Implementing Standardized Assessment for Bundled Payment and Other Integrated Service Approaches

Barbara Gage, PhD
Post-Acute Care Center for Research (PACCR)

Gerontological Society of America Annual Scientific Meeting
Economics of Aging Interest Group Symposium
November 7, 2014
The IMPACT Act: What does it do?

- Paves the way for “...standardizing post-acute care assessment data for quality, payment, and discharge planning, and for other purposes.”

- All PAC providers including HH, SNF, IRF and LTCH’s included

- Standardized collection on functional status, cognitive function, medical needs and conditions, impairments and other categories deemed necessary by Secretary

- Some data are already submitted by each PAC provider, but varies by type of provider, Act calls for replacing duplicative data collection

- Resource use data also collected to estimate per beneficiary spend

- Includes payment refinement provisions via report from MedPAC to Congress in 2016 based on PAC PRD data and report from CMS
Quality Measures
• Skilled Nursing Facilities, Inpatient Rehabilitation Facilities, and Long Term Care Hospitals begin October 2016
• Home Health Agencies begin January 2017

Payment Reform
• Medicare Payment Advisory Commission (MedPAC) Report to Congress on payment reforms June 2016
• Centers for Medicare & Medicaid Services, with input from MedPAC, submit to Congress a prototype PAC PPS
How are these timelines possible? Deficit Reduction Act of 2005

• Mandated a PAC Payment Reform Demonstration to understand costs and outcomes across different PAC sites.

• Three components:
  1. CARE: Standardized patient assessment instrument to measure severity in hospitals, PAC settings
  2. Secure, electronic, interoperable standards-based data system for multiple providers to share essential health information/improve transitions
  3. Data collection to analyze costs and outcomes across sites (acute, SNF, HHA, IRF, LTCH)
Standard language was needed to...

• Compare patients across settings
  ➢ Is the same patient treated in more than one type of licensed provider?
  ➢ If so, did both types of providers achieve equal outcomes?
  ➢ If so, were different types of PAC providers paid different amounts for treating similar patients - each PPS uses different items to measure the same concepts.

• Improve coordination of care – one set of terms to define pressure ulcer severity, functional impairment, cognitive impairment across providers.

• Improve data exchangeability – need standard language to transfer information between providers treating the case.
Continuity Assessment Record and Evaluation (CARE) Development

CARE Item Selection:
The Continuity Assessment Record and Evaluation item selection was based on input from the clinical and measurement communities serving PAC populations.

Consensus Input:
Over 25 national associations, including the AHA, AMRPA, NALTH, ALTHA, AHCA, Leading Age, NAHC, VNAA, APTA, AOTA, ASHA, ARN, ANA, CMAA and others provided input on item selection to measure medical, functional, cognitive status and social supports consistently across settings.
Existing Assessment Tools Vary by Setting

- All use at least one but:
  - Acute Hospitals → tools vary by hospital
  - Long-Term Care Hospitals → LTCH CARE required
  - Inpatient Rehabilitation Facilities → IRFPAI required
  - Skilled Nursing Facilities → MDS required
  - Home Health Agencies → OASIS required

Comparison of Current Instruments

**Similarities**
- Medical complexity
- Motor Functional status
- Cognitive status
- Social support and environmental factors

**Differences**
- Individual items that measure each concept
- Rating scales used to measure items
- Look-back or assessment periods
- Unidimensionality of individual items
Comparison of Tools: Functional Status

<table>
<thead>
<tr>
<th>Tools</th>
<th>No. of Functional Items</th>
<th>Rating Scale Levels</th>
<th>Assessment Periods</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRF - PAI</td>
<td>18</td>
<td>7</td>
<td>Past 3 days</td>
</tr>
<tr>
<td>MDS 3.0</td>
<td>11</td>
<td>2 codes (6 &amp;5)</td>
<td>Past 7 days</td>
</tr>
<tr>
<td>OASIS</td>
<td>8</td>
<td>Varies</td>
<td>Assessment day</td>
</tr>
<tr>
<td>CARE</td>
<td>25</td>
<td>6</td>
<td>2-3 day period</td>
</tr>
</tbody>
</table>
## Comparison of Functional Rating Scales

<table>
<thead>
<tr>
<th>IRF-PAI/FIM® Instrument</th>
<th>MDS 3.0 Coding on performance &amp; support provided</th>
<th>OASIS The scale varies in Meaning per item</th>
<th>CARE Tool</th>
</tr>
</thead>
<tbody>
<tr>
<td>7= Complete independence</td>
<td>0= Independent</td>
<td>0= independent</td>
<td>6= Independent with or without a device</td>
</tr>
<tr>
<td>6= Modified independence</td>
<td>0= no set-up</td>
<td>1= (this varies with item)</td>
<td></td>
</tr>
<tr>
<td>5= Supervision and Setup</td>
<td>1= Supervision</td>
<td>2= (this varies with item)</td>
<td>5= Setup and cleanup assistance</td>
</tr>
<tr>
<td></td>
<td>1= Set-up Assistance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4= Minimal Assistance</td>
<td>2= Limited Assistance</td>
<td>3= (this varies with item)</td>
<td>4= Supervision or touching assistance</td>
</tr>
<tr>
<td></td>
<td>2= 1 person assist</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3= Moderate Assistance</td>
<td>3= Extensive Assistance</td>
<td>4= (this varies with item or is not included as a coding choice)</td>
<td>3= Partial/ Moderate assistance</td>
</tr>
<tr>
<td></td>
<td>3= 2 + person assist</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2= Maximal Assistance</td>
<td>4= Total Dependence</td>
<td>5= (this varies with item or is not included as a coding choice)</td>
<td>2= Substantial</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1= Total Assistance</td>
<td>8= Activity NA</td>
<td>UK=Unknown</td>
<td>1= Dependent</td>
</tr>
<tr>
<td>0= Activity Did Not Occur</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Reliability of the Standardized CARE Items

• Most CARE items based on existing validated items currently used in the Medicare program; but few items had been used in multiple settings or across different levels of care.

• Two types of reliability tests were conducted to examine whether the items performed consistently across settings and across disciplines
  1) Traditional Inter-rater Reliability (pairs of assessors rate the same patient similarly)
  2) Video Reliability (cross disciplinary rating of standard video patients)
IRR Methods: Item Selection and Analysis

- CARE item analyses methods consistent with those used to evaluate existing CMS tools (MDS, OASIS, IRFPAI)
  - Categorical items: Kappa (for 2 levels), Weighted Kappa (for > 2 levels, Fleiss-Cohen weights)
    - Range: 0 poor, 0.01–0.20 slight, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial, and 0.81–1 almost perfect
  - Continuous items: Pearson Correlation
CARE IRR Results

- Majority of Kappas (weighted and unweighted) were above 0.60 (substantial)
- Prior functioning and history of falls: 0.69 - 0.863
- Functional status:
  - Core (self-care and mobility including walk and wheel): all above 0.6 except ‘tube feeding’ and ‘walk 150 feet’
  - Supplemental Self Care: all above 0.8 excluding letter codes all above 0.63 except ‘roll left and right’ including letter codes
  - Supplemental Mobility: all above 0.8 excluding letter codes all above 0.63 except ‘walk 50 feet with two turns’ ‘walk 10 feet on uneven surface’ and the wheel long and short ramps including letter codes
CARE IRR Summary

- IRR results indicate **substantial to almost perfect** agreement for the majority of items evaluated – most had already been found reliable in at least one setting
- The few lower kappa scores tend to be for low prevalence items
- IRR results for CARE items are in line with the majority of IRR results available for equivalent items on MDS, OASIS, and FIM
- CARE standardized items can be used reliably across settings
What Does High Reliability of Standardized Items Mean?

Standardized assessment items can be used to allow uniform terminology and definitions to measure patient acuity across settings.

Finding in PAC PRD Report to Congress:

Having uniform, reliable measures of patient acuity and outcomes is a positive step towards understanding differences in patient severity, needs, treatments, and outcomes in a consistent manner and helps foster better communication between providers.
Usefulness of Standardized Assessment Data

Allowed us to compare patients and providers on 3 constructs:

1. Resource Intensity/patient/setting
   • Routine Intensity: Nursing, case management, respiratory therapy, non-Part B services
   • Therapy Intensity: Physical therapy, occupational therapy, speech and language pathology

2. Outcomes/patient/setting
   • Physical Function: Self-Care
   • Physical Function: Mobility
   • Medical Status: Readmission within 30 days discharge from acute hospital

3. Discharge Destination comparisons
   • Characteristics of patients discharged to LTCH, IRF, SNF, HH as first sites of PAC under current policies

(See PAC PRD Final Report on CMS website, Gage et al, 2012.)
E- Specification of Standardized Items

Several CMS projects support the development of e-specified items and measure development from standardized items

- Care Prototype: Developed semantic and terminology for standardized CARE items as part of the developmental work
- E-measure specification projects: Build on e-specifying Federal measures
- CMS item library initiative: creating interoperable library of Federal items
- TEFT grants: testing the development of standardized CARE items for LTSS/HCBS initiatives
Other Related Resources:


CARE Item Development Related Reports

- 2 website URLs

- Continuity Assessment Record and Evaluation (CARE) Item Set: Additional Provider-Type Specific Interrater Reliability Analyses [PDF, 902KB]
- Continuity Assessment Record and Evaluation (CARE) Item Set: Video Reliability Testing. [PDF, 348KB]
Additional Information

Barbara Gage, PhD, MPA
Sr. VP, Scientific Research and Evaluation
bgage@paccr.org
(202) 697-3358

Post-Acute Care Center for Research
www.paccr.org
paccr@paccr.org