefit, substantial efforts to ensure intervention fidelity, and fielding a maximally intensive intervention. However, such strategies should be balanced with a design that can also yield generalizable findings that can be both

See also page 278

TABLE. Frameworks That Could Be Applied to Assessing Early Evidence

Modified USPSTF questions to assess certainty ⁵	Selected Bradford Hill criteria for causation ⁷
1. Does the study have the appropriate research design to answer the question?	1. Strength (effect size)—the larger the association, the more likely that it is causal
2. What is the quality/internal validity of the study?	2. Consistency (reproducibility)—consistent findings in different populations and settings strengthens the likelihood of an effect
3. To what populations and situations are the results applicable?	3. Specificity—causation is more likely if there are no other likely explanations
4. Are there other studies that address the question? How large are the studies?	4. Temporality—effects after the cause with the expected delay are more likely causal
5. How consistent are results across studies?	5. Biological gradient—greater exposure should generally lead to greater incidence of the effect
6. Are there additional factors that assist us in drawing conclusions (eg, dose-response effects, fit within a biological model)?	6. Plausibility—plausible mechanism between cause and effect are helpful
	7. Coherence—coherence between epidemiological, biological, and intermediate outcome findings increases the likelihood of an effect

replicated and put into practice. Early trial reports need to avoid overstating findings and recognize a need to replicate the results. Publications reporting results in both early and later stages of evaluating interventions should systematically report details about the population studied, intervention resourcing, and intervention intensity. These critical details are needed to understand the applicability of findings. Most importantly, findings that represent a substantial change in practice need to be replicated.

For evidence consumers, these findings demonstrate the need for caution when deciding whether to adopt early evidence. A body of evidence is needed to guide practice, and no one study can be taken out of context with that body of evidence. Although evidence-based methods can help to assess the internal validity and risk of bias for a study, the first study demonstrating benefit for a new intervention should most likely trigger further research rather than herald a significant practice change. An interesting challenge can occur if an early study is particularly well designed, has a large sample size, and utilizes an appropriate and generalizable intervention. This scenario is exemplified by the evidence for lung cancer screening with low-dose computed tomography, in which there are 4 randomized controlled trials with published evidence for the intervention and control groups.⁴ Three studies found no reduction in

mortality, but one study did find benefit—the National Lung Screening Trial—and it is 5 times the size of the others combined, dominates the evidence, and has compelling findings that should influence practice. If such a study were to be published as the sole initial evidence, it would be difficult to decide whether to adopt this practice change or wait for further evidence.

Useful frameworks for considering how consumers can assess evidence include the general approach used by the US Preventive Services Task Force for determining certainty^{5,6} and the Bradford Hill criteria for causation⁷ (Table). For interventions with more mature evidence, the size, quality, and consistency of findings from various studies are particularly important. For interventions that only have early evidence, consumers will need to rely more on ruling out other explanations for the observed findings, ensuring plausibility, and determining coherence with epidemiological data, biological findings, and studies that measure more intermediate outcomes to evaluate the evidence. Although these strategies are helpful for assessing causality, they do not replace repeating trials to replicate findings, and the reviewer should have more uncertainty when there are few trials evaluating an intervention.

In response to the findings reported by Alahdab and colleagues, I will use the sentence I try to avoid—clearly, we need more research to understand the Proteus effect. We need to

know how much research is needed before adopting new findings into routine practice, and we need to know how to balance findings from early and later trials to understand the overall and true effect.

ACKNOWLEDGMENTS

Dr Krist is a member of the US Preventive Services Task Force, but this article does not necessarily represent the views and policies of the US Preventive Services Task Force.

Alex H. Krist, MD, MPH

Department of Family Medicine and Population Health
Virginia Commonwealth University
Richmond, VA

Potential Competing Interests: The author reports no competing interests.

Correspondence: Address to Alex H. Krist, MD, MPH, Department of Family Medicine and Population Health, Virginia Commonwealth University, One Capital Sq,

Room 631, 830 E Main St, Richmond, VA 23219
(ahkrist@vcu.edu).

REFERENCES

1. Alahdab F, Farah W, Amasri J, et al. Treatment effect in earlier trials of patients with chronic medical conditions: a meta-epidemiologic study. *Mayo Clin Proc.* 2018;93(3):278-283.
2. Ioannidis J, Lau J. Evolution of treatment effects over time: empirical insight from recursive cumulative metaanalyses. *Proc Natl Acad Sci U S A.* 2001;98(3):831-836.
3. Pfeiffer T, Bertram L, Ioannidis JP. Quantifying selective reporting and the Proteus phenomenon for multiple datasets with similar bias. *PLoS One.* 2011;6(3):e18362.
4. Humphrey LL, Deffebach M, Pappas M, et al. Screening for lung cancer with low-dose computed tomography: a systematic review to update the US Preventive Services Task Force recommendation. *Ann Intern Med.* 2013;159(6):411-420.
5. Sawaya GF, Guirguis-Blake J, LeFevre M, Harris R, Petitti D; U.S. Preventive Services Task Force. Update on the methods of the U.S. Preventive Services Task Force: estimating certainty and magnitude of net benefit. *Ann Intern Med.* 2007; 147(12):871-875.
6. Krist AH, Wolff T, Jonas DE, et al. Update on the methods of the U.S. Preventive Services Task Force: methods for understanding certainty and net benefit when making recommendations. *Am J Prev Med.* 2018;54(1S1):S11-S18.
7. Hill AB. The environment and disease: association or causation? *Proc R Soc Med.* 1965;58:295-300.