January 19, 2022

Jennifer Bright, MPA
Executive Director
Innovation and Value Initiative
2 Bethesda Metro Center #850
Bethesda, MD 20814

Dear Ms. Bright:

The Partnership to Improve Patient Care (PIPC) is pleased to provide feedback on the Innovation and Value Initiative's (IVI) draft model protocol on major depressive disorder (MDD). We appreciate the process you have set forward providing transparency into and encouraging stakeholder feedback on your model and process. This type of transparency and robust stakeholder engagement, particularly from patients and providers, leads to stronger models that more accurately convey value to patients and society.

Upon reviewing the materials, PIPC would suggest you consider the following:

IVI should reframe how it incorporates mortality multipliers into the model to mitigate risk of underestimating the value of successful treatment.

IVI rightly states that there is strong empirical evidence that MDD patients have higher mortality rates than the general population.[1,2] The problem with the sources being used is that they are longitudinal in nature and have a sample of people defined as having been diagnosed with MDD at a single point in time. These estimates will therefore include people who are in any of the three states of response described in the model. As such, there is no distinction for what state a patient was in: non-response, partial or complete response.

This leads us to the more relevant question of how these mortality multipliers are applied in the model. Is the suggestion that the MDD multiplier be applied only to non-responders? If so, the three studies cited by IVI may be inappropriate sources for the requisite mortality multiplier(s) as each likely reports the average effect of MDD on mortality for patients distributed across all three states. The mortality multiplier is likely to be much higher among patients in the no-response state as compared to among patients in the complete or partial response states.

Alternatively, if we apply such an 'average' multiplier to all MDD states rather than to the no-response state alone, the model will observe no survival benefit associated with successful treatment. This would incorrectly imply the same probability of death for patients in the no-response state as for patients in the complete- and partial-response states.

[1] Pratt LA, Druss BG, Manderscheid RW, Walker ER. Excess mortality due to depression and anxiety in the United States: results from a nationally representative survey. General hospital psychiatry. 2016 Mar 1;39:39-45.

[2] Chiu M, Vigod S, Rahman F, Wilton AS, Lebenbaum M, Kurdyak P. Mortality risk associated with psychological distress and major depression: A population-based cohort study. Journal of Affective Disorders. 2018 Jul 1;234:117-23

With this in mind, we are concerned that applying the same 'average' multiplier to all MDD states will lead directly to underestimating the absolute health gain from any successful treatment in the model and would encourage IVI to consider how to appropriately handle this challenge.

The model as designed is unlikely to be able to address the issue of treatment heterogeneity. We would suggest broadening the question of how to estimate the accrual of 'marginal value' from new therapies.

Most treatments are effective for only a reasonable portion of all potential beneficiaries. Unfortunately, most current methods to estimate cost-effectiveness rely heavily on RCT data for estimates of average treatment effect (ATE). RCTs are designed to produce a mean population ATE and not to directly produce estimates of incremental effect of treatment for individuals.[3] As such they provide scarce information on the heterogeneity of treatment effect that is useful for translating what is mean efficacy of a new therapy into what is population specific effectiveness of a new therapy.[4] It is one of the many limitations of RCTs for informing practical health policy, and has been discussed and dissected at length in the literature.[5,6,7,8]

This means if a new treatment, and in particular a new mechanism of action or 'type' of treatment, is more efficacious in populations for which traditional therapies have previously been largely ineffective - even if the ATE across the entire population is no greater than that of current treatments - the model as designed would not allow for such nuances of health benefit for subpopulations to be teased out. As IVI values insight into a broader question of estimating the value of innovation in healthcare it should evolve away from the traditional cost-effectiveness methods' reliance on the RCT and ATE framework and develop approaches that provide insight into the value of introducing new types of therapy at the subpopulation level. This is an area in which IVI could add huge value to the current lexicon of value assessment methodologies.

IVI should not use the quality-adjusted life year (QALY) in its models.

IVI has made some very positive strides towards making value assessment more accurately represent value to patients, not just cost and value to payers. We are thankful for the efforts IVI has put into

---

[3] Stevens W, Normand C. Optimisation versus certainty: understanding the issue of heterogeneity in economic evaluation. Social science & medicine. 2004 Jan 1;58(2):315-20.

[4] Basu A, Grieve R, Pritchard D, Stevens W. One size does not always fit all in value assessment. Am J Manag Care. 2019 Nov 1;25(11):540-2.

[5] Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. Social Science & Medicine. 2018 Aug 1;210:2-1.

[6] Mustafa FA. Notes on the use of randomised controlled trials to evaluate complex interventions: Community treatment orders as an illustrative case. Journal of evaluation in clinical practice. 2017 Feb;23(1):185-92.

[7] Anjum RL, Copeland S, Rocca E. Medical scientists and philosophers worldwide appeal to EBM to expand the notion of 'evidence'. BMJ evidence-based medicine. 2020 Feb 1;25(1):6-8.

[8] Anjum RL, Copeland S, Rocca E. Rethinking causality, complexity and evidence for the unique patient: a CauseHealth Resource for healthcare professionals and the clinical encounter. Springer Nature; 2020.

moving value assessment in a patient-centered direction, that being said, to be truly patient-centered, IVI must stop using the discriminatory QALY in its models.

QALYs are discriminatory in design and implementation. For this reason, in 2019, the National Council on Disability, an independent federal agency advising Congress and the administration on disability policy, issued a report finding that use of the QALY would be contrary to United States civil rights and disability law.[9] The United States has a thirty-year, bipartisan track record of opposing the use of the QALY and similar discriminatory metrics and has established appropriate legal safeguards to mitigate their use. There is currently a ban on use of the QALY or similar metrics in Medicare decision-making.[10] In 1992, the U.S. Department of Health and Human Services established that Oregon's efforts to utilize a cost-effectiveness standard in Medicaid would violate the Americans with Disabilities Act.[11]

PIPC urges IVI to build on this precedent and cease using the QALY in its models. We encourage IVI instead to build on the strides it has made in patient-centric value assessment by investing in alternative metrics.

Additionally, we have compiled answers to several of the specific questions on which you have requested input:

6.1 Are there any other studies/data sources that will better represent the characteristics of the MDD population based on the target population of the model?

The IVI-MDD model protocol draft indicates that the target population for the model protocol includes adults diagnosed with MDD without diagnosis of other psychiatric and physiological comorbidities. In reality, the majority of MDD patients have comorbidities so limiting the population only to those without comorbidities risks building a model that will only represent a small proportion of the full population of need. Studies suggest that the majority of MDD patients have at least one other psychiatric disorder. With this in mind, in order to replicate a real-world population, we would recommend running the model for both the primary population outlined in the protocol, plus at least one other index patient group with a significantly common comorbidity.

6.5 and 8.2.3 Do you have some suggestions on studies/data sources/methods that we can reference in extrapolating the long-term efficacy inputs?

We have limited data on responses to treatments for some comparators from our literature review of meta-analyses (Table 3).

- Should we extract such inputs from clinical trials or observational studies?

---

[9] https://ncd.gov/sites/default/files/NCD_Quality_Adjusted_Life_Report_508.pdf

[10] 111th Congress of the United States of America. (2010). H.R. 3590 The Patient Protection and Affordable Care Act. *Section 1182*. Washington, DC.

[11] https://www.nytimes.com/1992/09/01/opinion/l-oregon-health-plan-is-unfair-to-the-disabled-659492.html

- If so, do you have any recommendation on data sources?

Delayed efficacy has long been a problematic aspect of traditional treatments for MDD, often being seen as a major driver of early discontinuation of treatment.[12] If you choose a cycle length of 3 months, a practical way to link short-term response data with long-term risk of relapse and retreatment would be to differentiate between fast and slow responders.

Several studies have shown that those that respond quickly in the first 2-6 weeks of treatment have significantly improved downstream outcomes, compared to those who are slow to respond to treatment in that early period. Multiple studies have estimated that slow responders in this period are between four to eight times more likely to relapse at multiple stages further into treatment; evidence suggest higher relapses rates at six, twelve, and eighteen months for slower responders.[13,14,15] As such, these two subtypes would have different transition matrices reflecting differing likelihoods of relapse and retreatment over time.

6.9.3.2 Is it reasonable to assume the same sets of model inputs (efficacy and safety) for the first and second lines of treatment?

In the absence of data for the key efficacy inputs for third and fourth lines of treatment, we intend to: (1) first use estimates based on the treatment-resistant depression (TRD) population as model inputs; and (2) if estimates based on TRD population do not exist, use a hazard rate approach where treatment efficacy rates will be proportional to efficacy rates used in the first and second lines.

- Do these assumptions seem reasonable to you?
- Do you have any suggestions for sources to derive model estimates for the third- and fourth-line treatments?

The most common source for efficacy across lines of therapy used in cost-effectiveness modeling in MDD has been the data from the Sequenced Treatment Alternatives to Relieve Depression (STARD) series of studies undertaken by the National Institute of Mental Health.[16,17] This paper has tables estimating rate of response up to four lines of therapy. This is quite old so it may only afford data for a subset of the available therapies you are looking at.

---

[12] Samples H, Mojtabai R. Antidepressant self-discontinuation: results from the collaborative psychiatric epidemiology surveys. Psychiatric Services. 2015 May 1;66(5):455-62.

[13] Lutz W, Stulz N, Köck K. Patterns of early change and their relationship to outcome and follow-up among patients with major depressive disorders. Journal of affective disorders. 2009 Nov 1;118(1-3):60-8.

[14] Roca M, Baca E, Caballero L, de Polavieja PG, Casillas M, Valladares A, Gilaberte I. Early response and remission as predictors of a good outcome of a major depressive episode at 12-month follow-up: a prospective, longitudinal, observational study. The Journal of clinical psychiatry. 2011 Oct 4;72(2):5328.

[15] Schlagert HS, Hiller W. The predictive value of early response in patients with depressive disorders. Psychotherapy Research. 2017 Jul 4;27(4):488-500.

[16] Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, Niederehe G, Thase ME, Lavori PW, Lebowitz BD, McGrath PJ. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR* D report. American Journal of Psychiatry. 2006 Nov;163(11):1905-17.

[17] Ross EL, Zivin K, Maixner DF. Cost-effectiveness of electroconvulsive therapy vs pharmacotherapy/psychotherapy for treatment-resistant depression in the United States. JAMA psychiatry. 2018 Jul 1;75(7):713-22.

One issue with the assumption needed to back up your hazard rate approach outlined above is that it instantly assumes that any new treatment will have the same relationships between initial efficacy and later effectiveness waning, as for other treatments. This distinction is important because two different treatments may have a very different structure to their long-term effectiveness. One may work initially with a high level of efficacy but have a high level of waning so that by the fourth or fifth line of therapy it is barely effective at all, whereas another may start with a lower level of initial efficacy (first line of therapy), but 'maintain' that level of effectiveness without waning over many treatment cycles. Given the importance of a component of long-term maintenance of efficacy, to assume the same rate of waning for all therapies would do a disservice to a treatment that may have a more robust maintenance, and also would discourage innovators from developing new therapies or approaches that achieve higher rates of maintenance of efficacy for long periods of time.

7.3 We have proposed two approaches to derive direct medical cost inputs in our model: a "top-down" approach (identify proportion of all-cause medical costs that can be attributed to MDD), or a "bottom-up" approach (identify individual resource requirements and unit costs; and sum across all resource use items).

- Is there one approach you would recommend over the other?
- Are you aware of any data sources/studies that we should look into for this issue?

Both top-down and bottom-up costing approaches have their own challenges. Bottom-up tends to be limiting in that you include only the costs you think are relevant. Top-down can include costs that are irrelevant and require more validation from multiple other sources of data. The better solution is to use a top-down source of cost that can allow potential hidden costs to be identified, by design. With this approach, you can simply compare an MDD population to a matched control, and any marginal difference can be allocated to MDD. Greenberg et al. use this method for MDD. These studies look at direct costs and all other healthcare costs as well as indirect costs, such as burden of suicides and work loss per patient.[18,19]

7.2 Of the possible data sources for utility inputs listed in Table 8, is there one we should prioritize? Are there other sources we should consider?

IVI has highlighted three potential sources for health state utility values in its protocol. It is important to try to tease out randomized clinical trial populations when estimating health state utility value (HSUV) in MDD populations, as the population is prone to strong Hawthorne effects,[20] which can lead to

---

[18] Greenberg PE, Fournier AA, Sisitsky T, Simes M, Berman R, Koenigsberg SH, Kessler RC. The economic burden of adults with major depressive disorder in the United States (2010 and 2018). Pharmacoeconomics. 2021 Jun;39(6):653-65.
[19] Greenberg PE, Fournier AA, Sisitsky T, Pike CT, Kessler RC. The economic burden of adults with major depressive disorder in the United States (2005 and 2010). The Journal of clinical psychiatry. 2015 Feb 25;76(2):0-5.
[20] Benedetti F, Carlino E, Piedimonte A. Increasing uncertainty in CNS clinical trials: the role of placebo, nocebo, and Hawthorne effects. The Lancet Neurology. 2016 Jun 1;15(7):736-47.

exaggerated utility values for baseline untreated and non-response states in RCTs.[21] If Brockbank[22] is chosen, any estimate should be limited to non-RCT sources for utility values. Alternatively, Revicki and Wood[23] would be the best source for the United States. Here the utility weight for no-treatment / no-response was 0.30 for example, whereas in RCT studies non-responders / no-treatment states can be artificially high with Brockbank suggesting between 0.5 and 0.7.

Conclusion

We are appreciative of the robust and transparent process you have put forward in which stakeholders may participate. Thank you for considering our input, and we are happy to provide additional comments as helpful.

Sincerely,

Tony Coelho
Chairman
Partnership to Improve Patient Care

---

[21] McCarney R, Warner J, Iliffe S, Van Haselen R, Griffin M, Fisher P. The Hawthorne Effect: a randomised, controlled trial. BMC medical research methodology. 2007 Dec;7(1):1-8.

[22] Brockbank J, Krause T, Moss E, Pedersen AM, Mørup MF, Ahdesmäki O, Vaughan J, Brodtkorb TH. Health state utility values in major depressive disorder treated with pharmacological interventions: a systematic literature review. Health and quality of life outcomes. 2021 Dec;19(1):1-7.

[23] Revicki DA, Wood M. Patient-assigned health state utilities for depression-related outcomes: differences by depression severity and antidepressant medications. Journal of affective disorders. 1998 Feb 1;48(1):25-36.