The Functional Assessment Specialists

Uniform Data System

for Medical Rehabilitation

Telephone
716.817.7800

Fax
716.568.0037

E-mail
info@udsmr.org

Web site
www.udsmr.org

Suite 300

270 Northpointe Parkway

Amherst, NY 14228

University at Buffalo
The State University of New York

June 17, 2019

Seema Verma
Administrator
Centers for Medicare and Medicaid Services
Department of Health and Human Services
Attention: CMS-1710-P
P.O. Box 8016
Baltimore, MD 21244-8016

Submitted via regulations.gov

Re: 42 CFR Part 412 (CMS-1710-P) Medicare Program: Inpatient Rehabilitation Facility Prospective Payment System for Federal Fiscal Year 2020 and Updates to the Inpatient Rehabilitation Facility Quality Reporting Program; Proposed Rule

Dear Administrator Verma,

On behalf of Uniform Data System for Medical Rehabilitation (UDSMR) and the more than nine hundred inpatient rehabilitation facilities we provide services to, we are pleased to present our comments on 42 CFR Part 412 Medicare Program: Inpatient Rehabilitation Facility Prospective Payment System for Federal Fiscal Year 2020 and Updates to the Inpatient Rehabilitation Facility Quality Reporting Program; Proposed Rule, which was published on April 24, 2019, in the *Federal Register*. With over thirty years of experience, UDSMR provides coding, clinical, and quality improvement services to IRFs and other postacute care (PAC) providers. UDSMR appreciates the opportunity to provide ongoing feedback to CMS and hopes to work with CMS toward solutions that meet the needs of IRF providers and patients.

Before proceeding with our comments, we present the following executive summary, which highlights our concerns and recommendations.

Executive Summary:

UDSMR appreciates CMS's consideration for the use of a weighted motor score in a new payment system. Although this was done as an attempt to produce something similar to the current payment system, CMS has not provided evidence that the calculated motor score is reliable and valid, nor has CMS provided access to a combined data set in order to allow IRFs to analyze and recommend alternatives. This forces IRFs into a compromised position with respect to payment for cases discharged on or after October 1, 2019. CMS is making IRFs choose between two poorly constructed payment models, both of which lack the evidence needed to instill confidence in the estimated effects or to help IRFs make an informed decision or recommend alternatives. UDSMR cannot support the implementation of the proposed payment system until such time as CMS provides a transparent process and evidence of a reliable and valid model.

UDSMR also appreciates CMS's desire to propose standardized patient assessment data elements (SPADEs) to satisfy the requirements of the IMPACT Act. Unfortunately, UDSMR cannot support these proposals either, as the SPADEs under consideration would

add a significant amount of administrative burden without any supporting evidence that SPADEs can be used to measure quality or differentiate case-mix for payment. CMS has not shared quantitative results from the National Beta Test prior to proposing these items for implementation. This limits IRFs' ability to determine whether the additional administrative burden will yield information useful for quality and/or payment purposes.

UDSMR supports regulatory relief efforts that would prevent Medicare contractors from inventing restrictive definitions or other criteria that go beyond Medicare law and regulations, ultimately denying access or payment to IRFs for patients who meet medical necessity standards and would benefit from the care provided by a rehabilitation physician. Erroneous denials are harmful to patients and to the inpatient rehabilitation infrastructure, and patients and providers should have confidence that any care provided consistent with the criteria for medical necessity under Medicare will be covered without argument or delay. We believe that these determinations are currently suited more for IRFs than for Medicare contractors, and we conditionally support the proposed amendments related to the definition of a rehabilitation physician. Our condition for support of these proposed amendments is for CMS to convene meetings or technical expert panels (TEPs) with industry stakeholders in order to consider providing more detailed guidance related to the "specialized training and experience in inpatient rehabilitation" necessary to be considered a rehabilitation physician.

In summary, UDSMR does not support CMS's proposals for payment and quality, which have failed to provide basic evidence to support that they will do what they are supposed to do and that they will produce reliable and valid results. We encourage CMS not to rush toward meeting self-imposed deadlines using poorly constructed solutions and to provide a more transparent process that allows industry stakeholders to work with CMS toward solutions that meet the needs of IRF providers and patients.

Concerns:

1. The proposed refinements to the case-mix classification and FY 2020 IRF PPS payment system

    a. lack sufficient research, testing, and analysis;

        i.   The proposed motor score has neither been tested for nor demonstrated reliability and validity.

        ii.   One item that was highly correlated with other items was removed, but other items that have greater correlation with one another were retained.

        iii. The calculation of weight-index values is based on highly questionable analytical practice.

        iv.   The use of the first two years of functional data collected alongside another similar but different set of functional data calls into question the integrity of the data being used for the payment model.

        v.   CMS still has not provided sufficient education, training materials, and supporting documentation about the functional items to support their use in developing a payment model.

    b. will artificially make IRF patients appear "less severe"; and

    c. will affect patient access to resources and an IRF level of care.

2. The collection of Transfer of Health Information data elements and the resulting quality measure values will not only fail to improve the quality of care provided, but also add unnecessary administrative burden to IRFs.

3. The proposed SPADEs do not provide sufficient evidence to justify the burden of collecting up to eighteen additional pages of information.

Recommendations:

1. UDSMR has the following recommendations with respect to the proposed IRF PPS payment model:

    a. Implement a one-year FY 2020 payment model based on an unweighted motor score that uses standardized patient assessment data elements that provide evidence to predict cost and that are not "highly correlated" with one another.

    b. During FY 2020:

        i. Provide a limited data set that matches patient-level IRF-PAI assessment data to claims/cost data in order to allow stakeholders to analyze and potentially model alternative recommendations.

        ii. Conduct monthly or quarterly stakeholder or technical expert panel (TEP) payment model meetings for the purposes of providing additional transparency and discussing and reviewing payment model–related analyses and information in order to prepare for a new proposed payment model for FY 2021.

        iii. Resolve issues related to the admission assessment guidelines by working with clinical industry experts to establish clear and concise examples and instructional materials that remove the need for "clinical judgement."

    c. Do not implement a payment system using weighted item values until two to three years of "clean" quality indicator data has been collected. In other words, no weighted motor score payment model should be adopted until two to three years of standardized patient assessment data elements have been collected without confusion from the collection of similar but slightly different items that measure the exact same construct.

    d. Provide evidence that any calculated value used for payment models can be proven to be reliable and valid. This should require that each of the following measures for reliability and validity are met:

        i. Test-retest reliability: This will address whether the calculated value (such as a motor score) is consistent across time. In order to assess this, CMS/RTI should examine the correlation between the motor score at two different time points (perhaps admission and discharge) to determine whether the measure value is consistent over time.

        ii. Internal consistency: This will address whether there is reliable consistency between responses to the items that make up a measure, such as a motor score. This can be accomplished by using a split-half correlation or providing a Cronbach's alpha value.

        iii. Construct validity: This will address whether the measure (such as a motor score) is capable of measuring what it claims to measure. CMS should provide evidence that the resulting measure is highly correlated with an existing measure of its intended construct.

        iv. Predictive validity: This will address whether the measure (such as a motor score) is capable on its own of predicting outcomes or other values of importance—in this case, providing evidence that the measure can predict length of stay or cost or other outcomes.

2. CMS should proceed with the proposal to amend §412.622 and resolve any instances where this issue has resulted in a denial of an IRF claim. CMS should also convene meetings or technical expert panels (TEPs) with industry stakeholders to consider providing more detailed guidance related to the "specialized training and experience in inpatient rehabilitation" necessary to be considered a rehabilitation physician.

3. CMS should delay implementation of the Transfer of Health Information measures until sufficient evidence is provided that these measures

   a. produce differentiation in performance between IRFs,

   b. reliably and validly capture that this information has been transferred,

   c. are predictive of future benefit, and

   d. account for the full amount of time needed to collect and transfer the information and to enter values on the IRF-PAI.

4. CMS should proceed with the proposal to remove baseline nursing facility discharges from the discharge-to-community measure and update historical quality measure information to reflect this change.

5. CMS should not finalize the implementation of SPADEs until the following recommendations are met:

   a. CMS must release additional quantitative information from the National Beta Test that provides evidence that the proposed SPADEs measure what they are intended to measure and can be used for quality and/or payment.

   b. CMS should evaluate whether SPADEs are already captured by other means, such as IRF-PAI assessment data or billing/claims data. Language in the IMPACT Act suggests that duplicative data elements are to be removed, and the proposed SPADEs may be unnecessary if they are duplicative.

   c. CMS should reduce data collection on elements that will not differ between admission and discharge.

The remainder of this letter addresses our concerns and recommendations in detail.

1. **The proposed refinements to the case-mix classification and FY 2020 IRF PPS payment system**

   a. **lack sufficient research, testing, and analysis;**

      i. **The proposed motor score has neither been tested for nor demonstrated reliability and validity.**

      ii. **One item that was highly correlated with other items was removed, but other items that have greater correlation with one another were retained.**

      iii. **The calculation of weight-index values is based on highly questionable analytical practice.**

      iv. **The use of the first two years of functional data collected alongside another similar but different set of functional data calls into question the integrity of the data being used for the payment model.**

      v. **CMS still has not provided sufficient education, training materials, and supporting documentation about the functional items to support their use in developing a payment model.**

   b. **will artificially make IRF patients appear "less severe"; and**

   c. **will affect patient access to resources and an IRF level of care.**

   As stated in section III of CMS-1710-P.

Before addressing each of the concerns noted above, as previously stated in this comment letter, UDSᴍʀ strongly recommends the following with respect to the proposed IRF PPS payment model:

1. Implement a one-year FY 2020 payment model based on an unweighted motor score that uses standardized patient assessment data elements that provide evidence to predict cost and that are not "highly correlated" with one another.

2. During FY 2020:

   a. Provide a limited data set that matches patient-level IRF-PAI assessment data to claims/cost data in order to allow stakeholders to analyze and potentially model alternative recommendations.

   b. Conduct monthly or quarterly stakeholder or technical expert panel (TEP) payment model meetings for the purposes of providing additional transparency and discussing and reviewing payment model–related analyses and information in order to prepare for a new proposed payment model for FY 2021.

   c. Resolve issues related to the admission assessment guidelines by working with clinical industry experts to establish clear and concise examples and instructional materials that remove the need for "clinical judgement."

3. Do not implement a payment system using weighted item values until two to three years of "clean" quality indicator data has been collected. In other words, no weighted motor score payment model should be adopted until two to three years of standardized patient assessment data elements have been collected without confusion from the collection of similar but slightly different items that measure the exact same construct.

4. Provide evidence that any calculated value used for payment models can be proven to be reliable and valid. This should require that each of the following measures for reliability and validity are met:

   a. Test-retest reliability: This will address whether the calculated value (such as a motor score) is consistent across time. In order to assess this, CMS/RTI should examine the correlation between the motor score at two different time points (perhaps admission and discharge) to determine whether the measure value is consistent over time.

   b. Internal consistency: This will address whether there is reliable consistency between responses to the items that make up a measure, such as a motor score. This can be accomplished by using a split-half correlation or providing a Cronbach's alpha value.

   c. Construct validity: This will address whether the measure (such as a motor score) is capable of measuring what it claims to measure. CMS should provide evidence that the resulting measure is highly correlated with an existing measure of its intended construct.

   d. Predictive validity: This will address whether the measure (such as a motor score) is capable on its own of predicting outcomes or other values of importance—in this case, providing evidence that the measure can predict length of stay or cost or other outcomes.

1a. The proposed refinements to the case-mix classification and FY 2020 IRF PPS payment system lack sufficient research, testing, and analysis.

The supporting research technical document from RTI related to the research, testing, and analysis of the proposed alternative payment model does not offer sufficient evidence that the selected items, the weighted motor score, and the resulting model are suitable replacements for the existing items and payment system. Although certain analyses were performed to determine the overall financial impact, the document fails to indicate whether each of the items chosen for the motor score is needed or can predict costs on its own. The analyses also do not indicate whether the resulting motor score measures the severity of patients as consistently as the weighted motor score index does. Without transparency and additional information about how the model was created and why certain selections were made, providers cannot evaluate whether the proposed payment model will provide the resources necessary to care for their patients.

   i. **The proposed motor score has neither been tested for nor demonstrated reliability and validity.**

   Although the individual items chosen for motor score were tested in the PAC PRD for reliability and validity, the resulting motor score has neither been tested for nor demonstrated its reliability and validity. CMS and its contractor, RTI International, have failed to provide the necessary information indicating that the motor score is capable of measuring what it is supposed to measure or is predictive on its own of cost or length of stay. Analyses of both the unweighted and weighted motor scores has shown little to no correlation with the weighted motor score currently in use and produces patient-severity levels that differ significantly from information that has been proven to be reliable and valid for over twenty years.

   The scatterplots below display the FY 2020 proposed weighted motor score and current or existing weighted motor score values for Medicare cases discharged in FY 2017 and

FY 2018, and they allow us to examine the amount of variability between the two scores and to identify any significant differences in the interpretation of patient severity.

FY 2017 Medicare Discharges:
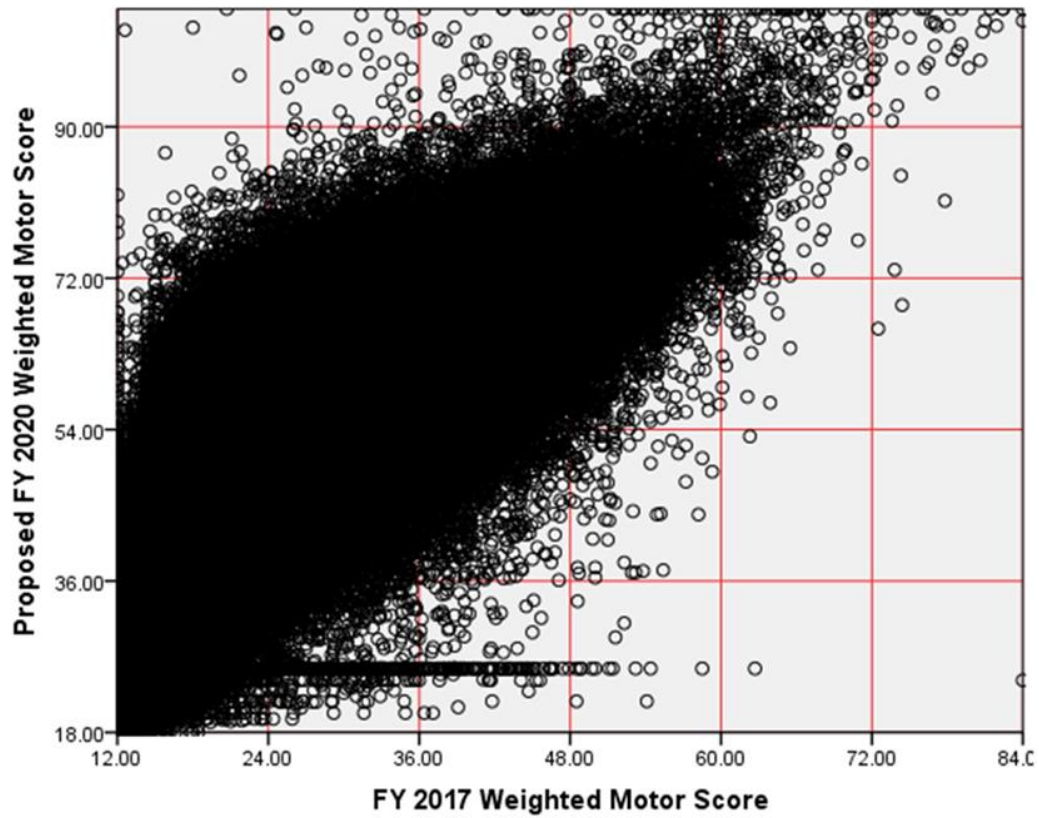
FY 2018 Medicare Discharges:
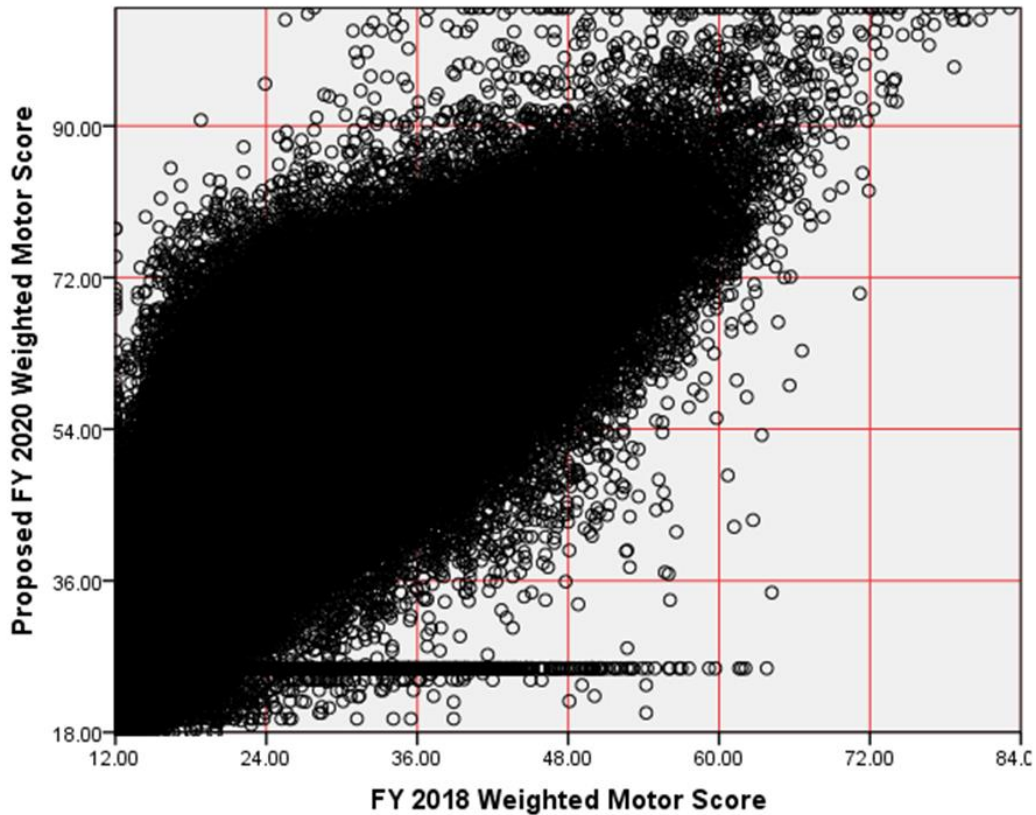
As you can see, there is a significant amount of variability in the weight motor score values. If CMS was attempting to develop a payment model that was similar to the existing payment model, one would expect that the proposed weighted motor score would be highly correlated with the existing reliable and valid weighted motor score. The Pearson's correlation values are .791 and .792 for FY 2017 and FY 2018, respectively. Typically, a correlation value of +.80 or greater is generally considered good. Additionally, analyzing whether the proposed FY 2020 weighted motor score is predictive of the existing reliable and valid weighted motor score, linear regression results indicate an adjusted R^2 value of .626. This suggests that the proposed FY 2020 weighted motor score accounts for less than 63% of the variance in the existing weighted motor score. If a model could explain 100% of the variance, the proposed values would always equal the current values. So while these statistics may be close, they do not support that the proposed measure can measure patient severity for payment in a reliable and valid manner.

Additional testing must be conducted in order to make sure that any proposed motor score is proven a reliable and valid measure for use in defining a payment model.

ii. **The removal of one item that was highly correlated with other items but continued use of other items that have greater correlation with one another is highly questionable.**

CMS/RTI removed the "Roll left and right" item from the motor score due to its being highly correlated with other items, but additional "highly correlated" items remained in the motor score. Examining the UDSMR® IRF database for Medicare cases discharged between FY 2017 and FY 2018 suggests that "Roll left and right" was not even the item with the highest inter-item correlation value. The following table displays the five items with the highest inter-item correlation values:

Item 1	Item 2	Inter-item correlation
GG0170B, Sit to lying	GG0170C, Lying to sitting on side of bed	0.900
GG0170D, Sit to stand	GG0170E, Chair/bed-to-chair transfer	0.834
GG0130G, Dressing lower body	GG0130G, Putting on/taking off footwear	0.759
GG0170A, Roll left and right	GG0170B, Sit to lying	0.755
GG0170A, Roll left and right	GG0170C, Lying to sitting on side of bed	0.753

This table shows that the highly correlated items produced essentially the same result or a similar one. Typically, the inclusion of highly correlated items in a calculated value would bias the result and bring into question the reliability and validity of the resulting model.

Was a threshold used to determine that the "Roll left and right" item should have been the only item removed? Was the requirement that higher correlation values were present on more than one item set? CMS/RTI did not provide sufficient information in the technical report to support this decision.

CMS/RTI should remove one of the highly correlated items, as they did with "Roll left and right," and should perform additional testing on the remaining items in order to make sure that any proposed motor score is proven a reliable and valid measure for use in defining a payment model.

iii. **The calculation of weight-index values is based on highly questionable analytical practice.**

The use of an average value from a highly correlated pair of items to create an equally distributed weight-index value across a pair of items is very questionable analytical practice. This practice also would bias the results of the weight values for other items by reducing the number of candidate items and values considered. Technically, removal of one of the highly correlated items or consideration for both items individually should have been used to create item-level weight indices.

It is also questionable analytical practice to use only the value from the "Walk 10 feet" item to produce weight values evenly distributed across this item and two other items ("Walk 50 feet with two turns" and "Walk 150 feet"). Each of these items may have different values at admission and, as such, should be considered individually for defining weight indices and creating a weighted motor score. Again, this not only negatively

affects the weighting of the three items, but also biases the results of the weight values for other items by reducing the number of candidate items and values considered.

CMS/RTI should perform additional analyses using well-established analytical processes to make sure that any proposed weight index values and resulting motor score are proven a reliable and valid measure for use in defining a payment model.

iv. **The use of the first two years of functional data collected alongside another similar but different set of functional data calls into question the integrity of the data being used for the payment model.**

The creation of the proposed weighted motor score and resulting CMGs was based on limited analyses that utilized the first two years of functional quality indicator data collected alongside another similar but different set of functional data. The integrity of the underlying data should be brought into question because this data was collected for quality purposes but the payment system was based on a different set of functional data. The duplicative nature of the functional data with different guidelines and scales has caused confusion among providers, and the resulting values may not properly represent patient severity.

We recommend that CMS not implement a payment system using weighted item values until two to three years of "clean" quality indicator data have been collected. In other words, no weighted motor score payment model should be adopted until two to three years of standardized patient assessment data elements have been collected without confusion from the collection of similar but slightly different items that measure the exact same construct.

v. **CMS still has not provided sufficient education, training materials, and supporting documentation about the functional items to support their use in developing a payment model.**

From the start of data collection on these quality indicator items, CMS has provided limited training opportunities and has provided multiple different responses or clarifications to industry questions. Additionally, CMS has published multiple different documents on varying websites attempting to address issues with the initial collection of this data. This has led to varying interpretations of CMS's guidelines and has resulted in data collection that is anything but "standardized."

CMS has stated that the admission codes recorded on the IRF-PAI should reflect a patient's "baseline functional abilities that occur soon after the patient's admission" and "prior to benefitting from treatment interventions." The training manual gives conflicting direction when it states, "If a patient performs the activity more than once during the assessment period, coding in section GG should be based on the patient's 'usual performance.'"

When asked how to rate specific GG items, CMS has issued multiple clarifications that refer to the use of "clinical judgment." But since such judgment involves a clinician's personal opinion, the results of assessments based on clinical judgment will often be inconsistent and unreliable from one clinician to another.

UDSMR recommends that CMS resolve issues related to the admission assessment guidelines by working with clinical industry experts to establish clear and concise

examples and instructional materials that remove the need for "clinical judgement." By establishing clear and concise guidelines, the industry can move forward with collecting and submitting truly standardized data that can be used to produce reliable and valid results.

1b. The proposed refinements to the case-mix classification and FY 2020 IRF PPS payment system will artificially make IRF patients appear "less severe."

Using the UDSᴍʀ® IRF database for Medicare Fee-for-Service cases discharged between October 2016 and September 2018 (containing over 80% of Medicare Fee-for-Service cases utilizing in the RTI International analysis), we replicated the proposed weighted motor score used for the proposed CMGs for FY 2020. We then compared these values to the existing weighted motor score index used for the current CMGs. As shown below, for FY 2017 and FY 2018 cases, we found that, on average, the proposed motor score for the FY 2020 CMGs indicated that patients were functioning at a higher percentage of total possible function than the current weighted motor score index did.

	Maximum Value	Mean (Average)	% of Max. Value
Proposed weighted motor score for FY 2020 CMGs	104	52.8	50.8%
Current weighted motor index for CMGs	84	28.1	33.5%

As the table shows, the proposed weighted motor score for FY 2020 CMGs suggests that the average IRF patient is functioning at 50% of maximum function at admission; by contrast, the current weighted motor index suggests that the average IRF patient is functioning at 33%. Such a difference causes us to question whether the proposed payment model will accurately reflect the resource needs of what appears to be a more severe population.

We believe that these differences are caused by two factors:

- **Assessment methodology:** The FIM® items utilize the lowest value over the three-day admission assessment period. This methodology is designed to capture the burden of care so that IRFs can provide the appropriate amount of resources to meet the patient's needs. The quality indicator items in section GG also have a three-day admission assessment period, but *The IRF-PAI Training Manual*'s instructions for choosing a value indicate that the assessment should occur prior to the start of therapy services, that it should be based on the patient's usual status, and that the patient's best and worst performances should not be recorded. Further instructions for assessment also state that the "assessment should occur prior to the patient benefitting from treatment interventions to capture the patient's true admission baseline status." Implementation of the section GG items created a great deal of confusion, as providers have not consistently adopted the assessment methodology noted. If therapy starts on the first day of the stay, providers may only assess a patient once prior to the start of therapy, and that day may be the worst or best performance over the three-day admission assessment period. In this scenario, providers are confused about whether to use the assessment prior to therapy or the usual performance result. In addition, the value used by providers who document only one assessment of these items may not be consistent with the value used by those who document multiple assessments. Either way, this difference in the assessment

methodology appears to be causing patients to appear less severe. We ask that CMS and RTI clarify the assessment methodology further and identify whether this methodology captures the burden of care for the purposes of payment.

- **Scale/code differences:** Differences in the item scales/codes can make patients appear less severe. The table below provides details about the scale/code options.

Score	FIM® Level	SPADE Section GG Levels
1	Total assistance (subject performs 25% or less independently or requires assistance from two or more helpers)	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*
0	*Activity does not occur (admission only)*	—

In terms of making patients less severe, we are most concerned about the differences in level 1, which is the most severe functional level. For the FIM® instrument, a patient who performs 25% or less of an activity or requires assistance from two or more helpers is assigned a level 1 value. For the section GG items, level 1 is assigned only when a patient does not perform any of the effort to complete the activity or when the patient requires assistance from two or more helpers to complete the activity. For a patient who performs 1% to 24% of an activity, the section GG items require that the patient be coded level 2, along with patients who perform 25% to 49% of an activity. So if patients attempt and complete only a small portion of an activity yet still require a significant amount of assistance, they are now moved from a value of 1 to a value of 2. In terms of the percentage of maximal function value, this change takes these patients from 14.3% of maximal function value to 33.3%. From a resource perspective, these patients will still require a great deal of time and resources, but the scale options are producing a result that may make these patients appear less severe. We ask that CMS and RTI evaluate whether their scale items appropriately capture the burden of care for these patients and whether the scale values represent the actual amount of time and resources needed for these patients.

For bladder management and bowel management, the proposed replacement items are measuring a different construct using a different scale on a different admission assessment period. The FIM® ratings for Bladder Management and Bowel Management are each the result of two function modifiers that assess both the frequency of accidents over the last seven days (the three-day admission assessment period and the four days prior) and the level of assistance provided during the three-day admission assessment period. In contrast, the SPADE bladder and bowel continence items only assess the presence of incontinent episodes over the three-day admission assessment period. Because the SPADE items do not capture the amount of assistance provided to the patient and do not capture incontinent episodes prior to admission, patient severity appears significantly different for these items.

Based upon data in the UDS_MR® IRF database for Medicare Fee-for-Service cases discharged between October 2016 and September 2018, significantly more patients are coded as "always continent" on the SPADE items, which is recoded to the highest functional level for the purposes of the proposed payment model. Additionally, significantly fewer patients are coded as "always incontinent," which is recoded to the lowest functional level for the purposes of the proposed payment model. In comparison to the FIM® items, it appears that patient severity has completely changed because of the change in items and what they measure. This is causing patients to appear less severe from a payment model standpoint.

We also note that alternative bladder and bowel items were tested as part of the national beta test for additional SPADE items previously mentioned in this comment letter. We question the applicability of the existing bladder and bowel items for payment purposes when alternative items have been tested and potentially considered for implementation. We ask that CMS evaluate which bladder and bowel items are most appropriate and which best capture patient severity and burden of care for the purposes of payment.

Finally, in regard to the proposed payment model making patients appear less severe, we would like to point out issues with the chosen items that differ from the current FIM® items or have no comparable FIM® items. These items typically involve "smaller" activities or tasks, and we believe that these items were added as SPADEs at the request of LTCHs and SNFs in order to allow these settings to demonstrate functional change for more severely impaired patients. As previously stated, CMS and RTI have not provided additional information related to an individual item's ability to predict cost on its own, and, as such, some of the choices they have made for inclusion in the motor score significantly affect the level of severity of IRF patients.

- CMS is proposing to replace the existing FIM® item 39L, Locomotion: Walk, Wheelchair, which is based on the patient's ability to walk or wheel 150 feet, with three walking items, each of which assesses a different distance and functional ability. Although item GG0170K, Walk 150 feet, has the most similar definition and produces a very similar average functional ability for IRF patients, CMS and RTI also included item GG0170I, Walk 10 feet, and item GG0170J, Walk 50 feet with two turns. As shown previously, these two items have a higher average value and represent a higher functional level for IRF patients. The inclusion of these two additional walking items in the payment model motor score makes the patient look less severe because the values of these two items outweigh the value measured by item GG0170K, Walk 150 feet. We ask CMS and RTI to provide evidence related to the inclusion of these two additional walking items

and to determine whether they capture an accurate burden of care and patient severity for the purposes of predicting cost and resulting payments.

- CMS is proposing to replace the existing FIM® item 39M, Locomotion: Stairs, which is based on the patient's ability to go up and down twelve to fourteen stairs, with item GG0170M, 1 step (curb), which is based on the patient's ability to step over a curb or up and down one step. For quality purposes, IRFs are required to collect and submit data on a SPADE item that is equivalent to the existing FIM® item and is based on twelve steps, but CMS and RTI did not choose that item for the purposes of the payment model. When evaluating the twelve-step item, we have noted that average patient performance is more closely aligned to the existing FIM® item. By choosing a one-step item to replace a twelve-step item, as shown previously, average performance on the one-step SPADE item is higher than performance on the twelve-step FIM® item and represents a higher functional ability for IRF patients. CMS and RTI offered little information as to why the twelve-step item was not used in the proposed payment model motor score, suggesting that "excluded elements are typically too challenging for patients on admission to IRFs and therefore are less likely to be assessed on admission. The frequency with which these items are not assessed on admission in the IRF decreases their relevance for predicting costs." We ask that CMS and RTI reevaluate their choice of a replacement for FIM® item 39M, Locomotion: Stairs, and provide evidence that any chosen items capture the true burden of care and patient severity for the purposes of predicting cost and resulting payments.

- CMS is proposing the inclusion of three items in the weighted motor score that have no comparable FIM® instrument items:

  - GG0130B, Sit to lying

  - GG0130C, Lying to sitting on side of bed

  - GG0130D, Sit to stand

  Based upon data in the UDSMR® IRF database for Medicare Fee-for-Service cases discharged between October 2016 and September 2018, these items average at or around level 3 on the SPADE section GG scale, which represents "Partial/moderate assistance— Helper does LESS THAN HALF the effort." In addition, a significant percentage of IRF patients appear to be coded at level 3 or 4. We believe that these items are duplicative in nature, as these smaller activities/tasks are part of the assessment of other SPADE items, such as item GG0170E, Chair/bed-to-chair transfer. The inclusion of these items in the motor score will outweigh other item values, which produce a much lower average value and represent the level of severity of an IRF patient. We ask that CMS and RTI provide evidence related to the inclusion of these items and whether they capture the true burden of care and patient severity for the purposes of predicting cost and resulting payments.

For all the reasons we identify above, we believe that the payment model and weighted motor score value that CMS and RTI are proposing are making IRF patients appear less severe. We want to work with CMS and RTI to resolve these issues and produce a payment model and patient classification system that not only accurately reflect the severity of IRF patients, but also provide the appropriate amount of resources to meet the burden of care.

1c. The proposed refinements to the case-mix classification and FY 2020 IRF PPS payment system will affect patient access to resources and an IRF level of care.

After replicating the proposed motor score using the UDSᴍʀ® IRF database for Medicare Fee-for-Service cases discharged between October 2016 and September 2018, we proceeded to map these patients into the proposed CMGs for FY 2020 in order to evaluate potential changes in payment and other factors influencing care. We found significant differences in the CMG groupings that may affect patient access to resources and an IRF level of care.

Below, we display the distribution of tier A stroke cases for FY 2018.

			FY 2020 CMG	A0101	A0102	A0103	A0104	A0105	A0106	A0107
			FY 2020 Rel. Wgt.	0.8152	1.0253	1.2175	1.4289	1.7146	1.8676	2.1805
			FY 2020 CMS Avg. LOS	10	12	14	16	19	21	24
FY 2018 CMG	FY 2018 Rel. Wgt.	FY 2018 CMS Avg. LOS	FY 2020 FPP	$13,510.31	$16,992.30	$20,177.63	$23,681.16	$28,416.07	$30,951.73	$36,137.43
A0101	0.6435	8	$10,664.73	989	173	20	5	2	1	2
A0102	0.8080	10	$13,390.98	1,164	914	209	20	7	0	3
A0103	0.9136	11	$15,141.09	249	272	103	8	4	0	2
A0104	0.9800	12	$16,241.54	809	1,791	1,078	185	38	3	15
A0105	1.1404	13	$18,899.85	256	1,001	1,365	433	141	3	21
A0106	1.2630	15	$20,931.70	107	518	1,259	776	414	9	32
A0107	1.4103	16	$23,372.90	43	228	814	833	869	26	86
A0108	1.7922	20	$29,702.13	5	65	193	352	1,094	1,973	0
A0109	1.6170	19	$26,798.54	24	87	309	518	1,027	0	340
A0110	2.1083	24	$34,940.86	12	51	190	472	2,322	0	6,364

The cases in the red-shaded cells indicate patients who will see a decrease in both payment and CMS average LOS. It appears that over three thousand patients who are currently assigned to the most severe and highest-paying stroke CMG (A0110) would be projected to have their payment and LOS values significantly reduced. These changes are based solely on the differences in the functional assessment and the use of SPADE items, which, as we noted previously, make IRF patients appear less severe and have not been shown to be reliable and valid. How will providers handle stroke patients who are clearly severe in today's payment model but who will appear less severe and be paid less in the proposed model? Will these patients be admitted to IRFs for care if CMS does not make adequate resources available? Will these patients receive a full course of IRF-level care, or will they instead be discharged to a lower level of care for continued services? And will patients in CMG 0101 be targeted for payment denials, as they have been in the past? We are very concerned that this proposed payment model is based on information that makes patients appear less severe and places them into payment categories that may not provide an adequate amount of services or may prohibit them from receiving IRF care. We ask that CMS and RTI analyze these differences and projected effects to produce a payment model that does not affect patient access to resources and care.

To summarize these comments, UDSᴍʀ does not support the proposed refinements to the case-mix classification system. Although we believe all of the previously noted issues must be addressed prior to implementing a new payment system, we recognize that CMS plans to remove

the FIM® instrument from the IRF-PAI and benefit from the removal of administrative burden. This forces IRFs to choose between two flawed payment models and determine which one poses the lesser risk of limiting the resources available for patient care.

UDSₘᵣ and our subscribers would appreciate the opportunity to work with CMS to identify a way of ensuring that a new payment system accurately accounts for patient severity and provides the necessary resources for patient care.

2. **The collection of Transfer of Health Information data elements and the resulting quality measure values will not only fail to improve the quality of care provided, but also add unnecessary administrative burden to IRFs.**

   As stated in section VIII.D.1-2 of CMS-1710-P.

Although the Transfer of Health Information proposed measures address one of the domains of the IMPACT Act, UDSₘᵣ believes that the collection of this information and the resulting quality measure values will not only fail to improve the quality of care provided, but also add unnecessary administrative burden to IRFs.

These process measures collect information only to determine whether an IRF did what it was supposed to do; they do not suggest that the process was beneficial to the patient, the patient's family, a caregiver, or a subsequent provider. Solely transferring the information does not incorporate understanding of the information, nor does it evaluate the ability of the patient, the patient's family, a caregiver, or a subsequent provider to use the transferred information.

Although CMS has provided guidelines regarding what should be included in the transfer of medication information, data collection on this measure does not require that these guidelines be met. Does CMS intend to audit IRFs in order to ensure that the resulting measure values are consistent with the information being shared?

Additionally, although only a few items are being added to the IRF-PAI, a significant amount of time and resources will be necessary to compile the suggested information for the transfer. This additional burden has not been captured by CMS's estimates and will therefore result in a much larger administrative burden than that currently identified by CMS.

UDSₘᵣ recommends that CMS delay implementation of these measures until sufficient evidence is provided that these measures:

- produce differentiation in performance between IRFs,

- reliably and validly capture that this information has been transferred,

- are predictive of future benefit, and

- account for the full amount of time needed to collect and transfer the information and to enter values on the IRF-PAI.

3. The proposed standardized patient assessment data elements (SPADEs) do not provide sufficient evidence to justify the burden of collecting up to eighteen additional pages of information.

As stated in section VIII.F of CMS-1710-P.

According to materials shared by CMS during Special Open Door Forums and National Beta Test results meetings, the evaluation of SPADEs for implementation should be based on four main characteristics and their underlying criteria:

1. Potential for improving quality

    a. Improve care transitions, person-centered care, and care planning

    b. Improve care practices and patient safety

    c. Use for quality comparisons, including value-based payment models

    d. Supports clinical decision making and care coordination

2. Validity and reliability

    a. Inter-rater reliability (consensus in ratings by two or more assessors)

    b. Validity (captures the construct being assessed)

3. Feasibility for use in PAC

    a. Potential to be standardized and made interoperable across settings

    b. Clinically appropriate

    c. Relevance to work flow

4. Utility for describing case mix

    a. Potential use for payment models

    b. Measures differences in severity levels related to resource needs

CMS has provided some information suggesting proposed SPADEs meet the "validity and reliability" and "feasibility for use in PAC" characteristics noted above. National Beta Test documents have provided inter-rater reliability testing results indicating that the information is reliable and valid. CMS has also cited current use in other PAC settings, as well as the qualitative results of National Beta Test surveys, as demonstrating that the proposed SPADEs are "clinically useful" or may be feasible for data collection; however, CMS has not provided any evidence that the proposed SPADEs have the "potential for improving quality" or "utility for describing case mix," which are two key characteristics for evaluating SPADEs for implementation.

With respect to the "potential for improving quality," CMS has failed to supply quantitative evidence that the results of the National Beta Test offer the ability to meet any of the criteria noted above. Even simple frequencies of responses would allow IRFs and other PAC providers to determine whether these SPADEs be collected for quality comparisons or to support clinical decision making. The failure to supply even this basic level of information suggests the possibility of underlying issues preventing this data from meeting this consideration.

In terms of the "utility for describing case mix," CMS has again failed to provide any quantitative information suggesting that the proposed SPADEs are capable of meeting these criteria for consideration. As we evaluate a proposed payment model in this proposed rule, we note that CMS and RTI International have provided frequencies of responses at a minimum, but even this level of detail has not been provided for the National Beta Test and the items being proposed for implementation. How can IRFs evaluate whether the proposed SPADEs have the potential for use in payment models or have the ability to measure differences in patient severity without seeing how these SPADEs tested on IRF patients?

CMS must release additional quantitative information from the National Beta Test that provides evidence that the proposed SPADEs measure what they are intended to measure and can be used for quality and/or payment.

Below are additional concerns related to the proposed implementation of SPADEs:

- Brief Interview for Mental Status (BIMS) data elements are currently being collected by IRFs at admission and have not shown evidence that they can predict costs or differentiate case-mix. CMS has not provided any additional evidence to suggest that the BIMS is capable of being utilized for quality purposes to support the collection of these data elements at discharge.

- Although IRFs are seeking additional information related to cognitive status, it is unclear as to whether the addition of the CAM and PHQ 2 to 9 data elements will identify differences in cognitive status or measure changes during the stay resulting from therapeutic interventions. Additionally, there are concerns about the reliability and validity of the PHQ 2 to 9 data elements and their patient-reported outcomes for patients with severe cognitive deficits, prior mental health issues, or noncommunicative conditions.

- The proposed collection of certain elements at both admission and discharge adds unnecessary administrative burden for patient characteristics that are not anticipated to change from admission to discharge. For example, the preferred language, need for an interpreter, and lack of transportation data elements should be assessed only once during a patient's stay.

- Although potentially useful in determining resource needs, the collection of the Special Treatments, Procedures, and Programs data elements at both admission and discharge creates additional administrative burden when these circumstances may already be present on claims data or are capable of being determined through data elements already captured on the IRF-PAI (e.g., comorbidity ICD-10 codes).

- Timing estimates are significantly understated. IRFs that participated in the National Beta Test indicated that time studies were based only on the time needed to enter a value on a tablet and did not include the time to evaluate the patient on each item.

UDSᴍʀ has the following recommendations with respect to the implementation of the proposed SPADEs:

1. CMS should evaluate whether SPADEs are already captured by other means, such as IRF-PAI assessment data or billing/claims data. Language in the IMPACT Act suggests that duplicative data elements are to be removed, and the proposed SPADEs may be unnecessary if they are duplicative.

2. CMS should reduce data collection on elements that will not differ between admission and discharge.

3. Until the prior recommendations are met, CMS should not finalize the implementation of SPADEs. This decision will help IRFs avoid unnecessary administrative burden and will adhere to principles of the Meaningful Measures and Patients Over Paperwork initiatives.

Additional comments/recommendations on proposals not already addressed:

4. Proposal to amend §412.622 to state that an IRF determines whether a physician qualifies as a rehabilitation physician

As stated in section VII of CMS-1710-P.

UDSᴍʀ generally supports regulatory relief efforts that would prevent Medicare contractors from inventing restrictive definitions or other criteria that go beyond Medicare law and regulations, ultimately denying access or payment to IRFs for patients who meet medical necessity standards and would benefit from the care provided by a rehabilitation physician. Erroneous denials are harmful to patients and the inpatient rehabilitation infrastructure alike, and patients and providers should have confidence that any care provided consistent with the criteria for medical necessity under Medicare will be covered without argument or delay.

We also understand there are concerns related to the definition of a rehabilitation physician and the lack of understanding as to what constitutes "specialized training and experience in inpatient rehabilitation." Although we support efforts to further clarify this definition, we caution that doing so may continue to affect access to IRF care. We also would like to ensure that clarifications will meet the geographical and specialty training variations present in our industry while ensuring that the quality of care is maintained. With considerations to offset the erroneous denials of access and payment that continue to occur, we support the notion that CMS convene meetings or technical expert panels (TEPs) with industry stakeholders in order to consider providing more detailed guidance related to the "specialized training and experience in inpatient rehabilitation" necessary to be considered a rehabilitation physician.

Currently, UDSᴍʀ has no evidence to support or believe that IRFs are acting irresponsibly in their current role of designating who qualifies as a "rehabilitation physician." UDSᴍʀ continuously monitors IRF outcomes, and historical trends suggest that the quality of care delivered by IRFs significantly exceeds that of SNFs, LTCHs, and HHAs—venues in which the role of a rehabilitation physician is significantly different or not required at all. By contrast, we have evidence suggesting that Medicare contractors are erroneously denying access or payment to IRFs for reasons that are inconsistent with Medicare law and regulations, specifically with respect to their determinations of what constitutes a rehabilitation physician.

Because of this, we believe that these determinations are currently better suited for IRFs than for Medicare contractors, and we conditionally support the proposed amendments. We also recommend that CMS convene meetings or technical expert panels (TEPs) with industry stakeholders in order to consider providing more detailed guidance related to the "specialized training and experience in inpatient rehabilitation" necessary to be considered a rehabilitation physician.

5. **Proposed Data Reporting on Patients for the IRF Quality Reporting Program Beginning with the FY 2022 IRF QRP**

   As stated in section VIII.H.5 of CMS-1710-P.

UDSᴍʀ is concerned with CMS's proposal to "expand the reporting of IRF–PAI data used for the IRF QRP to include data on all patients, regardless of their payer, beginning with patients discharged on or after October 1, 2020 for the FY 2022 IRF QRP and the IRF–PAI V4.0, effective October 1, 2020." Although we recognize and support prior feedback suggesting that IRF quality should be measured on all patients regardless of payer source, we are concerned that CMS has not provided sufficient details to evaluate the ability for IRFs to supply the requested information. Additionally, although CMS has suggested that this data collection on patients from all payers would be used for quality purposes, we are concerned that transmission of IRF-PAI data (including data elements not used for quality purposes) would allow for use of data outside the scope of quality measurement. For example, pages 1 and 2 of the IRF-PAI include a great deal of information, such as therapy information, that is not being used for quality measurement purposes. Would CMS require the transmission of this data for purposes outside of quality measurement—and, if so, how would this information be utilized?

Until CMS provides further details related to what data elements will be required for non-Medicare patients, we cannot support this proposal.

———

We appreciate both the opportunity to comment on this proposed rule and CMS's careful consideration of the concerns and issues raised in this letter. With over thirty years of experience providing coding, clinical, and quality improvement services to IRFs and other PAC providers, UDSᴍʀ welcomes the opportunity to work with CMS to provide ongoing feedback regarding the selection and implementation of standardized and interoperable quality indicators and standardized patient assessment data. If you have any questions about these comments or require additional information, please contact us at 716-817-7800.

Sincerely,

Kathy Dann
Executive Director/CEO

Cc: Pawel Wieczorek, Chief Operating Officer / Director, Information Technology
     Fran Hagerty, Executive Vice President / Manager, Marketing
     Kathleen Conboy, Manager, Sales and Client Services
     Jack Falsone, Manager, Legal Services
     Brigid Greenberg, Manager of Postdischarge Services and Appeals
     Carol Harper, Manager of Clinical Education, Training, and Consultation
     Troy Hillman, Manager, Analytical Services / Government Relations Associate
     Tammy Schneider, Manager, UDS-PRO® (IRF) Product and Services