



Response to: *Comment Request Federal Register / Vol. 86, No. 23 / Friday, February 5, 2021 P. 8362 [Document Identifiers CMS–10203, CMS– 2088–17, CMS–1763, and CMS–1696]*

The SNP Alliance is a nonprofit leadership organization that focuses on beneficiaries with the most complex physical, mental health, and social risk individuals who are also primarily low income, dually-eligible for Medicare and Medicaid.

The Alliance strongly supports obtaining member input around health plans actions affecting outcomes of care.

However, our analysis of the Medicare Health Outcomes Survey (HOS) and the potential value in comparison to burden reveals opportunity for substantial improvement. We recommend that the Department of Health and Human Services and Centers for Medicare and Medicaid Services invest in such improvement.

Background

While widely used and researched, the Medicare HOS was originally developed using a veteran population of primarily older white males. To date, the survey has not been re-validated in diverse populations including those for whom English is not their primary language, those with low health literacy, those with chronic or progressive physical or intellectual disabilities. Furthermore, the HOS is currently available in just three languages (English, Spanish and Mandarin). When sub-populations are not adequately represented and/or survey questions do not reflect the experiences of the individuals being surveyed, the risk is that resultant data and assumptions or conclusions from that data may be skewed.

Analysis

Over the last few years the SNP Alliance has surveyed health plans, analyzed peer-reviewed literature, and gathered information from researchers and experts to better understand the strengths and limitations around use of the Medicare Health Outcome Survey (HOS) in the Medicare Stars rating and quality measurement system for Medicare Advantage plans. There is compelling evidence indicating need for improvement. The challenges are particularly pronounced for special needs plans as they serve diverse, low-income, disabled, and chronic care, complex, or advanced-illness populations. These populations and plans are more likely to be negatively impacted by the limitations in HOS instrument, methods, and scoring.

Limitations discovered include:

- *Diversity of Medicare Population is not fully Considered*
- *Instrument Limitations as Applied to People with Special Needs*
- *Inadequate Methods of Administration*

- *Information Decay, Time Lag*
- *Proxy Bias*
- *Results May be Affected by Underlying Characteristics of Enrollment*
- *Attribution, Context Issues*
- *Inadequate Models and Adjustments*
- *Cautions Regarding Analysis, Interpretation, and Reporting*

The SNP Alliance offers potential solutions and strategies for addressing limitations, from those that would require only a change in analysis and reporting to those that would require more extensive modification. These include:

1. Level the measurement field
2. Use additional exceptions and exclusions
3. Discontinue the longitudinal design
4. Use better predictive models and data sources
5. Use alternative instrument(s)
6. Re-test HOS with disabled and diverse groups
7. Improve methods of administration and accommodation
8. Conduct analysis at a finer level and report on results
9. Provide greater transparency and consumer education
10. Convene stakeholder and expert panels

In our remarks to this PRA regulation we provide in-depth analysis on one measure generated from the HOS to illustrate the limitations, burden, and shine a light on accuracy, validity, and utility issues—the Physical Functioning Activities of Daily Living measure.

SNP ALLIANCE ANALYSIS & POSITION on HOS – PFADL Measure Focus as Example

We focus on the potential burden as compared to value of the measure entitled: *Physical Functioning Activities of Daily Living* (PFADL). This is a longitudinal change measure¹ derived from the Health Outcome Survey (HOS). It measures, at the contract level, the change over two years in the physical functioning of beneficiaries enrolled in MA contracts.

FOOTNOTE:

¹ The PFADL combines two physical functioning questions (limitations in moderate activities and climbing stairs) with the six activities of daily living questions to create a Likert-type scale. PFADL scores are created from responses to the baseline and the two-year follow-up questions. Contract-level change scores are on a 0-100 scale, with 100 equivalent to all MA beneficiaries retaining 100% of baseline function over two years and 0 corresponding to every beneficiary in the MA contract experiencing maximum decline. The PFADL change measure score for an MA plan is its mean change score rather than the proportion of individuals passing the measure.

CMS has indicated they will use a methodology to correct population characteristic differences by grouping PFADL scores into four baseline groups from lowest to highest and comparing the average scores for each group two years later. They will use a case-mix linear regression model that sets predictive rates for the second PFADL score.

The SNP Alliance has long supported a focus on functional status as an important characteristic of people with complex chronic, behavioral, and long-term care needs. Understanding a person's functional status is a very important piece of the puzzle in crafting a tailored response to care. However, this measure does not focus on how a health plan is helping to address the physical functioning of a beneficiary to address activities of daily living functional limitations. We explain further below.

It is important to understand how the PFADL measure is derived. It comes from HOS using the questions about abilities such as moving tables, or hobbies such as bowling or playing golf. We have frequently discussed the problems with this wording, that is sometimes considered insensitive or lacking relatability to cultural/ethnic diverse or disabled individuals, and is potentially confusing to the respondent.

The PFADL measure is longitudinal. It will compare two snapshots of functional status at two-year intervals for a sample of people enrolled in the plan. Health plan contracts with a large proportion of members who are frail, have complex chronic, degenerative, progressive conditions, have high social risk factors and other life events that impact their physical health and their ability to perform activities of daily living—are much more likely to have a random sample of members (HOS sample) with poor physical functioning in ADLs. It is predictable given the characteristics of the population.

We have concerns regarding value of the measure toward accurately measuring change for specific beneficiaries who may bear the burden of the survey without value regarding the utility of the results. This is because of the wording of the questions used to derive the PFADL measure. The wording seems to set up a pre-emptive bias against beneficiaries who are disabled or have permanent limitations in functions and against the plans/providers who care for these people. Such individuals would report that they have substantial limitations and cannot perform these ADLs *without help*. The way the question about six ADLs to derive the PFADL is worded, the beneficiary responding would be limited in reporting only what he/she/they can do on their own. The question reads:

*Because of a health or physical problem, do you have any difficulty doing the following activities **without special equipment or help from another person**? (bathing, dressing, eating, getting in and out of chairs, walking and using the toilet)*

A person responding who is already permanently disabled or has substantial ADL limitations and cannot perform one or more of these ADLs without help starts at the lowest end of the scale (“Unable to do”).

If the person is in a SNP or other health plan that provides long term services and supports, adaptive equipment, puts in grab bars, provides a shower bench, lift chair for the stairwell, special shoe-horn, button-hole tool, personal care assistant for ADL assistance—it can result in change where the person is able to perform some of these functions *with help*. However, the HOS question asks the person about performing the task “without special equipment or help from another person”—so two years later that individual would again say “Unable to do.”

Even though this person would have improved in ADLs with assistance - where his or her level of function in bathing, dressing, eating or moving is better *with help*—the measure would not capture this change. The point is not that the individual still can't do the task alone – but that the person can now do something they couldn't before *because they now receive help*.

The difference between being unable to dress or bathe alone and being able to do some of these daily activities with the help of a personal care assistant, shower bench, grab bar, or special shoe-horn—can be life changing and is an important difference to note.

We are concerned that the PFADL measure specification using the wording of the HOS questions may close the door to showing improvement in function – among people who have substantial limitations—as these individuals use assistance. This calls into question the utility of the measure for special needs populations where many members have chronic limitations of activities of daily living and/or are permanently disabled. Thus, the administrative and beneficiary burden associated with conducting the survey and calculating this measure is high without providing benefit. This is an inequity.

It will not really matter that the HOS sample average PFADL scores is separated into four groups based on functioning at baseline. If the person cannot perform ADLs without help, or if decline occurs from baseline to 2 years later—the average for that group falls and the health plan would be rated lower (negatively) on performance of the PFADL measure. This would be the case even if the groups actually improve in functional ability *with help*. This is counter-productive and may be a dis-incentive for health plans to invest in adaptive equipment, innovations such as home visits that help change the home environment or add home support services to the individual's care plan and service package—since these things will not be “counted” when the PFADL questions are asked of the respondent. We do not believe this is CMS' intent for this measure.

Another point with the current HOS question is that there are only three response options in the Likert scale around ADLs (“No, I do not have any difficulty;” “Yes, I have difficulty,” or “I am unable to do this activity”) this limits the opportunity to mark improvement. A five-point or seven-point scale would show more change.

We are most concerned that this measure misses a very important point: If a health plan serves primarily functionally-impaired or disabled individuals--*the opportunity to demonstrate improvement on this measure will not be present or severely hampered*.

An alternative self-report measure on ADL functional change achieved *with help*--and how the health plan actions helped to bring this change to effect—is necessary. There are other measures, instruments and methods focusing on function which may be better suited to the purpose of performance evaluation in this area.

Recommendations

- Re-test HOS in ethnically diverse and non-English speaking populations, who are dually eligible.
- Re-test HOS in a complex, chronically ill population with high behavioral/mental health conditions and high social determinant of health risk factors

- Explore alternatives to HOS and the measures derived from HOS, such as PROMIS measures
- Address attribution issue – that is connect the measurement results to actions under the health plan control
- Add in contextual information that the beneficiary can provide to help guide toward understanding and quality improvement efforts
- Make changes in the HOS to include achievement of functional status improvement with or without help in the PFADL measure

We recommend a re-check around HOS to ensure that the principles for good measurement systems are upheld when a measure is to be used for performance evaluation and comparison-- that is: (1) accuracy, (2) validity, (3) utility, (4) feasibility, (5) equity/fairness.

We make these recommendations to re-balance the measurement burden on special needs health plans and their enrolled beneficiaries so that the burden of conducting the HOS and calculating the measure offers value.

Thank you for considering this analysis is your determination of burden and equity around measurement and the use of the HOS. We also offer our White Paper on HOS for further information: <https://www.snpalliance.org/wp-content/uploads/2020/04/snpa-paone-hos-white-paper-final-dec-2018-1.pdf>

Sincerely,



Cheryl Phillips, MD
President & CEO
SNP Alliance
cphillips@snpalliance.org
202-960-1688

Deborah Paone, DrPH, MHSA
Performance Evaluation Lead
SNP Alliance
dpaone@snpalliance.org