Federation of American Hospitals®

Charles N. Kahn III
President and CEO

June 17, 2019

The Honorable Seema Verma, Administrator
Centers for Medicare and Medicaid Services
Department of Health and Human Services
Hubert H. Humphrey Building, Room 445-G
200 Independence Avenue, SW.
Washington, D.C. 20201

RE: CMS-1710-P, Medicare Program; Inpatient Rehabilitation Facility Prospective Payment System for Federal Fiscal Year 2020 and Updates to the IRF Quality Reporting Program

Dear Administrator Verma,

The Federation of American Hospitals (FAH) is the national representative of more than 1,000 investor-owned or managed community hospitals and health systems throughout the United States. Our members include teaching and non-teaching hospitals in urban and rural America, as well as inpatient rehabilitation, psychiatric, long-term acute care, and cancer hospitals. The FAH appreciates the opportunity to comment to the Centers for Medicare & Medicaid Services (CMS) about the referenced Notice of Proposed Rulemaking Medicare Program; Inpatient Rehabilitation Facility Prospective Payment System for Federal Fiscal Year 2020 and Updates to the IRF Quality Reporting Program.

Proposed Refinements to the Case-Mix Classification System Beginning with FY 2020

Summary

In its Fiscal Year (FY) 2020 Inpatient Rehabilitation Facility (IRF) proposed rule, CMS has proposed a significant overhaul to the current law case-mix group (CMG) system. These changes alter how individual cases are scored under the motor index, grouped into CMGs and paid, which ultimately changes the operational, clinical and financial meaning of the CMGs.

The FAH is concerned with these changes for a number of reasons. First, data and technical documentation published by CMS have been inadequate to fully understand and engage in a constructive dialogue around these changes. Second, rehabilitation experts and stakeholders were not engaged to review and offer their perspectives on the appropriateness of the motor index weighting approach despite major methodological departures from the current system. Third, the

implementation of a weighting index that has not benefited from their review of the clinical and operational implications could introduce unintended and unforeseen consequences into the IRF PPS that may ultimately reduce IRF PPS accuracy in the short term and affect the accuracy of rate setting and payment determination in a future unified PAC PPS implementation.

Ultimately, the FAH recommends that CMS delay the implementation of the Section GGH weighted motor score until further knowledge of the impact of changed GGH items, written methodology for constructing the weighted score, and stakeholder engagement are more robust.

CMG Changes

The FY2020 proposed rule includes the use of IRF-PAI assessment Section GG and H items (referred to here as Section GGH) rather than FIM® assessment data; this is consistent with CMS's FY2019 final rule. While we are concerned that the Section GGH data are new and perhaps less accurately coded than the established and familiar FIM® data, we acknowledge the effort to align patient assessment data across post-acute care settings and that CMS has already finalized this decision in the FY2019 IRF PPS rulemaking cycle.

New to the FY2020 proposed rule is the addition of a weighting approach to the Section GGH motor score using a novel methodology. This weighting approach – implemented as a key input to rate setting – translates to an overhaul of the CMG definitions and cut points, patient grouping criteria, payment weights and CMG average length of stay values. When compared with the current law FIM® item weights, key differences are seen in the magnitude of the motor score weights for certain items. For instance, the eating assessment item carries a motor score weight of 2.7, while the similar current law item carries a much lower weight of 0.6; with weights centered around 1, this is in effect a reversal of current payment policy.[1] The FY2020 proposed rule also removes one Section GGH item from the motor score index (roll left and right); this item was included in the FY2019 proposed motor score index.

- **Dobson | DaVanzo Impact Analyses of Proposed Changes**

The data needed to fully assess the proposed changes and model possible alternatives is only available to CMS and its contractor, RTI. As such, analytic contractor Dobson DaVanzo and Associates (Dobson | DaVanzo) obtained the best available data to gain a better understanding of the proposed changes. Uniform Data System for Medical Rehabilitation (UDSMR) provided Dobson | DaVanzo with a de-identified case-level database under a Data Use Agreement including about 84% of IRF discharges from FY2017 and FY2018, though that database lacked relevant information such as claims data and links to facility information that would enable accurate case cost estimates and the development of alternative weighting proposals.

UDSMR and Dobson | DaVanzo collaborated to program and validate rule-relevant variables, including developing unadjusted payments for the current law system, the FY2019 proposed rule and the FY2020 proposed rule using a consistent base rate (FY2020 proposed) to best match the analytic set as described in the RTI technical report accompanying this year's rulemaking.[2]

[1] Morley, Melissa, Benjamin Silver, Anne Deutsch, Nicole Coomer, Allison Dorneo, Laura Coots Daras, Melvin Ingber. "Analyses to Inform the Use of Standardized Patient Assessment Data Elements in the Inpatient Rehabilitation Facility Prospective Payment System" Prepared by RTI International for CMS, March 2019. https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/InpatientRehabFacPPS/Downloads/IRFPPSAnalysis2019RTI032219.pdf.

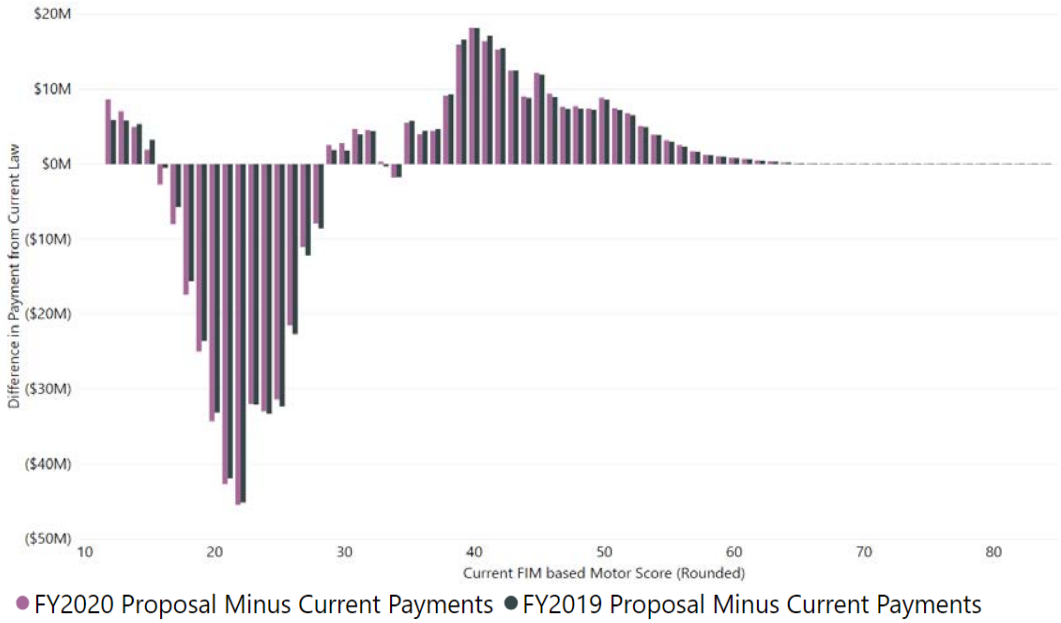
[2] Morley, Melissa, Benjamin Silver, Anne Deutsch, Nicole Coomer, Allison Dorneo, Laura Coots Daras, Melvin Ingber. "Analyses to Inform the Use of Standardized Patient Assessment Data Elements in the Inpatient

While both the current law and recently proposed systems have comparable predictive performance (overall model adjusted R2) in aggregate, their treatment of individual cases is quite different. This is likely due to the fundamental differences between FIM® and Section GGH data, as described in Dobson | DaVanzo's report on the FY2019 IRF PPS proposed rule, attached as an appendix[3]. However, the shift from FIM® to Section GGH data (which were introduced to the IRF-PAI in FY2016 and are still relatively unfamiliar to clinicians), may be further exacerbated by prematurely weighting the motor index.

Examining cases for payment differences, contractor Dobson | DaVanzo found the FY2020 proposal creates significantly different payments for the same patients compared to the current law system, particularly for those patients with relatively poor function as rated by the FIM®. Many cases rated with lower functionality in the FIM®-based weighted motor index appear to experience substantial decreases in payments in both the FY2019 and FY2020 proposed systems.

Exhibit 1 demonstrates payment differences from the current FIM®-based CMG system to both the FY2019 and FY2020 proposals. Cases with poor functional status at admission (measured by a low score on the FIM® motor index, located on the left side of the chart) tend to have the greatest reduction in payments under the proposed overhaul. This chart suggests that the FY2019 and FY2020 proposals have introduced a new concept of which cases require what level of resources.

Exhibit 1: *Total Payment Differences (FY2020 and FY2019 Proposals Compared to Current Law) by FIM® Motor Index Score (Rounded to Nearest Integer)*



Rehabilitation Facility Prospective Payment System" Prepared by RTI International for CMS, March 2019. https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/InpatientRehabFacPPS/Downloads/IRFPPSAnalysis2019RTI032219.pdf.

[3] Dobson, Allen, Alex Hartzman, Kimberly Rhodes, Sung Kim, and Joan DaVanzo, "Proposed Refinements to the IRF PPS Case Mix Classification System and Removal of FIM® Items from the IRF PAI: Impact Analysis and Potential Considerations" Dobson DaVanzo and Associates. June 2018.

Data Availability

As discussed in the preliminary FAH comment letter submitted in response to the FY2020 proposed rule, CMS has not made sufficiently detailed data available to stakeholders and the public to meaningfully comment on this proposal. Further, the CMS impact file was extremely difficult to interpret as it did not include a set of variable definitions.

As mentioned above, only CMS and its contractor RTI have the data required to evaluate the proposed changes and to test alternative proposals of the motor index or CMG system.

This critical data links the following information at the case level:

1) Section GGH functional status information,
2) Other claims and beneficiary characteristics, and
3) Imputed case-level costs.

Without these linked case-level data elements, outside analysts are handicapped in their attempts to engage in a productive dialogue with CMS to improve the IRF PPS' accuracy and clinical appropriateness. Data availability is a particularly important issue for this rulemaking cycle in which CMS has departed from the methods used in the current law payment system without robust justification, detailed technical description, or expert clinical review.

Methodological Transparency and Stakeholder Input

The technical report provided by CMS and its contractor RTI is lacking important details, which limits the cogency of stakeholder commentary to the proposed rule. First, though the proposed rule mentions testing of multiple alternative motor index weighting schemes,[4] only one was described in the technical report.

Second, the method used to develop the FY2020 proposed weighted score differs substantially from that used in current law. From the technical report descriptions and limited data, the public cannot know whether this difference may yield substantive differences in motor score weights even if applied to the same data. Specifically, the change from ordinary least squares cost prediction across all cases (RTI/FY2020[5]) to averaging and normalizing an "optimal" set of weights within each RIC (RAND/FY2006-FY2019[6]) may significantly affect the resulting motor index weights.

CMS and contractor RTI did not justify this departure from the current motor score weight setting framework nor describe how the proposed system differs from an "optimally weighted" one.

[4] FY2020 IRF PPS Proposed Rule, CMS, page 17250.
[5] Morley, Melissa, Benjamin Silver, Anne Deutsch, Nicole Coomer, Allison Dorneo, Laura Coots Daras, Melvin Ingber. "Analyses to Inform the Use of Standardized Patient Assessment Data Elements in the Inpatient Rehabilitation Facility Prospective Payment System" Prepared by RTI International for CMS, March 2019. Page 2-5. https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/InpatientRehabFacPPS/Downloads/IRFPPSAnalysis2019RTI032219.pdf.
[6] Relles, Daniel A., Greg Ridgeway, Grace M. Carter, and Melinda Beeuwkes Buntin, "Possible Refinements to the Construction of Function-Related Groups for the Inpatient Rehabilitation Facility Prospective Payment System" Prepared for CMS by RAND Corporation, 2005. Page 24. https://www.rand.org/pubs/technical_reports/TR207.html.

Clearly some of the weight values now represent inverse values relative to the current law system, is clinically and operationally concerning. During this comment period, stakeholders have not been able to determine whether these results are logical given unfamiliarity with the Section GGH items for payment and a lack of case cost imputation data. As a particular example, it is not clear whether it is clinically reasonable that some of the self-care items are now more highly weighted while some mobility items now possess relatively lower weights compared to the current motor score weight values. This represents a major shift in the clinical conceptualization of the CMGs as the proposal would shift clinical importance from mobility to self-care. It is unlikely the true relationship of function to resource use has changed and the lack of justification for the methodological departure or expert oversight of the results is troubling.

Finally, no outside technical experts or stakeholders were engaged to inform the system's development with clinical and operational input. Seeking stakeholder (particularly clinician) expertise and input is a key step in proposing payment system changes as these individuals can help review the reasonableness of model outputs and interpret them in clinical and operational contexts. As such, clinicians and other stakeholders were unable to offer potentially useful and important insights or to assure the soundness of CMS's motor score weighting proposal.

Recommendation

The proposed CMG changes from current law are significant and substantially alter how individual cases are rated and paid, which ultimately changes the operational and clinical meaning of the CMGs. We caution against moving forward with the proposed CMG changes given that data and technical documentation to examine and engage in a dialogue around the changes have been inadequate. Furthermore, stakeholders and rehabilitation experts were not engaged during the development of the proposed CMG changes which is concerning due to methodological departures from the current system and significantly different motor score weight values.

Implementation of a clinically and operationally unvetted weighting index could introduce unintended consequences into the IRF PPS that may ultimately reduce IRF PPS accuracy in the short term and affect rate setting and payment determination in an eventual unified PAC PPS implementation, based in part on Section GGH data. FAH recommends that CMS delay the implementation of the Section GGH weighted motor score until further knowledge of the GGH items, methodology for constructing the weighted score, and stakeholder engagement is more robust.

Proposed Amendments to § 412.622 to Clarify the Definition of a Rehabilitation

The FAH appreciates CMS's proposal to clarify and codify that rehabilitation hospitals ultimately determine whether a physician qualifies as a rehabilitation physician as so defined, and we urge the agency to finalize this proposal. While board-certified physiatrists play a crucial caregiver and leadership role in rehabilitation hospitals, they are not alone in doing so. Physicians representing other specialties can and do also display the leadership and caregiving skills and experience that clearly qualify them as a rehabilitation physician. The rehabilitation hospital is best positioned to make that determination, and CMS has, through its welcome and needed proposal, made that clear.

Facility-Level Adjustment Factors

The FAH strongly recommends that the CMS monitor and report on these factors annually and adjust them if a material change is noted. CMS should provide as part of its annual rulemaking a detailed analysis of the Agency's review justifying either a continued freeze or an update to the adjustment factors. Further, CMS should consider responding to comments expressing concerns with the specific adjustment factors, as their inclusion at any level in the proposed rule arguably reflects a policy proposal subject to notice and comment and the rules governing that public process. Finally, we ask that CMS adjust all three factors at a minimum once every three years in order to maintain payment accuracy. This will help ensure a dynamic and accurate IRF payment system that recognizes and responds to change in the cost of care and promotes the delivery of efficient and effective IRF services.

Proposed Update to the IRF Wage Index to Use Concurrent FY IPPS Wage Index Beginning with FY 2020

We support the CMS proposal to help level the playing field among providers who often compete for the same labor pool by applying to IRFs a wage index concurrent with other post-acute care settings as well as acute care hospitals, namely the same current year pre-classified wage index value as acute care hospitals rather than prior year values. This is consistent with past comments the FAH has submitted. In addition, we recommend that CMS adopt other wage index policies that apply to IPPS hospitals such as geographic reclassifications, and that any changes to wage index policy that CMS may finalize in the FY 2020 IPPS also apply to IRFs in FY 2020. Finally, the FAH opposes a wage index exclusively for IRFs. Among other concerns such as data validity, unlike IPPS acute care hospitals, IRFs are not located evenly across geographic areas, which could well distort the data on which wage index values are derived.

Proposed Update to Payments for High-Cost Outliers under the IRF PPS for FY 2020

The FAH supports an outlier pool no greater than three percent and that CMS use the latest available data in finalizing the FY 2020 fixed loss threshold. The FAH notes that in recent years CMS appears to have paid out the full amount indicating that the fixed loss threshold has achieved its target when viewed across all hospitals. The FAH also supports including in the calculation of the fixed loss threshold historical outlier reconciliation dollars as has been proposed in the FY 2020 IPPS. Finally, according to an analysis of CMS FY 2020 rate-setting files the top ten decile IRFs will receive some 57 percent of IRF outlier payments, a disproportionately high amount. In addition, there appears to be a disconnect between patient acuity, efficiency, and outlier payments, all of which suggests that CMS consider imposing a cap on outlier payments a hospital could receive.

Proposed Revisions and Updates to the IRF Quality Reporting Program (QRP)

New Measures for FY 2022

- Transfer of Health Information to the Provider -- PAC Measure

This proposed measure would assess whether a current reconciled medication list is given to the subsequent provider when a patient is discharged or transferred from his or her current PAC setting.

The FAH supports the inclusion of this measure as it focuses on promoting care coordination and may lead to reductions in patient harm and overuse of services. The FAH also appreciates that the measure allows facilities the flexibility to implement this transfer of a current reconciled medication list based on their current capabilities. The FAH encourages CMS to submit the measure to NQF for endorsement review in the near future.

- Transfer of Health Information to the Patient -- PAC Measure

This related proposed new measure would assess whether a current reconciled medication list was provided to the patient, family, or caregiver when a patient was discharged from a PAC setting to a private home/apartment, board or care home, assisted living, group home, transitional living, or home under care of a home health service organization or hospice.

The FAH supports the inclusion of this measure as it focuses on promoting care coordination and may lead to reductions in patient harm and overuse of services. The FAH also appreciates that the measure allows facilities the flexibility to implement this transfer of a current reconciled medication list based on their current capabilities. The FAH encourages CMS to submit the measure to NQF for endorsement review in the near future.

- Update to the Discharge to Community PAC Measure

CMS proposes to update the specifications for this measure to remove baseline nursing facility residents. The FAH supports the removal of baseline nursing facility residents from inclusion in this measure. This revision serves to further improve the validity and usefulness of the measure and FAH appreciates CMS' responsiveness to this issue.

Removal of Topped Out Measures

The FAH believes that three measures currently reported in the IRF QRP are topped out. These measures are:

- Changes in Skin Integrity Post-Acute Care: Pressure Ulcer/Injury
- Application of Percent of Residents Experiencing One or More Falls with Major Injury (Long Stay) (NQF #0674)
- Application of Percent of Long-Term Care Hospital Patients with an Admission and Discharge Functional Assessment and a Care Plan That Addresses Function (NQF #2631)

All three measures demonstrate little to no variation across facilities and do not provide information that enables end users (e.g., facilities, patients, caregivers) to distinguish high quality of

care. The FAH encourages CMS to proposal these measures for removal to reduce data collection and reporting burdens for IRFs.

Standardized Patient Assessment Data Reporting Beginning with FY 2022

CMS proposes that IRFs report SPADEs, most of which are the same or modifications of the SPADEs that were previously proposed and not finalized. Reporting would begin during FY 2022.

The FAH believes that CMS should not add to the burden of IRFs unless it is truly necessary to track and report on the quality of care provided by these facilities. Regardless, the inclusion of the SPADES items must align with the goals of "Patients over Paperwork" and Meaningful Measures" initiatives. The FAH is concerned that the current proposed expansion does not achieve these goals and we urge CMS to reconsider the expansion of SPADES to all of these items. The FAH strongly encourages CMS to conduct a thorough analysis of the SPADE items currently collected and determine whether any could potentially be retired, particularly if the item is not needed for any of the quality measures.

The FAH notes that the Brief Interview for Mental Status (BIMS) and the Confusion Assessment Method (CAM) items under Cognitive Function and Mental Status are redundant to what is currently collected at admission and does not provide any additional useful information in light of the additional collection burden. The Intravenous (IV) Medications and High-Risk Drug Classes under Special Services, Treatments and Interventions are duplicative of other items within the SPADES and the Conditions of Participation, respectively. The items under Social Determinants of Health are also not likely to change during a patient's length of stay and should only be assessed at the time of admission. The FAH encourages CMS to ensure that a standardized approach is undertaken to ensure consistency of the items such as race/ethnicity as patients are transferred across settings and over time. The FAH urges CMS to prioritize consistency of the data and avoid duplication of data collection requirements to the greatest extent possible.

In addition, CMS should carefully consider whether any of these proposed additions could be captured via claims, which would serve as a reliable and valid source for these data. FAH identified several items that would be readily available and potentially more accurate through this alternative data source, particularly under Special Services, Treatments and Interventions. The National Beta Test report demonstrated that there was significant disagreement across abstractors leading to low interrater reliability kappa statistics and were items that testing participants identified as difficult and complex to collect (Rand, 2018). Use of administrative claims for some of these items would reduce data collection burden and improve the data quality.

****

The FAH appreciates the opportunity to submit these comments. If you have any questions, please contact me at 202-624-1534, or Steve Speil, Executive Vice President, Policy at sspeil@fah.org or 202-624-1529.

Sincerely,

APPENDIX

Proposed Refinements to the IRF PPS Case Mix Classification System and Removal of FIM® Items from the IRF PAI

Impact Analysis and Potential Considerations

Submitted to:
Federation of American Hospitals

Allen Dobson, Ph.D.
Alex Hartzman, M.P.A, M.P.H
Kimberly Rhodes, M.A.
Sung Kim
Joan DaVanzo, Ph.D., M.S.W.

Tuesday, June 26, 2018 — *Final Report*

INTRODUCTION

This report summarizes analyses of the potential impact of proposed changes to the Inpatient Rehabilitation Facility Prospective Payment System (IRF PPS) functional status categorization case-mix system as proposed in the technical report, "Analyses to Inform the Potential Use of Standardized Patient Assessment Data Elements in the Inpatient Rehabilitation Facility Prospective Payment System"[1] (referred to herein as "RTI technical report" or "technical report") and the FY2019 IRF PPS proposed rule. We found that while the proposed system is budget neutral and leaves many facets of the current case mix system in place, it would fundamentally alter the clinical foundation of the IRF PPS through changing the functional assessment items and not weighting them. We note several data and interpretation issues with the new items and characterize the effect of the changes through additional analyses of payments and CMG changes from the current to the proposed system. Dobson DaVanzo and Associates (Dobson | DaVanzo) was commissioned by the Federation of American Hospitals (FAH) on behalf of its membership to conduct this analysis.

CMS has not made the assessment data (required to fully analyze the impacts of the proposed changes) available to date. Although IRF functional assessment data are generally available for use more broadly, the data for the proposed assessment items in question are not yet included in publicly available data sets. Furthermore, an impact file was not made available with the publication of the proposed rule. For purposes of these analyses, Uniform Data System for Medical Rehabilitation (UDSMR) provided Dobson | DaVanzo with a deidentified case-level database contains roughly 84% of IRF cases in the relevant period (FY2017).[2] With these data, in cooperation with UDSMR, we were able to conduct an independent impact assessment of the proposed functional status scoring system.

If implemented without first addressing several important data and interpretation issues pointed to in this report, we anticipate a variety of contradictory financial incentives that may affect the payment accuracy and long-term financial stability of the IRF field.

[1] Morley, Melissa, Benjamin Silver, Anne Deutsch and Melvin Ingber, "Analyses to Inform the Potential Use of Standardized Patient Assessment Data Elements in the Inpatient Rehabilitation Facility Prospective Payment System" RTI, April 2018, page 1-5 https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/InpatientRehabFacPPS/Downloads/IRFPPSAnalysis2018RTI.pdf.

[2] The data for this study was obtained and used with permission from the Uniform Data System for Medical Rehabilitation, a division of UB Foundation Activities, Inc. The service marks and trademarks associated with the FIM® instrument are all owned by Uniform Data System for Medical Rehabilitation, a division of UB Foundation Activities Inc.

Overview of Proposed Changes to the Inpatient Rehabilitation Facility (IRF) Prospective Payment System (PPS) Case-Mix System

Current Case-Mix System

The current law IRF PPS case-mix system uses Functional Independence Measure (FIM®) motor and cognitive assessment items to predict case resource use, set rates and determine payments. The motor scores are derived from 12 functional assessment items (listed in Exhibit 1) and weighted to reflect their relative contribution to the costs of care.

Exhibit 1: FIM® Motor Score Functional Assessment Items

Item	IRF PAI Number
Eating	39Aa
Grooming	39Ba
Bathing	39Ca
Dressing, upper body	39Da
Dressing, lower body	39Ea
Toileting	39Fa
Bladder management	39Ga
Bowel management	39Ha
Transfers, bed/chair/wheelchair	39Ia
Transfers, toilet	39Ja
Walk/wheelchair	39La
Stairs	39Ma

Motor scores are used in conjunction with the patients' rehabilitation impairment category (RIC), and sometimes age and the sum of the cognitive score items to place patients into Case Mix Groups (CMGs). Each CMG carries a set of relative payment weights; the exact weight for a given patient is selected based on his or her comorbidity tier (0-3). Each CMG and comorbidity tier combination has an associated average length of stay which is used as part of the payment determination for early transfer cases to another institutional setting. If a patient's actual length of stay was less than the CMG-comorbidity tier average and the patient was transferred to another institutional setting, the case is paid on a per diem basis (calculated by multiplying the standard payment amount by the case's relative payment weight and dividing the product by the CMG-comorbidity tier's average length of stay. This per diem amount is then multiplied by the case length of stay plus 0.5 days). The CMG is the primary mode by which case payments are differentiated by patient in the IRF PPS.

Proposed Case-Mix System

The proposed case-mix and patient classification system would replace FIM® items with standardized patient assessment data (SPAD) elements from the Quality Indicators section of the IRF-PAI. These SPAD elements proposed for the new case-mix system are a subset

of standardized assessment items that were implemented to align functional outcome measurement across IRFs, SNFs, LTCHs and HHAs as mandated by the Improving Medicare Post-Acute Care Transformation (IMPACT) Act of 2014. IRF clinicians began collecting patient data on these items for the first time starting in October 2016.

In the proposed system, a collection of 19 items (listed in Exhibit 2) are summed to create an unweighted functional score. Cognitive items were examined for inclusion but not used in the proposed system.

Exhibit 2: Standardized Patient Assessment Data (SPAD) Elements Motor Score Functional Assessment Items

Item	IRF PAI Number
Eating	GG0130A1
Oral hygiene	GG0130B1
Toileting hygiene	GG0130C1
Shower/bathe self	GG0130E1
Upper-body dressing	GG0130F1
Lower-body dressing	GG0130G1
Putting on/taking off footwear	GG0130H1
Roll left and right	GG0170A1
Sit to lying	GG0170B1
Lying to sitting on side of bed	GG0170C1
Sit to stand	GG0170D1
Chair/bed-to-chair transfer	GG0170E1
Toilet transfer	GG0170F1
Walk 10 feet	GG0170I1
Walk 50 feet with two turns	GG0170J1
Walk 150 feet	GG0170K1
One-step curb	GG0170M1
Bladder continence	H0350
Bowel continence	H0400

Neither the proposed rule nor the RTI technical report provide detail on the rationale for selecting these particular 19 items as opposed to others. And while the RTI technical report does explain that assessment items that were not included "are more challenging and less likely to be assessed on admission",[3] neither document elaborates as to why specific items

[3] Morley, Melissa, Benjamin Silver, Anne Deutsch and Melvin Ingber, "Analyses to Inform the Potential Use of Standardized Patient Assessment Data Elements in the Inpatient Rehabilitation Facility Prospective Payment System" RTI, April 2018, page 1-5 https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/InpatientRehabFacPPS/Downloads/IRFPPSAnalysis2018RTI.pdf.

were excluded, such as those pertaining to wheelchair locomotion, which are not necessarily more challenging items than the included items related to walking.

The unweighted functional score, along with the patients' RIC and sometimes age, are used to assign patients to a CMG. As with the current system, each CMG carries a set of relative weights which are selected for each patient depending on his or her comorbidity tier, and each CMG and comorbidity tier combination has an associated average length of stay. Payment calculations for early transfer cases remain the same as in the current system, but the average length of stay values are different given the changes to the CMG grouper.

If finalized, these proposed changes would be implemented in FY2020 (discharges on or after October 1, 2019). In both the technical report and proposed rule, CMS has indicated that changes will be implemented in a budget neutral manner indicating the projected aggregate payment amount to all IRFs for a given case mix would not change. As such, the effect of changes manifests on the distribution of payments across CMGs as well as in the distribution of cases in each CMG, CMG payment rate and average length of stay.

Differences between Proposed SPAD and Current FIM® Items

While the assessment window for the SPAD and FIM® items is the same (within the first 3 days of IRF admission), SPAD items measure patient functionality in a clinically and quantitatively different way, changing the basic clinical implications (if not treatment implications) of the CMGs. These changes are likely exacerbated by the high prevalence of non-scaled SPAD responses (assessment item responses which are subsequently rescaled to be the most impaired assessment response for the purposes of calculating a functional score for CMG assignment), discussed further below.

- SPAD assessment items are based on patients' "usual performance" on the given item within the assessment period. In comparison, FIM® are based on patients' lowest performance, or highest burden of care, on a given item. Both measures attempt to capture the intensity of assistance patients require, but from different perspectives. While it is not yet clear whether this proposed change to items measuring "usual performance" is inherently problematic, we note numerous implementation issues with the SPAD assessment items.
    - We have received anecdotal reports from FAH members that "usual performance" thus far is not adequately defined, leading to potential inconsistency and problems in inter-rater reliability and stability over time. FAH members reported audits of medical records and found that in more than half of cases the technical reviewer would revise SPAD item responses substantially, suggesting that the inter-rater reliability on SPAD items is currently low. This is particularly concerning given the

high prevalence of non-scaled SPAD assessment responses used in analyses supporting the proposed rule as they substantially affect rate setting and payment determination but may represent coding confusion rather than reasonable non-assessment.

  - Further, CMS repeatedly issued significant updated guidance on the definitions of SPAD items in relation to "usual performance" and other aspects during FY2017, the period in question. As SPAD response meaning has shifted over the first year since implementation to reflect updated guidance, the data used in this proposed payment system change cannot be representative of final or consistent coding practices by definition.

- 6-point scale on SPAD items vs. 7-point scale on FIM®. SPAD items characterize a narrower range of functional impairment but are definitionally broader to attempt to capture a similar range of impairment. See Exhibits 3 and 4, below.
- Both the FIM® and the SPAD assessment items contain possible responses which are subsequently rescaled to be the most impaired assessment response for the purposes of calculating a functional score for CMG assignment. In this report, we are calling such responses "non-scaled responses." However, there are four possible non-scaled responses on the SPAD items while there is just one non-scaled response on FIM®. For SPAD items, these additional non-scaled responses are "patient refused", "not applicable", "not attempted due to safety concerns" and blank; for the FIM® items, the non-scaled response is "activity did not occur". See Exhibit 3 for descriptions of each response items.
- Unlike the FIM®, SPAD motor score items are not weighted, and cognitive items are not present in the proposed CMG grouper.

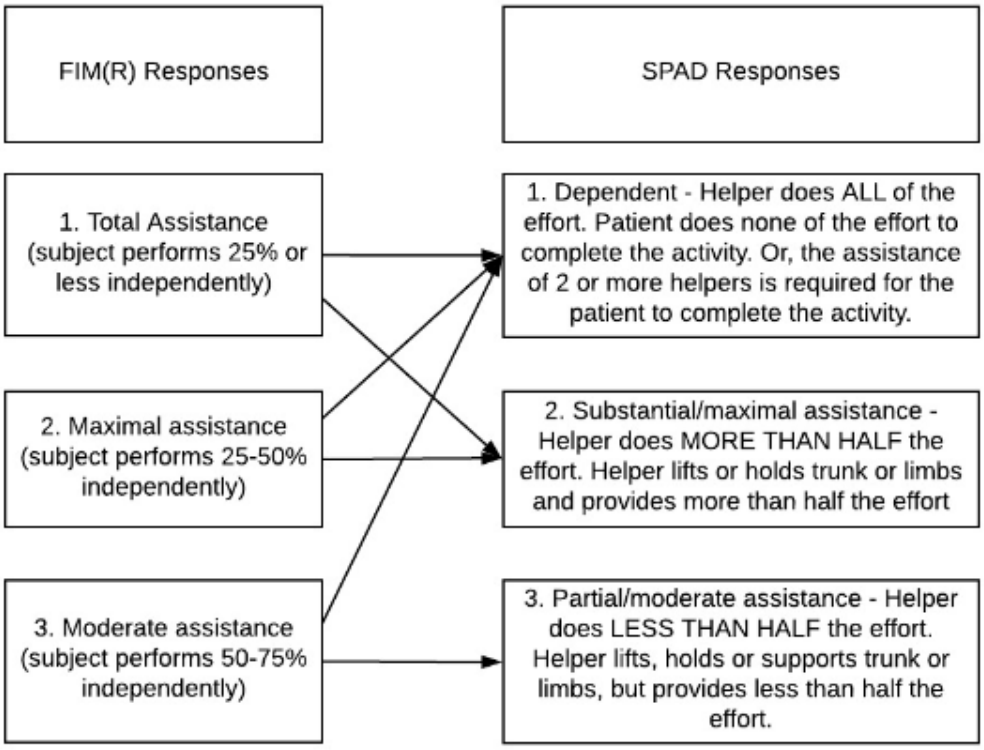
Exhibit 3: Comparison of FIM® and SPAD Item Scales

Non-scaled responses are in italic font.

Score	FIM® Levels	SPAD Levels
0	*Activity does not occur (admission only)*	
1	Total assistance (subject performs 25% or less independently)	Dependent - Helper does ALL of the effort. Patient does none of the effort to complete the activity. Or, the assistance of 2 or more helpers is required for the patient to complete the activity.
2	Maximal assistance (subject performs 25-50% independently)	Substantial/maximal assistance - Helper does MORE THAN HALF the effort. Helper lifts or holds trunk or limbs and provides more than half the effort
3	Moderate assistance (subject performs 50-75% independently)	Partial/moderate assistance - Helper does LESS THAN HALF the effort. Helper lifts, holds or supports trunk or limbs, but provides less than half the effort.
4	Minimal assistance (subject performs 75% or more independently)	Supervision or touching assistance - Helper provides VERBAL CUES or TOUCHING/STEADYING assistance as patient completes activity. Assistance may be provided throughout the activity or intermittently.
5	Supervision (subject performs 100% independently)	Setup or clean-up assistance - Helper SETS UP or CLEANS UP; patient completes activity. Helper assists only prior to or following the activity.
6	Modified independence (device used)	Independent - Patient completes the activity by him/herself with no assistance from a helper.
7	Complete independence	*Patient refused*
88		*Not attempted due to medical concern or safety issue*
9		*Not applicable*

Leaving aside that the FIM® and SPAD items measure fundamentally different constructs of patient impairment (worst versus "usual" performance), we note that the scales used to describe impairment correspond across assessment instruments in ways which may have contributed to case payment redistributions. In Exhibit 4, we examine the three scaled items indicating most severe functional impairment under FIM® and SPAD and indicate how FIM® ratings may be reinterpreted under the broader SPAD response scale. For example, the highest severity SPAD response can be inclusive of a broad range of FIM® performance because of the clause "including patients requiring two or more helpers"; further, the SPAD response of 2 spans the range of impairment (50% or more work done by assistant) included in FIM® responses 1 and 2.

Exhibit 4: Comparison of FIM® and SPAD Responses on Highest Severity Ratings



Resulting Changes to Payment System

To incorporate selected SPAD items into CMGs, CMS and RTI took a relatively similar approach to grouping patients as in the original IRF PPS CMG setting. That is, CMGs are defined by clinical condition and combined motor score (sometimes accounting for patient age as well). This process resulted in a different set of CMGs with updated definitions reflecting the SPAD functional assessment items and updated combined motor score. The proposed system would have fewer CMGs (88 rather than the current 92). This would alter the clinical and operational interpretation of each group as relative payment weights and average group lengths of stay also changed to reflect the new CMGs.

Key aspects of the payment system would not change under this proposal. Definitions and usage of rehabilitation impairment categories (RICs), comorbidity tier groups and age were not altered in the proposed rule.

CMS Rationale for Proposed Changes

In the proposed rule and technical report, CMS offered two arguments in support of the proposed change:

- Reduce IRF administrative burden by removing the FIM® items and associated functional modifiers; and
- Support the broader movement to align data collection across PAC settings.[4]

CMS did not include a specific rationale for changing the clinical meaning of the CMGs or elaborate on how removing the FIM® would support the alignment of the SPAD items across settings.

Dobson | DaVanzo's Replication of the Current and Proposed Case-Mix Groupings and Payments

Given that an impact file was not made available by CMS, Uniform Data System for Medical Rehabilitation (UDSMR) provided us with a deidentified case-level database that accounts for approximately 84% of traditional Medicare IRF cases. The database included the current and proposed assessment items, patients' age, along with the RIC, comorbidity tier, and proposed and current CMG, CMG average length of stay and unadjusted case payment amounts. Unadjusted case payment amounts are defined as payments that were not adjusted for outlier payments, wage adjustment, rural vs. urban status, low-income status or teaching status. Using unadjusted payments in both the current and proposed systems creates an 'apples to apples' comparison and allows us to analyze the effects of the proposed system changes. Similarly, the impact analyses presented in the RTI report are based on unadjusted payments.[5]

For the purposes of this study, we restricted the dataset to FY2017 cases (patients discharged between October 1, 2016 and September 30, 2017). This is the same period used in the setting of relative weights for the FY2020 proposed CMGs and in the proposed rule and technical report impact assessments.

Using the assessment item, patient age, comorbidity tier and length of stay variables provided in the UDSMR dataset, Dobson | DaVanzo replicated the current and proposed CMG groupings and unadjusted payments. Dobson | DaVanzo also verified that the CMG

[4] Proposed Rule: Medicare Program; Inpatient Rehabilitation Facility Prospective Payment System for Federal Fiscal Year 2019 (83 FR 20972 - 21015 // CMS-1688-P). Pages 20988 - 20989. Centers for Medicare & Medicaid Services. May 8, 2018. https://www.federalregister.gov/d/2018-08961.

[5] Morley, Melissa, Benjamin Silver, Anne Deutsch and Melvin Ingber, "Analyses to Inform the Potential Use of Standardized Patient Assessment Data Elements in the Inpatient Rehabilitation Facility Prospective Payment System" RTI, April 2018, page 3-1. https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/InpatientRehabFacPPS/Downloads/IRFPPSAnalysis2018RTI.pdf.

average length of stay variables in the UDSMR dataset matched the FY2020 values provided in the FY2019 proposed rule[6] and the FY2017 values in the FY2017 final rule.[7]

To replicate the CMG groupings, Dobson | DaVanzo used the UDSMR dataset variables and the FY2017 CMG relative weights from the FY2017 final rule[8] and the FY2020 CMG relative weights from the FY2019 proposed rule.[9]

With this robust sample size and independent replication of current and proposed CMG groupings and payments, we are confident the following findings are generalizable to IRFs generally.

FINDINGS

Budget Neutrality

Budget neutrality between the current and proposed IRF case mix system has been asserted by CMS and contractor RTI through the proposed rule and preceding technical report. However, CMS did not make data publicly available to verify neutrality or to examine the impacts of proposed changes. As such, independent analysts assessed budget neutrality and redistributive effects via samples and data subsets without a complete database.

Within our sample, we find $332M in case payment reductions and $304M in case payment increases, for a net change of negative $28M (-0.46%) in revenue on a basis of the UDSMR dataset of 310,175 discharges and $6,067M revenue in FY2017. In combination with results shared by eRehab (which is representative of much of the remainder of the industry), we conclude payment simulation results are consistent with the CMS statement of a budget neutral case mix system change. However, as discussed below, there are legitimate questions about the budget neutrality of this system moving forward if implemented as outlined in the proposed rule.

Though budget neutrality is a typical condition of payment system changes unless otherwise explicitly altered by legislation, it is not a wholly sufficient measure of the appropriateness of a proposed system change. We also examine case payment and facility

[6] Proposed Rule: Medicare Program; Inpatient Rehabilitation Facility Prospective Payment System for Federal Fiscal Year 2019 (83 FR 20972 - 21015 // CMS-1688-P). Table 9 on pages 20992 – 20994. Centers for Medicare & Medicaid Services. May 8, 2018. https://www.federalregister.gov/d/2018-08961.

[7] Final Rule: Medicare Program; Inpatient Rehabilitation Facility Prospective Payment System for Federal Fiscal Year 2017 (42 CFR 412 pages 52055-52141). Table 1 on pages 52063-52070. Centers for Medicare & Medicaid Services. August 5, 2016. https://www.federalregister.gov/d/2016-18196.

[8] Final Rule: Medicare Program; Inpatient Rehabilitation Facility Prospective Payment System for Federal Fiscal Year 2017 (42 CFR 412 pages 52055-52141). Table 1 on pages 52063-52070. Centers for Medicare & Medicaid Services. August 5, 2016. https://www.federalregister.gov/d/2016-18196.

[9] Proposed Rule: Medicare Program; Inpatient Rehabilitation Facility Prospective Payment System for Federal Fiscal Year 2019 (83 FR 20972 - 21015 // CMS-1688-P). Table 9 on pages 20992 – 20994. Centers for Medicare & Medicaid Services. May 8, 2018. https://www.federalregister.gov/d/2018-08961.

revenue change as these are indicative of the potential redistribution of revenue and may be useful in predicting industry stress and other threats to beneficiary access when considering proposed system changes. Regardless of the assessment outcome of the proposed change effect on total or average facility revenue, proposed system changes must align with clinical practice and be based on accurate measures – tests which the proposal does not seem to pass.

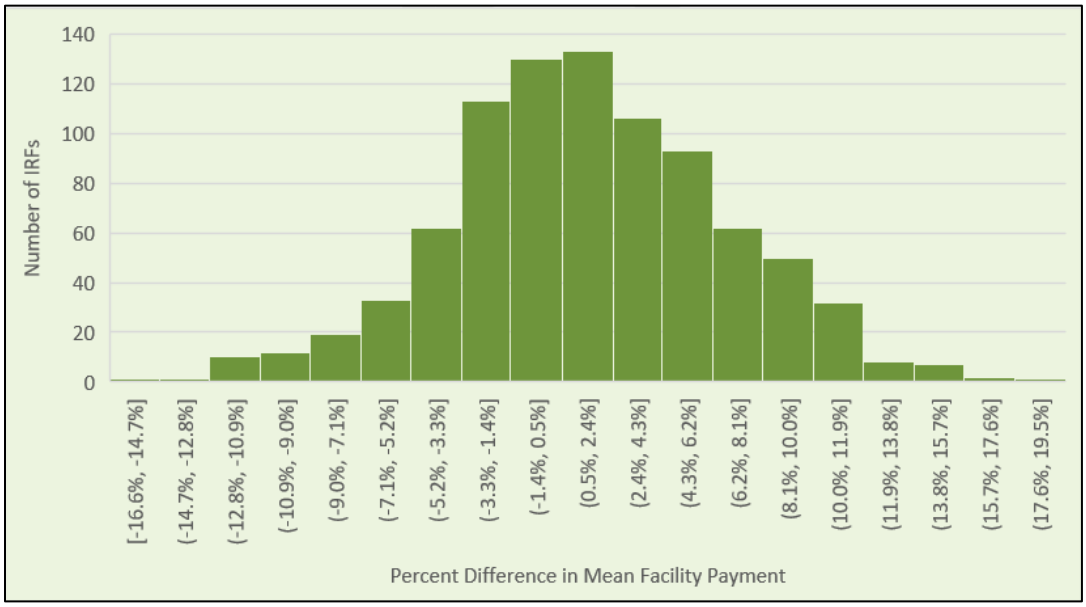
Redistribution of Cases and Case Payments

Though the proposed system change appears to be budget neutral, we found the proposed rule would redistribute a substantial portion of payments as:

- Case Mix Group definitions are changed, and thus
- CMG payment weights and length of stay are revised to reflect the new groupings.

Given this, we detected substantial changes in overall facility revenue with 54% of facilities in the sample experiencing a +/-3% change; the full distribution of facility percent change in revenue is displayed in Exhibit 5, below.

Exhibit 5: Distribution of Facility Mean Percent Change in Facility Payment



Source: Dobson DaVanzo analysis of UDSMR FY2017 cases.

Exhibit 6, below, demonstrates the regrouping of beneficiaries from current law CMG (down the left) to proposed CMG (across the top) for stroke cases. In general, the group makeups shift appreciably, suggesting incongruence between the current and proposed sys-

tems. That is, a case that is currently CMG 105 could move to CMG 102 or 103 in the proposed system. In general, cases have migrated towards the middle CMGs in the proposed system, suggesting less variation than is observed under the current system. While to some extent this is expected given the proposed removal of CMGs 107-110, the amount of case CMG movement from one system to the next is significant.

Exhibit 6: Regrouping of Beneficiaries between Current and Proposed CMGs (RIC 1, Stroke, n = 60,956)

		FY2020 Proposed Stroke CMGs						
		101	102	103	104	105	106	Total
FY2017 Current Stroke CMGs	101	1,358	297	22	1			1,678
	102	1,762	1,578	328	7	1	3	3,679
	103	458	465	114	6		2	1,045
	104	1,187	3,183	2,156	60	2	15	6,603
	105	424	1,901	3,300	257	5	24	5,911
	106	195	917	3,531	811	27	96	5,577
	107	84	432	2,760	1,737	80	347	5,440
	108	8	80	783	1,289	4,415		6,575
	109	36	173	1,397	1,783		1,108	4,497
	110	25	115	1,136	2,957		15,718	19,951
	Total	5,537	9,141	15,527	8,908	4,530	17,313	60,956

Source: Dobson DaVanzo analysis of UDSMR FY2017 cases.

This change in grouping cases redistributes payments in ways that are predictable from a data perspective, but perhaps not from a clinical one. That is, intuitively beneficiaries that have a higher severity should be reimbursed at higher levels (so far as this is predictive of resource use); however, it is not intuitive why a beneficiary who falls into one CMG under the current system may fall into a very different functional grouping under the proposed system. See Exhibit 7, which shows the aggregate payment change associated with the regrouping of stroke CMGs as an example.

Exhibit 7: Average Case Payment Change ($) in (FY2020 – FY2017) by Regrouped CMG (RIC 1, Stroke, n = 60,956)

FY2017 Current Stroke CMGs	FY2020 Proposed Stroke CMGs						
	101	102	103	104	105	106	Total
101	2,896.23	6,188.27	9,772.13	11,890.21			3,574.42
102	310.29	3,605.54	7,654.18	12,660.77	516.01	23,796.57	2,421.14
103	(1,638.94)	1,563.83	5,197.25	9,409.72		19,560.39	635.99
104	(2,662.95)	658.07	4,533.90	8,340.42	7,370.91	7,788.76	1,414.63
105	(5,024.98)	(1,655.53)	2,316.59	6,972.70	4,732.77	9,796.79	747.38
106	(7,180.49)	(3,700.71)	327.51	4,952.71	5,212.52	10,421.43	272.64
107	(9,251.57)	(5,836.38)	(1,738.79)	3,108.92	4,179.50	8,852.05	130.27
108	(15,199.82)	(10,904.35)	(6,920.04)	(1,752.42)	878.80		(728.72)
109	(12,504.03)	(8,841.12)	(4,450.84)	582.14		6,425.65	(8.87)
110	(20,946.46)	(16,457.65)	(11,577.77)	(6,477.76)		393.19	(1,430.66)
Total	(873.23)	(328.99)	(495.48)	(955.26)	969.96	1,030.12	(59.61)

Source: Dobson DaVanzo analysis of UDSMR FY2017 cases.

Non-Scaled Assessment Responses and their Impact on Payment

Missing, non-applicable, and other non-scaled assessment responses are included in both the FIM® (response = 0) and proposed SPAD (response = 7, 88, 9 or blank) motor scores which are used in the current and proposed CMG systems. Missing or otherwise non-scaled responses are automatically given the highest severity rating for the measure, which tends to receive higher payments once aggregated in the combined functional score and assigned a CMG.

Though missing items recoded as high severity scores are included in both assessments' composite score approaches, we found them to be much more prevalent in the SPAD-based motor scores. In general, we found:

- 86% of cases were missing at least one SPAD item compared to 56% of cases missing at least one FIM® motor item.
- 69% of cases had at least two missing SPAD items compared to 10% missing at least two FIM® items.
- 81% of cases had a greater number of missing SPAD items than FIM® items; 2% had more missing FIM® items.

We also analyzed missing scores at the item level and found that when we attempted to compare similar items across both systems (current vs. proposed), certain items that had low prevalence in missing items in the FIM® had much higher prevalence of missing items in the SPAD. The largest discrepancies between the share of missing items from the FIM® to SPAD were in the locomotion category. For instance, we compared the FIM® Locomotion Walk/Wheelchair item to the SPAD Walk 150 feet item, as shown in Exhibit 8 below. We believe that the most appropriate one-to-one comparison with the FIM® Walk/Wheelchair item would be 'Admission Walk 150 feet (GG0170K)', given that the IRF-PAI manual indicates that the FIM® Walk distance of interest is 150 feet, which mirrors the selected equivalent SPAD item distance.

Exhibit 8: Share of Non-Scaled Item Responses (at the Case-Level) for Walk Items in the Current and Proposed Systems

Percent of Cases Missing Admission FIM® Locomotion - Walk/Wheelchair (39L) (Where item = "0")	Percent of Cases Missing Admission Walk 150 feet (GG0170K) (Where item = "7", "88","9", or blank)
6.3%	75.5%

Source: Dobson DaVanzo analysis of UDSMR FY2017 cases.

The existence of proportionally more missing items in the SPAD motor score indicates potential 'blind spots' in the proposed system; if an item was able to be scored under FIM® for a particular patient, but the 'same' item on the SPAD is missing a response, this could be a signal for gaps in patient assessment.

Cases with more missing SPAD items tended to receive higher payments under the proposed system.

- 15% of cases had 5 or more items with a non-scaled SPAD response. These cases had an average 4% increase in payment under the proposed system.

- Exhibit 9 demonstrates that, on average, with increasing number of missing SPAD items, proposed FY2020 payments increase relative to current law payments.

Exhibit 9: Distribution of Payment Changes by Count of Non-Scaled SPAD Item Responses

Number of Non-Scaled SPAD Item Responses	Percent of Cases	Average LOS	Sum of FY2017 Unadjusted FPP	Sum of Difference in Payment (FY2020 minus FY2017 Unadjusted FPP w/FY2017 Base)	Percent Change of Difference in Payment (FY2020 minus FY2017 Unadjusted FPP w/FY2017 Base)
0--3	66.8%	11.6	$ 3,832,670,636	$ (80,270,777)	-2.10%
4--8	31.5%	14.7	$ 2,163,404,614	$ 46,790,388	2.16%
9--17	1.7%	8.5	$ 71,135,196	$ 5,434,895	7.86%
Total	100%	12.5	$ 6,067,210,447	$ (28,045,495)	-0.46%

Source: Dobson DaVanzo analysis of UDSMR FY2017 cases.

Thus, CMGs, payment weights, average CMG length of stay, and payment determinations appear to be influenced by the large portion of items with non-scaled responses. This may be indicative of SPAD data quality issues and warrants closer examination due to the potential for adding instability to the system (payments could decrease as measurement practices improve) or conversely inadvertent incentives to the IRF PPS. **Overall, it is not clear that the FY2017 SPAD data reflecting a significant proportion of non-scaled responses is appropriate for inclusion in the IRF PPS as proposed.**

Fundamentally, the SPAD non-scaled responses represent a broader variety of reasons (compared to the FIM non-scaled response) for why a clinician may not have assigned a scalable response to a given assessment item, including patient refusal or inability to perform the activity. However, these responses are also used much more often under SPAD and it is not clear if the results created by these non-scaled responses are adequate reflections of the patients' functional abilities. As in the FIM® -based CMGs, this tends to compress the effect of sets of patients with somewhat differing functional limitations. If the high level of SPAD non-scaled responses is indeed correct, it may indicate that the 6-point response scale inadequately captures the range of functional impairment, compressing on the end of the least functional beneficiaries. Further, it is also apparent that the greater number of items proposed to be included in the combined SPAD motor score (19 measures, up from 12 FIM® measures in the current motor score) also increases the likelihood of including non-scalable items in rate setting and payment determinations.

Implementing SPAD-based CMGs at this time may incentivize the use of non-scaled assessment responses as items scored as non-responsive tend to result in higher payments. As there are four such responses per SPAD assessment question, there are more

opportunities to compress beneficiaries into groups in unexpected ways (when interpreting the array of possible responses which could lead to the same CMG).

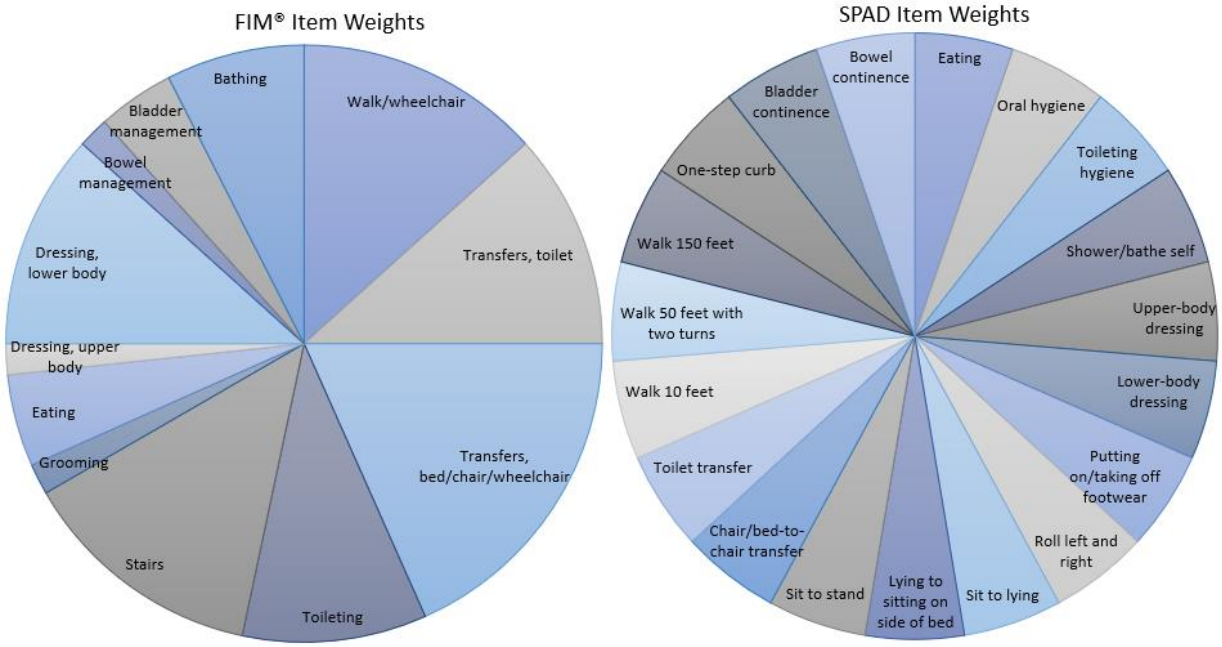
The ability of the SPAD items and motor score to predict resource usage may be eroded over time if the incentive to code items as missing or choosing a non-scaled response remains. As payment weights are reassessed over time, the payment system could become less sensitive to beneficiary impairment should a greater preponderance of non-scaled responses be used (as incentivized by the payment system). This would be detrimental to long-term data quality and make it difficult to assess patient improvement or conduct accurate risk adjustment for quality measures and Alternative Payment Models. Should the assessment data quality decrease, it would also undermine CMS's long-term goal of being able to compare patient functional status across PAC settings.

Conversely, if the prevalence of non-scaled item responses decreases through improved understanding of SPAD ratings over time, it would likely raise combined motor scores (indicating lower impairment) which could then systematically lower payments for providers under the proposed CMGs. The presence of these non-scaled responses and the associated uncertainty of how the prevalence of these responses will change over time is indicative of a system that may not be ready for implementation. Given this, if the proposed system changes were implemented we would anticipate the need for future changes to both the CMG weights and perhaps the base rate as well to assure continued budget neutrality over time.

Implications of an Unweighted Motor Score

CMS has not proposed to weight the new motor score at this time. As outlined in the proposed rule, the motor score assessment items would each have equal weight in contrast to the current motor score, which is weighted according to each items' associated burden of care. Exhibit 10 depicts the current weighting of the motor score compared to the proposed, unweighted motor score composition.

Exhibit 10: FIM® vs. SPAD Item Weights



We have been unable to locate related documentation or analyses to identify whether CMS carefully evaluated the potential implications of removing the weighting system tied to the motor score. In the past, considerable research has gone into identifying and applying optimal motor score weights. This research (led by RAND for CMS) found improvements in explanatory power of models to predict actual costs when applying weights to the motor FIM® items, thus adding accuracy to the payment system.[10] Regardless, given that the burden of care represented by some items (oral hygiene, for instance) do not carry the same weight or burden as others (such as chair/bed to chair transfer, or stairs), the exploration of an item-level weighting system would be beneficial.

It is unclear whether CMS plans to implement a weighting system for the proposed motor score items in the future, as the proposed rule states that CMS is "not proposing to apply a weighting methodology to the motor score *at this time*" [Emphasis added].[11] However, if CMS were to implement a weighting system, the impacts of this would flow through the entire payment system and thus should be carefully considered.

[10] Relles, Daniel, Gregory Ridgeway, Grace Carter, and Melinda Beeuwkes Buntin. "Possible Refinements to the Construction of Function-Related Groups for the Inpatient Rehabilitation Facility Prospective Payment System" RAND Health supported by the Centers for Medicare and Medicaid Services. 2005. https://www.rand.org/pubs/technical_reports/TR207.html.

[11] Proposed Rule: Medicare Program; Inpatient Rehabilitation Facility Prospective Payment System for Federal Fiscal Year 2019 (83 FR 20972 - 21015 // CMS-1688-P). Page 20990. Centers for Medicare & Medicaid Services. May 8, 2018. https://www.federalregister.gov/d/2018-08961.

Differences in Evaluating Patient Functional Status

The items and scale on which patients would be assessed in the proposed system fundamentally changes the way patient functionality is measured. One of the effects this change has is that it redistributes which patients are considered the most functionally impaired. This could be potentially harmful to patients considered highly impaired under the current system as it would reduce treatment budgets for these patients. As the change in measurement set and motor score were not well-justified clinically as well as the prevalence of numerous data reporting problems, we cannot say whether the change in payment for the most severely impaired patients currently under FIM® is appropriate.

In a sense, the proposed system may be biased against the most severe patients, who are most in need of the care that IRFs provide. Exhibit 11 shows how payments are redistributed from many of the most impaired beneficiaries to less impaired beneficiaries, as rated by the weighted FIM® composite score.

Exhibit 11: Distribution of Payment Changes by Case Composite Weighted FIM® Motor Score

Case Composite Weighted FIM® Motor Score	Percent of Cases	Average LOS	Sum of Difference in Payment (FY2020 minus FY2017 Unadjusted FPP w/FY2017 Base)	Average of Difference in Payment (FY2020 minus FY2017 Unadjusted FPP w/FY2017 Base)
12	5.10%	16.35	$ 1,057,508	$ 401
14	7.04%	15.75	$ 84,158	$ 62
16	7.86%	15.25	$ (470,442)	$ (327)
18	7.31%	14.86	$ (1,378,710)	$ (1,118)
20	6.97%	14.43	$ (2,186,542)	$ (1,908)
22	6.67%	13.73	$ (1,852,793)	$ (1,718)
24	6.35%	13.18	$ (1,597,472)	$ (1,570)
26	6.13%	12.52	$ (533,189)	$ (559)
28	5.88%	11.94	$ 136,690	$ 141
30	5.77%	11.53	$ 271,276	$ 297
32	5.43%	11.19	$ 60,653	$ 67
34	5.08%	10.80	$ 277,638	$ 350
36	4.76%	10.44	$ 353,509	$ 480
38	4.35%	9.91	$ 783,082	$ 1,162
40	3.90%	9.52	$ 911,382	$ 1,472
42	3.27%	9.21	$ 671,089	$ 1,287
44	2.47%	8.89	$ 559,016	$ 1,448
46	1.89%	8.45	$ 433,272	$ 1,436
48	1.31%	8.09	$ 407,749	$ 2,006
50	0.96%	7.59	$ 389,160	$ 2,540
52	0.56%	7.15	$ 271,456	$ 2,923
54	0.37%	6.68	$ 176,066	$ 2,907
56	0.22%	6.30	$ 104,167	$ 2,640
58	0.13%	5.86	$ 56,857	$ 2,404
60	0.09%	5.41	$ 43,803	$ 2,376
62	0.05%	5.06	$ 22,889	$ 2,134
64	0.03%	4.49	$ 12,068	$ 1,940
66	0.02%	4.02	$ 5,871	$ 1,467
68	0.01%	4.16	$ 2,735	$ 1,466
70	0.01%	2.84	$ 2,536	$ 966
72	0.01%	3.33	$ 3,419	$ 1,576
74	0.00%	3.29	$ 5,186	$ 2,420
76	0.00%	2.90	$ 840	$ 691
78	0.00%	2.43	$ 1,270	$ 809
80	0.00%	1.40	$ 339	$ 339
82	0.00%	2.00	$ 679	$ 339
Grand Total	100.00%	12.53	$ (28,045,495)	$ (90)

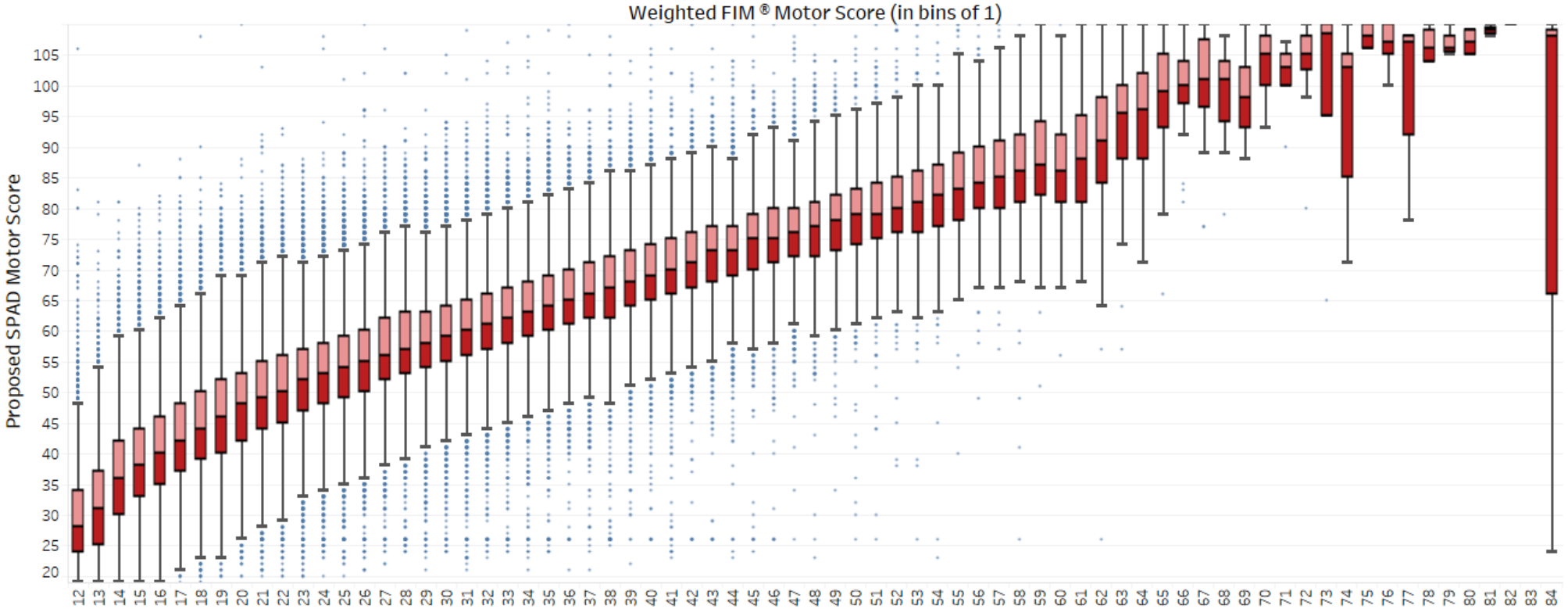
Source: Dobson DaVanzo analysis of UDSMR FY2017 cases.

Exhibit 12 shows the wide variability in patients' SPAD motor scores as compared to their weighted FIM® motor score. We also see overlapping interquartile ranges of SPAD motor scores from one FIM® score to the next. Both of these findings suggest that the SPAD items and their associated motor function scale blur the distinction between functional status levels, which are clearly distinguished from one another in the current FIM® scoring system. Ultimately, it appears that the SPAD scale has some amount of "noise" and makes it difficult to clearly discern different levels of functionality, reducing the distinctness of one SPAD level from another.

We also looked at the same distribution restricted to stroke (RIC 1) cases (not shown) and compared it to the FIM® motor score thresholds for CMGs 102 and 103. We found that a broad range of SPAD scores fit into the relatively narrow range of motor score cut points under the current CMG system.

Exhibit 12: Distribution of FY2017 Cases by Weighted FIM® Motor Score and Proposed SPAD Motor Score

Each "box" shows the 25th and 75th percentile SPAD motor scores for each FIM® motor score (truncated to closest integer); the midpoint of each box (where the light and dark red meet within each box) shows the 50th percentile of the SPAD motor score for each IRF cases' FIM® motor score. The "whisker" (lower and upper-most bound of each box plot) extend to values within 1.5 times the inter-quartile range. The dots represent cases that fall outside of the range defined by + or – 1.5 times the interquartile range and show the minimum and maximum values.



Source: Dobson DaVanzo analysis of UDSMR FY2017 cases.

Differences in Assessing Change in Patient Functional Status

While the admission functional assessment scores are the only ones used for payment determination purposes, we also analyzed the discharge assessment scores to get a sense for how patients' change in functionality is captured in the current and proposed systems. To do this, we constructed admission and discharge motor functional scores and calculated the percent change in functionality from admission to discharge for each case. We used the unweighted motor FIM® score in order to see the raw change in functionality from admission to discharge, allowing for the most direct "apples to apples" comparison from the current to the proposed scores.

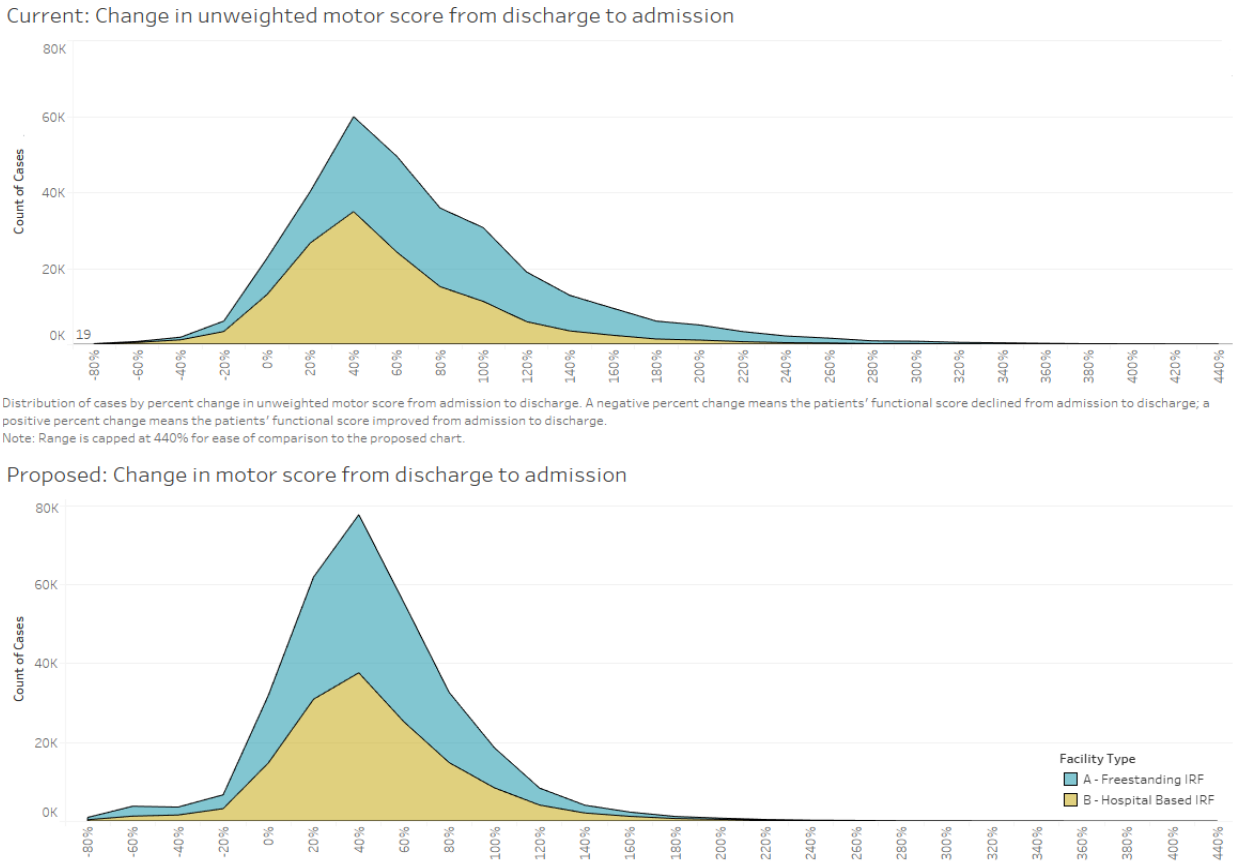
Based on this analysis, we found that the proposed motor SPAD items and scale may dilute measurement of "true" changes in patient functionality. If the new system causes patients to appear as less functionally impaired compared to the results of the case FIM® assessment, then less improvement over time is likely to be measured.

Exhibit 13 shows the distribution of cases by their percent change in functional status as measured by the unweighted motor FIM® score (top graph) and by the proposed motor SPAD score (bottom graph).

A slightly wider and flatter distribution is observed in the FIM® motor score distribution compared to the SPAD motor score distribution, suggesting that there is more pre-post (admission to discharge) variation (or sensitivity) in the FIM® than in the SPAD.

The narrower, more concentrated distribution of the proposed SPAD motor scores indicates that patients may appear to be less severe in the proposed system than they show in the current FIM® system, and that changes in functionality from admission to discharge may not be captured with as much granularity in the SPAD as they are in the FIM®. This could affect future quality measurement (and potentially payment), which focuses on the outcomes of therapy and the provision of other IRF services.

Exhibit 13: Distribution of FY2017 Cases by Percent Change in Motor Score from Discharge to Admission: Current (Unweighted FIM®) vs. Proposed (SPAD)



Source: Dobson DaVanzo analysis of UDSMR FY2017 cases.

Differences between FIM® and SPAD Response Meanings

The proposed functional status items are not direct equivalents to the current FIM® items and in some cases may make patients look less functionally impaired on similar items. For example:

- "Easier" items were chosen.
    - FIM® walk/wheelchair item was replaced with three walking items.
    - Locomotion 12 step FIM® item was replaced with a 1 step SPAD item.
    - Bowel and bladder FIM® items captured more nuance in acuity than SPAD items.

Other changes to the motor score include:

- Additional assessment items that do not replace existing FIM® items:
    - Roll left and right, sit to lying, lying to sitting on side of bed, and sit to stand
    - The removal of cognitive assessment items.
    - Removal of weights in the motor score calculation.

Last, the core concept of what is being measured is fundamentally different. SPAD items measure "usual performance" at admission as opposed to FIM® which captures worst admission performance. This is a change to the concept of how patient need is considered in relation to resource needs. This change, as well as the inclusion of more assessment items to inform the motor score, inherently alters which patients are viewed as having any given level of functional impairment. Exhibit 14a shows the redistribution of beneficiary scores from the FIM® to SPAD on the similar eating measure.

Exhibit 14a: Distribution of Cases by their Current and Proposed Assessment Item Score

		Admission SPAD Eating										
		1	2	3	4	5	6	7	9	88 (missing)		Total
Admission FIM® Eating	0	0.01%	0.00%	0.01%	0.02%	0.03%	0.02%	0.02%	0.01%	0.16%	0.01%	**0.28%**
	1	2.34%	1.05%	0.58%	0.67%	0.90%	0.32%	0.03%	0.23%	1.07%	0.01%	**7.20%**
	2	0.06%	0.69%	0.36%	0.34%	0.50%	0.12%	0.01%	0.01%	0.03%	0.00%	**2.11%**
	3	0.02%	0.13%	1.04%	0.52%	0.81%	0.23%	0.01%	0.01%	0.03%	0.00%	**2.79%**
	4	0.03%	0.07%	1.99%	3.02%	3.18%	1.08%	0.04%	0.03%	0.08%	0.00%	**9.51%**
	5	0.04%	0.09%	0.67%	8.49%	37.64%	15.64%	0.18%	0.16%	0.41%	0.01%	**63.32%**
	6	0.00%	0.00%	0.03%	0.14%	1.18%	5.88%	0.03%	0.03%	0.05%	0.00%	**7.34%**
	7	0.00%	0.00%	0.02%	0.10%	0.70%	6.46%	0.07%	0.03%	0.07%	0.00%	**7.45%**
	Total	**2.51%**	**2.04%**	**4.69%**	**13.28%**	**44.93%**	**29.74%**	**0.38%**	**0.51%**	**1.89%**	**0.03%**	**100.00%**

Source: Dobson DaVanzo analysis of UDSMR FY2017 cases.

This redistribution of scores has several consequences for scoring. For example, as Exhibit 14b (derived from 14a) shows, a 13.2-percentage point difference exists between the average percentage of maximal function in eating when looking at the FIM® item versus the SPAD eating item, causing the same cases to appear measurably more functionally independent in the SPAD item.

Exhibit 14b: Percentage of Maximum Functional Score of Eating Measures (Average Functional Score divided by most Independent Scale Item)

FIM®	SPAD
67.3%	80.5%

Source: Dobson DaVanzo analysis of UDSMR FY2017 cases.

Altered Incentives for Providers and Coding Clinicians

The proposed system changes the average reimbursement and expected average length of stay for many cases. If implemented, the new case-mix groupings would reshape resource usage as these CMG characteristics effectively set the case payment amount budget. Thus, the same patient may be associated with a higher or lower level of resource use and a longer or shorter expected length of stay under the new system, as Exhibit 15 shows.

Exhibit 15: Average Change in CMG Average Length of Stay for Stroke Cases between the Current and Proposed CMGs

FY2017 Current Stroke CMGs	FY2020 Proposed Stroke CMGs									
	101	102	103	104	105	106	5001	5103	5104	Grand Average
101	1.72	3.70	6.59	10.00						2.14
102	(0.00)	2.00	4.98	8.86	10.00	14.00				1.33
103	(1.97)	0.03	3.03	7.00		11.50				(0.46)
104	(1.98)	0.03	3.00	7.02	8.50	11.33				0.73
105	(3.95)	(1.97)	1.00	5.00	6.20	9.67				(0.10)
106	(5.00)	(3.00)	(0.03)	3.95	5.33	8.63				0.06
107	(7.00)	(4.99)	(2.01)	2.00	3.39	6.62				(0.41)
108	(10.50)	(8.38)	(5.38)	(1.39)	0.01					(1.02)
109	(8.69)	(6.57)	(3.56)	0.45		5.06				(0.00)
110	(14.00)	(12.19)	(9.20)	(5.26)		(0.61)				(1.87)
5001							11.38			11.38
5103								11.72		11.72
5104									0.42	0.42
Grand Average	(0.89)	(0.82)	(0.86)	(0.90)	0.12	(0.03)	11.38	11.72	0.42	(0.41)

Source: Dobson DaVanzo analysis of UDSMR FY2017 cases.

This could be problematic if the new scale is in some ways inaccurate and/or misleading. That is, if the new clinical groupings are not appropriate, this may introduce instability and perverse incentives in the system – as measurement improves, composite functional scores may decrease leading to potential system underpayments. As described above, new clinical groups represent a different view of patient functional status that is somewhat incongruent with the current assessment methods. As proposed CMGs consist of different groupings of beneficiaries compared to the current CMGs, case budgets and expected lengths of stay will change. This may alter how resources are apportioned – the course of care – across patients.

Data Availability and Transparency of Analytic Methods Behind CMS' Impact Analysis

The technical report lacked key details about analytic approach, clinical advice, and rationale for broad changes to the clinical interpretation of the case mix system. These

details would have been reassuring that the process was conducted with the same rigor as the preliminary IRF PPS rate setting technical report or the highly detailed proposed update to the SNF PPS (which included a facility-specific impact file) and could ensure that relative weight and length of stay values were not drastically altered. Furthermore, we are not aware that CMS or RTI conducted a technical expert panel with IRF clinicians to gain input from the field on the potential clinical and operational impacts of these changes.

Though the technical report does include some descriptions of the process for revising CMG definitions, it excludes information about the CART algorithm, clinical input and results. In particular, there were numerous clinical decisions we have found here which were not addressed in the report, such as the removal of item weights, the addition of new items to the motor score, etc. In the initial IRF PPS technical report conducted by RAND that was published in 2002,[12] numerous combined assessment score models were attempted – each supported by published literature – and results including goodness of fit statistics were shared for each alternative. This due diligence may have been conducted by RTI, but the lack of prior rigorous evaluation of the SPAD items may have made this more challenging. In any case, the RTI report did not attempt to justify the change in scale or change in the number and types of assessment items selected beyond describing some as burdensome.

CONCLUSION

The FY2019 proposed system changes for FY2020 would fundamentally change the way IRF patients' functional abilities and impairments are determined. The change in assessment items requires updates to the CMGs which, now representing different groupings of the same cases, change. As patients are recategorized and incentives realigned to meet this categorization, payments will be redistributed across case types relative to current law. This may cause unpredictable changes in how patients are cared for in inpatient rehabilitation facilities as well as in other settings, such as home health and home with a family or friend caregiver.

Further, there are implementation challenges inherent in using the first operational period of data from a new clinical assessment tool to reset a payment system. The performance of the new SPAD items has not yet been widely evaluated for benchmarking and comparative purposes and seem to have numerous coding and clinical interpretation problems. If the proposed system changes are implemented without first ensuring the data assessment tool performs adequately, CMS may introduce unintended incentives and year-to-year payment

[12] Carter, Grace, Melinda Beeuwkes Buntin, Orla Hayden, Jennifer Kawata, Susan Paddock, Daniel Relles, Gregory Ridgeway, Mark Totten and Barbara Wynn. Analyses for the Initial Implementation of the Inpatient Rehabilitation Facility Prospective Payment System. RAND Health prepared for the Centers for Medicare and Medicaid Services. 2002. https://www.rand.org/content/dam/rand/pubs/monograph_reports/MR1500/MR1500.pdf.

system instability as providers seek to understand the new case mix system. Ultimately, it is the beneficiaries who will face the consequences of CMG, length of stay and payment confusion and instability as the proposal would alter both how patients are perceived and how they access care as providers seek to adapt to these changes.

There remain essential unanswered questions both clinically and operationally. Data and CMG construction considerations indicate this model may not be an appropriate step at this time. Finalizing the proposed change at this time may introduce payment instabilities or perverse incentives that could weaken various aspects of functional assessment measurement as well as create an unbalanced system that may require further updates.

Ultimately, we have found evidence suggesting that the proposed system is not ready for implementation as of FY2020. First, the data on which the proposed system is new (as of October 2016) and through FAH member studies has shown evidence of low inter-rater reliability. Second, budget neutrality would be compromised if the non-scaled item responses (now set to be the least independent value) were more frequently coded at a different value than the default value. Third, as clinicians better understand the differences between usual and worst performance, this will have an impact of how they assess patients, and thus their patients' eventual assessment score. Fourth, the impact of the exclusion of key elements of the current system (namely, cognitive items and motor score item weights) is yet to be determined. As the new system reflects the above considerations, the functional status score for any given individual is likely to change – the implications are that any patient will be given different treatment for their care and that this care will be paid for differently. If these elements are not addressed, unintended consequences, including potential harm to IRF patients, could result.