Review of Department of Veterans Affairs Monograph on the Economic Impact and Cost Effectiveness of Service Dogs for Veterans with Post Traumatic Stress Disorder

Committee on a Review of Department of Veterans Affairs Monograph on the Economic Impact and Cost Effectiveness of Service Dogs for Veterans with Post Traumatic Stress Disorder

Institute for Laboratory Animal Research

Division on Earth and Life Studies

A Consensus Study Report of
The National Academies of
SCIENCES • ENGINEERING • MEDICINE

THE NATIONAL ACADEMIES PRESS
Washington, DC
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, NW Washington, DC 20001

This activity was supported by a contract between the National Academy of Sciences and the Department of Veterans Affairs under Contract No. 36C24E20C0006. Any opinions, findings, conclusions, or recommendations expressed in this publication do not necessarily reflect the views of any organization or agency that provided support for the project.

International Standard Book Number-13: 978-0-309-67470-6
International Standard Book Number-10: 0-309-67470-0
Digital Object Identifier: https://doi.org/10.17226/26187

Additional copies of this publication are available from the National Academies Press, 500 Fifth Street, NW, Keck 360, Washington, DC 20001; (800) 624-6242 or (202) 334-3313; http://www.nap.edu.

Printed in the United States of America

Suggested citation: National Academies of Sciences, Engineering, and Medicine. 2021. *Review of Department of Veterans Affairs Monograph on the Economic Impact and Cost Effectiveness of Service Dogs for Veterans with Post Traumatic Stress Disorder.* Washington, DC: The National Academies Press. https://doi.org/10.17226/26187.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

The **National Academy of Sciences** was established in 1863 by an Act of Congress, signed by President Lincoln, as a private, nongovernmental institution to advise the nation on issues related to science and technology. Members are elected by their peers for outstanding contributions to research. Dr. Marcia McNutt is president.

The **National Academy of Engineering** was established in 1964 under the charter of the National Academy of Sciences to bring the practices of engineering to advising the nation. Members are elected by their peers for extraordinary contributions to engineering. Dr. John L. Anderson is president.

The **National Academy of Medicine** (formerly the Institute of Medicine) was established in 1970 under the charter of the National Academy of Sciences to advise the nation on medical and health issues. Members are elected by their peers for distinguished contributions to medicine and health. Dr. Victor J. Dzau is president.

The three Academies work together as the **National Academies of Sciences, Engineering, and Medicine** to provide independent, objective analysis and advice to the nation and conduct other activities to solve complex problems and inform public policy decisions. The National Academies also encourage education and research, recognize outstanding contributions to knowledge, and increase public understanding in matters of science, engineering, and medicine.

Learn more about the National Academies of Sciences, Engineering, and Medicine at **www.nationalacademies.org**.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Consensus Study Reports published by the National Academies of Sciences, Engineering, and Medicine document the evidence-based consensus on the study's statement of task by an authoring committee of experts. Reports typically include findings, conclusions, and recommendations based on information gathered by the committee and the committee's deliberations. Each report has been subjected to a rigorous and independent peer-review process and it represents the position of the National Academies on the statement of task.

Proceedings published by the National Academies of Sciences, Engineering, and Medicine chronicle the presentations and discussions at a workshop, symposium, or other event convened by the National Academies. The statements and opinions contained in proceedings are those of the participants and are not endorsed by other participants, the planning committee, or the National Academies.

For information about other products and activities of the National Academies, please visit www.nationalacademies.org/about/whatwedo.

COMMITTEE ON A REVIEW OF DEPARTMENT OF VETERANS AFFAIRS MONOGRAPH ON THE ECONOMIC IMPACT AND COST EFFECTIVENESS OF SERVICE DOGS FOR VETERANS WITH POST TRAUMATIC STRESS DISORDER

Members

SUSAN BUSCH (*Co-Chair*), Yale University, New Haven, Connecticut
BISAKHA SEN (*Co-Chair*), University of Alabama at Birmingham
ANIRBAN BASU, University of Washington, Seattle
MARISA ELENA DOMINO, University of North Carolina at Chapel Hill
DOMINIC HODGKIN, Brandeis University, Waltham, Massachusetts
MIREILLE JACOBSON, University of Southern California
EVAN MAYO-WILSON, Indiana University, Bloomington
MARGUERITE E. O'HAIRE, Purdue University, West Lafayette, Indiana
TODD OLMSTEAD, The University of Texas at Austin
KOSALI SIMON, Indiana University Bloomington

Staff

ANDREA HODGSON, Study Director
ROBIN SCHOEN, Co-Study Director
SARAH KWON, Senior Program Assistant
TERESA SYLVINA, Director, Institute for Laboratory Animal Research Council

INSTITUTE FOR LABORATORY ANIMAL RESEARCH COUNCIL

ROBERT DYSKO (*Chair*) (*after March 2021*), University of Michigan Medical School, Ann Arbor
MARGARET S. LANDI (*Chair*) (*until February 2021*), GlaxoSmithKline, Brentford, United Kingdom
CORY BRAYTON, Johns Hopkins University, Baltimore, Maryland
BARBARA NATTERSON HOROWITZ, University of California, Los Angeles
SONNET S. JONKER, Oregon Health & Science University, Portland, Oregon
ANNE MAGLIA, University of Massachusetts Lowell
SUZAN MURRAY, Smithsonian Conservation Biology Institute, Front Royal, Virginia
GUY PALMER (NAM), Washington State University, Pullman
CHRISTINE A. PETERSEN, The University of Iowa, Iowa City, Iowa
ROSALIND ROLLAND, New England Aquarium, Boston, Massachusetts
CAROLINE ZEISS, Yale University, New Haven, Connecticut

Staff

TERESA SYLVINA, Director
ANGELA (ELLA) BLUE, Program Assistant

Reviewers

This Consensus Study Report was reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise. The purpose of this independent review is to provide candid and critical comments that will assist the National Academies of Sciences, Engineering, and Medicine in making each published report as sound as possible and to ensure that it meets the institutional standards for quality, objectivity, evidence, and responsiveness to the charge. The review comments and draft manuscript remain confidential to protect the integrity of the process.

We thank the following individuals for their review of this report:

CATHY BRADLEY, University of Colorado Denver
ERIKA FRIEDMANN, University of Maryland
SUSAN GERHART, Catch the Vision
ERINN HADE, The Ohio State University
DAVID HOWARD, Emory University
EDWARD NORTON, University of Michigan
ANKUR PANDYA, Harvard University

Although the reviewers listed above provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations of this report, nor did they see the final draft before its release. The review of this report was overseen by **ALICIA L. CARRIQUIRY (NAM),** Iowa State University, and **RICHARD G. FRANK (NAM),** Harvard Medical School. They were responsible for making certain that an independent examination of this report was carried out in accordance with the standards of the National Academies and that all review comments were carefully considered. Responsibility for the final content rests entirely with the authoring committee and the National Academies.

Contents

Summary

The Department of Veterans Affairs (VA) was directed to conduct a study to assess the potential therapeutic and economic benefits of using service dogs for Veterans with posttraumatic stress disorder (PTSD) by the National Defense Authorization Act of 2010 (NDAA). In addition to the study, the NDAA mandated that the VA commission the National Academies of Sciences, Engineering, and Medicine (the National Academies) to review the study, as reported in two monographs. The first VA monograph describing the therapeutic outcomes of the trial is now publicly available, as are the reports from the National Academies documenting an ad hoc committee's review of the iterative drafts. At the completion of that process, a new ad hoc committee, which authored this report, was convened to begin a review of a second draft monograph that presents the economic outcomes requested by the NDAA and a cost-effectiveness analysis (CEA). The VA commissioned the Institute for Clinical and Economic Review to design and conduct the CEA.

This report serves as the committee's review of the VA's second draft monograph. The review focused on assessing the completeness and accuracy of reporting; rigorousness of the study design, conduct, and data analysis; and scientific validity of the conclusions presented within the draft monograph. The committee identified a number of positive elements of the draft monograph recognizing the importance of taking on such a complex randomized controlled trial; nonetheless, the committee identified limitations with several aspects of the analyses and interpretation, which prevent this draft monograph from adequately addressing the main research questions. This report describes the committee's approach to the review process, provides a general assessment of the draft monograph and identifies major deficiencies and recommendations for addressing the committee's concerns.

The committee notes that the generalizability of the results of the trial is limited because the difference between the emotional support dogs used for this study and those used in practice (e.g., purpose breeding; genetic and phenotypic selection for physical health, behavior, and temperament; advanced obedience training). To most accurately interpret the study findings, the draft monograph should include a thorough description and acknowledgment of these differences as limitations of the study and how they may impact the generalizability of the results.

With respect to the study design and randomization, similar to the first monograph, this draft monograph does not follow standard approaches for analyzing clinical trials. For the draft monograph's primary analysis, the committee strongly recommends that the authors evaluate the effect of "assignment to intended treatment" (reanalyzing the trial as designed, with data on all participants randomized to service and emotional dogs, irrespective of whether they were paired with a dog or if they completed the study); the "per protocol analysis" could be an appropriate sensitivity analysis. The use of the intent-to-treat (ITT) analysis would also be useful for the CEA as the current analysis excludes the cost of the added time that participants spent waiting to be paired with a dog. The committee also found that the post hoc pre- versus post-period analysis of the outcomes for the pooled set of participants is not a scientifically rigorous and reasonable approach to answer the research questions, and that the findings of such an analysis do not provide objective answers to the research questions. The committee strongly recommends that the pre-post analyses be dropped altogether from the draft monograph (both chapters) and the inability to compare to usual care (i.e., no dog) should be acknowledged as a limitation of the comparator chosen and the trial design.

Chapter 1 contains a great many regressions with different outcomes, including types of care, types of medications, productivity outcomes, all crossed with assessment periods, etc. The draft monograph should address multiple hypothesis testing in order to provide a coherent and accurate account of the main messages that are important to communicate.

The CEA described in Chapter 2 relied exclusively on a single secondary outcome measure from the trial (self-reported PTSD [PCL-5] scores), despite the existence of two primary outcome measures (World Health Organization Disability Assessment Schedule 2.0 [WHODAS 2.0] and Veterans RAND 12 Item Health Survey [VR-12]) and Clinician Administered PTSD Scale for DSM-5 (CAPS-5) scores. The committee agrees with including PCL-5 scores as one outcome considered in the CEA; however, absent any strong justification, also including the WHODAS 2.0, VR-12, and CAPS-5 scores for the ITT population is recommended.

From the outset, the authors should acknowledge that the CEA was conducted using post hoc methods of analysis. It is unclear how many analyses were undertaken and whether the results in the draft monograph are representative of all of the results obtained. The committee relies on the published best practices guidelines for conducting the analyses and thus notes a number of design issues related to the CEA that need to be addressed, such as the aggregation of the data, forecasting assumptions, and the perspectives chosen for the cost analysis.

The discussions of the major deficiencies and recommendations summarized above are followed by the identification of relatively minor issues and recommendations for addressing them.

1

Introduction

The Department of Veterans Affairs (VA) was directed to conduct a study to assess the potential therapeutic and economic benefits of using service dogs[1] for Veterans with posttraumatic stress disorder (PTSD) by the National Defense Authorization Act of 2010.[2] The NDAA specifically stated:

> "The Secretary shall conduct a scientifically valid research study of the costs and benefits associated with the use of service dogs for the treatment or rehabilitation of veterans with physical or mental injuries or disabilities. The matters studied shall include the following:
>
> (1) The therapeutic benefits to such veterans, including the quality of life benefits reported by the veterans partaking in the study.
> (2) The economic benefits of using service dogs for the treatment or rehabilitation of such veterans, including—
> (A) savings on health care costs, including savings related to reductions in hospitalization and reductions in the use of prescription drugs; and
> (B) productivity and employment gains for the veterans."

In addition to the study,[3] the NDAA mandated that the VA commission the National Academies of Sciences, Engineering, and Medicine (the National Academies) to review the study. In 2014, the VA initiated a randomized controlled trial to "examine the impact of service dogs on disability, quality of life, and PTSD-related symptoms in Veterans with PTSD … also required an economic evaluation, such as the effect on hospitalizations and prescription drug use, as well as productivity and employment" (quoted from the draft monograph[4]). The trial was designed to specifically compare the health and economic outcomes of Veterans who were assigned to receive a service dog compared to those who were assigned to receive an emotional support dog. The dogs used in the trial were intended to be as comparable as possible but for the additional training given to service dogs to perform tasks that mitigate a disability. At the conclusion of the trial the VA produced two monographs that have been submitted to the National Academies for review. In 2020, a National Academies committee began an iterative review of the first monograph, which described the therapeutic outcomes of the trial. The National Academies reports on the review of the first monograph are publicly available (NASEM, 2021a,b), as is the final version of the first VA monograph (VA, 2021). At the completion of that process, the committee authoring this report was convened to begin review of the second draft monograph.[5] This draft monograph presents the economic outcomes requested by the NDAA,

[1]Under the Americans with Disabilities Act (ADA), service animals are defined as dogs that are individually trained to do work or perform tasks for people with disabilities. In comparison, the sole function of emotional support dogs is to provide comfort or emotional support and thus they do not qualify as service animals under the ADA.

[2]National Defense Authorization Act (NDAA) for Fiscal Year 2010. 2009. Pub. L. No. 111-84, 123 Stat. 2190.

[3]For clarity, throughout this report, "the study" will always refer to the clinical trial conducted by the VA. Similarly, reference to "authors" refers to the VA and Institute for Clinical and Economic Review authors of the draft monograph.

[4]The "draft monograph" will always refer to the second VA monograph on economic outcomes, as opposed to the first monograph on therapeutic effects of which a final, public version is available.

[5]This draft monograph is the main reference in this report. Any mention of page numbers, figures, or tables throughout this report refers to the document received from the VA and provided to the committee for review. This draft and all subsequent drafts will be made available to the public upon completion of the committee's review through the public access file.

as well as a cost-effectiveness analysis (CEA). The VA commissioned the Institute for Clinical and Economic Review (ICER) to design and conduct the CEA.

This report serves as the committee's review of the VA's draft monograph. The review was focused on assessing the completeness and accuracy of reporting; rigorousness of the study design, conduct, and data analysis; and scientific validity of the conclusions presented within the draft monograph (the full Statement of Task is provided in Box 1-1). Similar to the review committee of the first monograph, this committee was specifically tasked to review only this trial and the outcomes presented in the VA's draft monograph and not with reviewing the use of service dogs writ large or the wider fields of PTSD, clinical trials, cost analysis, and the CEA. However, the committee draws on the best practices from those fields in addition to its own knowledge and expertise to inform its review.

BOX 1-1 Statement of Task

An ad hoc committee of the National Academies of Sciences, Engineering, and Medicine (the National Academies) will conduct an independent review of a Department of Veterans Affairs (VA) draft monograph related to health economic outcomes of a clinical trial that examined the efficacy of service dogs versus emotional support dogs in mitigating symptoms of posttraumatic stress disorder (PTSD) in Veterans. Specifically, the committee will review Monograph No. 2, which reports the comparative results of health care utilization data and other health economic benefits of the trial and analyzes the cost effectiveness of providing service dogs versus emotional support dogs, along with suggested pricing benchmarks at the population level of service dogs for Veterans with PTSD. The committee will prepare a consensus report that critiques the monograph and addresses the following questions:

- Are the research design and methods well documented, scientifically rigorous, and do they offer reasonable approaches to answer the research questions?
- Does the data analysis systematically apply appropriate statistical and sound reasoning techniques to evaluate the health economic data or cost effectiveness and price benchmarking for the provision of service dogs and emotional support dogs for Veterans with PTSD?
- Do the findings thoroughly report the data analysis and provide factual and objective answers to the research questions?
- Do the findings present original scholarship and discuss principal outcomes of primary research with reliable credibility in a factual and objective way in relation to the research question and existing knowledge?
- Does the draft monograph provide a coherent and cohesive written account and description of the main messages that are important to communicate?
- Does the draft monograph provide clear, appropriate, and accurate graphics of the research results?
- What other significant improvements, if any, might be made in the draft monograph?

The consensus report for the monograph will be subject to the National Academies' external peer-review process. Subsequent revisions of the monograph by the VA may also be reviewed by the committee using the above criteria. These additional reviews may result in an additional consensus report or statement that will also be subject to peer review.

At the conclusion of the committee's review process of the monograph, a statement of completion of review will be provided to the VA, indicating the committee's assessment of whether the monograph is consistent with accepted scientific principles.

COMMITTEE'S APPROACH TO THE TASK AND ORGANIZATION OF THE REPORT

To address the Statement of Task, the committee held a public meeting on January 15, 2021, to discuss the task with VA and ICER representatives. Following this meeting, the committee independently reviewed the draft monograph and subsequently held several virtual closed meetings to discuss the monograph, create a plan to address the task, and come to a consensus on all aspects of the review. These discussions were informed by the draft monograph and all of the associated documents provided by the VA for the committee's review (including all documents provided to the committee that reviewed the first monograph) and for referencing and discussing the relevant literature. A complete list of all of the documents provided to the committee is in Appendix B.

Chapter 2 of this report provides the committee's review. The chapter begins with a general assessment of the study and the draft monograph. This includes some overarching considerations for the generalizability of the findings of the clinical trial. The committee next discusses five major deficiencies related to the study design (in particular the intent-to-treat analysis and the accounting for time variability within the CEA), the analyses included in the draft monograph comparing the pre and post outcomes, multiple hypothesis testing, the choice of outcomes in the CEA, and the overall CEA study design and assumptions. These discussions are then followed by discussion of minor issues specific to each chapter of the draft monograph. Additional minor points are listed in a table in Appendix A.

REFERENCES

NASEM (National Academies of Sciences, Engineering, and Medicine). 2021a. *Letter Report on Review of Department of Veterans Affairs Monograph on Potential Therapeutic Effects of Service and Emotional Support Dogs on Veterans with Post Traumatic Stress Disorder*. Washington, DC: The National Academies Press. https://doi.org/10.17226/26039.

NASEM. 2021b. *Review of Department of Veterans Affairs Monograph on Potential Therapeutic Effects of Service and Emotional Support Dogs on Veterans with Post Traumatic Stress Disorder*. Washington, DC: The National Academies Press. https://doi.org/10.17226/26039.

VA (Department of Veterans Affairs). 2021. *First results of VA study on PTSD and service dogs*. https://www.research.va.gov/ptsd-service-dogs.cfm.

2

Committee's Review

GENERAL ASSESSMENT

In its review, the committee identified a number of positive elements of the draft monograph and study process, as well as a few overarching comments about the generalizability of the clinical trial results that warrant discussion early in this report. Similar to the review committee for the first monograph, this committee recognizes the many challenges associated with conducting a study where the intervention in question is a living being. The Department of Veterans Affairs (VA) research team had to overcome supply, breeding, and training challenges in order to conduct the study effectively. This created variable timelines across all trial participants between the randomization of participants to when they would be paired with either a service dog or an emotional support dog. Additionally, a multi-site trial requires additional logistics in terms of the standardization of practices and timelines as much as possible across all sites. Given the importance of finding additional tools to help Veterans with posttraumatic stress disorder (PTSD), it is commendable that the VA was still able to conduct this milestone trial. The committee appreciates the VA research team's attempt to address a question of deep social relevance using a randomized controlled trial (RCT) with a multifaceted and challenging intervention, the complexity of working with sensitive populations such as military Veterans with PTSD, and commends the appropriate care taken to collect clinical and psychosocial assessments to inform cost-effectiveness analysis (CEA). The committee would also like to applaud the VA for its inclusion of and attention to stakeholder viewpoints and potential contextual considerations in its work. Other positive attributes of the study and the draft monograph include the survey of the study participants, use of administrative data to inform the analyses, inclusion of an impact inventory in the CEA, inclusion of a number of sensitivity analyses used to assess the robustness of the results, and extrapolation of the results across the lifetimes of the participants.

Nonetheless, the committee identified limitations with several aspects of the analyses and interpretation, which prevent this monograph from adequately addressing the main research questions.

Generalizability of the Trial Results

Regarding the generalizability of the trial results to the "real world," the committee notes that there are key differences between the emotional support dogs used in the trial versus emotional support dogs in practice. For the trial, the VA sourced emotional support dogs from service dog providers that provided purpose-bred canines. Both emotional support dogs and service dogs came from the same pool of dogs that were bred over generations to select for traits most suitable to assistance dog roles. Targeted breeding included rigorous selection for physical health as well as behavior and temperament. From birth to roughly 1.5-2 years of age, both groups received the same treatment and basic training. Both groups were required to pass tests for superior physical health, behavior, and temperament. Both groups were trained and passed testing for a variety of advanced obedience tasks. Service dogs were separated out from emotional support dogs with the addition of five PTSD-relevant trained tasks and the right to access any public places as protected under the American with Disabilities Act (ADA). Outside of these distinctions, the two groups were nearly identical, with similar breeding, histories, screening, and preparation. While the first VA monograph states the similarities between the dog groups were intentional given the VA's prior experience, this approach differs significantly from emotional support dogs used in the real world (non-trial settings).

Typically, emotional support dogs are sourced from the same places as companion animals (e.g., rescues, shelters, breeders, unknown sources) and can be obtained at any age. Emotional support dogs are not provided with pre-existing training and do not need to pass any physical, behavioral, or temperament screenings. In usual practice, the only difference between an emotional support dog and a companion animal or "pet" is the need of the *human*, who must obtain documentation from a human health professional of their need for canine comfort or support. Because emotional support dogs are not protected under the ADA, they are restricted from many public spaces and are not required to adhere to any behavior, training, or other standards.

The similarities across the two dog types in this trial and the clear delineation of differences (i.e., the five PTSD-relevant trained tasks and public access rights) are useful for understanding the mechanism behind any estimated differences in outcomes across groups. That is, the comparison in this study might clarify "how" or "why" service dogs' ability to perform disability-related tasks affect (or does not affect) the outcomes of interest. Controlling for non-specific treatment components in this way can be especially important in later-stage research, especially research conducted after efficacy has been established. Controlling for non-specific treatment components can undermine the goals of trials that aim to evaluate "whether" assignment to one intervention is better than the other. That is, the comparison in this study did not answer the pragmatic, policy-relevant research question that it was intended to address, and it likely has limited external validity because it does not reflect the real-world differences between these two groups in practice (e.g., purpose breeding; genetic and phenotypic selection for physical health, behavior, and temperament; advanced obedience training). To most accurately interpret the study findings, the committee recommends including in the draft monograph a thorough description and acknowledgment of these differences as limitations of the study and how they may impact the generalizability of the findings of the study.

STUDY DESIGN AND RANDOMIZATION

Intent-to-Treat Analysis of Chapter 1 of the Draft Monograph

Similar to the first monograph, this draft monograph does not follow standard approaches for analyzing clinical trials. The analyses and results presented in the draft monograph are not sufficiently rigorous and do not reasonably answer the research question. For the monograph's primary analysis, the committee strongly recommends that the authors evaluate the effect of "assignment to intended treatment"; the "per protocol analysis" could be an appropriate sensitivity analysis.

As noted in the review of the first monograph, an intent-to-treat (ITT) analysis, which considers outcomes for all randomized participants, is the gold standard approach to inference in superiority trials such as this one (Ranganathan et al., 2016; Schulz et al., 2010). Randomized trials minimize selection bias in the groups as assigned, and post-randomization exclusions from the assigned groups can introduce selection bias. Thus, it is a basic principle of trial design that outcomes should be collected for all assigned participants—regardless of intervention receipt or responsiveness to treatment—to obtain an unbiased estimate of the effect of intervention assignment (i.e., by comparing the assigned groups). The analytical approach described in the draft monograph breaks the initial random assignment in some ways; thus, this approach does not fully adhere to random assignment as stated on page 19 of the draft monograph. In fact, the analyses ignore a substantial number of participants who were assigned to each group but left the study prior to pairing with a dog. Attrition prior to pairing has the potential to bias all subsequent outcomes. Of the 227 participants randomized to a dog, only 181 were paired. Importantly, the attrition differed across the two groups. Pairing occurred for 85% (97 out of 114) of participants randomized to a service dog but only 74% (84 out of 113) of participants randomized to an emotional support dog. The higher pre-pairing attrition for the emotional support dogs is especially noteworthy given the longer training period and thus wait times for a service dog. This suggests that those paired with a service dog were more dedicated to staying in the study—even though they incurred a higher "cost" in terms of wait time. This also suggests that attrition did not happen completely at random. By conditioning on pairing, a mediating variable in the

causal pathway, the analyses as presented in the monograph are likely to be biased. These differences were exacerbated by post-pairing terminations: 77% (88 out of 114) of participants randomized to a service dog completed the study but only 58% (65 out of 113) of participants randomized to emotional support dogs completed the study. Different amounts of missing data, and different reasons for missing data, suggest that between-group differences (e.g., in symptoms and quality of life) might be biased because of differential attrition.

It is a limitation that the draft monograph uses administrative data only for participants who received the intervention (i.e., the "per protocol" population), which is not an appropriate technique to evaluate the cost effectiveness of the intervention. Unlike the missing self-reported measurements, administrative data should be available for all of the assigned participants. Those data could be used to obtain valid estimates for the true ITT population. To adhere to the planned design of this study, and to produce valid estimates of the effect of the assignment, the authors should analyze the actual values for all outcomes for all randomly assigned participants whenever possible. At present, the draft monograph tries to address post-pairing attrition using monotonic multiple imputation models. A true ITT would include all randomized participants using the observed data whenever available, thus accounting for both pre- and post-pairing attrition. While the committee recognizes that incorporating administrative data for the full sample might require amending an existing institutional review board protocol, this effort would enable the team to substantially improve the validity of the analyses. With the per protocol analysis, one can easily construct plausible scenarios that raise questions about the findings. For example, one of many possible scenarios that could generate a null result in the per protocol analysis is if the likelihood of pairing were positively related to improvements in health and thereby health care utilization and spending. In this scenario, the differential attrition for emotional support dogs could bias the per protocol analysis toward zero (i.e., no difference in health care utilization or spending for the two groups) even if service dogs pairing actually improved participant outcomes relative to emotional support dogs. While VA administrative data are not a panacea because they would not capture care received outside of the VA, a true ITT analysis would help put this type of speculation to rest and add considerable credibility to the study. In addition, the draft monograph indicates that the vast majority of health service use is within the VA. If is not possible to obtain identifiable data, an alternative is to request a limited dataset with health care utilization and spending as well as treatment assignment but no identifying information. Given that these types of limited datasets are routinely used by health services researchers at the VA, this seems like a feasible way to address bias.

The committee cannot overstate the value of reanalyzing the trial as designed, with data on all participants randomized to service and emotional dogs, irrespective of whether they were paired with a dog or if they completed the study. An unpublished study protocol[1] includes appropriate methods for causal inference and for dealing with threats to validity because of missing data. Specifically, the protocol indicates that the purpose of this study was to evaluate "the effect of assignment to intended treatment" and that the analyses would include all randomized participants. Moreover, the protocol indicates that missing data would be imputed for this population, which would properly account for uncertainty due to missing data. As with the first monograph, the methods used in this draft monograph to deal with missing data are not appropriate because there is a risk of selection bias in the data collected and because post hoc changes to the analyses increase the risk of bias because of the selective non-reporting of results (NASEM, 2021).

Refocusing the CEA to Better Account for Time in Chapter 2 of the Draft Monograph

The ITT approach recommended for Chapter 1 of the draft monograph would also be useful for the CEA. At present, the CEA focuses on the effects of pairing with service dogs versus emotional support dogs with the analysis beginning at the point of pairing, thus excluding the cost of the added time that participants spent waiting to be paired with a dog. The committee strongly recommends that the draft monograph provide the distribution of time to pairing under both arms of the trial. The first VA monograph

[1]All references to the study protocol in this report refer to the protocol provided to the committee by the VA. See Appendix B for a complete list of the documents provided to the committee.

indicated that there is a range of times that participants waited, but this was not broken out by treatment assignment. If sourcing and supply constraints led to greater delays in pairing with a service dog compared to an emotional support dog, then the overall evaluation should account for the differences in time and begin at the point of randomization (i.e., when the demand for a dog is made). For the same reason, a randomized participant who is never paired with a dog should *not* be dropped from the analysis. The availability of baseline values of health outcomes at randomization and the estimates of health care utilization during the many months prior to pairing should make such an analysis feasible. This change in focus of the CEA is recommended because, theoretically, a program under which a less efficacious intervention can be accessed early and often may become cost effective compared to a program where a more efficacious intervention is accessed late or never used. In general, both demand- and supply-side constraints should be part of economic evaluations (Vassall et al., 2016; Wright et al., 2019).

Other Related Minor Points on Randomization and Pairing

This draft monograph is relatively light on details of the sample and study design. Importantly, it does not include the standard flow diagram for a randomized clinical trial. These details are important for considering any potential bias due to exclusions. This is true in general and particularly in the case of a complex study such as this one. The committee recommends that the revised draft monograph include Figure I from the first monograph and the related discussion of the reasons for missing data.

Related to data exclusions, the draft monograph refers to the main analysis on page 19 as a "modified intent to treat" or a per protocol analysis (sometimes written as "per protocol"). While the term "modified intent to treat" has been used in publications of randomized trials, the term is ambiguous and has been used to describe a wide range of exclusions and deviations from protocol (Abraha, 2010; Ioannidis et al., 2004; Montedori et al., 2011). The analysis is better described throughout as a "per protocol analysis" where pairing with a service (N=97) or emotional support (N=84) dog is the protocol. As stated above, the committee considers this analysis inappropriate for the primary analysis and strongly recommends using the ITT approach or the main analysis; the per protocol analysis would be an appropriate sensitivity analysis.

NON-RANDOMIZED SECTION COMPARING PRE AND POST OUTCOMES

The findings from the RCT were accompanied by a post hoc pre- versus post-period analysis of outcomes for the pooled set of participants. The committee finds that this is not a scientifically rigorous and reasonable approach to answer the research questions, and that the findings of such an analysis do not provide objective answers to the research questions. The committee strongly recommends that the pre-post analyses be dropped altogether from the draft monograph (both chapters) and the inability to compare to usual care (i.e., no dog) should be acknowledged as a limitation of the comparator chosen and the trial design. This could be identified as a critical future research direction to address the needs of Veterans with PTSD.

The committee was tasked with reviewing a CEA based on the results of an RCT, and this pre-post study is outside the scope of the RCT. Including the pre-post analyses alongside an RCT (comparing the assignment of a service dog versus an emotional support dog) can give the false impression that the pre-post study design provides the same level of evidence regarding causal inference as the RCT. Because dogs helping Veterans with PTSD to attain better health would be an emotionally compelling story, there is the risk that the pre-post analyses could be taken out of context and exaggerated in spite of any disclaimers the authors may provide. The fact that the authors devote much more space in the Discussion section of the draft monograph to the pre-post analyses than to the results of the RCT further exacerbates this risk. The summary of Chapter 2 notes that the "no dog study" is hypothetical but it is not explicit that there was no randomized "no dog group" in the VA trial. Again, equal space is devoted to the results from the RCT and the results from the hypothetical no dog analysis. The standard caveat about not taking the results as causal does not dispel these concerns.

As noted in the review of the first monograph (NASEM, 2021), RCTs are designed to assess the effect of intervention assignment by estimating the difference between assigned groups. RCTs are not designed to estimate within-group differences (e.g., between baseline and post-treatment). This study was designed as an RCT to compare the benefits of the assignment of a service dog versus an emotional support dog. The authors speculate that both service dogs and emotional support dogs may provide health benefits, but in the absence of a randomized "usual care" control group, it would be impossible to imply any causal interpretation of this finding.

The committee emphasizes the removal of the pre-post analyses from this draft monograph because it falls outside the scope of the RCT for the reasons stated above. Such analyses would have their own sets of requirements, including acquiring additional data (i.e., administrative data for individuals not included in the trial that will serve as a valid "usual care" control group). If the VA decides to create a *separate* manuscript with the pre-post analyses, then the study authors should consider the following points:

- The study should be framed as an observational study. Furthermore, making better use of the 18 months of "pre-pairing data" might help inform whether costs were already trending downward pre-pairing. For example, 90-day graphs similar to Figure 1.1 and 1.2 (p. 28, Section 1.3.2), but pooling the service dog and emotional support dog groups, could be used to show trends in costs at 90-day intervals for the 18 months in the pre-period.
- Ideally, the estimates for a matched no dog control group would include study participants who were never paired, rather than from participants who were eventually paired with a dog. This is because the anticipation of being paired in the near future could influence health and health behaviors, which prevents such participants from being a proper control group.
- If the time-to-pairing analysis shows considerable variation in time across participants and that variation is described to be exogenous (e.g., potentially as a function of the sourcing and supply of dogs and not correlated with the severity of PTSD symptoms), then the staggered implementation could be leveraged to assess changes in health outcomes over time without a dog (among those who are assigned a dog, acknowledging that there may be anticipation effects). One concern with this approach is that the long-term trajectory among beneficiaries waiting for a dog may not generalize to the population with no dog in a real-world setting.
- Include a discussion of the well-known phenomenon of "regression to the mean," where high-cost, severely ill patients may, on average, improve over time without additional interventions (e.g., Linden, 2013).
- Include a discussion of improvements that may occur with "usual care." All of the study participants were enrolled in mental health care throughout the study period, and more discussion of the types and amounts of care received would be appropriate. Again, without a no dog comparison group, it is impossible to decipher whether a comparable decline would have occurred if the Veterans had received just the usual care. Prior studies, including studies focused on Veterans, have shown that while special interventions *appear* to improve health outcomes and reduce costs over time, these changes are actually not statistically different from outcomes experienced by a control group receiving the usual care (e.g., Finkelstein et al., 2020; Peikes et al., 2009; Zulman et al., 2017). Improvements seen in the pre-post analyses may be entirely independent from the receipt of a dog, or receipt of the dog might even be harmful.
- The authors currently report using linear regression models with person-level random effects. Because the pre-post analyses do not involve randomization, the use of a random effects model requires the assumption that the random effect in the error term is uncorrelated with other included variables in the model, which appears to be a strong assumption. The Hausman test does not address whether this assumption holds. Consider a fixed effects model as a more acceptable method for controlling for individual-level, unobserved confounders.

MULTIPLE HYPOTHESIS TESTING

Chapter 1 contains a great many comparisons, including types of care, types of medications, productivity outcomes, all crossed with assessment periods, etc. However, the committee did not find any attempts made to adjust for multiple comparisons. Without such adjustments, one might give undue attention to results that are "statistically significant" purely by chance due to a large number of outcomes and models being estimated. It is important to account for multiple hypothesis testing even if all of the outcomes are fully registered, which is not the case in this study as described in the review of the first monograph (NASEM, 2021, pp. 21-22). The draft monograph should address multiple hypothesis testing in order to provide a coherent and accurate account of the main messages that are important to communicate. One often used method for such situations that could be applied is to adjust the p-values of each analysis using the Bonferroni correction. Vickerstaff et al. (2019) provides a discussion of this and other formalized adjustment methods. Regardless of whether such correction is used, the draft monograph should indicate the number of analyses conducted and results calculated, and make available the complete results in tables or in an appendix. The interpretation of findings should consider the totality of the evidence and not give undue consideration to "statistically significant" results, especially if the threshold to determine "significance" does not account for multiple comparisons or was developed post hoc.

CHOICE OF OUTCOMES USED IN THE CEA

The choice of outcomes for the CEA, and the use of a per protocol analysis, do not reflect sound reasoning techniques to evaluate the cost effectiveness of service dogs and emotional support dogs for Veterans with PTSD. NASEM (2021) highlighted the need for clarity concerning changes to the outcomes and other differences between the trial registration and protocol(s) and the final report (NASEM, 2021). The CEA described in Chapter 2 relied exclusively on a single *secondary* outcome measure from the trial (self-reported PTSD [PCL-5] scores), despite the existence of two *primary* outcome measures (World Health Organization Disability Assessment Schedule 2.0 [WHODAS 2.0] and Veterans RAND 12 Item Health Survey [VR-12]) and Clinician Administered PTSD Scale for DSM-5 (CAPS-5) scores. The committee agrees with including PCL-5 scores as one outcome considered in the CEA (but see concerns provided below); however, absent any strong justification, also including the WHODAS 2.0, VR-12, and CAPS-5 scores for the ITT population is recommended for the following reasons:

- WHODAS 2.0 and VR-12: As noted in the Abstract in the first monograph, the objective of the trial was to "determine whether *overall disability and quality of life* for Veteran participants in treatment for PTSD were improved." Moreover, as noted in Section D (Outcome Measures) of the first monograph, "[t]he research team designed the study so that outcomes would be assessed in terms of impacts on *overall mental, social, and psychosocial function*." The two primary outcome measures selected for the trial were scores on WHODAS 2.0 (to measure overall disability) and VR-12 (to measure health-related quality of life). Thus, it is important to also include these measures in the CEA, regardless of whether the trial found significant differences between groups on either of these measures. In general, it is appropriate to include outcomes with non-significant differences in the CEA (Briggs and O'Brien, 2001; Glick et al., 2015).
- CAPS-5: CAPS-5, like PCL-5, is used to assess the severity of PTSD symptoms and to diagnose PTSD. Unlike PCL-5 (which is self-reported), CAPS-5 is administered by a clinician. CAPS-5 was administered twice during the trial—at screening and month 15—and no differences were observed between groups in changes from screening to 15 months in CAPS-5 scores (in total or for any of the subscales).

Thus, it appears that the evidence is mixed regarding the relative impact of assignment to receive a service dog compared to an emotional support dog on quality-adjusted life-years (QALYs) and symptoms of PTSD for the per protocol population, and it is unclear what the evidence might be for the ITT population.

Recognizing that service dogs are already unlikely to be cost effective compared to emotional support dogs based on changes in PCL-5 scores, and that including VR-12 and CAPS-5 would make it even less likely, it is important to include these additional outcomes in the CEA. Doing so would avoid giving the misleading impression that the evidence is unequivocal regarding the relative effectiveness of receiving a service dog (versus an emotional support dog) with respect to QALYs and PTSD symptoms. And, as noted above, the committee thinks it is important to include WHODAS 2.0 scores inasmuch as it was one of only two primary outcome measures in the trial and the only primary measure to address overall disability.

Finally, the limitations section should acknowledge that the study used an ad hoc method to convert changes in PCL-5 scores to changes in utility weights. Specifically, the study (a) relied on an analysis reported in Freed et al. (2009) that regressed PCL-4 scores against the 36-Item Short Form Survey (SF-36) derived utility weights (which were, in turn, estimated using an algorithm developed to convert SF-36 responses to utility weights), and then (b) based on written communication from K. Magruder (p. 64 of the draft monograph), multiplied Freed et al.'s finding by 0.85 to account for differences between PCL-5 and PCL-4. It is not clear to the committee that this ad hoc method of translating changes in PCL-5 scores to changes in utility weights is valid, and this limitation should be noted.

CEA STUDY DESIGN AND ASSUMPTIONS

From the outset, the authors should acknowledge that the CEA was conducted using post hoc methods of analysis. To the committee's knowledge, no formal analysis plan was developed, published, or submitted to a public registry before seeing the data or carrying out the analyses. Consequently, it is unclear how many analyses were undertaken and whether the results in the draft monograph are representative of all of the results obtained. The committee is not able to evaluate the presence or absence of selective reporting. Therefore, all of the comments from the committee in this section will rely on published best practices guidelines for conducting the analyses (Drummond et al., 2015; Sanders et al., 2016) and not necessarily on the choice of specific outcomes studied.

There are a number of design issues related to the CEA. These include decisions about the aggregation of the data, forecasting assumptions, and the perspectives chosen for the cost analysis. This section describes each of these flaws with suggestions for addressing each.

The Use of Individual-Level Data

Although not explicitly stated, the CEA appears to take a decision-analytic approach by collapsing study data into parameters reflecting key components, such as mean component costs, mean outcomes, etc. While this is not necessarily an incorrect method of conducting a back-of-the-envelope estimate of costs and benefits, it misses an opportunity for greater precision of the estimates through information on the multi-collinearity of variables at the individual level, using a statistical approach based on individual-level data. For example, in the section on time horizon and extrapolation below, the oversimplification of the service life of a dog is described, estimated at 7 years. While this may be true on average, individual-level data on study participants would include the actual length of service with one or more dogs during the study period, which would provide more information on the relationship between the length of dog service and other study outcomes. Because these data had to be created for trial records, the committee suggests that the CEA is conducted using these regression-based estimates using individual-level data, recognizing that this may alter the needed data-sharing arrangements for the CEA team.

Cost Analysis Perspectives Used

Chapter 2 of the draft monograph reflects analyses using three separate perspectives: a comprehensive health system perspective, a VA system perspective, and a modified societal perspective. Establishing different perspectives for a CEA is important because it guides the question of "costs to whom?" recognizing that some costs are outside the domain of different decision makers. Two of these perspectives are reflected in Table 2.A1: Impact Inventory (Prosser et al., 2016), which helpfully indicates which cost components are used for which of the two perspectives reflected in the table (health care sector and societal). The third perspective (VA system) was not included in the Impact Inventory but should be.

The terms in this section could be better defined, identifying which analysis is from which perspective. It is also strongly recommended that the authors drop the "modified" societal perspective and conduct the analysis from a full societal perspective, even if such analysis remains incomplete due to the unavailability of certain data. An incomplete analysis from the societal perspective may still be useful if there is a clear description of the implications. The committee recommends that the authors (a) use the impact inventory to catalog all relevant components from the societal perspective, (b) come as close as possible to including all of these components (either by using existing data that were excluded from the CEA [e.g., health care utilization, work productivity] and/or estimating new components [e.g., participant time and travel costs for services]), explicitly noting any limitations of the data used, and (c) discuss how excluding these components would affect the incremental cost-effectiveness ratio. Recommendations for each perspective are discussed below, in turn.

***The societal perspective*:** This perspective is the gold standard for the CEA as it encompasses a broad range of inputs that could be affected by an intervention. The authors use a "modified" societal perspective, but do not indicate how this differs from a true societal perspective. The authors should attempt to come as close as possible to a societal perspective, even if that means estimating factors that do not currently seem to be included, such as participant time and travel costs for services, including dog training and pairing, and out-of-pocket costs of service use. In addition, the societal perspective should include the opportunity costs of "producing" each type of dog, even if the dogs are acquired through donations. Currently, training costs (i.e., the 1-2 week onsite pairing and training period) incurred by study participants and their caregivers are not included in the analysis, but should be included in the societal perspective. In addition, following the societal perspective would allow for greater opportunities to correct sources of measurement error (e.g., the actual costs of replacement dogs are not described, but could be factored in). Did the dog costs include the cost of replacement dogs for the nine patients that required replacement dogs? Were dogs ever re-assigned? If yes, were the fixed dog costs redistributed across multiple patients? At the end of the dog's working life, or in the event of unpairing, were any additional variable costs (i.e., insurance costs) incurred? The modified societal perspective analysis (and the health care sector perspective) currently excludes both health care utilization costs and work productivity domains (apparently on the grounds that the trial did not find a significant difference in either domain between groups) and concludes that the modified societal perspective resulted in the same incremental findings as the primary analysis. The committee disagrees with the approach of netting out these costs and recommends both economic domains be included in calculating the incremental cost-effectiveness ratio and conducting the probabilistic sensitivity analyses from the societal perspective. The CEA should not arbitrarily drop components of costs when they are measured even if there is not a statistically significant difference between intervention arms for several reasons. First, the study may not be powered to look at individual components to assess significance. Therefore, even if the standard errors for each component are consistently estimated, there is not enough power to reject the null hypothesis. Second, different components of costs may be correlated and therefore dropping some components may lead to incorrect estimates of the standard errors for the mean total costs. In fact, in order to better capture these covariances and obtain consistent estimates of the standard errors for mean total costs, the authors should use individual-level data available to them from the trial and use the appropriate study characteristics, such as clustering, as suggested above.

***The health system perspective*:** This is a perspective that reflects costs related to only a typical health system, including costs beyond those covered by the VA. Therefore, a clear delineation is needed as to what specific costs are included in the health system perspective. The committee recommends including the costs of acquiring, training, and maintaining the dogs, regardless of whether such costs are incurred in practice by the VA.

***The VA system perspective*:** This perspective was adopted only in the post hoc pre versus post period CEA for the pooled set of participants. However, for reasons discussed in detail in the section above on pre and post outcomes, the committee recommended that the pre-post analyses be dropped altogether from the report. In addition, to be consistent with the trial design, the authors are limited to conducting a CEA from a hypothetical VA perspective that likely deviates in important ways from a realistic VA perspective (more below). By conducting a CEA from a hypothetical VA system perspective that relies on the costs and outcomes of *both groups as observed in the trial* (along with appropriate extrapolations over time) the authors would be assisting VA decision makers on whether to extend the same insurance coverage to psychiatric service dogs that other types of service dogs currently receive. With this analysis, the authors would clearly delineate those costs incurred by the VA (as opposed to society or the health care system) in the trial. The committee leaves it up to the VA researchers to properly identify the relevant costs from the VA's perspective. Recognizing that the hypothetical VA system perspective may deviate in important ways from a realistic VA perspective (e.g., in the real world, it is unlikely that (a) the quality of dogs in both arms of the trial would be as high as in the trial [see the discussion on generalizability at the start of the chapter], (b) the VA would pay for insurance benefits for emotional support dogs, and (c) emotional support dogs would be paid for by a third party), clearly stating the limitations and potential biases of such a hypothetical analysis is recommended.

Time Horizon and Extrapolation/Forecasting Assumptions

The committee applauds the authors' attempt to forecast costs and outcomes beyond the 18 month duration of the trial. However, the committee recommends a modification of this approach by separating: (1) an analysis limited to the duration of the trial and (2) a lifetime time horizon. The decision to stop at 7 years, which also requires assumptions about extrapolating beyond the trial, seems arbitrary because the costs of maintaining dogs after the end of their work life is a real cost from the societal perspective and also from a health care system perspective. An appropriate time horizon (Basu and Maciejewski, 2019) for such an analysis must be long enough to capture the intended and unintended benefits and harms of the intervention. The committee has three main recommendations regarding assumptions underlying the forecasting methodology that would be relevant to a lifetime time horizon, which perhaps could be addressed via sensitivity analyses.

First, when carrying forward the difference in outcomes (comparing assignment to receive a service dog versus an emotional support dog) from 18 months to the end of the working life of the dog, instead of using the *difference over time* observed during the trial period (measured using the per protocol sample), the committee recommends using the *difference observed at month 18*, measured using the ITT sample.

Second, the authors assume no difference in PCL-5 scores (comparing assignment to receive a service dog versus an emotional support dog) beyond the working life of the dog. It is unclear what happens after the service dogs are retired. Inasmuch as arguments can be made that such a difference may increase or decrease (and in any case continue) beyond the working life of the dog, sensitivity analyses to explore these possibilities are recommended. If a replacement service dog is not obtained, PTSD and quality of life may maintain, decline, or improve. The rationale for differential changes over time is related to debate over reliance versus growth from a service dog partnership (see Box 2-1).

Third, when extrapolating health care costs, Table 2.4.3 states that the annual baseline medical cost ($21,522) was applied to all interventions and *across all model years*. This assumption seems incorrect—annual health care costs would be expected to increase as study participants age in the model.

BOX 2-1 Safety Behaviors and Task Reliance

Safety Seeking Behaviors: Safety seeking behaviors are common features of posttraumatic stress disorder (PTSD). These behaviors are enlisted to protect from perceived danger or avoid fearful situations. They can be maladaptive, particularly when danger does not exist and when they prevent individuals from overcoming their fears (Blakey et al., 2020). If service dogs are enlisted to protect against perceived danger that does not exist, they could lead Veterans to maintain false beliefs about the safety of the world around them, instead of learning to trust and engage in society. After service dog retirement, symptoms might return and functional improvements could decline.

Pathway to Societal Engagement: Many Veterans with PTSD report debilitating fear and anxiety that prevent them from leaving their home. If service dogs provide initial feelings of safety that enable Veterans to leave the house to go participate in evidence-based treatments or engage in work and society, then this initial support could be valuable where a Veteran otherwise might not be able to do so. Once the initial fears are overcome by the Veteran being exposed to successful public outings, the Veteran may gradually reduce their reliance on these service dog tasks (Rodriguez et al., 2020). After service dog retirement, Veterans might continue to enjoy gains made during the treatment period.

Heterogeneity and Impact: Due to the heterogeneity of safety seeking behaviors (e.g., type, timing) and their potential differential impact on treatment outcomes (Goodson and Haeffel, 2018), the committee recommends sensitivity analyses to model differential outcomes beyond the duration of the trial.

It's important to note that in this analysis, a lifetime time horizon (based on the trial participants' perspective) would consist of three separate time phases: (1) first 18 months corresponding to the duration of the trial, (2) beyond the 18 months and up to the working life of the dog, and (3) beyond the working life of the dog to the lifetime of the individual. The second and third phases require extrapolation beyond the trial data. The third phase extrapolation is more uncertain given limited data related to any potentially sustained effects beyond the working life of the dog. These uncertainties can be addressed using sensitivity analyses.

MINOR ISSUES FROM CHAPTER 1 OF THE DRAFT MONOGRAPH

Statistical Power of Analyses and Value of Information Analyses

Results for different outcomes are presented without a discussion on whether the study was designed to estimate effects on particular outcomes and of specified magnitudes. Whether an estimate is found to be statistically significant when the true treatment effect is non-zero is an important factor to address in a study, as it informs readers about the strength of a null finding as well as how likely a non–null finding is due to chance. While there are methods for ex-ante simulated power analyses that can be performed for studies such as this one (Burlig et al., 2020; Gelman and Carlin, 2014; Lewis and Rao, 2015), given the other issues raised in this review regarding the methods, the committee does not suggest conducting those exercises. However, the committee believes that a post hoc value of information (VOI) calculation would be useful. Such analysis can help inform the VA of the value of a future study that tries to build on this one and answers the comparative question more precisely. In VOI analyses, the final results from this study, the point estimate and the uncertainty around it serve as a guide to the value added for future study. An ex-ante simulated power analysis can then be performed for a future study, to assess the sample sizes needed to be able to capture effects of different sizes. The VOI approach goes beyond the power analysis. It also compares the expected losses associated with the adoption decision of a particular type of dog within the VA system that would be made with new information versus the information available from the current study. For the purpose of this report, the authors can perform an expected value of perfect information (EVPI) calculation, which would be a trivial calculation using the probabilistic sensitivity analysis results from their cost-effectiveness model. A small EVPI value would indicate that the current study generated

sufficient information such that the value of more precise information arising out of a future study is low as it is less likely to change decisions made based on current results. See Basu et al. (2019), Claxton (1999), Fenwick et al. (2020), Meltzer (2001), and Rothery et al. (2020) for more details.

Medication Adherence Measurement

Medication adherence is reported to have been measured using the proportion of days covered (PDC) for four psychiatric medication classes (specifically antidepressants, antipsychotics, benzodiazepines, and other hypnotics and sedatives). As the authors note, this measure has been endorsed by the National Quality Forum and the Pharmacy Quality Alliance, and it is widely used in research. It is of course a proxy worth noting, because it uses information on medication prescription fills, and not every prescription that is filled is necessarily consumed.

In this study, PDC is the number of days supplied for a prescription during the observation period divided by the number of days in the observation period. This value is multiplied by 100 to obtain a percentage. The draft monograph notes that the PDC is capped, but because it identifies actual days of dispensed medications, it should never exceed 100%. The description of the PDC in the draft monograph as the sum of the days supplied appears to be a related measure, the Medication Possession Ratio, which can exceed 100%. The authors should clarify which measure of adherence was used for the draft monograph.

The draft monograph states (p. 20) that "Proportions were computed by drug class and were only computed if the participant had at least one refill in the drug class." It is not clear why the proportion was limited to participants with at least one refill, as this excludes participants with one fill (but no refills). The "at least one refill" measure has been used in other studies to track "Early-Stage Persistence," a particular type of adherence (Karter et al., 2009), but is not required for the PDC. The draft monograph does not provide a rationale for the non-standard use of the PDC. In addition, if the rationale for the examining of the PDC is that adherence was considered as a possible channel through which the intervention could have affected economic outcomes, then that should be clearly explained.

MINOR ISSUES FROM CHAPTER 2 OF THE DRAFT MONOGRAPH

The Use of $100,000 per QALY Threshold

In several places, the CEA refers to a commonly used threshold of $100,000 per QALY to indicate an important metric that is often cited to determine whether an intervention yields a pay-off in QALY terms that is worth the investment. The committee encourages the authors to revisit this comparison for a number of reasons. As Neumann et al. (2014) write in "The Curious Resilience of the $50,000 per QALY Threshold," any threshold is by its very nature arbitrary. Recent theoretical and empirical work suggests estimates of the CEA threshold in the United States to be around $100,000 per QALY (Phelps, 2019; Vanness et al., 2021). These thresholds are intended to compare societal willingness to pay for a one QALY improvement to the societal investment in achieving the one QALY gain. However, the United States does not have a decision-making body that has adopted a specific threshold to determine whether an intervention is cost effective and thus the use of a single threshold may be less relevant here than in countries where such a national approach exists, such as in the United Kingdom. Given that the costing perspectives incorporated here do not incorporate all of the elements of a full societal approach, using a societal willingness to pay threshold for comparison will lead to misleading results. Calculating a cost-effectiveness acceptability curve, which reflects the probability that an intervention is cost effective at a range of values, could be one viable alternative to a single comparison.

Choice of Decision Analytic Model

What is the "value add" of using a Markov model as opposed to a simpler decision analytic approach to estimate the incremental cost effectiveness of a service dog compared to an emotional support dog? For example, in the base case (Section 2.4.6), the incremental cost ($9,822) of a service dog compared to an emotional support dog is due solely to differences in the one-time fixed costs of procurement and pairing that occur at the start of the trial. In addition, it appears that the incremental QALYs (0.039) can be calculated by discounting the average annual treatment effect (3.7*0.85*0.002) at 3% for 7 years (0.0392 QALYs). Is the Markov model perhaps necessary for one or more of the sensitivity or scenario analyses? If not, a more parsimonious model as described above is recommended.

REFERENCES

Abraha, I. 2010. Modified intention to treat reporting in randomized controlled trials: Systematic review. *BMJ* 340:c2697. doi: 10.1136/bmj.c2697.

Basu, A., and M. L. Maciejewski. 2019. Choosing a time horizon in cost and cost-effectiveness analyses. *JAMA* 321(11):1096-1097. doi: 10.1001/jama.2019.1153.

Basu, A., D. L. Veenstra, J. J. Carlson, W.-J. Wang, K. Branch, and J. Probstfield. 2019. How can clinical researchers quantify the value of their proposed comparative research? *American Heart Journal* 209:116-125.

Blakey, S. M., A. C. Kirby, K. E. McClure, E. B. Elbogen, J. C. Beckham, L. L. Watkins, and J. D. Clapp. 2020. Posttraumatic safety behaviors: Characteristics and associations with symptom severity in two samples. *Traumatology* 26(1):74-83. https://doi.org/10.1037/trm0000205.

Briggs, A., and B. O'Brien. 2001. The death of cost-minimization analysis? *Health Economics* 10:179-184.

Burlig, F., L. Preonas, and M. Woerman. 2020. Panel data and experimental design. *Journal of Development Economics* 144:102458.

Claxton, K. 1999. The irrelevance of inference: A decision-making approach to the stochastic evaluation of health care technologies. *Journal of Health Economics* 18(3):341-364.

Drummond, M. F., M. J. Schulpher, K. Claxton, G. L. Stoddart, and G. W. Torrance. 2015. *Methods for the economic evaluation of health care programmes, 4th ed.* Oxford, UK: Oxford University Press.

Fenwick, E., L. Stueten, S. Knies, S. Ghabri, A. Basu, J. F. Murray, H. Koffijberg, M. Strong, G. D. S. Schimdler, and C. Rothery. 2020. Value of information analysis for research decisions—an introduction: Report 1 of the ISPOR value of information analysis emerging good practices task force. *Value in Health* 23(2):139-150.

Finkelstein, A., A. Zhou, S. Taubman, and J. Doyle. 2020. Health care hotspotting—a randomized, controlled trial. *New England Journal of Medicine* 382(2):152-162. doi: 10.1056/NEJMsa1906848.

Freed, M. C., D. E. Yeager, X. Liu, K. L. Gore, C. C. Engel, and K. M. Magruder. 2009. Preference-weighted health status of PTSD among veterans: An outcome for cost-effectiveness analysis using clinical data. *Psychiatric Services* 60(9):1230-1237.

Gelman, A., and J. Carlin. 2014. Beyond power calculations: Assessing type S (Sign) and type M (Magnitude) errors. *Perspectives on Psychological Science* 9(6):641-651. https://doi.org/10.1177/1745691614551642.

Glick, H. A., J. A. Doshi, S. S. Sonnad, and D. Polsky. 2015. *Economic evaluation in clinical trials, 2nd ed.* (pp. 206-208). Oxford, UK: Oxford University Press.

Goodson, J. T., and G. J. Haeffel. 2018. Preventative and restorative safety behaviors: Effects on exposure treatment outcomes and risk for future anxious symptoms. *Journal of Clinical Psychology* 74(10):1657-1672. https://doi.org/10.1002/jclp.22635.

Ioannidis, J. P. A., S. J. W. Evans, P. C. Gøtzsche, R. T. O'Neill, D. G. Altman, K. Schulz, D. Moher, and CONSORT Group. 2004. Better reporting of harms in randomized trials: An extension of the CONSORT statement. *Annals of Internal Medicine* 141(10):781-788.

Karter, A. J., M. M. Parker, H. H. Moffet, A. T. Ahmed, J. A. Schmittdiel, and J. V. Selby. 2009. New prescription medication gaps: A comprehensive measure of adherence to new prescriptions. *Health Services Research* 44:1640-1661.

Lewis, R. A., and J. M. Rao. 2015. The unfavorable economics of measuring the returns to advertising. *Quarterly Journal of Economics* 130(4):1941-1973.

Linden, A. 2013. Assessing regression to the mean effects in health care initiatives. *BMC Medical Research Methodology* 13(1):1-7. https://doi.org/10.1186/1471-2288-13-119.

Meltzer, D. 2001. Addressing uncertainty in medical cost-effectiveness analysis—implications of expected utility maximization for methods to perform sensitivity analysis and the use of cost-effectiveness analysis to set priorities for medical research. *Journal of Health Economics* 20(1):109-129.

Montedori, A., M. I. Bonacini, G. Casazza, M. L. Luchetta, P. Duca, F. Cozzolino, and I. Abraha. 2011. Modified versus standard intention-to-treat reporting: Are there differences in methodological quality, sponsorship, and findings in randomized trials? A cross-sectional study. *Trials* 12:58. http://www.trialsjournal.com/content/12/1/58.

NASEM (National Academies of Sciences, Engineering, and Medicine). 2021. *Letter report on review of Department of Veterans Affairs monograph on potential therapeutic effects of service and emotional support dogs on veterans with post traumatic stress disorder*. Washington, DC: The National Academies Press. https://doi.org/10.17226/26039.

Neumann, P. J., J. T. Cohen, and M. C. Weinstein. 2014. Updating cost-effectiveness—the curious resilience of the $50,000-per-QALY threshold. *New England Journal of Medicine* 371(9):796-797. https://doi.org/10.1056/nejmp1405158.

Peikes, D., A. Chen, J. Schore, and R. Brown. 2009. Effects of care coordination on hospitalization, quality of care, and health care expenditures among Medicare beneficiaries: 15 randomized trials. *JAMA* 301(6):603-618. doi: 10.1001/jama.2009.126.

Phelps, C. E. 2019. A new method to determine the optimal willingness to pay in cost-effectiveness analysis. *Value Health* 22(7):785-791.

Prosser, L. A., P. J. Neumann, G. D. Sanders, and J. E. Siegel. 2016. Reporting cost-effectiveness analysis. In *Second panel on cost-effectiveness analysis in health and medicine*, edited by Neumann et al. Oxford, UK: Oxford University Press.

Ranganathan, P., C. Pramesh, and R. Aggarwal. 2016. Common pitfalls in statistical analysis: Absolute risk reduction, relative risk reduction, and number needed to treat. *Perspectives in Clinical Research* 7(1):51-53.

Rodriguez, K. E., M. R. LaFollette, K. Hediger, N. Ogata, and M. E. O'Haire. 2020. Defining the PTSD service dog intervention: Perceived importance, usage, and symptom specificity of psychiatric service dogs for military veterans. *Frontiers in Psychology* 11:1638. https://doi.org/10.3389/fpsyg.2020.01638.

Rothery, C., M. Strong, H. Koffijberg, A. Basu, S. Ghabri, S. Knies, J. F. Murray, G. D. S. Schimdler, L. Stueten, and E. Fenwick. 2020. Value of information analytical methods: Report 2 of the ISPOR value of information analysis emerging good practices task force. *Value in Health* 23(3):277-286.

Sanders, G. D., P. J. Neumann, A. Basu, D. W. Brock, D. Feeny, M. Krahn, K. M. Kuntz, D. O. Meltzer, D. K. Owens, L. A. Prosser, J. A. Salomon, M. J. Sculpher, T. A. Trikalinos, L. B. Russell, J. E. Siegel, and T. G. Ganiats. 2016. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: Second panel on cost-effectiveness in health and medicine. *JAMA* 316(10):1093-1103. doi: 10.1001/jama.2016.12195.

Schulz, K. F., D. G. Altman, and D. Moher. 2010. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *BMJ* 340:c332.

Vanness, D. J., J. Lomas, and H. Ahn. 2021. A health opportunity cost threshold for cost-effectiveness analysis in the United States. *Annals of Internal Medicine* 174(1):25-32.

Vassall, A., L. Mangham-Jefferies, G. B. Gomez, C. Pitt, and N. Foster. 2016. Incorporating demand and supply constraints into economic evaluations in low-income and middle-income countries. *Journal of Health Economics* 25(Suppl 1):95-115.

Vickerstaff, V., R. Omar, and G. Ambler. 2019. Methods to adjust for multiple comparisons in the analysis and sample size calculation of randomised controlled trials with multiple primary outcomes. *BMC Medical Research Methodology* 19:129. https://doi.org/10.1186/s12874-019-0754-4.

Wright, S. J., W. G. Newman, and K. Payne. 2019. Accounting for capacity constraints in economic evaluations of precision medicine: A systematic review. *Pharmacoeconomics* 37(8):1011-1027.

Zulman, D. M., C. Pal Chee, S. C. Ezeji-Okoye, J. G. Shaw, T. H. Holmes, J. S. Kahn, and S. M. Asch. 2017. Effect of an intensive outpatient program to augment primary care for high-need veterans affairs patients: A randomized clinical trial. *JAMA Internal Medicine* 177(2):166-175. doi: 10.1001/jamainternmed.2016.8021. PMID: 28027338.

Appendix A

Additional Minor Points

	GENERAL
Comment	Why control for gender but not race ethnicity in the modeling? No rationale is presented despite having the data.
	CHAPTER 1
p. 20	Which consumer price index was used for 2018 (general, medical care, etc.)? Please be more specific.
p. 20	Please justify the assumption that the Department of Veterans Affairs (VA) unit costs apply to non-VA services (and in any case, do sensitivity analyses on this assumption).
p. 20	In response to the committee question to the VA about Table 1.7-1.9, the VA responded: The computation of proportion of days covered (PDC) was done separately for the pre- and post-periods. The PDC starts on the day the prescription was dispensed and then was only computed for prescriptions with a refill in the time period. It is possible for a person to have a PDC in one time period, but not in the other (i.e., an unbalanced panel). *The committee recommends that the VA include this explanation in its monograph.*
p. 20	For self-reported measures, were assessments really administered at exactly 90-day intervals for every participant? In practice, trials often use time windows and it is common for assessments to be off by 1 week or more. If the latter occurred, how were non-uniform time periods handled (e.g., self-reported counts collected during an 83-day period could be prorated to 90 days)?
p. 22	Clarify how you handled participants who worked in some 90-day periods and not in others.
p. 22	Why cap reported hours worked at all (and why at 50)?
p. 23	If follow-up period was 18 months, why was no discounting used in Chapter 1 (Chapter 2 assumes 3% discount rate for costs and outcomes)?
p. 23	Sensitivity analysis did not examine alternative assumptions about unit costs for non-VA utilization.
p. 26	Why present results of total costs (VA and non-VA) in Table 1.4 before presenting results of non-VA costs in Section 1.3.3 that follows?
p. 26	Second sentence of Section 1.3.2 states that service dogs did not significantly affect the VA health care utilization with the exception of outpatient substance abuse treatment (Table 1.3). But is inpatient length of stay also significantly affected?
p. 30	Last sentence of p. 26 states that Table 1.4 reports the sum of VA and non-VA costs, but the title of Table 1.4 is Total VA Costs. Are non-VA costs included? Or perhaps the VA pays for non-VA utilization anyway?

p. 30	Notes to Table 1.4 might add month 3 as a reference category.
p. 31	Why is non-VA hospitalization expressed as a percentage instead of length of stay or total number of days?
p. 34	First sentence states that Table 1.10 shows the unadjusted economic outcomes. Add "for workplace productivity" for clarity.
p. 38	First paragraph inconsistently refers to utilization and costs. For example, "In the analysis of VA utilization over time, the linear models … showed that the SERV [service dog] group experienced significantly greater costs…". Which was it—costs or utilization?
	CHAPTER 2
Figures	The titles of the figures in this section are not adequately informative. Figures should be self-explanatory enough regarding type of analysis, dependent, and independent variables so that readers do not need to refer to the monograph narrative text. For example, a title such as "Tornado Diagram" (Figure 2.4.2) is not adequate for this purpose. Instead the title would be "Tornado Diagram of the Cost per Quality of Life Year (QALY) Attributable to…".
p. 54	Section 2.2 Stakeholder Perspectives. Confusing subheader: "perspectives" in an economic evaluation has a specific meaning, therefore something else should be used (e.g., Feedback from Stakeholders".
p. 61	In Figure 2.4.1 Model Structure, why list economic outcomes (e.g., work productivity) that were not included in the model? May as well list all of the clinical outcomes that were ignored by the model as well.
p. 72	In alternative PCL-5 utility mapping, why assign 0.0038 if this is not the target population?

Appendix B

Documents Reviewed by the Committee

In order to complete its review, the committee was provided with several documents that are listed below and are available through the National Academies' Public Access File. In order to review these documents, email paro@nas.edu for more information.

1. The final version of the monograph titled *A Randomized Trial of Differential Effectiveness of Service Dog Pairing Versus Emotional Support Dog Pairing to Improve Quality of Life for Veterans with PTSD*.
    a. The committee also viewed the cover letters and additional documents from the Department of Veterans Affairs (VA) that were part of the review of the first monograph.
2. A link to the VA's website that houses:
    a. Contract of Statement of Work for procuring dogs for the study.
    b. The complete set of study forms used to conduct the study.
    c. The dog care course and assessment taken by Veterans prior to receiving a dog.
3. The draft second monograph titled *The Economic Impact and Cost Effectiveness of Service Dogs for Veterans with Post Traumatic Stress Disorder.*
4. The study protocol.
5. Two additional text documents providing written responses to questions posed by the committee to the VA and to the Institute for Clinical and Economic Review (ICER).
6. The presentation provided by the VA and ICER at the public meeting.

Appendix C

Acronyms and Abbreviations

ADA	Americans with Disabilities Act
CAPS-5	Clinician Administered PTSD Scale for DSM-5
CEA	cost-effectiveness analysis
EVPI	expected value of perfect information
ICER	Institute for Clinical and Economic Review
ITT	intent to treat
NDAA	National Defense Authorization Act of 2010
PCL-5	self-reported PTSD
PDC	proportion of days covered
PTSD	posttraumatic stress disorder
QALY	quality-adjusted life-year
RCT	randomized controlled trial
SF-36	36-Item Short Form Survey
VA	Department of Veterans Affairs
VOI	value of information
VR-12	Veterans RAND 12 Item Health Survey
WHODAS 2.0	World Health Organization Disability Assessment Schedule 2.0

Appendix D

Committee Biographies

Susan Busch (*Co-Chair*) is a Professor in the Department of Health Policy and Management at the Yale School of Public Health and has a Ph.D. in Health Policy (2000) from Harvard University. Dr. Busch is a health economist with 20 years of experience studying innovations in health care payment, financing, and delivery related to behavioral health. Using administrative databases and federally sponsored survey data, Dr. Busch has conducted multiple studies of the effects of both policy changes and health care interventions on access to mental health and substance use disorder treatments as well as their cost and quality.

Bisakha "Pia" Sen (*Co-Chair*) is a tenured Professor and the first woman Blue Cross Blue Shield of Alabama Endowed Chair in Health Economics at the Department of Health Care Organization & Policy in the School of Public Health at The University of Alabama at Birmingham. Dr. Sen receive her Ph.D. in economics from The Ohio State University in 1992. She specializes in health policy and health economics, and her research focuses on the impact of policies, programmatic interventions, and socio-ecological factors on health outcomes of vulnerable populations. She has published extensively in leading peer-reviewed academic journals and she has received funding support from the National Institutes of Health (NIH), the Centers for Disease Control and Prevention (CDC), the Alabama Department of Public Health, the Robert Wood Johnson Foundation, and other sources. Dr. Sen has served as an expert reviewer for NIH, examples of which include the Social Sciences and Population Studies study section, the Health, Behavior, and Context subcommittee. She has also served as expert reviewer for Active Living Research. She routinely serves as a journal reviewer for numerous leading peer-reviewed journals.

Anirban Basu is a Professor of health economics and the Stergachis Family Endowed Director of The CHOICE Institute at the School of Pharmacy at the University of Washington (UW), Seattle. He holds joint appointments with the Departments of Health Services and Economics at UW, is a Research Associate at the U.S. National Bureau of Economic Research, and an elected Fellow of the American Statistical Association. His work sits at the intersection of microeconomics, statistics, and health policy. His research focuses on understanding the economic value of health care through scientific disciplines of applied economic theory, comparative and cost-effectiveness analyses, causal inference methods, program evaluation, and outcomes research. He served on the Second Panel on Cost-effectiveness Analysis in Health and Medicine and serves on the Editorial Advisory Board for the *Value in Health* journal. He is a past recipient of the ISPOR Methodology Awards and the Bernie O'Brien New Investigator Award. He received his master's degree in biostatistics from the University of North Carolina at Chapel Hill and a Ph.D. in public policy studies from the University of Chicago.

Marisa Elena Domino is a Professor of health economics in the Department of Health Policy and Management in the Gillings School of Global Public Health at the University of North Carolina (UNC) at Chapel Hill. She is also the Director of the Program on Mental Health and Substance Abuse Systems and Services Research at the Cecil G. Sheps Center for Health Services Research. Dr. Domino is deeply interested in vulnerable populations and has created a research agenda throughout her career that examines the efficiency of health care policies in low-income and disabled populations. She has substantial expertise in applied econometric analyses and has worked extensively on large administrative databases from a variety of health insurance programs. She has considerable experience extracting measures of medication use and adherence, quality of care, utilization, and costs from a large variety of data sources. Dr. Domino's

work has focused on the effects of Medicaid program design on a variety of populations and outcomes, especially related to behavioral health and chronic illness. She has received funding from the National Institute on Drug Abuse, National Institute of Mental Health, the Agency for Healthcare Research and Quality, the Robert Wood Johnson Foundation, and the Brain & Behavior Research Foundation to examine the effect of policy changes on the use of mental health and medical services, prescription medications, and other measures of health services use, quality, and costs. She is the recipient of the 2013 ISPOR Award for Excellence in Application of Pharmacoeconomics and Health Outcomes Research, the 2017 Edward G. McGavran Award for Excellence in Teaching from the UNC Gillings School of Global Public Health, and the 2021 Willard Manning award in Mental Health Policy and Economics.

Dominic Hodgkin is a Professor at the Heller School for Social Policy and Management at Brandeis University. He is also a Research Director of the Brandeis/Harvard National Institute on Drug Abuse Center to Improve System Performance of Substance Use Disorder Treatment, located at the Heller School's Schneider Institute for Behavioral Health. Dr. Hodgkin is a health economist with more than 25 years of experience in health policy analysis and research in academic settings, who has served as a Principal Investigator on four National Institutes of Health (NIH)-funded studies. Most of his research focuses on the effects of different organizing and financing approaches in health care, particularly addressing mental and substance use disorders (SUDs). These studies have applied a variety of econometric methods to datasets ranging from insurance claims and other administrative data to individual and organizational surveys. Dr. Hodgkin's studies have evaluated the use of substance abuse services under Medicaid carve-out programs in Michigan and Massachusetts; the effects of the Substance Abuse and Mental Health Services Administration's Access to Recovery program on SUD treatment and utilization; and the use of patient incentives and care navigators to connect detox patients to substance use specialty SUD treatment. He is a standing member of the NIH grant review panel for organization and delivery of health services. Dr. Hodgkin received his Ph.D. in economics from Boston University in 1995 and his B.A. in economics from Manchester University, England, in 1985.

Mireille Jacobson is an Associate Professor in the Davis School of Gerontology, the Co-Director of the program on aging and cognition at the University of Southern California's Leonard D. Schaeffer Center for Health Policy & Economics, and a Research Associate in the Health Care Program at the National Bureau of Economic Research. Dr. Jacobson holds a Ph.D. (2001) and an M.A. (1998) in economics from Harvard University and was a National Institute of Mental Health Post-Doctoral Fellow at the Harvard Medical School from 2001 to 2002. Dr. Jacobson has a diverse portfolio of research united by an interest in understanding how health care policies affect well-being. Much of her work focuses on the supply-side of U.S. health care markets, analyzing the effects of direct supply changes (e.g., hospital closures) on access to care and the impact of Medicare reimbursement policy on treatment and outcomes. Other work focuses on the demand side, assessing the risk-protective value of health insurance for consumers and the impact of health insurance on the mental health of seniors.

Evan Mayo-Wilson is an Associate Professor in the Department of Epidemiology and Biostatistics at the School of Public Health at Indiana University Bloomington. His research focuses on methods for conducting, reporting, and synthesizing health and social intervention research. His primary area of interest is in ways to increase transparency and reproducibility in clinical trials and systematic reviews, such as trial registration and data sharing. Dr. Mayo-Wilson received his B.A. from Columbia University in psychology; his M.P.A. from the University of Pennsylvania; and his M.S. and D.Phil. (Ph.D.) from the University of Oxford Department of Social Policy and Intervention. Prior to his current appointment, Dr. Mayo-Wilson worked as a Departmental Lecturer at the University of Oxford, a Senior Research Associate in the Department of Clinical, Educational and Health Psychology at University College London, and an Associate Scientist in the Department of Epidemiology at the Johns Hopkins Bloomberg School of Public Health.

Marguerite E. O'Haire is an Associate Professor of human–animal interaction in the Purdue University College of Veterinary Medicine with courtesy appointments in the Department of Human Development and Family Studies and the Department of Psychological Sciences. She received her BA in psychology from Vassar College, New York, and her Ph.D. in psychology from The University of Queensland, Australia. Her expertise and training focus on the psychology of human–animal interaction, with an emphasis on quantifying the psychophysiological effects of service dog interventions. She runs a National Institutes of Health (NIH)-funded research lab in the Purdue Center for the Human-Animal Bond. She has acquired NIH funding from three different centers/institutes, including the National Institute of Child Health and Human Development, the National Center for Advancing Translational Sciences, and the National Center for Complementary and Integrative Health. She has produced several seminal publications on human–animal interaction strategies for mental health in international peer-reviewed journals. Most recently, she led a registered, nationwide clinical trial to quantify the efficacy and role of service dogs for Veterans with posttraumatic stress disorder. She is the past recipient of career and achievement awards from the International Society for Anthrozoology and the Purdue University Showalter Faculty Scholars.

Todd Olmstead is an Associate Professor of public affairs at the Lyndon B. Johnson (LBJ) School of Public Affairs at The University of Texas at Austin. At the LBJ School, Dr. Olmstead conducts economic evaluations of a wide variety of health care programs and teaches courses in health economics, management science, and empirical methods. He is the health economist on several large grants funded by the National Institutes of Health, primarily in the area of behavioral health. Current research projects include estimating the cost effectiveness of (a) providing mental health services to low-income pregnant and parenting women living in public housing systems, (b) using technology to improve retention in and adherence to addiction treatment for individuals with opioid use disorders, and (c) providing early childhood obesity prevention programs to children attending Head Start centers. Dr. Olmstead holds a B.S. and an M.S. in industrial engineering from State University of New York at Buffalo (1987, 1989), an M.S. in operations research from the University of North Carolina at Chapel Hill (1994), and a Ph.D. in public policy from Harvard University (2000). In addition to his work in health care, Dr. Olmstead has published in the areas of intelligent transportation systems, highway safety, and administrative rulemaking.

Kosali Simon is an endowed Professor at the Indiana University Bloomington O'Neill School of Public and Environmental Affairs and the Associate Vice Provost of health sciences for the campus. She is a nationally known health economist who studies the impacts of state and federal health policies, and is active in leadership and mentoring roles in organizations with national prominence in economics and health policy; she is the Editor of the *Journal of Health Economics*, the Co-Editor of the *Journal of Human Resources*, and the Vice President of the Association for Public Policy Analysis and Management. She is a Research Associate of the National Bureau of Economic Research. Her work has appeared in health services and health policy journals (including *Health Affairs* and *Health Services Research*), medical journals (including *The New England Journal of Medicine* and *Annals of Emergency Medicine*), and economics journals (including *American Economic Journal* and *Journal of Health Economics*) and has been featured by many major media outlets. She received a Ph.D. in economics in 1999 from the University of Maryland. She serves as a member of the National Academies of Sciences, Engineering, and Medicine's Roundtable on Population Health Improvement.