Office of Research and Development
Veterans Health Administration
Department of Veterans Affairs

December 6, 2021

Susan Busch, PhD
Bisakha "Pia" Sen, PhD
Co-Chairs, VA PTSD/Service Dog Study
National Academies of Sciences, Engineering, and Medicine

Dear Drs. Busch and Sen:

We sincerely thank the committee for its final letter report received 10/1/21 addressing the VA study team's 8/16/21 revisions to the monograph entitled "The Economic Impact and Cost Effectiveness of Service Dogs for Veterans with Post Traumatic Stress Disorder." The final report acknowledged the VA study team's efforts to strengthen the monograph by making key changes such as moving the Chapter 1 pre-post analysis to the appendix, adding information related to the first monograph, discussing key differences between emotional support dogs used in the study and in practice, and adding a cost-effectiveness acceptability curve to characterize the probability that service dogs are cost effective when compared to emotional support dogs at a range of cost-effective thresholds. We have further modified the monograph in an attempt to address the committee's remaining recommendations.

We understand the committee's concerns about the lack of an intent to treat analysis, and we have incorporated the committee's suggestion to conduct sensitivity analyses to assess the effect of missing data. There were no significant changes in the results across these additional analyses. To address the committee's concern about the lack of a full societal perspective in the cost-effectiveness analysis, we conducted a new threshold analysis from the societal perspective that solves for the added work hours needed to meet common cost-effectiveness thresholds. We thank the committee for these recommendations and believe these additional analyses have further strengthened the monograph.

Attached you will find our response to the reviewers' specific comments and the revised final version of the monograph. We sincerely appreciate the committee's extensive review.

Sincerely,

Joan T Richerson
Study Chair

Todd H. Wagner
Lead VA Health Economist

Jonathan D Campbell
SVP, ICER

Response to the NASEM Committee Letter Report on Revised Monograph 2

Major Concern: Intent-to-Treat Analysis

NASEM Comment: The committee acknowledges the challenges faced by the VA and ICER research teams in analyzing health economic information from this clinical trial. To prevent selection bias, the usual practice is to ask all trial participants before enrollment to consent for follow-up assessment regardless of their intervention assignment and intervention uptake. In this trial, such consent was not obtained. The research team later asked the VA's Institutional Review Board (IRB) to approve the use of additional data from Veterans who dropped out of the trial after randomization and prior to pairing with a dog, but the IRB denied the request. The IRB's decision prevented the research team from using the most appropriate comparison data set—the true intent-to-treat (ITT) population—in drawing conclusions about the relative health costs and cost effectiveness of providing a service dog versus an emotional support dog to Veterans with PTSD.

Instead of the ITT, the researchers used per protocol analysis (called the "modified intent to treat" analysis in the first monograph). The per protocol analysis is inappropriate because participant dropout after randomization might introduce bias and there is no guarantee that the treatment arms in the per protocol analysis will be balanced as randomized. Although the committee appreciates that the authors caveat the interpretation of the per protocol results in many places, it encourages the authors to state the limitations of "per protocol" in the extended Abstract and in the Preface so that the casual reader will understand this important limitation.

Without the data for all randomized participants, an appropriate statistical analysis is needed to control for confounding and bias. The committee recommended that the investigators use statistical methods such as imputation to deal with missing data, but the research team asserted that the imputation method "generates a more biased [estimate] than a complete case analysis when data are not missing completely at random" (p. 22). The committee disagrees with this interpretation of the literature. If missingness is completely at random, then both complete case analysis and imputation should give unbiased estimates. If missingness is not at random, then imputation can mitigate bias and provide better estimates of uncertainty compared with per protocol analysis.

In addition, imputation is a common method to deal with missing data, but it is not the only method. An alternative strategy could be the use of inverse probability weighting. The point is that the missingness could be informative in nature and while more involved methods would be required to model missingness not at random (e.g., some form of pattern-mixture models), the researchers can at least try to allay some concerns by presenting the observed trends in outcomes from pairing among those who completed the trial versus those who did not.

The committee's concern is supported by Table K on page 63 of Monograph 1[2] (describing the health outcomes of the trial), which shows that participants who were paired (N = 181) were different than those who were not (N = 46) across several dimensions of baseline characteristics. Thus, it cannot be said that using the pairing date would provide "unbiased information for the analysis," as claimed on page 22 of the revised monograph. The committee strongly recommends the inclusion of a table that reports all of the baseline characteristics reported in Table K of Monograph 1 by treatment for the randomized group, and separately for the paired group. This will help readers understand to what extent any new imbalance in baseline characteristics was generated in the per protocol analysis. Table 1.2 of the revised monograph does not seem complete with all of the characteristics from Table K.

The committee also strongly recommends that the researchers conduct sensitivity analysis to assess possible mechanisms and effects of missing data (e.g., bias) and to explore if the results of the per protocol analysis would change if different assumptions were made about the missing data. For example, do the data suggest the results are biased toward finding no difference when actually one group seems to improve outcomes relative to another? Putting the findings in perspective beyond saying the estimates may be biased would be helpful. Any observed differences, and lack of appropriate statistical controls, raise concerns about the validity of the results and also about the generalizability of the per protocol analysis, which was not discussed.

Finally, the committee was surprised to see the sentence "The average time to pairing for those randomized to EMOT was 158 days whereas the average time to pairing for those randomized to SERV was also 158 days" (p. 68). This seems to contradict the sentence on page 19 that "Time between unblinding and pairing was approximately 2 weeks for the emotional support dog group and 6 weeks for the service dog group." Although the 3-month training period before pairing mitigates the use of per protocol versus ITT, the authors should clarify the time to pairing from unblinding the randomization for each arm of the study. This is important because differences in time to pairing after unblinding further heighten the concerns about differential drop out before pairing.

Author Response: We have edited the abstract and preface to note the limitations of the "per protocol" analysis. We have also created new table providing the comparison between baseline characteristics, as requested (section 1.3.7).

We conducted two additional analyses. First, we used multiple imputation to estimate follow-up costs for those participants who were randomized but not paired with a dog. We reshaped the data as a wide dataset and then used chained equations to impute 10 replicants of the data. We then reshaped the data into a long dataset to continue with the panel data models. The results, as shown in section 1.3.7., found no significant differences between those randomized to SERV or EMOT. The multiple imputation models generally resulted in slightly larger SE estimates.

Second, we used baseline information to compute the probability of being successfully paired with a dog. We examined the common support across the two groups; two people in the paired group shared no common support with those who dropped out of the study. We used the probability as an inverse probability weight in analyses. In addition, we also ran the weighted analysis excluding the 2 paired individuals who did not share any common support. There were no significant changes in the results across these additional analyses.

The average time from randomization to pairing was 158 days for both study groups. Randomization occurred during the 3+ month observation period before pairing (this 3-month observation period included dog education training, as mentioned above, but also encompassed additional study activities, including baseline assessments and clearing visits). Chapter 1 had initially presented the time from unblinding, which occurred later than randomization. We have edited Chapter 1 to present the time from randomization to pairing to be more consistent with Chapter 2.

Major Concern: Societal Perspective

NASEM Comment: The committee notes several concerns with the cost-effectiveness analysis (CEA) from the societal perspective (see Section 2.4.5). The authors changed the name of this analysis section from "modified societal perspective" to "societal perspective" in response to the committee's recommendation for an analysis reflecting a *full* societal perspective. The authors ultimately did not provide a full societal perspective, citing missing evidence from several domains such as patient time costs, unpaid caregiver time costs, and transportation costs. The committee disagrees with the decision to not include a full societal analysis and offers several additional suggestions.

First, on page 71 of the revised monograph, the authors say that the societal perspective inputs are identical to those of a government payer perspective, which the committee does not believe is accurate. A government payer perspective would exclude costs to study participants (e.g., for time receiving services and travel time). These costs would be included in a societal perspective.

Second, although domains related to patient time costs and out-of-pocket costs were not solicited from study participants, it is possible to do a back-of-the-envelope approach using external sources of information to incorporate these costs into the analysis. For example, based on the participants' age, education, and region, an approximate hourly wage can be estimated from census data and then used to multiply by the number of services received and the approximate time per service. Caregiver time costs are potentially more problematic to impute, and could be substantially different between the two modalities, if there are any positive effects on the management of PTSD. Information could be derived from the literature on the percent of individuals with PTSD who rely on a caregiver, for example.

The revised monograph also assumes that domains such as patient out-of-pocket costs and patient time costs for training are likely comparable for those with a service dog versus those with an emotional support dog, but provides no external evidence to support this claim. Patients with a service dog had to fly to the vendor and spend 1-2 weeks in training. Given the relatively small difference in total costs between the two arms, the differential patient training time costs should be included in the analysis even if they have to be estimated.

Additional trainer costs to work with the patients post pairing with a service dog should also be included. Page 53 of the revised monograph indicates that trainers spent 1-2 days for an emotional support dog and 1 week for a service dog. It is not clear whether these differential trainer time costs were included in the bundled payment. It would be helpful if the components of this payment were briefly
described.

Finally, the committee still thinks that a more elaborate formal impact inventory table that includes information on these components would be useful to convey what dimensions were considered, why relevant components were not included, and whether their absence

may influence the results substantively. The goal would not be to invalidate the current analysis, but to inform a discussion about how these omissions may change the incremental cost-effective ratio

Author Response: We thank NASEM for providing suggested solutions to address a societal perspective cost-effectiveness analysis. In the final monograph, we conducted a new threshold analysis from the societal perspective that solves for the added work hours (per week, SERV vs. EMOT) needed to meet common cost-effectiveness thresholds. The number of work hours was not statistically significantly different in Chapter 1, but trended toward increases for SERV. Therefore, we cannot be confident in these findings, but if increases in work hours were confirmed through another study, SERV would likely be cost effective from a broader societal perspective.

Major Concern: Bias in the Choice of Methods and the Reporting in the Cost-Effective Analysis

NASEM Comment: The committee notes, and the revised monograph acknowledges, that the design of the cost effectiveness study was post hoc to the clinical study. It is incumbent on the researchers to be conscious of potential bias in selecting methods of analysis and assumptions and characterizing their results. In several instances, as noted below, the committee found a lack of balance in the presentation of the results and in the representation of the level of uncertainty of the measures.

Threshold Price Analysis (Comprehensive Health System Payer Perspective)

The committee recommends estimating and reporting the uncertainty around the key results of the threshold price analysis (see Section 2.4.6), including the estimated incremental benefit of service dogs relative to emotional support dogs (approximately 14 days [~0.039 quality-adjusted life years (QALYs)]) and the percent cost reduction in the service dog intervention (14%) that would meet a threshold of $100,000 per QALY gained. The current lack of details surrounding the uncertainty in these key results may inadvertently lead readers to be overconfident in the accuracy of the findings. The uncertainty (e.g., 95% confidence interval) should be reported in both the Abstract and the text. In addition, the committee notes that the reported 14% reduction in the service dog intervention cannot be attained by reducing insurance costs (i.e., the trial was run with equivalent insurance costs incurred by both service dogs and emotional support dogs), which should be acknowledged in the monograph.

Author Response: Uncertainty in findings (95% intervals generated from probabilistic sensitivity analyses) are now presented in the abstract and in the cost-effectiveness results.

Major Concern: Bias in the Choice of Methods and the Reporting in the Cost-Effective Analysis

NASEM Comment: *Threshold Health Improvement Analysis (the VA Perspective)*

The committee agrees with the VA threshold health improvement conclusion (see Section 2.4.8) that Veterans would need a 15.8 point improvement in the self-reported PTSD (PCL-5) total score to meet the $100,000 QALY threshold. The next sentence of the conclusion ("Unadjusted pre-post trial analyses yielded –15.4 points on the PCL-5 total score.") should be removed from the Abstract. The committee also recommends either deleting it from the Conclusion or adding the caveat that the research design used to estimate 11.7 of the 15.4 point improvement in the "unadjusted pre-post trial analyses" does not permit a causal interpretation of results; said differently, the observed improvements over time could also happen at least in part, due to regression to the mean or due to continued receipt of usual care for PTSD symptoms that all of the study participants were receiving.

Author Response: We added language in the abstract and conclusions that the pre-post analysis should not be interpreted as causal.

Major Concern: Bias in the Choice of Methods and the Reporting in the Cost-Effective Analysis

NASEM Comment: *Excluding Individual-Level Data from Analyses*

The committee disagrees with the authors' decision to exclude individual-level data, as the committee recommends, on the grounds that "differences were not anticipated in the deterministic or the uncertainty analyses using the adjusted summary statistics versus the regression-based individual-level data" (VA Response document). To begin, relying exclusively on adjusted summary statistics loses information on potential correlation across variables that, at a minimum, will affect the precision of model estimates. Moreover, using individual-level data to model economic domains (e.g., health care costs, productivity) in a CEA, even when no significant difference exists between the groups on such domains, is considered best practice. Excluding these domains artificially decreases the uncertainty in the results.

Author Response: We respectfully disagree that our choice of methods introduces bias into the cost-effectiveness analysis.

Major Concern: Bias in the Choice of Methods and the Reporting in the Cost-Effective Analysis

NASEM Comment: *Presentation of How the Differences in Emotional Support Dogs Used in Practice Versus Emotional Support Dogs Used in the Trial Would Affect the Cost-Effectiveness Outcome*

The committee commends the authors for explaining the differences in emotional support dogs used in practice versus those used in the trial (pp. 17-18) and agrees that these differences likely bias *patient outcomes* in favor of emotional support dogs. The committee recommends including a counterpart statement in the cost-effectiveness chapter (Chapter 2) noting that these same differences would likely bias *costs* in favor of service dogs (because emotional support dogs in practice are likely to be much less expensive than those used in the trial), and so the net effect on cost effectiveness is unclear. Said differently, emotional support dogs in practice would be expected to result in both lower scores and lower costs than those used in the trial, and so the net effect on the incremental cost-effectiveness ratio is indeterminate.

Author Response: We did not further address the generalizability of EMOTs used in real practice (versus the trial) as the directionality of the cost-effectiveness findings could go either way. We note text in Section 2.4.7 that speaks to this uncertainty on directionality for SERVs and we do not feel that adding further description around EMOTs would be helpful.

Major Concern: Bias in the Choice of Methods and the Reporting in the Cost-Effective Analysis

NASEM Comment: *Choice of Outcomes in Cost-Effectiveness Analysis*

The committee commends the authors for including the Veterans RAND 12 Item Health Survey (VR-12) as an outcome in the CEA. However, the committee recommends the cost-effectiveness section acknowledge (1) there were no significant differences between the groups on the Clinician Administered PTSD Scale for DSM-5 (CAPS-5), (2) CAPS-5 focuses exclusively on PTSD symptoms (like the PCL-5), (3) CAPS-5 was excluded from the formal CEA because no mapping exists from CAPS-5 scores to utility scores, and (4) if such a mapping existed, the incremental cost-effectiveness ratio of service dogs relative to emotional support dogs would increase (i.e., the incremental cost-effectiveness ratio would be less favorable for service dogs).

Author Response: We did not further address the CAPS-5 statement in the report as the directionality of the cost-effectiveness findings remains unknown, even if CAPS-5 did have a mapping algorithm. Stated differently, because there was a lack of a statistically significant finding on the CAPS-5 need not suggest that the changes in utility scores must be less than that of the PCL-5.

Minor Concerns:

- Pre-Post Analysis in Supplement 3
- Medication Adherence Measurement

Author Response:

- We have removed the discussion of the pre-post analysis from the abstract.

- We did not obtain any PDC values that exceeded 100%. We updated this methods section to include the following text “The maximum value of PDC was one, and this value was multiplied by 100 to obtain a percent.”

  The Raebel citation provides a framework for other types of adherence and explains the rationale for having a refill. They refer to this as secondary adherence and note that “most [studies] estimate adherence only among individuals with secondary adherence.” We have edited this section in the report to make it clearer. We have also clarified in the discussion that this is “adherence among those with a prescription” and is only measuring one type of adherence. We have added a more obvious callout to this footnote in Table 1.7.

  We have edited the variable label in Table 1.7 for clarity.