



U.S. Department of Health and Human Services
Assistant Secretary for Planning and Evaluation
Office of Disability, Aging and Long-Term Care Policy

DEVELOPMENT OF QUALITY MEASURES FOR INPATIENT PSYCHIATRIC FACILITIES:

FINAL REPORT

February 2015

Office of the Assistant Secretary for Planning and Evaluation

The Office of the Assistant Secretary for Planning and Evaluation (ASPE) is the principal advisor to the Secretary of the Department of Health and Human Services (HHS) on policy development issues, and is responsible for major activities in the areas of legislative and budget development, strategic planning, policy research and evaluation, and economic analysis.

ASPE develops or reviews issues from the viewpoint of the Secretary, providing a perspective that is broader in scope than the specific focus of the various operating agencies. ASPE also works closely with the HHS operating agencies. It assists these agencies in developing policies, and planning policy research, evaluation and data collection within broad HHS and administration initiatives. ASPE often serves a coordinating role for crosscutting policy and administrative activities.

ASPE plans and conducts evaluations and research--both in-house and through support of projects by external researchers--of current and proposed programs and topics of particular interest to the Secretary, the Administration and the Congress.

Office of Disability, Aging and Long-Term Care Policy

The Office of Disability, Aging and Long-Term Care Policy (DALTCP), within ASPE, is responsible for the development, coordination, analysis, research and evaluation of HHS policies and programs which support the independence, health and long-term care of persons with disabilities--children, working aging adults, and older persons. DALTCP is also responsible for policy coordination and research to promote the economic and social well-being of the elderly.

In particular, DALTCP addresses policies concerning: nursing home and community-based services, informal caregiving, the integration of acute and long-term care, Medicare post-acute services and home care, managed care for people with disabilities, long-term rehabilitation services, children's disability, and linkages between employment and health policies. These activities are carried out through policy planning, policy and program analysis, regulatory reviews, formulation of legislative proposals, policy research, evaluation and data planning.

This report was prepared under contract #HHSP2332010016WI between HHS's ASPE/DALTCP and Mathematica Policy Research. Additional funding provided by the HHS Centers for Medicare and Medicaid Services. For additional information about this subject, you can visit the DALTCP home page at http://aspe.hhs.gov/office_specific/daltcp.cfm or contact the ASPE Project Officer, D.E.B. Potter, at HHS/ASPE/DALTCP, Room 424E, H.H. Humphrey Building, 200 Independence Avenue, S.W., Washington, D.C. 20201. Her e-mail address is: D.E.B.Potter@hhs.gov.

DEVELOPMENT OF QUALITY MEASURES FOR INPATIENT PSYCHIATRIC FACILITIES: Final Report

Randall Blair
Junqing Liu
Miriam Rosenau
Michael Brannan
Natalie Hazelwood
Kelsey Farson Gray

Jonathan Brown
Eric Morris
Alyssa Hart
Kenneth Jackson
Angela Schmitt
Katherine Sobel

Mary Barton
Milesh Patel
Allison Siegwarth
Xiao Barry
Stephanie Rodriguez

Mathematica Policy Research

February 4, 2015

Prepared for
Office of Disability, Aging and Long-Term Care Policy
Office of the Assistant Secretary for Planning and Evaluation
U.S. Department of Health and Human Services
Contract #HHSP2332010016WI

The opinions and views expressed in this report are those of the authors. They do not necessarily reflect the views of the Department of Health and Human Services, the contractor or any other funding organization.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	vii
ABSTRACT	viii
ACRONYMS	ix
EXECUTIVE SUMMARY	xii
I. PROJECT RATIONALE	1
A. Project Overview	2
B. Report Roadmap	3
II. SPECIFICATION OF MEASURES	4
A. Evidence Review	4
B. Reviewing Specifications of Similar Measures	4
C. Defining Data Sources, Denominators, and Numerators	6
III. RESEARCH METHODS	12
A. Quantitative Approach to Chart-Based Measure Testing	13
B. Quantitative Approach to Claims-Based Follow-Up Measure Testing	16
C. Qualitative Approach to Measure Testing	21
IV. TESTING RESULTS FOR SCREENING MEASURES	23
A. IPFs and Denominator Population	23
B. Quantitative Testing Results	27
C. Qualitative Testing Results	47
D. Summary of Findings and Proposed Revisions	51
V. TESTING RESULTS FOR THE FOLLOW-UP AFTER IPF HOSPITALIZATION MEASURE	53
A. Characteristics of Medicare Beneficiaries Who Used IPFs	53
B. Performance by Numerator Options	54
C. Impact of Measure Exclusions on Follow-Up Measure Performance	58
D. Follow-Up after IPF Hospitalization by Beneficiary Characteristics and Geographic Location	59
E. Reliability Analysis	63
F. Stakeholder Feedback on the Follow-Up Measure	64
G. Summary and Revisions to Follow-Up Measure Specification	65
VI. FOLLOW-UP MEASURE PERFORMANCE USING MERGED MEDICARE-MEDICAID CLAIMS	67

VII. COMPARISON OF A CHART VERSUS CLAIMS-BASED APPROACH TO THE FOLLOW-UP AFTER IPF HOSPITALIZATION MEASURE	70
A. Methods.....	70
B. IPFs' Efforts to Encourage and Track Follow-Up Care	71
C. IPF Perspectives on a Chart-Based Approach to Measuring Follow-Up Care	73
D. Analysis of Insurance Coverage, Patient Demographics, and Sample Sizes.....	74
E. Conclusion.....	78
VIII. CONCLUSIONS AND LESSONS	80
REFERENCES	81
APPENDICES	
APPENDIX A. IPF Technical Expert Panel Members.....	A-1
APPENDIX B. Screening Measure Specifications.....	A-2
APPENDIX C. Follow-Up after IPF Hospitalization Measure Specifications	A-6
APPENDIX D. Supplemental Tables for Screening Measures.....	A-15
APPENDIX E. Summary of State Selection Process for Follow-Up Measure Analysis with Medicaid Claims	A-18

LIST OF FIGURES AND TABLES

FIGURE IV.1. Average Screening Performance Rates across 6 IPFs	27
FIGURE IV.2. Suicide Risk Screening Performance across 6 IPFs	29
FIGURE IV.3. Violence Risk Screening Performance across 6 IPFs	31
FIGURE IV.4. Substance Use Screening Performance across 6 IPFs	36
FIGURE IV.5. IPF Screening Performance, by Measure Specification	38
FIGURE IV.6. Metabolic Screening Performance across 6 IPFs	41
FIGURE IV.7. Performance on Screening Measures, by Primary Payer	44
FIGURE V.1. 30-Day Follow-Up Rates by IPF Diagnosis, among Non-Dual FFS Medicare Beneficiaries	60
FIGURE V.2. 30-Day Follow-Up Rate by Patient Ethnicity, among Non-Dual FFS Medicare Beneficiaries.....	61
FIGURE V.3. 30-Day Follow-Up after IPF Hospitalization by Region, among Non-Dual FFS Medicare Beneficiaries.....	63
FIGURE VII.1. Primary Payer for IPF Stays in 6 IPFs, 2013.....	75
<hr style="border: 0.5px solid black;"/>	
TABLE ES.1. Measures Tested, Performance	xiii
TABLE I.1. Measures Developed for the IPFQR Program, 2012-2014	2
TABLE I.2. Timeline of IPFQR Program Measure-Development Activities	3
TABLE II.1. Comparison of New Measures and Similar Existing Measures.....	5
TABLE III.1. Quantitative Analyses for Chart-Based Measures	13
TABLE III.2. IPFs Represented in Measure Testing, 2014	14

TABLE III.3.	Quantitative Analysis of Follow-Up after IPF Hospitalization Measure.....	18
TABLE IV.1.	Number of Discharged and Sampled Patients, by IPF	24
TABLE IV.2.	Characteristics of Denominator Population for Screening Measures, by IPF.....	25
TABLE IV.3.	Screening Performance Rates, by IPF	27
TABLE IV.4.	Suicide Risk Screening Performance, by IPF.....	30
TABLE IV.5.	Violence Risk Screening Performance, by IPF	32
TABLE IV.6.	Substance Use Screening Performance, by IPF	35
TABLE IV.7.	Performance Rates across Screening Measures	37
TABLE IV.8.	1-Day versus 3-Day Measure Performance	39
TABLE IV.9.	Metabolic Screening Performance, by IPF.....	40
TABLE IV.10.	Screening Measure Exclusions	42
TABLE IV.11.	Screening Measures Performance Before and After Exclusions	43
TABLE IV.12.	Performance on Screening Measures across Demographic Characteristics	45
TABLE IV.13.	Inter-Rater Agreement on Screening Measures	47
TABLE IV.14.	Summary of Testing Results, Stakeholder Feedback, and Proposed Revisions for Screening Measures.....	52
TABLE V.1.	Characteristics of FFS Medicare Beneficiaries with At Least One Mental Health IPF Hospitalization in Calendar Year 2008.....	54
TABLE V.2.	Numerator Options for Follow-Up Measure	56
TABLE V.3.	Follow-Up within 7 and 30 Days of IPF Hospitalization, among Non-Dual FFS Medicare Beneficiaries	56
TABLE V.4.	Performance among Numerator Options: Follow-Up within 30 Days of IPF Hospitalization, among Non-Dual FFS Medicare Beneficiaries	57
TABLE V.5.	Proportion of Eligible Discharges Excluded from the Follow-Up Measure Denominator, among Non-Dual FFS Medicare Beneficiaries.....	58

TABLE V.6.	Impact of Measure Exclusions on Follow-Up Rates, among Non-Dual FFS Medicare Beneficiaries	59
TABLE V.7.	Follow-Up after IPF Hospitalization by Patient Characteristics, among Non-Dual FFS Medicare Beneficiaries	60
TABLE V.8.	Follow-Up after IPF Hospitalization by State, among Non-Dual FFS Medicare Beneficiaries.....	61
TABLE V.9.	Follow-Up after IPF Hospitalization by Number of Discharges per Facility, among Non-Dual FFS Medicare Beneficiaries	64
TABLE V.10.	Testing Results, Stakeholder Feedback, and Proposed Revisions to the Follow-Up after IPF Hospitalization Measure.....	66
TABLE VI.1.	Facility Performance by Numerator Option: Follow-Up within 30 Days of IPF Hospitalization among Dual and Non-Dual Eligible Beneficiaries	68
TABLE VI.2.	Follow-Up within 30 Days of IPF Hospitalization among All Medicare Beneficiaries, by Data Source	69
TABLE VII.1.	IPFs Represented in Focus Groups and Debriefing Sessions, 2014	70
TABLE VII.2.	Commonly Cited Facilitators to Tracking Follow-Up Care	73
TABLE VII.3.	Commonly Cited Constraints to Tracking Follow-Up Care.....	73
TABLE VII.4.	Comparison of Patient Demographics for FFS Medicare IPF Discharges versus All IPF Discharges	76
TABLE VII.5.	Quarterly Sample Sizes for the Follow-Up after IPF Hospitalization Measure.....	78
TABLE VII.6.	Advantages and Disadvantages of Chart and Claims-Based Approaches to the Follow-Up after IPF Hospitalization Measure.....	79
TABLE A.1.	IPF Technical Expert Panel Members	A-1
TABLE B.1.	Measure Specifications: Screening for Risk of Suicide.....	A-2
TABLE B.2.	Measure Specifications: Screening for Risk of Violence.....	A-3
TABLE B.3.	Measure Specifications: Screening for Substance Use	A-4
TABLE B.4.	Measure Specifications: Metabolic Screening	A-5

TABLE C.1.	Codes to Identify IPF Discharges.....	A-7
TABLE C.2.	Codes to Identify Principle Mental Health Diagnosis	A-8
TABLE C.3.	Codes to Identify Acute Care Facilities.....	A-8
TABLE C.4.	Codes to Identify Admission to Non-Acute Care	A-8
TABLE C.5.	Codes to Identify Patient Deaths and Transfer/Discharge to Another Institution.....	A-9
TABLE C.6.	Codes to Identify Outpatient Visits, Intensive Outpatient Encounters, and Partial Hospitalizations	A-11
TABLE C.7.	Codes to Identify Mental Health Practitioners in Medicare	A-12
TABLE C.8.	Codes to Identify Mental Health Practitioners in Medicaid.....	A-13
TABLE C.9.	Codes to Identify Outpatient Visits	A-14
TABLE C.10.	Additional Resource: HEDIS Definition of Mental Health Practitioner	A-14
TABLE D.1.	Average Performance on Screening Measures, Including and Excluding IPF.....	A-15
TABLE D.2.	Number and Proportion of Patients Excluded from Screening Measures, by IPF.....	A-16
TABLE D.3.	Percent Agreement on Screening Measures, by IPF.....	A-17
TABLE E.1.	Results of State Selection Process for Merged Medicare-Medicaid Analysis	A-20

ACKNOWLEDGMENTS

Mathematica Policy Research and the National Committee for Quality Assurance prepared this report under contract to the Office of the Assistant Secretary for Planning and Evaluation (ASPE), U.S. Department of Health and Human Services (HHS) (HHSP23320100019WI/HHSP23337001T). The authors appreciate the guidance of D.E.B. Potter, Joel Dubenitz, and Kirsten Beronio (ASPE), Elizabeth G. Ricksecker and Jeff Buck (HHS Centers for Medicare and Medicaid Services [CMS]), and Lisa Patton (HHS Substance Abuse and Mental Health Services Administration [SAMHSA]).

The views and opinions expressed here are those of the authors and do not necessarily reflect the views, opinions, or policies of ASPE, CMS, SAMHSA, HHS, or the technical expert panel. The authors are solely responsible for any errors.

ABSTRACT

Summary: As part of its National Quality Strategy, the U.S. Department of Health and Health Services (HHS) Office of the Assistant Secretary for Planning and Evaluation (ASPE) and the HHS Centers for Medicare and Medicaid Services (CMS) are committed to developing and implementing measures that can be used for behavioral health care quality improvement. To further the implementation of such measures, and as mandated in Section 3401, Subsection 10322 of the Patient Protection and Affordable Care Act of 2010, CMS developed the Inpatient Psychiatric Facility (IPF) Quality Reporting (IPFQR) program, a pay-for-reporting program that went into effect for fiscal year 2014. Under this program, IPFs must report their performance on a set of quality measures or face a two percentage point reduction to the update of their Medicare standard federal rate for that year. Funded through an inter-agency agreement between ASPE and CMS, the goal of this project was to develop and test measures that may be incorporated into the IPFQR program; these included four chart-based measures that assess screening for risk of suicide, risk of violence, substance use, and metabolic conditions, and one claims-based measure that assesses whether Medicare beneficiaries receive follow-up care after IPF hospitalization.

Major Findings: Among the six inpatient psychiatric facilities (IPFs) that piloted the chart-based measures, performance was generally high on the suicide, violence, and substance use screening measures. In contrast, there was wide variation in metabolic screening rates across IPFs. All chart-based measures demonstrated good inter-rater reliability and had moderate to strong stakeholder support. The claims-based follow-up measure demonstrated wide variation across IPFs and very strong reliability, but received mixed stakeholder support.

Purpose: This project developed measures that may be incorporated into the IPFQR program, including four chart-based screening measures (risk of suicide, risk of violence, substance use, and metabolic conditions) and a claims-based measure to assess whether individuals discharged from the IPF receive follow-up care. The measures were tested using quantitative and qualitative methods to assess attributes consistent with National Quality Forum endorsement criteria -- importance, feasibility, usability, and scientific acceptability (reliability and validity).

Methods: This project first reviewed existing measures and gathered input from consumers, IPFs, IPFQR program vendors, state agencies, and performance measurement experts to identify opportunities for new measures. Based on the evidence to support measure concepts, measure specifications were developed and pilot tested. The follow-up measure was tested using Medicare claims data for over 1,600 IPFs. The chart-based measures were piloted at six IPFs. Quantitative testing for all measures involved calculating performance rates to examine variation across IPFs,

differences in performance among subpopulations, and reliability. For all measures, qualitative data collection included focus groups with a range of stakeholders to get input on the measure specifications and understand whether the measures yield findings that could be used to inform quality improvement efforts. A technical expert panel provided input throughout the project.

ACRONYMS

The following acronyms are mentioned in this report and/or appendices.

ADA	American Diabetes Association
APA	American Psychiatric Association
ASPE	HHS Office of the Assistant Secretary for Planning and Evaluation
ASSIST	Alcohol, Smoking and Substance Involvement Screening Test
AUDIT	Alcohol Use Disorders Identification Test (screen)
AUDIT-C	Alcohol Use Disorders Identification Test Consumption (screen)
BETOS	Berenson-Eggers Type of Service
BMI	Body Mass Index
BVC	Broset Violence Checklist
CAGE	Cut down, Annoyed, Guilty, and Eye-opener (questionnaire)
CAH	Critical Access Hospital
CHP	Child Psychiatrist
CMS	HHS Centers for Medicare and Medicaid Services
CHN	Child Neurologist
CPT	Current Procedural Terminology
CRAFFT	Car, Relax, Alone, Forget, Friends, Trouble (screen)
DAST-10	Drug Abuse Screening Test-10 items
DO	Doctor of Osteopathy
DUA	Data Use Agreement
EDB	Enrollment Data Base
ER	Emergency Room
FACTYP	Facility Type
FFS	Fee-For-Service
FUH	Follow-up After Hospitalization for Mental Illness
FY	Fiscal Year
G-MAST	Geriatric version-MAST
GAIN-SS	Global Appraisal of Individual Needs Short Screener
HbA1c	Glycated Hemoglobin
HBIPS	Hospital-Based Inpatient Psychiatric Services
HCPCS	Healthcare Common Procedure Coding System
HEDIS	Healthcare Effectiveness Data and Information Set

HHA	Home Health Agency
HHS	U.S. Department of Health and Human Services
HMO	Health Maintenance Organization
ICD-9-CM	International Classification of Disease, 9th revision, Clinical Modification
ICF	Intermediate Care Facility
IPF	Inpatient Psychiatric Facility
IPFQR	Inpatient Psychiatric Facility Quality Reporting
IQR	Interquartile Range
MAST	Michigan Alcohol Screening Test
MAX	Medicaid Analytic eXtract (files)
MCO	Managed Care Organization
MD	Medical Doctor
MSIS	Medicaid Statistical Information System
MSSI	Modified Simple Screening Instrument
NCQA	National Committee for Quality Assurance
NPI	National Provider Identification
NQF	National Quality Forum
OP	Outpatient
OT	Medicaid Other Services/Therapy File
PA	Physician Assistant
POS	Place of Service
SAFE-T	Suicide Assessment Five-Step Evaluation and Triage
SAMHSA	HHS Substance Abuse and Mental Health Services Administration
SBQ-R	Suicide Behaviors Questionnaire-revised
SNF	Skilled Nursing Facility
SSI-SA	Simple Screening Instrument for Substance Abuse
SSRI	Suicide Screening Risk Inventory
SUB	Substance Use
TAP	Technical Advisory Panel
TEP	Technical Expert Panel
TJC	The Joint Commission
TOS	Type of Service (in Medicaid files)
TPBBEG	Part B Enrollment -- Beginning Date
TPBEND	Part B Enrollment -- End Date
TWEAK	Tolerance Worried Eye-opener Amnesia K/cut down
TYPSVC	Type of Service (in Medicare files)

UB	Uniform Billing
VERDICT	Veterans Evidence-based Research, Dissemination, and Implementation Center
V-RISK	Violence Risk Screening

EXECUTIVE SUMMARY

Despite improvements in behavioral health treatments, gaps remain between evidence-based care and the care provided to millions of individuals living with mental health problems (Institute of Medicine 2006). As part of its National Quality Strategy, CMS is committed to reducing this gap by developing and implementing measures that can be used for quality improvement within inpatient psychiatric facilities (IPFs). To further the implementation of such measures, and as mandated in Section 3401, Subsection 10322 of the Patient Protection and Affordable Care Act of 2010, CMS developed the Inpatient Psychiatric Facility Quality Reporting (IPFQR) program, a pay-for-reporting program that went into effect for fiscal year 2014. Under this program, IPFs must report their performance on a set of quality measures or face a 2 percentage point reduction to the update of their Medicare standard federal rate for that year.

Over 1,800 IPFs (both freestanding psychiatric hospitals and psychiatric units of general hospitals) reported their performance on several measures in the first year of the IPFQR program. These measures include six chart-based process measures that address patient safety, care coordination, and medication use.¹ Although the six measures currently included in the IPFQR program provide a strong foundation for improving the quality of inpatient behavioral health care, gaps in measurement persist.²

In September 2012, the Office of the Assistant Secretary for Planning and Evaluation, with support from CMS, modified an existing contract with Mathematica Policy Research and its subcontractor -- the National Committee for Quality Assurance (NCQA) -- to develop measures for the IPFQR program. The goal of this new component of the project was to develop and test four chart-based measures that assess screening for risk of suicide, risk of violence, substance use, and metabolic conditions, and one claims-based measure that assesses whether Medicare beneficiaries receive follow-up care after IPF hospitalization.

The first phase of work under this contract involved conducting a targeted review of evidence to support the selected measure concepts; this review was completed in late 2012. Next, the team held several meetings with IPF staff and other subject matter experts to obtain input and guidance on the technical specifications of these measures. In September 2013, the team presented draft specifications for the five measures to a technical expert panel (TEP), and the TEP provided the team with useful feedback on ways to further refine and strengthen the specifications prior to measure testing.

¹ These are the Hospital-Based Inpatient Psychiatric Services (HBIPS) measures 2-7, developed by The Joint Commission (TJC) and endorsed by the National Quality Forum.

² Another TJC measure, Alcohol Use Screening (SUB-1) will be incorporated into the IPFQR in 2015, along with the follow-up measure presented in this report.

In early 2014, the team pilot tested the chart-based measures at six IPFs and began testing the claims-based measure using Medicare claims data. Starting in April 2014, Mathematica and NCQA staff also gathered qualitative feedback on the performance and usability of the measures through debriefing sessions with IPFs that participated in testing, as well as focus groups with state policymakers, consumer and advocacy groups, measure experts, IPFQR program vendors, and additional IPF staff. The results of quantitative measure testing are summarized in Table ES.1.

TABLE ES.1. Measures Tested, Performance			
Measure	Variation in Measure Performance Across IPFs (number of IPFs) ¹	Mean Measure Performance ¹	Reliability ²
Screening for risk of suicide	67.6-99.4% (6 IPFs)	93.4%	0.65
Screening for risk of violence	47.7-99.1% (6 IPFs)	89.0%	0.63
Screening for substance use	51.4-96.4% (6 IPFs)	85.8%	0.49
Metabolic screening	6.2-98.6% (6 IPFs)	41.5%	0.93
Follow-Up after IPF hospitalization (30 days) ³	0-100% (1,669 IPFs) 25 th percentile: 42.3 75 th percentile: 67.3	53.5%	0.93
NOTES:			
1. Expressed as the proportion of patients who met the measure requirement.			
2. Reliability for the follow-up measure was calculated using beta-binomial statistic (score of 0.7 or higher indicates that the measure can reliably discriminate performance between IPFs). Reliability for all other measures is the agreement between 2 chart abstractors (inter-rater agreement) for the numerator of the measure, calculated using Cohen's kappa statistic. A kappa of 0.21-0.40 indicates fair agreement; a kappa of 0.41-0.60 indicates moderate agreement; a kappa of 0.61-0.80 indicates substantial agreement; a kappa of 0.81 or higher indicates almost perfect agreement.			
3. The follow-up measure has 2 rates: 7-day and 30-day follow-up. 30-day rates are reported in this table for the sake of simplicity; there was also wide variation in the 7-day follow-up rates.			

Measure Testing Results

Admission Screening Measures. The three admission screening measures -- screening for risk of suicide, risk of violence, and substance use -- require that IPF staff collect information on core screening elements within one day of patient admission. Performance was quite high across IPFs on these measures, with average performance on the measures ranging from 86 percent (in the case of substance use) to 93 percent (in the case of suicide). Reliability was moderate for the substance use measure and substantial for the suicide and violence measures. Stakeholders were generally supportive of the measures and thought they represented an improvement over existing screening measures used in an inpatient psychiatric setting, including HBIPS-1: Admission Screening for Violence Risk, Substance Use, Psychological Trauma History and Patient Strengths, a TJC measure reported by a large portion of IPFs throughout the country.

Regarding changes to measure specifications, stakeholders generally recommended that the final specification of the substance use, violence, and suicide screening measures use a three-day time frame to allow for complete and accurate screenings. Obstacles to performing accurate screenings within one day of admission include staff shortages, patient uncooperativeness, and lack of patient lucidity. Some stakeholders noted that the suicide and violence measures should be conducted within a one-day time frame, given the clinical importance of obtaining that information quickly. Based on this feedback, the research team recommends changing the time frame for the substance use screening measure from one day to three days, and keeping the suicide and violence screening specifications at one day (as specified prior to testing). The additional two days for the substance use measure will facilitate the capture of complete and accurate information regarding patients' alcohol and drug use, without compromising the need to capture important information on suicide and violence risk in the first day of admission.

Metabolic Screening Measure. The metabolic screening measure requires that the following four screenings are documented in the patient record for all individuals discharged on antipsychotic medications: (1) body mass index (BMI); (2) blood pressure; (3) glucose or glycated hemoglobin (HbA1c); and (4) a full lipid panel. Performance on the metabolic screening measure was low, on average, across the six IPFs. The measure's average performance rate of 42 percentage points highlights a sizable performance gap on the measure. The metabolic screening measure also demonstrated non-trivial variation in performance among IPFs as well as by patient characteristics. In addition, it demonstrated near-perfect agreement between chart abstractors (kappa of 0.93 for the measure numerator).

Overall, stakeholders found the metabolic screening measure to be important for addressing a notable gap in psychiatric care. However, focus group participants and TEP members were divided over whether to keep the requirement of a full lipid panel, as some felt that blood pressure, BMI, and glucose/HbA1c tests were sufficient screening requirements. In particular, three of nine TEP members expressed concern that the measure might inadvertently encourage IPFs and other clinicians to conduct unnecessary tests -- namely a full lipid panel in instances in which there is no clinical need. However, given the preponderance of clinical evidence supporting a full lipid panel on an annual basis for patients taking regularly prescribed antipsychotic medications, we suggest that the full lipid panel remain a screening element in the metabolic screening measure.

Follow-Up Measure. The claims-based follow-up measure calculates the proportion of patients that had an outpatient visit with a mental health practitioner within seven and 30 days following IPF hospitalization. The measure demonstrated strong quantitative performance; there was good variation in measure performance across IPFs and among demographic subgroups. In addition, IPFs' low average performance on the measure on a national scale highlights room for improvement. The measure also had very good reliability (beta-binomial statistic of 0.93 for the 30-day measure).

Stakeholder support for the follow-up measure was mixed. Three of the six IPFs involved in testing, and at least 11 of the 28 focus group participants expressed concern that the measure may inappropriately hold IPFs solely accountable for follow-up care, despite the range of community-level factors that may influence performance on the measures. However, at least five focus group participants -- primarily policymakers and measurement experts -- noted that this measure could help to drive innovative partnerships between facilities, community mental health agencies, health plans, and providers to improve follow-up care for IPF patients. Likewise, TEP members were divided in their support for the follow-up measure. Two TEP members were concerned that the measure would unfairly hold IPFs accountable for factors outside of their control, whereas two other TEP members expressed strong support for the follow-up measure, arguing that it could identify opportunities for quality improvement among facilities with low rates of follow-up care.

I. PROJECT RATIONALE

Despite improvements in behavioral health treatments, gaps remain between evidence-based care and the care provided to millions of individuals living with mental health problems (Institute of Medicine 2006). As part of its National Quality Strategy, the U.S. Department of Health and Human Services (HHS) Centers for Medicare and Medicaid Services (CMS) is committed to reducing this gap by developing and implementing measures that can be used for quality improvement within inpatient psychiatric facilities (IPFs). To further the implementation of such measures, and as mandated in Section 3401, Subsection 10322 of the Patient Protection and Affordable Care Act of 2010, CMS developed the Inpatient Psychiatric Facility Quality Reporting (IPFQR) program, a pay-for-reporting program that went into effect for fiscal year (FY) 2014. Under this program, IPFs must report their performance on a set of quality measures or face a 2 percentage point reduction to the update of their Medicare standard federal rate for that year.

Over 1,800 IPFs (both freestanding psychiatric hospitals and psychiatric units of general hospitals) reported measures in the first year of the IPFQR program. These measures include performance rates on six chart-based process measures that address patient safety, care coordination, and medication use.³ Although the six measures currently included in the IPFQR program provide a strong foundation for improving the quality of inpatient behavioral health care, gaps in measurement persist.⁴

In September 2012, the HHS Office of the Assistant Secretary for Planning and Evaluation (ASPE), with funding from CMS, modified an existing contract with Mathematica Policy Research and its subcontractor, the National Committee for Quality Assurance (NCQA), to support the development of measures for the IPFQR program. Prior to the modification, the contract supported the development of behavioral health quality measures for health plans with funding from ASPE and the HHS Substance Abuse and Mental Health Services Administration. The goal of this new component of the project was to develop and test four chart-based measures that assess screening for risk of suicide, violence, substance use, and metabolic conditions, and one claims-based measure that assesses whether Medicare beneficiaries receive follow-up care after IPF hospitalization (Table I.1).

Under a separate contract in 2012, a technical expert panel (TEP) prioritized these screening measure concepts as the most clinically relevant and feasible to measure among a set of nearly 20 promising measurement concepts. These new measures were intended to strengthen the standards of existing psychiatric inpatient screening

³ These are the Hospital-Based Inpatient Psychiatric Services (HBIPS) measures 2-7, developed by The Joint Commission (TJC) and endorsed by the National Quality Forum (NQF).

⁴ Another TJC measure, Alcohol Use Screening (SUB-1) will be incorporated into the IPFQR in FY 2015, along with the follow-up measure presented in this report.

measures by requiring specific screening elements that were recommended by the TEP and supported by the evidence, and requiring screening within one day of admission (as opposed to within three days of admission). In late 2012, CMS and ASPE prioritized the adaptation of an existing Healthcare Effectiveness Data and Information Set (HEDIS) follow-up measure (NQF #0576) for use in IPFQR program.

TABLE I.1. Measures Developed for the IPFQR Program, 2012-2014	
Measure Concept	Primary Data Source*
Screening for risk of suicide	Medical record
Screening for risk of violence	Medical record
Screening for substance use	Medical record
Screening for metabolic conditions	Medical record
Follow-up after IPF hospitalization	Medicare claims
* The data for the screening measures primarily derive from medical records. However, other data sources could be used to populate the measure numerator, such as administrative data or laboratory data that are not integrated with the medical record.	

A. Project Overview

Building on work that Mathematica and NCQA completed under a previous contract with CMS (HHSP23320100019WI), the first phase of work under this contract involved conducting a targeted review of evidence to support the selected measure concepts; this review was completed in late 2012. Mathematica and NCQA staff then used that evidence to develop measure specifications throughout 2013. In addition, the team held several meetings with IPF staff and other subject matter experts to obtain input and guidance on the technical specifications of these measures. The team presented draft specifications for the five measures to a TEP at its September 2013 meeting, and the TEP provided the team with useful feedback on ways to further refine and strengthen the specifications prior to measure testing. (A full list of TEP members is provided in Appendix A.)

In early 2014, the team pilot tested the chart-based measures at six IPFs and began testing the claims-based measure using Medicare claims data. Starting in April 2014, Mathematica and NCQA staff also gathered qualitative feedback on the performance and usability of the measures through debriefing sessions with IPFs that participated in testing and focus groups with state policymakers, consumer and advocacy groups, measure experts, IPFQR program vendors, and additional IPF staff. In late 2014, the team revised the measure specifications based on qualitative and quantitative testing results, as well as input from the final TEP meeting. (Full measure specifications for the screening measures and the follow-up measure are provided in Appendix B and Appendix C, respectively.)

Table I.2 provides a timeline of testing activities performed under this contract. At the time of this report, CMS has not made final decisions about the measure specifications or inclusion of these measures in IPFQR program. As of late 2014, these measures have not been submitted to the NQF for endorsement.

TABLE I.2. Timeline of IPFQR Program Measure Development Activities	
Date	Activities
September 2012	TEP meeting to receive feedback on measure concepts
December 2012	Updated evidence review focused on selected measure concepts
Early to mid-2013	Specified measures
September 2013	Obtained TEP input on draft measure specifications.
January-April 2014	Tested chart-based measures in 6 IPFs and obtained input through stakeholder focus groups
January-July 2014	Conducted analysis of claims-based follow-up measure and obtained input through stakeholder focus groups
October 2014	Obtained final TEP input on measures

B. Report Roadmap

This report presents final testing results for the four chart-based screening measures and the claims-based follow-up measure developed and tested under this contract. Chapter II describes the process for specifying the measures; Chapter III describes the methods used to test the measures; and Chapter IV, Chapter V, and Chapter VI present the findings. The final chapter (Chapter VII) offers a summary of findings and lessons learned from this project that may be applicable to future measure-development and implementation efforts for inpatient psychiatric populations.

II. SPECIFICATION OF MEASURES

The specification of the measures consisted of three overarching steps: (a) conducting an evidence review; (b) reviewing specifications of similar measures; and (c) identifying feasible data sources that could be used to construct the denominator and numerator for each measure. These steps are discussed below.

A. Evidence Review

In early 2013, Mathematica and NCQA staff updated evidence reviews that were completed under the previous contract with CMS (HHSP23320100019WI) for all five measure concepts in development. In all evidence reviews, the team attempted to assess whether there was clear guidance to specify the denominator and numerator of each measure. The evidence reviews also addressed a critical component of NQF review -- the importance of each measure, including the evidence base supporting the measure and the extent to which it reflects a high-impact aspect of the national health care system. The reviews drew on clinical guidelines, systematic reviews (including meta-analyses), and the recommendations of authoritative government agencies and task forces, including the U.S. Preventive Services Task Force, the HHS Centers for Disease Control and Prevention, and others.

The review identified several guidelines that informed measure specifications, particularly regarding necessary tests to screen for metabolic conditions, as well as the frequency with which metabolic screening should occur.⁵ However, the review provided no clear guidance regarding the content of screenings for suicide, substance use, and violence -- that is, the screening elements that constitute a high quality suicide, violence, or substance use screening. For this reason, the team conducted an analysis of screening elements in validated screening tools, and identified a core set of screening elements that appeared across screening tools. This analysis of screening elements informed the specifications for these three screening measures.

B. Reviewing Specifications of Similar Measures

Next, the team reviewed specifications of similar screening measures to determine potential areas for improvement. Several screening measures developed as part of this project are conceptually similar to existing screening measures currently reported by

⁵ These guidelines include the American Diabetes Association (ADA) and American Psychiatric Association (APA) Consensus Development Conference on Antipsychotic Drugs and Obesity and Diabetes (2004); the APA practice guideline on assessment and treatment of patients with schizophrenia (2002); the APA practice guideline on assessment and treatment of patients with bipolar disorder (2010); and the APA practice guideline on assessment and treatment of patients with major depressive disorder (2010).

IPFs. Specifically, three of the four chart-based screening measures are similar to two existing screening measures developed by TJC for inpatient populations (see Table II.1). Below is a comparison between the measures in development under this contract and similar TJC measures.

TABLE II.1. Comparison of New Measures and Similar Existing Measures			
New Measure	Description	Similar To:	Key Difference
Suicide risk screening	<p>Percentage of admissions for which a detailed screening for risk of suicide was completed within 1 day of admission.</p> <p>Screening must include inquiry into: (1) suicidal ideation; (2) plans or preparations; (3) intent; (4) past suicidal behavior; and (5) risk and protective factors.</p>	HBIPS-1 Violence Risk to Self screening component	<p>The new measure requires screening elements, whereas HBIPS-1 only requires documentation of a screening.</p> <p>The new measure requires a completed screening within 1 day of admission versus 3 days for HBIPS-1.</p>
Violence risk screening	<p>Percentage of admissions for which a detailed screening for risk of violence was completed within 1 day of admission.</p> <p>Screening must include inquiry into: (1) threats of violence; and (2) lifetime history of violent episodes.</p>	HBIPS-1 Violence Risk to Others screening component	<p>The new measure requires screening elements, whereas HBIPS-1 only requires documentation of a screening.</p> <p>The new measure requires a completed screening within 1 day of admission versus 3 days in the case of HBIPS-1.</p>
Substance use (alcohol and drug) screening	<p>Percentage of admissions for which a detailed screening for drug use was completed within 1 day of admission.</p> <p>Screening must include inquiry into: (1) type, frequency, and amount of alcohol <i>and</i> substance use in the past 12 months; (2) adverse effects of this use (if use is reported); (3) dependence upon these substances (if use is reported); and (4) any lifetime history of drug/alcohol abuse.</p>	<p>HBIPS-1 Substance Use screening component</p> <p>SUB-1: Alcohol Use Screening</p>	<p>The new measure requires specific alcohol and drug screening elements, whereas HBIPS-1 only requires documentation of drug and alcohol screening. SUB-1 screens only for alcohol use.</p> <p>The new measure requires a completed screening within 1 day of admission versus 3 days for HBIPS-1 and SUB-1.</p> <p>SUB-1 is specified for a general inpatient population, whereas the new measure and HBIPS-1 are specified for an inpatient psychiatric population.</p> <p>SUB-1 requires screening with a validated instrument, whereas HBIPS-1 and the new measure do not.</p>

Three of the screening measures under development in this contract are similar to components of TJC’s HBIPS measure titled “Admission Screening for Violence Risk,

Substance Use, Psychological Trauma History and Patient Strengths” (HBIPS-1).⁶ TJC-accredited IPFs -- more than one-fourth of all IPFs included in the IPFQR program -- currently report HBIPS-1 to TJC (NRI 2012). HBIPS-1 reports whether screenings for suicide, violence, and substance use (among other assessments) were completed; IPFs earn credit on the measure only if all screenings were completed within three days of patient admission. In this contract, Mathematica and NCQA developed and tested individual screening measures for suicide, violence, and substance use. Based on the guidance from the TEP, these new measures were intended to strengthen the screening standards of HBIPS-1 by requiring specific screening elements that were recommended by the TEP and supported by the evidence, and requiring screening within one day of admission (as opposed to within three days in the case of HBIPS-1).

In addition, the newly developed substance use screening measure is conceptually related to TJC’s SUB-1⁷ (Alcohol Use Screening), which will be included the IPFQR program in FY 2015.⁸ The primary distinction between SUB-1 and the substance use screening measure developed under this contract is that the new measure requires the documentation of specific drug and alcohol screening elements recommended by the TEP (for example, inquiry into negative consequences of alcohol use), whereas SUB-1 requires an alcohol screening with a validated instrument, but no drug screening. The new substance use screening measure also requires that the screening be completed within one day of admission, versus within three days in the case of SUB-1.⁹

As described in Chapter III, during measure testing, the research team compared IPFs’ performance on the newly specified screening measures to their performance on the relevant components of HBIPS-1 and SUB-1. The goal of these comparisons was to better understand how alternate specifications for screening measures affect IPF performance. We present these results in Chapter IV.

C. Defining Data Sources, Denominators, and Numerators

Next, the research team determined the appropriate data sources for the: (1) admission screening measures; (2) metabolic screening measure; and (3) follow-up measure, discussed below.

⁶ This measure reported by TJC-accredited psychiatric inpatient hospitals. HBIPS-1 was endorsed by NQF in April 2014.

⁷ SUB stands for substance use. The two measures, recently recommended by Measure Applications Partnership for the IPFQR, SUB-1 and SUB-2, are part of a set of four linked measures related to alcohol and substance use developed for an acute care setting.

⁸ SUB-1 appeared on the Measures Under Consideration list published on December 1, 2012. Under the current schedule, facilities would be required to begin reporting SUB-1 in FY 2015.

⁹ Starting in 2014, SUB-1 required alcohol use screening within three days of admission, similar to HBIPS-1. However, in 2013 and in previous years, SUB-1 required screening prior to patient discharge.

1. Admission Screening Measures

Identification of Data Sources. Admission screening measures include the suicide, violence, and substance use screening measures. Based on feedback from stakeholder focus groups and our TEP, we determined that patient record review was necessary to accurately capture the numerator of the admission screening measures, given that claims and administrative data would not have complete information on individual screening elements. However, as described below, administrative data were used to identify the measure denominator and exclusions, as they provide reliable information regarding patients' length of stay and age.

Defining the Denominator. We sought screening measures that would be broadly applicable to all IPFs and their full patient populations. As such, we defined the denominator for admission screening measures as all discharged IPF patients. In the interest of comparing measure performance with existing measures, we aligned these measures' denominators with existing HBIPS sampling methods (TJC 2012), which use IPF administrative data (not claims) to draw sufficiently large sample sizes across five age groups to generate performance rates for each of these groups. This sampling approach slightly oversamples patients under 18 and patients over 64, but largely yields a random sample of at least 20 percent of each IPF's entire patient population on a monthly or quarterly basis.

Refining Admission Screening Numerator Time Frame to Require Completion of Screening within One Day of Admission. We specified the admission screening measures to require completion of each screening within one day of admission. We made this decision based on initial input from the TEP, which reasoned that these screenings must occur within one day, given that screening results -- particularly for suicide and violence -- are necessary early in the inpatient stay to inform subsequent care. This approach differs from similar measures (including HBIPS-1 and SUB-1), which require that screening be completed within three days of admission.¹⁰ As part of measure testing, the research team documented whether all screening elements were completed within 1-3 days of admission, to compare performance.

Strengthening Numerator Requirements to Reflect a Higher Standard of Quality. Based on stakeholder feedback, guidance from the TEP, and the evidence review and subsequent analysis, the admission screening measures require the documentation of specific screening elements. For example, a patient record must include documentation on the presence or absence of suicidal ideation, plans, intent, history of suicidal behavior, and risk and protective factors in order for the facility to receive credit for completing the suicide screen. We identified these screening elements through a systematic review of evidence, analysis of validated screening tools, and

¹⁰ Based on conversations with TJC staff, the rationale for requiring screening within three days of admission -- as opposed to within one or two days of admission -- is the lack of availability of qualified staff to complete screenings on weekends (particularly in small IPFs) as well as patients' possible reluctance to provide complete and accurate information immediately upon admission, particularly with respect to drug and alcohol use.

consultation with the TEP. Generally, these core elements reflect screening elements that are common across validated screening tools and relevant clinical guidelines.¹¹

As discussed above, this element-centered approach differs from similar measures, which require only documentation that screening was completed (in the case of HBIPS-1) or that screening was completed using a validated instrument (SUB-1).¹² The TEP and other stakeholders perceived that requiring core screening elements represented a higher standard of quality than merely documenting presence or absence of a completed screening (as in the case of HBIPS-1). Furthermore, the TEP and other stakeholders asserted that requiring a core set of screening elements for each measure would have more clinical value than requiring the use of a validated screening tool. However, the TEP and stakeholders reported that the use of validated instruments that contain the specific screening elements was acceptable in order for an IPF to receive credit for the measure.

2. Metabolic Screening Measure

Identification of Data Sources. Based on feedback from stakeholder focus groups and the TEP, we determined that patient record review was necessary to accurately capture the numerator of the metabolic screening measure, which requires a series of measurements and tests. This is because: (1) data elements to examine all aspects of metabolic screening (like a blood pressure measurement or a full lipid panel) are not captured in claims; and (2) only a portion of all IPF stays are captured in claims. Similar to the admission screening measures, administrative data from the IPFs were also vital to determining the metabolic screening measure denominator and exclusions -- including patients' length of stay and whether patients were discharged from IPFs on antipsychotic medications.¹³

¹¹ The suicide measure screening elements are cited in the following guidelines and screening tools: APA Guideline for Treatment of Patients with Major Depressive Disorder (2010), the Suicide Behaviors Questionnaire-revised (SBQ-R), Osman et al. (2001), Veterans Evidence-based Research, Dissemination, and Implementation Center (VERDICT), Suicide Assessment Five-Step Evaluation and Triage (SAFE-T), and Modified Simple Screening Instrument (MSSI). The substance use measure screening elements are cited in the following guidelines and screening tools: APA Guideline for the Treatment of Patients with Substance Use Disorders (2007), Alcohol Use Disorders Identification Test (AUDIT), the Global Appraisal of Individual Needs Short Screener (GAIN-SS), the Simple Screening Instrument for Substance Abuse (SSI-SA), and the Drug Abuse Screening Test (DAST-10). The violence measure screening elements are cited in the following guidelines and screening tools: the Violence Risk Screening-10 (V-RISK-10), the Broset Violence Checklist (BVC).

¹² These tools include the AUDIT; the Alcohol Use Disorders Identification Test Consumption (AUDIT-C); Alcohol, Smoking and Substance Involvement Screening Test (ASSIST); Tolerance Worried Eye-opener Amnesia K/cut down (TWEAK); the Car, Relax, Alone, Forget, Friends, Trouble (CRAFFT); Michigan Alcohol Screening Test (MAST); and Geriatric Version-MAST (G-MAST), but may not include the Cut down, Annoyed, Guilty, and Eye-opener (CAGE). However, this list is not exhaustive.

¹³ Patient records were often the primary data source regarding whether patients were discharged on antipsychotics. However, most IPFs involved in testing migrated this information to their administrative data systems, primarily to report the quality measure HBIPS-4: Patients discharged on multiple antipsychotic medications.

Defining the Denominator. Consistent with clinical research (Marder et al. 2004), TEP input, and an existing HEDIS diabetes screening measure,¹⁴ we defined the denominator as all patients discharged on antipsychotic medications. We selected patients discharged on any antipsychotic medication -- as opposed to patients discharged on second-generation antipsychotic medications -- for the measure denominator because there is evidence that both first-generation and second-generation antipsychotics can contribute to weight gain, dyslipidemia, and type 2 diabetes (Marder et al. 2004; ADA 2006; Roohafsza et al. 2013).¹⁵ In light of this risk, relevant consensus statements recommend a full metabolic screening for patients discharged on any antipsychotic medication. Most notably, in a consensus statement on antipsychotic drugs and obesity and diabetes, the ADA, the APA, the American Association of Clinical Endocrinologists, and the North American Association for the Study of Obesity stated, "The panel recommends that baseline screening measures be obtained before, or as soon as clinically feasible after, the initiation of any antipsychotic medication (ADA-APA 2004)."

In addition, the selection of *any antipsychotic medication* for the measure denominator was influenced by feasibility concerns, given that TJC-accredited IPFs currently track the number of patients discharged on any antipsychotic medication for HBIPS-4: Patients discharged on multiple antipsychotic medications. Drawing the distinction between first-generation and second-generation antipsychotic medications for this metabolic screening measure would require IPFs to construct new data elements, which IPFs described as quite burdensome. Therefore, basing this measure denominator on the HBIPS-4 denominator was an appropriate option. No distinction is made in the measure specifications between patients who initiated antipsychotic treatment during the IPF stay versus those who continued an antipsychotic treatment regimen during the IPF stay, as guidelines state that a full metabolic screening is necessary for both populations (ADA-APA 2004).

Defining the Numerator. The metabolic screening measure requires that the following four screenings were documented in the patient record: (1) body mass index (BMI); (2) blood pressure; (3) glucose or glycated hemoglobin (HbA1c); and (4) a full lipid panel. These requirements were largely based on clinical guidelines for individuals taking antipsychotic medications, as well as data elements included in the HEDIS diabetes screening measure and similar measures designed for alternate health care settings and populations. Experts agree that the combination of these tests, as opposed to any individual test, provides more accurate information about patients' risk for diabetes and cardiovascular disease (ADA-APA 2004).

Related to the high risk of diabetes among individuals on antipsychotics, an HbA1c or glucose test plays a vital role in assessing diabetes risk before and after initiation of

¹⁴ Diabetes Screening for People with Schizophrenia or Bipolar Disorder who are Using Antipsychotic Medications (NQF #1932).

¹⁵ Some studies and consensus statements, including ADA-APA consensus statement, point to the higher risk of second-generation antipsychotic medications relative to first-generation medications. However, both medication types pose risks to metabolic functioning, particularly within 12 weeks of initiation of medication.

an antipsychotic medication regimen. One evidence review (Marder et al. 2004) states, “A baseline measure of plasma glucose should be collected for all patients before starting a new antipsychotic. Measurement of the fasting plasma is preferred, but measurement of HbA1c is acceptable if a fasting plasma glucose tests is not feasible.” A full lipid panel is also an integral component of metabolic screening, given that antipsychotics may be associated with hyperlipidemias, which can increase the risk of coronary heart disease (Marder et al. 2004; ADA-APA 2004; Casey 2004). In addition, at least one guideline supported regular blood pressure and BMI measurement for individuals with serious mental illness, due to these measurements’ low cost and high utility in identifying hypertension and obesity, respectively (Marder et al. 2004).

Defining the Numerator Time Frame. Consistent with the ADA-APA guideline (2004) and TEP input, we determined that the measure should require a complete metabolic screening at least once a year for all patients discharged on antipsychotics.¹⁶ To receive credit for the screening, each component must be completed during the index IPF stay or in previous IPF stays or outpatient visits in the 12 months preceding the IPF discharge.¹⁷ If completed at IPFs, this screening could serve as a baseline for patients that began antipsychotics during the IPF stay, or it could serve to monitor for metabolic conditions among patients who were taking antipsychotic medications at the time of IPF admission.

3. Follow-up after IPF Hospitalization Measure

The follow-up measure calculates the proportion of patients that received outpatient mental health care within seven and 30 days following IPF discharge. This measure is calculated using only Medicare claims data. The follow-up measure specification was modeled on the NQF-endorsed Follow-up After Hospitalization for Mental Illness (FUH) measure (NQF #0576), for which NCQA is the steward.

Identification of Data Source. Although the other measures we tested rely mostly on chart data, claims data were the only suitable data source for the follow-up measure at this time. This is because IPFs have very limited access to information regarding their patients’ follow-up care, either in patient charts or administrative data. As such, claims data offer more complete information on patients’ follow-up. However, the primary limitation of claims data is that they are available only for Medicare beneficiaries, who comprise a subset of IPF patients that may not be representative of all patients.

Defining the Denominator. We sought for this measure to be broadly applicable to IPF patients. Based on feedback we received from CMS and an analysis of claims

¹⁶ The ADA-APA guideline mandates that all four tests should be completed before or right after initiation of antipsychotic medications, and completed again at 12 weeks and/or 12 months after initiation.

¹⁷ Screening components could be completed over the course of multiple stays, as long as all components were completed in the 12 months prior to discharge.

data, we limited the denominator to IPF stays with a principal mental health diagnosis.¹⁸ These mental health diagnosis codes are fully aligned with the HEDIS FUH measure. We excluded dually eligible Medicare and Medicaid (“dual”) beneficiaries from the denominator, because Medicaid claims data are not available on a timely basis in each state to examine Medicaid-reimbursed follow-up services for this population, and there may be systematic differences in their access to follow-up care relative to non-dual beneficiaries. However, as a sensitivity analysis for the specification that uses only Medicare data, we tested an alternate version of the measure that tabulates dual beneficiaries’ receipt of follow-up care using merged Medicare and Medicaid data from calendar year 2008.

Defining the Numerator. The numerator for the measure requires an outpatient or partial hospitalization visit with a mental health practitioner, and specifies both a seven-day follow-up rate and a 30-day follow-up rate for each IPF. However, in specifying this measure as a Medicare claims-based measure, we identified and tested alternate numerator options, including an outpatient visit or partial hospitalization with a mental health diagnosis. Testing these alternate numerator specifications allowed us to examine the extent to which IPF performance would change using different numerator options.

¹⁸ IPF stays with a mental health diagnosis comprised 71 percent of all IPF discharges. As such, it was reasonable to limit the measure denominator to these stays, given that they comprised a majority of all fee-for-service (FFS) Medicare-paid stays. Non-mental health diagnoses included Alzheimer’s (8 percent), psychosis (4 percent), senility (3 percent), and other diagnoses (14 percent).

III. RESEARCH METHODS

The testing protocol was designed to assess the psychometric properties and performance of the measures and to gather information to inform their eventual implementation. Moreover, the goal of testing was to gather information about the importance, scientific acceptability, usability, and feasibility of the measures, as defined in the following NQF measure criteria:

1. **Importance.** Strength of evidence supporting the measure concept that promotes high quality care and allows for differentiation in performance.
2. **Scientific Acceptability.** Verification that the psychometric properties of the measure -- validity and reliability -- are strong enough to justify its use to assess quality of care.
 - **Validity.** The correct data elements are included in the measure, and the final measure score promotes correct conclusions regarding measured entities' quality of care.
 - **Reliability.** The ability of measure specifications to promote consistency in data collection and aggregation to ensure that variability in measure score reflects actual variation in performance.
3. **Usability.** The value of the measure in informing quality improvement activities.
4. **Feasibility.** The availability of data elements required for the calculation of the measure, whether the measure is susceptible to human error, and the level of effort involved in collecting and calculating the measure.

The following overarching questions guided measure testing:

- Do the measures assess quality of care and do they address a priority condition? Is there room for improvement and are there gaps in care?¹⁹ (*importance*)
- As specified, can the data elements and measures be calculated consistently (reliability) and capture the intended information? (*validity*)
- Are measure exceptions or exclusions necessary and appropriate? (*validity*)
- Can the measures be calculated accurately and without undue burden? (*feasibility*)

¹⁹ Gaps in care are defined as variation in performance among IPFs or overall less than optimal performance.

- Can stakeholders use performance results for quality improvement and decision making? (*usability*)

In addition to these overarching questions, measure testing answered more specific questions about the denominator and numerator specifications, as described in Table III.1.

TABLE III.1. Quantitative Analyses for Chart-Based Measures			
NQF Criterion	Testing Question(s)	Data Source	Data Analysis
Importance/performance gap	<ul style="list-style-type: none"> • Is there room for improvement on this measure? • Are there differences in performance across IPFs? • Are there differences in performance related to diagnosis and age? 	<ul style="list-style-type: none"> • Abstracted medical records 	<ul style="list-style-type: none"> • Descriptive analysis (for example, mean, median, range) of IPF performance • Tests of differences in general IPF performance and performance for subgroups
Feasibility*	<ul style="list-style-type: none"> • Are the data needed to define the denominator, numerator, and exclusions available? 	<ul style="list-style-type: none"> • Administrative and medical records 	<ul style="list-style-type: none"> • Descriptive analysis
Reliability (inter-rater)	<ul style="list-style-type: none"> • Are the data required for data element and measure calculation comparable when collected by 2 different chart abstractors? 	<ul style="list-style-type: none"> • Data abstracted by 2 abstractors 	<ul style="list-style-type: none"> • Agreement using kappa statistic
Validity (content)*	<ul style="list-style-type: none"> • How does IPF performance vary when screening is required within 1 day versus 3 days? • How does IPF performance vary at the level of screening elements? 	<ul style="list-style-type: none"> • Abstracted medical records 	<ul style="list-style-type: none"> • Analyses to explore the impact of different numerator specifications on performance
Validity (content)*	<ul style="list-style-type: none"> • Do measure exclusions affect performance rates in a substantive way? 	<ul style="list-style-type: none"> • Abstracted medical records 	<ul style="list-style-type: none"> • Sensitivity analyses to explore the impact of calculating the measure without exclusions
* The research team also assessed measure feasibility and validity with qualitative methods, as discussed in the next chapter.			

We used qualitative and quantitative methods to test the four chart-based screening measures and the claims-based measure of follow-up after IPF hospitalization. Quantitative data collection largely informed our analyses of measure validity, reliability, and importance -- namely gaps in care -- whereas qualitative data collection largely informed our analyses of measure validity, feasibility, and usability. Below is a brief summary of these methods.

A. Quantitative Approach to Chart-Based Measure Testing

Quantitative testing of the measures focused on demonstrating the importance of the measures, based on the evidence of performance gaps and disparities in care,

reliability between chart abstractors in obtaining data from patient records, and validity of the specifications, especially the measure numerator and exclusions.

Quantitative testing was divided into five phases: (1) site selection; (2) developing data collection instruments and protocols; (3) IPF staff training; (4) chart-abstraction and data collection; and (5) analysis.

1. **Site Selection.** We recruited a total of six IPFs to participate in testing the measures; this was the maximum number of facilities the project could support while offering a sample size that would allow the detection of variation in measure performance across IPFs. Each IPF was offered \$25,000 as an honorarium to participate.²⁰ Potential partner IPFs were identified through conversations with the CMS and ASPE, existing relationships with IPFs, and other data sources, including HBIPS performance statistics compiled by TJC. After a list of potential partner IPFs was compiled, we attempted to select IPFs that represented a mix of facility types and facility ownership. This included a combination of freestanding facilities and psychiatric wards, as well as public and private IPFs. Site outreach activities occurred from August 2013 to October 2013. The six IPFs selected for testing included three freestanding facilities (two public and one private) and three psychiatric wards (all private); see Table III.2.

TABLE III.2. IPFs Represented in Measure Testing, 2014		
	Freestanding Facilities	Psychiatric Wards
Private facilities	1	3
Public facilities	2	0

2. **Developing Data Collection Instruments and Protocols.** Parallel to conducting IPF recruitment, we developed a data collection tool for chart-abstraction in participating IPFs. We developed a Microsoft Access-based tool that contained all of the necessary data elements to calculate measure performance. The tool had pre-programmed skip logic and error checking to ease the burden of data collection while ensuring high quality data. Abstractors followed the instructions included in the tool's user interface to review each patient record and answer a set of questions about the information provided in it. Abstractors completed one electronic form per patient, which populated a back-end spreadsheet. In addition to collecting data with the chart-abstraction tool, IPFs also extracted administrative data on patient demographics, insurance status, length of stay during the visit selected for abstraction, number of stays during the past year, and other relevant data elements.

All IPFs obtained the appropriate authorizations (in some cases institutional review board approval) to participate in measure testing. IPFs submitted all relevant administrative and abstracted data using a secure password-protected encrypted website accessible only to immediate project staff. Mathematica and NCQA did not

²⁰ One IPF received an additional \$5,000 to offset their time providing additional assistance in the formative stages of designing chart-abstraction materials.

have any direct access to patient medical records and did not receive any personally identifiable information such as patient birthdates, Social Security numbers, or insurance identifiers. Rather, IPFs generated and employed random patient numbers that were not linked to other identifiers. All data were housed on Mathematica's secure servers.

3. ***IPF Staff Training.*** Before chart-abstraction began, all IPFs participated in training sessions that presented the testing methodology, introduced the measures, reviewed the structure and process for completing data collection instruments, and informed IPFs of the global testing timeline. A Mathematica senior researcher who was familiar with IPF services, chart-abstraction, testing protocol, and data collection instruments led the sessions. Participants in the training included chart abstractors, quality improvement staff, and any necessary administrative staff. Abstractors had multiple options to ask for clarification and additional guidance. The project team held biweekly check-in calls with all IPFs or with individual facilities as needed.
4. ***Chart-Abstraction and Data Collection.*** Chart-abstraction took place from January 2014 to April 2014 in all six participating IPFs. During this time, experienced chart abstractors from each IPF abstracted at least 115 patient charts corresponding to one month of discharges in larger facilities and three months of discharges in smaller facilities. Patient charts were randomly sampled from a universe of discharges from previous months, corresponding to October, November, and December 2013 for facilities participating in testing. Following HBIPS-1 sampling procedures, sampling was stratified by the four age strata, which were: (1) younger than age 13; (2) ages 13-17; (3) ages 18-64; and (4) ages 65 and older. This sampling approach was sufficient to detect differences in performance on the measures between the IPFs participating in testing, as well as differences in performance by age group and other patient characteristics.

Two staff at each testing site conducted chart-abstraction: a primary abstractor, who collected data from all sampled charts, and a secondary abstractor, who collected data from a subset of ten charts to allow for assessment of inter-rater reliability.²¹ During the first week of testing at all six IPFs, primary and secondary abstractors each abstracted ten charts and then reviewed them with the research team. During this review, the team discussed any discrepancies between the primary and secondary abstractors, and reached consensus regarding the correct abstraction of records. This review allowed us to ensure that the abstractors understood the specifications and data collection protocol before proceeding with full record abstraction.

5. ***Chart-Based Measure Analysis.*** We used the data from chart-abstraction and administrative data sources to summarize the demographic characteristics of the population, analyze IPF performance on each measure, examine performance

²¹ There was no abstractor turnover of during the course of testing, with the exception of one clinical resident who could not abstract all sampled charts and was assisted by another trained abstractor.

rates for subgroups, determine the sensitivity of performance rates to alternate numerator specifications and exclusions, and calculate inter-rater reliability. Each analysis was designed to investigate one or more issues related to the importance, reliability, and validity of the measures. We discuss these analyses in more depth below:

- *Validity.* To estimate the measures' validity, we conducted tests to determine whether measure exceptions altered performance rates, and how alternate numerator specifications altered IPF performance rates. In particular, we tested how performance varied across individual screening elements. In addition, we disaggregated alcohol and drug use screening rates within the substance use screening measure to assess potential differences in IPF performance between rates. We also conducted some validity and reliability tests for the alcohol and drug components separately; we report the results of these tests in Chapter IV.
- *Room for Improvement/Performance Gap.* We calculated a score for each IPF, as well as the average across IPFs, to determine if there was room for improvement and variation in performance rates. We also explored performance rates by patient diagnosis and age to determine if disparities in screening for specific subpopulations were present.
- *Measure Reliability.* As described above, this analysis answered the question of whether the data collected by two abstractors at the same IPF were comparable. To do this, we generated kappa statistics, or indicators of the measures' inter-rater reliability.
- *Comparison with Existing Measures.* During measure testing, we also compared IPFs' performance on the newly specified screening measures to their performance on the relevant components of HBIPS-1 and SUB-1. The goal of these comparisons was to determine the extent to which stronger requirements regarding screening elements and screening time frames (within one day versus three days of admission) would affect IPFs' performance on violence, suicide, and substance use screening, and the extent to which the requirement of screening with a validated instrument would affect IPFs' performance on alcohol screening.

B. Quantitative Approach to Claims-Based Follow-Up Measure Testing

Our primary follow-up measure analysis used Medicare claims to assess facility performance and measure reliability. However, we also assessed facility performance using a file that contained Medicare and Medicaid Analytic eXtract (MAX) claims data. Although the IPFQR program is a Medicare quality reporting program, over half of patients discharged from IPFs are eligible for both Medicare and Medicaid, according to

information from MedPac (2012). Because these “dual eligible” beneficiaries could access additional outpatient mental health services through Medicaid, using only Medicare data to assess measure performance may undercount these beneficiaries’ receipt of follow-up care.²² Therefore, we created a file that linked Medicare and MAX data at the beneficiary level to enable more accurate calculation of the measure for dual eligible beneficiaries.

Quantitative testing for the follow-up measure was divided into five components: (1) preparation, including development of a data use agreement (DUA); (2) receipt and preparation of claims data; (3) descriptive analyses and detailed data review; (4) performance and reliability analyses; and, (5) supplementary analysis of a chart-based versus claims-based approach to measuring follow-up care.

1. **Preparation.** We initiated a DUA to obtain 2008 Medicare claims and MAX data. We used data from 2008 because this was the latest year of MAX data available at the contract start date. We obtained DUA approval and received these data in early 2013.
2. **Preparation and Analysis of Claims Data.** We linked the Medicare and MAX claims at the beneficiary level. We followed the linking protocol developed by Prael et al. (2009) to link Medicaid and Medicare databases for dual eligible beneficiaries. This protocol used health insurance claim numbers, patient gender, and date of birth.

In addition, we investigated the completeness of MAX data to determine whether any states should be excluded from the dual eligible beneficiary analysis due to potential missing data. MAX data contains information on FFS Medicaid encounters in all states and managed care Medicaid encounters in some states. This analysis of dual eligible beneficiaries is limited to FFS Medicaid data because managed care encounters are not reliably captured in MAX data for every state.²³ Thus, if a substantive proportion of dual eligible beneficiaries in a state are enrolled in Medicaid managed care, these analyses would likely underestimate the receipt of outpatient care. To avoid this potential bias, states with more than 25 percent of dual eligible beneficiaries enrolled in Medicaid managed care were excluded from the analyses. In addition, states that did not have complete 2008 MAX data -- generally related to the availability of data elements to identify mental health practitioners -- were excluded from the dual eligible beneficiary analysis. In total, we excluded 24 states from our analysis of MAX claims, leaving an analysis sample of 26 states for the dual eligible

²² Medicare Part B provides limited coverage of outpatient mental health services. In contrast, depending on the state, Medicaid covers long-term support services and community-based mental health care.

²³ Although the variables in MAX are standardized to create comparable measures of service use across states, the data on beneficiaries enrolled in managed care is often incomplete; this is because data submitted by MCOs do not undergo the same review as data submitted for FFS beneficiaries.

beneficiary analysis.²⁴ Additional details of this state selection process are provided in Appendix E.

TABLE III.3. Quantitative Analysis of Follow-Up after IPF Hospitalization Measure			
Analysis	Testing Question(s)	Data Source	Data Analysis
Facility performance	<ul style="list-style-type: none"> • Is there is room for improvement in measure performance? • Does measure performance indicate gaps in care, as measured by high variation in performance? • Are there meaningful differences in measure performance related to patient or facility characteristics? • Do alternate approaches to calculating the measure numerator or exclusions generate substantially different performance rates? 	<ul style="list-style-type: none"> • Medicare and Medicaid claims data 	<ul style="list-style-type: none"> • Use claims data to conduct descriptive analyses of measure performance and explore the impact of alternate numerators and measure exclusions. • Numerical and graphical summaries of variation in performance according to beneficiary and facility characteristics • Summaries of variation in performance according to alternate numerator calculation and exclusion schemes

3. **Performance Analyses.** After generating descriptive statistics and analyzing MAX data completeness, we calculated measure performance for each IPF -- expressed as a rate (percentage) for each facility -- and tested the reliability of the follow-up measure. Table III.3 presents the details of each of these analyses, including relevant testing questions and a brief description of the approach.

- *Testing Alternative Numerators.* We explored IPF performance on the measure using four alternative methods of calculating the numerator:
 - The first method follows the original HEDIS specification, which defines an outpatient mental health visit as a visit to a mental health practitioner (specified using provider specialty or National Provider Identification [NPI] codes) for a specific mental health treatment (specified using Current Procedural Terminology [CPT] and Healthcare Common Procedure Coding System [HCPCS] codes).
 - The second method defines an outpatient mental health visit as the presence of a designated mental health CPT/HCPCS code in combination with a primary mental health diagnosis code. Unlike the first calculation method, visits with non-mental health providers count toward the numerator under this method.
 - The third method defines outpatient mental health care as *any* outpatient visit with a primary mental health diagnosis code, regardless of the provider or specific procedural code.

²⁴ States with more than 25 percent of dual eligible beneficiaries enrolled in Medicaid managed care included Arizona, Colorado, Hawaii, Iowa, Kansas, Maine, Minnesota, Montana, Oregon, Pennsylvania, Rhode Island, Tennessee, Utah, and Vermont. States with data completeness issues included Alabama, District of Columbia, Georgia, Idaho, Kentucky, Mississippi, Nevada, Ohio, South Dakota, Washington, and Wyoming.

- The fourth method defines follow-up care as any outpatient visit, regardless of diagnosis. This option is not a viable measure of outpatient mental health care; rather, it was calculated to provide context for the other numerator options.

These numerators were developed in consultation with CMS, ASPE, and claims measurement experts. The primary rationale for the second numerator option was to determine the viability of identifying follow-up care using patient diagnosis, as opposed to provider specialty. The primary rationale for the third numerator option was to test specifications of follow-up care that could be accurately measured with Medicare as well as Medicaid claims, given the high proportion of dual eligible beneficiaries using IPFs. All of these options are feasible with Medicare data, but the third option may be the most feasible with Medicaid data, given the inconsistency of Medicaid data across states.²⁵

- *Sensitivity Tests, Disparities, and Room for Improvement.* Next, we tested the sensitivity of measure performance to proposed exclusions. Most of the proposed exclusions are related to admission to an IPF or another inpatient setting within the follow-up period, as this admission could preclude beneficiaries' access to outpatient care. In addition, we examined facility performance according to various facility and beneficiary characteristics including geographic location and size of IPFs, as well as patients' principal diagnosis, type of insurance coverage, age, gender, and race.
- *Comparing Medicare-Only Rates to Merged Medicare-Medicaid Rates.* Another key sensitivity test was determining the extent to which performance rates were affected by excluding Medicaid claims from the calculation. For this test, we compared performance rates calculated using linked Medicare and Medicaid claims to performance rates calculated using only Medicare claims data.²⁶ The extent to which these rates differed provided insight into the additional follow-up care captured through Medicaid claims.
- *Testing Variation in Performance.* In addition, we examined the distribution of facilities' performance. We calculated the minimum, maximum, median, mean, and interquartile range (IQR) for the follow-up measure. The IQR is demarcated by the values at the 25th and 75th percentiles of a distribution. Generally speaking, measures with a broader IQR are preferable to those with a narrowly distributed IQR or those with an IQR at the very low or very high end of the distribution. Based on our past experience with quality

²⁵ Notably, there is heterogeneity in Medicaid data in states' use of CPT/HCPCS codes as well as provider codes. Because options 1 and 2 rely on these codes, using Medicaid claims alone would likely underestimate the receipt of follow-up care in states that do not use these codes in a manner that is consistent with Medicare claims.

²⁶ This analysis used a fixed population of Medicare beneficiaries to ensure that measured differences in performance were the result of supplementing Medicare claims with Medicaid claims.

measure testing, we consider measures with an IQR of at least 10 percentage points to have the strongest evidence of importance for quality measurement purposes.

4. **Reliability Analyses.** Next, we tested the reliability of the follow-up measure. This involved a beta-binomial test and an analysis of the stability of IPF performance across quarters.
 - *Beta-Binomial Test.* We conducted a beta-binomial test to examine how well the measure as specified can distinguish performance between IPFs (the ratio of signal to noise). The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in IPF performance. The beta-binomial approach is appropriate for measures like this one, where each denominator event represents a binary opportunity to pass or fail the measure (Adams 2009). The approach assumes that the performance measure score (pass/fail rate) across IPFs has a flexible beta distribution, characterized by a signal variance. Based on the performance measure score, the observed data (number of passes/failures) for each IPF has a binomial distribution, which provides the noise (measurement error) variance. From the beta-binomial model, the signal and noise variances are used to calculate reliability as follows: $\text{Signal variance} / (\text{signal} + \text{noise variance})$.
 - *Stability of IPF Performance Across Quarters.* We also examined the stability of facility performance over three quarters during 2008. We compared each facility's performance quartile in the first quarter with its performance in the other quarters, and examined whether facilities remained in the same quartile throughout all three quarters. In addition, we examined these changes in performance by facility size in an effort to determine whether large facilities were less likely to experience shifts in performance from one quarter to the next.
5. **Supplementary Analysis of the IPF Follow-Up Measure.** In March 2013, ASPE and CMS expressed interest in exploring the relative merits and drawbacks of a claims-based approach versus a chart-based approach to measuring follow-up care after IPF hospitalization. This included an analysis of patient characteristics, sample sizes, IPF capacity, and data availability related to patient follow-up. To explore this issue in more depth, we used administrative data provided by the six IPFs that participated in measure testing to conduct some preliminary quantitative analysis on the potential implications of insurance coverage, patient demographics, and sample size constraints associated with a chart-based versus claims-based approach to follow-up care. We present the results of this analysis in Chapter VII.

C. Qualitative Approach to Measure Testing

Qualitative testing spanned chart-based and claims-based measures, and comprised four components: (1) debriefing with IPFs; (2) focus groups with stakeholders; (3) TEP consultation; and (4) consultation with TJC.

1. **Debriefings with IPFs.** In mid-2014, we held a debriefing meeting with each of the six IPFs after we calculated facility performance on the measures. These debriefing sessions provided us with an opportunity to share preliminary results with IPFs, assess the total amount of time and effort associated with collecting the data, and to document the IPFs' final conclusions and perspectives on the measures. We also used these debriefings as an opportunity to gather input for the supplementary follow-up analysis discussed above. During these conversations, we gathered stakeholder feedback on IPFs' efforts to promote and track follow-up care, including factors that facilitate recording accurate data on patients' care following their IPF stay. In addition, we asked stakeholders about the feasibility of a chart-based approach to the follow-up measure, including data collection and reporting burden, as well as infrastructure and resources that would be necessary to support reporting.
2. **Focus Groups with Stakeholders.** In late 2014, we held focus groups with additional stakeholders including quality measure experts, consumers/advocacy organizations, state policymakers, and IPFs that did not participate in our chart-abstraction work. Below is a brief description of the stakeholder groups involved in focus groups and their value in providing feedback on the measures.
 - **Quality Measurement Experts.** Measurement experts provided feedback on the measure specifications, strength of evidence supporting the measures, and practical considerations in implementing the measures.
 - **Consumers and Advocacy Organizations.** Consumers and behavioral health advocacy organizations provided feedback on the saliency of measure concepts and the usefulness of performance on the measures for decision making and improving the quality of care.
 - **State Policymakers.** Although IPFQR program is a Medicare quality reporting program, the existing IPFQR measures (and those under development in this contract) are reported for both Medicare and Medicaid beneficiaries; therefore, state Medicaid agencies have an interest in performance on these measures. State mental health and substance abuse agencies also have an interest in the performance, given that IPFs play a central role in the state mental health system and some IPFs are state-operated. These stakeholders provided insight into the importance and usability of the measures within the larger context of the mental health service system.

- *IPF Representatives.* In part due to the diversity of IPF resources and services at the national level, IPF representatives have varying perspectives on the proposed measures. Holding separate focus groups for different types of IPFs provided critical insight into the feasibility, usability, and importance of each measure from the perspective of each type of facility. We recruited a mix of freestanding facilities and psychiatric units within general medical hospitals for these focus groups. During these focus groups, we asked IPF staff about the feasibility of a chart-based approach to the follow-up measure, including data collection and reporting burden, as well as infrastructure and resources that would be necessary to support reporting.
 - *IPFQR Program Vendors.* Vendors contracted by IPFs to assist with IPFQR program reporting have an in-depth understanding of reporting burden and IPF capacity to sample patients, abstract patient records, and aggregate relevant information for measure reporting.
3. ***TEP Consultation.*** We met with the TEP throughout the testing process. Three key TEP meetings took place from 2012 to 2014 with regard to measure testing. In late 2012, we met with the TEP to secure their approval for the screening measure concepts and their guidance regarding the measure specifications. In late 2013, we shared our specifications with the TEP to obtain their approval on all measures before testing began. In late 2014, we shared our final performance and reliability estimates with the TEP once testing was complete. This meeting allowed the TEP to conduct a final analysis of the measures' properties -- including their importance, reliability, validity, usability, and feasibility -- with the full set of testing results. (See Appendix A for a list of all TEP members.)
 4. ***Consultation with TJC.*** Similar to consultations with the TEP, we consulted with TJC staff at three key points in the testing process. First, we shared our specifications with TJC in an August 2013 meeting, before pilot testing began. Next, we shared our final performance and reliability estimates with TJC staff once analysis was complete in late 2014. In addition, we shared testing results with TJC's technical advisory panel (TAP) for the HBIPS measure set in September 2014. This ongoing communication with TJC and its TAP was intended to share testing results among interested parties, particularly given commonalities between the new screening measures and TJC's HBIPS-1 and SUB-1 measures.

IV. TESTING RESULTS FOR SCREENING MEASURES

This chapter summarizes the quantitative and qualitative results of measure testing for the admission screening measures -- suicide risk, violence risk, and substance use screening -- as well as the metabolic screening measure. Section A describes how the measure denominator population was selected, and summarizes patient characteristics in the IPFs that participated in testing. Section B and Section C summarize measure performance, inter-rater reliability, and stakeholder feedback on the measures. Section D discusses changes to final measure specifications based on testing results.

A. IPFs and Denominator Population

The six IPFs that participated in measure testing are diverse in both structure and size. Three IPFs are private psychiatric wards with fewer than 50 patient beds, two are public freestanding facilities with over 100 beds, and one is a private freestanding facility with 400 beds. The IPFs are located throughout the country, with representation in the Mid-Atlantic, Northeast, Mid-West, South, and West. Four of the six IPFs employ a vendor to tabulate and report quality measures for the IPFQR program.

A total of 1,857 patients were discharged from the six IPFs between October 1, 2013, and December 31, 2013 (with the exception of one large IPF [IPF 6], for which only one month's data were necessary, corresponding to December 2013).²⁷ We implemented an approach that mirrored the HBIPS sampling procedure, sampling at least 20 percent of patients in each of four age strata, with a minimum of 120 sampled patients in each IPF. This resulted in a total of 825 patients selected for manual chart-abstraction across the six IPFs, with sample sizes ranging from 120 patients in IPFs 1 and 2 to 176 patients in IPF 3.

As illustrated in Table IV.1, the suicide, violence, and substance use screening measures used the full denominator of 825 sampled patients. In contrast, the metabolic screening measure included only sampled patients discharged on antipsychotic medications, resulting in a smaller denominator for the six sites combined (506 patients; 61 percent of patients in the full sample).

²⁷ This IPF supplied only one month of discharge data because one month's data were sufficient to meet the minimum chart-abstraction requirement of 120 patient records.

TABLE IV.1. Number of Discharged and Sampled Patients, by IPF					
	Total Discharges	Final Sample Size for Suicide, Violence, and Substance Use Screening		Final Sample Size for Metabolic Screening	
		N	% of Total Discharges	N	% of Total Discharges
All IPFs	1857	825	44.4	506	27.2
IPF 1 (private psychiatric unit)	172	120	69.8	56	32.6
IPF 2 (private psychiatric unit)	382	176	46.1	118	30.9
IPF 3 (private psychiatric unit)	409	120	29.3	85	20.8
IPF 4 (public freestanding hospital)	152	120	78.9	79	52.0
IPF 5 (public freestanding hospital)	272	118	43.4	78	28.7
IPF 6* (private freestanding hospital)	470	171	36.4	90	19.1
SOURCE: Administrative data and medical records from 6 IPFs, corresponding to all discharges from October 1, 2013, to December 31, 2013.					
* Discharges in IPF 6 are all discharges from December 1 to December 31, 2013.					

The overall patient population used for measure testing was diverse in terms of age, gender, race, diagnosis, and type of insurance coverage (Table IV.2). The majority of patients were between the ages of 27 and 64 at discharge, with approximately 17 percent of patients over age 64. Roughly half were female. The most common primary diagnoses were bipolar disorder (42 percent across IPFs) and schizophrenia (19 percent across IPFs). Patients were most likely to be insured with Medicare FFS coverage (23 percent across IPFs), private or commercial insurance (21 percent across IPFs), or state or county insurance (17 percent across IPFs).

We observed important differences in patient characteristics between IPFs, particularly with respect to patient length of stay, ethnicity, diagnosis, and age. Notably, there were pronounced differences in length of stay between psychiatric units (IPFs 1, 2, and 3) and freestanding hospitals (IPFs 4, 5, and 6) involved in measure testing. Average length of stay at freestanding hospitals ranged from 15 to 82 days, whereas average length of stay at psychiatric units ranged from five to nine days.²⁸ In addition, over 80 percent of patients in two psychiatric units were White, compared to less than 50 percent of patients at two public freestanding hospitals. Also, a larger share of psychiatric units' patients had primary diagnoses of bipolar disorder (38-76 percent) relative to freestanding hospitals (18-25 percent). Also notable is that over 75 percent of patients in IPF 4 were male; a large portion of these patients were transferred directly from correctional facilities whose populations were predominantly male.

²⁸ Longer lengths of stay among freestanding hospitals (compared to psychiatric units) likely reflects these facilities' functions and patient populations. In general, freestanding hospitals serve patients with chronic mental health conditions who are admitted for an extended period of time, whereas psychiatric units have a larger proportion of patients who require short-term stabilization.

TABLE IV.2. Characteristics of Denominator Population for Screening Measures, by IPF

Patient Characteristics	All IPS		Psychiatric Units						Freestanding Hospitals					
			IPF 1 (private)		IPF 2 (private)		IPF 3 (private)		IPF 4 (public)		IPF 5 (public)		IPF 6 (private)	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Age at the Time of Discharge														
Average (in years)	42.1		38.5		51.7		50.3		36.7		38.0		35.6	
Younger than 13	23	2.8	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	23	13.5
13-17	39	4.7	0	0.0	0	0.0	0	0.0	6	5.0	0	0.0	33	19.3
18-26	146	17.7	22	18.3	32	18.2	23	19.2	27	22.5	23	19.5	19	11.1
27-44	276	33.5	59	49.2	35	19.9	28	23.3	55	45.8	58	49.2	41	24.0
45-64	203	24.6	36	30.0	44	25.0	29	24.2	30	25.0	35	29.7	29	17.0
65 and older	138	16.7	3	2.5	65	36.9	40	33.3	2	1.7	2	1.7	26	15.2
Primary Diagnosis														
Bipolar disorder	346	41.9	76	63.3	133	75.6	45	37.5	22	18.3	27	22.9	43	25.1
Schizophrenia	154	18.7	14	11.7	11	6.3	44	36.7	38	31.7	38	32.2	9	5.3
Alcohol/drug	68	8.2	18	15.0	2	1.1	3	2.5	14	11.7	15	12.7	16	9.4
Psychosis	61	7.4	2	1.7	3	1.7	10	8.3	22	18.3	15	12.7	9	5.3
Major depressive disorder	45	5.5	2	1.7	5	2.8	3	2.5	3	2.5	6	5.1	26	15.2
Alzheimer's/dementia	34	4.1	0	0.0	17	9.7	7	5.8	2	1.7	0	0.0	8	4.7
Delusional disorder	2	0.2	0	0.0	1	0.6	0	0.0	0	0.0	1	0.8	0	0.0
Other	86	10.4	8	6.7	2	1.1	8	6.7	19	15.8	15	12.7	34	19.9
Missing	29	3.5	0	0.0	2	1.1	0	0.0	0	0.0	1	0.8	26	15.2
Race/Ethnicity														
Caucasian	541	65.6	106	88.3	145	82.4	75	62.5	56	46.7	50	42.4	109	63.7
African American	217	26.3	4	3.3	23	13.1	13	10.8	61	50.8	63	53.4	53	31.0
Other	23	2.8	3	2.5	8	4.5	5	4.2	1	0.8	5	4.2	1	0.6
Missing	44	5.3	7	5.8	0	0.0	27	22.5	2	1.7	0	0.0	8	4.7
Gender														
Female	415	50.3	65	54.2	113	64.2	74	61.7	29	24.2	50	42.4	84	49.1
Male	410	49.7	55	45.8	63	35.8	46	38.3	91	75.8	68	57.6	87	50.9
Insurance Coverage														
Medicare FFS	192	23.3	20	16.7	73	41.5	35	29.2	33	27.5	15	12.7	16	9.4
Private/commercial insurance	172	20.8	46	38.3	63	35.8	14	11.7	6	5.0	0	0.0	43	25.1
State/county	139	16.8	0	0.0	0	0.0	0	0.0	66	55.0	0	0.0	73	42.7
Medicaid FFS	53	6.4	0	0.0	12	6.8	19	15.8	0	0.0	22	18.6	0	0.0
Self-pay	41	5.0	0	0.0	25	14.2	9	7.5	0	0.0	0	0.0	7	4.1
Medicare managed care	36	4.4	0	0.0	0	0.0	11	9.2	0	0.0	0	0.0	25	14.6
Medicaid managed care	32	3.9	0	0.0	0	0.0	20	16.7	12	10.0	0	0.0	0	0.0
Other	46	5.6	23	19.2	3	1.7	12	10.0	1	0.8	0	0.0	7	4.1
Missing*	114	13.8	31	25.8	0	0.0	0	0.0	2	1.7	81	68.6	0	0.0

TABLE IV.2 (continued)														
Patient Characteristics	All IPS		Psychiatric Units						Freestanding Hospitals					
			IPF 1 (private)		IPF 2 (private)		IPF 3 (private)		IPF 4 (public)		IPF 5 (public)		IPF 6 (private)	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Length of Stay														
Average (in days)	29.0		5.1		9.1		7.9		81.8		70.8		15.2	
0 or 1 days	41	5.0	12	10.0	3	1.7	13	10.8	2	1.7	8	6.8	3	1.8
2 or 3 days	125	15.2	40	33.3	13	7.4	38	31.7	8	6.7	20	16.9	6	3.5
4 to 7 days	249	30.2	45	37.5	60	34.1	38	31.7	18	15.0	40	33.9	48	28.7
8 to 14 days	216	26.2	20	16.7	75	42.6	15	12.5	14	11.7	30	25.4	62	36.3
15 to 21 days	62	7.5	2	1.7	20	11.4	4	3.3	9	7.5	7	5.9	20	11.7
22 to 30 days	34	4.1	1	0.8	5	2.8	5	4.2	9	7.5	1	0.8	13	7.6
>30 days	98	11.9	0	0.0	0	0.0	7	5.8	60	50.0	12	10.2	19	11.1
Total Patients	825	100.0	120	100.0	176	100.0	120	100.0	120	100.0	118	100.0	171	100.0
SOURCE: Administrative data and medical records from 6 IPFs, corresponding to all discharges from October 1, 2013, to December 31, 2013.														
* The high proportion of missing data on patient insurance in IPF 5 was due to hospital staff's inability to fully merge administrative files with financial records.														

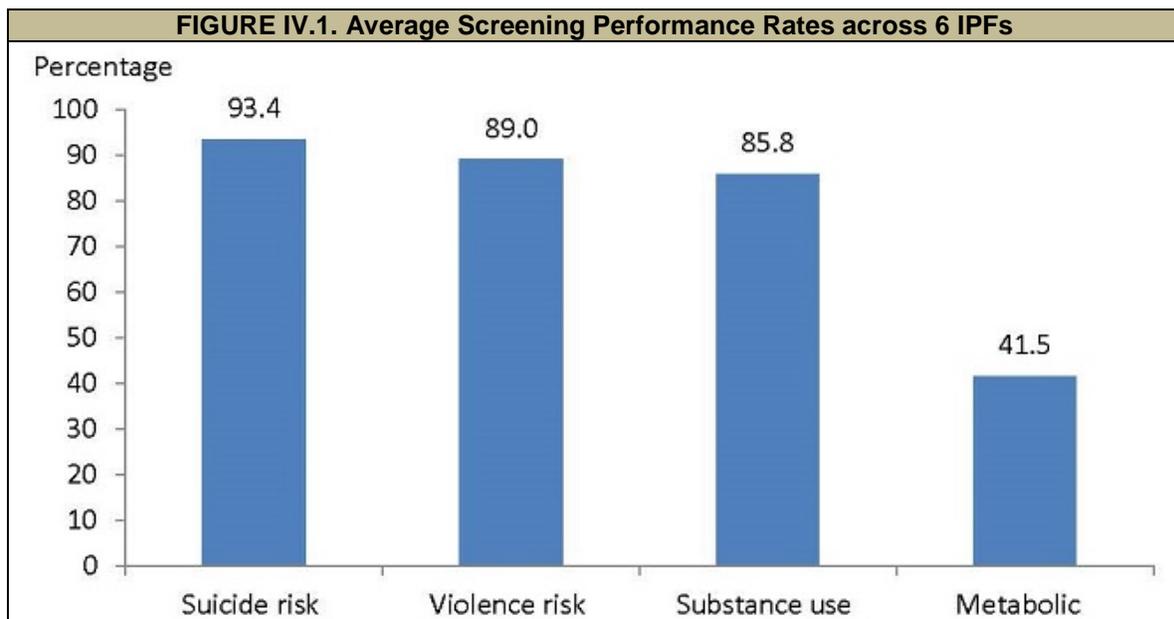
B. Quantitative Testing Results

This section provides quantitative testing results for the suicide, violence, substance use, and metabolic screening measures. In addition to presenting overall performance statistics, we describe any variation in performance across facilities, as well as IPF performance on individual screening elements. Performance rates discussed below are the proportion of patients that met each measure numerator requirement, after accounting for exclusions.

Screening Measure	All IPS	Psychiatric Units			Freestanding Hospitals		
		IPF 1 (private)	IPF 2 (private)	IPF 3 (private)	IPF 4 (public)	IPF 5 (public)	IPF 6 (private)
		%	%	%	%	%	%
Suicide risk	93.4	67.6	97.1	98.1	98.2	94.9	99.4
Violence risk	89.0	47.7	97.1	86.8	99.1	94.2	98.8
Substance use	85.8	51.4	89.6	80.4	94.4	94.8	96.4
Metabolic	41.5	16.3	13.9	51.7	98.6	6.2	59.8

SOURCE: Administrative data and medical records from 6 IPFs, corresponding to all discharges from October 1, 2013, to December 31, 2013.

Table IV.3 and Figure IV.1 provide an overview of IPF performance on the screening measures. As illustrated, IPF performance on the admission screening measures (suicide, violence, and substance use) was generally strong, with average performance above 85 percent on all three measures. In contrast, performance on the metabolic screening measure was generally poor, at 42 percent across IPFs on average.



Admission Screening Measures. In general, IPFs that performed well on one admission screening measure also performed well on the other two measures. One private freestanding facility (IPF 6) had the highest or second-highest performance on

all three admission screening measures, with performance ranging from 96 percent to 99 percent. Four other facilities, representing a mix of freestanding hospitals and psychiatric units, did consistently well on admission screening measures, with performance in the 80 percent to 99 percent range for all three measures. One private psychiatric unit (IPF 1) had the lowest performance on all three measures, with performance ranging from 48 percent to 68 percent. If this IPF were excluded from the analysis, average performance across the remaining five IPFs would be 97.7 percent for suicide, 95.7 percent for violence, and 91.5 percent for substance abuse. (See Table D.1 in Appendix D for additional details.)

Metabolic Screening Measure. IPF performance on the metabolic screening measure exhibited similar trends, in that IPF 6 had relatively high performance (at 60 percent) and IPF 1 had relatively low performance (at 16 percent). However, a public freestanding IPF (IPF 5) had the lowest performance (at 6 percent) and another public freestanding IPF (IPF 4) had the highest performance (at 99 percent). Consistent with these findings, clinical staff from the lowest-performing site noted that they do not perform a full metabolic screening on their patients as a general practice, whereas staff from the highest performing IPF noted that they provide all their patients -- regardless of their use of antipsychotic medications -- with a full metabolic screening at least once during their stay.

1. Suicide Risk Screening

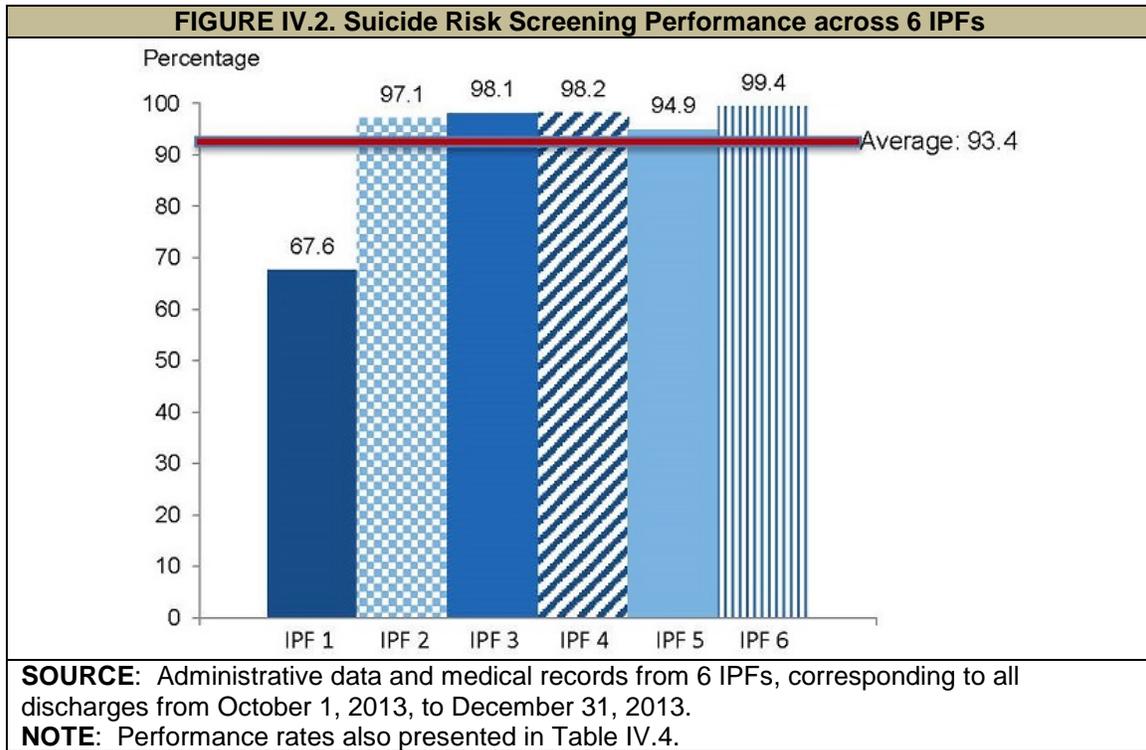
In order for the IPF to meet the numerator requirement for the suicide screening measure, the medical record must contain documentation that all five screening elements were completed within one day of admission -- either through use of a validated screening tool or a non-validated assessment protocol.²⁹ These screening elements are listed below.

Suicide Risk Screening Numerator Requirement
The medical record must provide documentation that all five of the following screening elements were completed within one day of admission:
<ol style="list-style-type: none">1. Presence or absence of suicidal ideation.2. Extent of plans or preparation (if ideation is reported).3. Intent to act on those plans (if plans are reported).4. Any history of suicidal behavior.5. Risk and protective factors related to suicide.

Overall Performance. Five of the six IPFs had relatively high performance on the suicide screening measure, with performance rates ranging from 95 percent to 99 percent of patients screened within one day of admission (Figure IV.2 and Table IV.4). However, one psychiatric unit (IPF 1) had a relatively poor performance rate of 68 percent. In a debriefing session following chart-abstraction, clinical staff from this

²⁹ Documented full completion of the SAFE-T constituted a full assessment, as it covered all five screening elements.

psychiatric unit noted that they consistently ask patients about each of these suicide screening elements upon admission. However, patient records do not systematically reflect the information provided by patients on these elements, thus generating relatively poor performance on the measure. Aggregating performance among all six IPFs, average performance on the measure was 93 percent.



Performance on Individual Elements. Average performance on individual screening elements was relatively uniform and high across IPFs, ranging from 97 percent completion for past suicidal behavior to 98 percent completion for suicidal ideation. However, one IPF (IPF 1) had relatively low performance on two screening elements: 82 percent completion of past suicidal behavior, and 90 percent completion of intent to act on suicide plans. These element-level performance rates contributed to the IPF’s relatively low performance rate on the measure of 68 percent.

Use of Standard Screening Tools. Standard instruments were used to conduct suicide risk screening for only 5 percent of patients across all IPFs. For these patients, IPFs generally used the SAFE-T or the Suicide Screening Risk Inventory (SSRI) tool.

TABLE IV.4. Suicide Risk Screening Performance, by IPF

	All IPS		Psychiatric Units						Freestanding Hospitals					
			IPF 1 (private)		IPF 2 (private)		IPF 3 (private)		IPF 4 (public)		IPF 5 (public)		IPF 6 (private)	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Denominator (after exclusions)	761	100.0	108	100.0	172	100.0	106	100.0	109	100.0	99	100.0	167	100.0
Overall rate (met numerator requirement)	711	93.4	73	67.6	167	97.1	104	98.1	107	98.2	94	94.9	166	99.4
1 element missing	26	3.4	21	19.4	1	0.6	2	1.9	1	0.9	0	0.0	1	0.6
2 or more elements missing	24	3.2	14	13.1	4	2.4	0	0.0	1	0.9	5	5.0	0	0.0
Screening elements met within 1 day of admission^a														
Suicidal ideation	749	98.4	103	95.4	168	97.7	106	100.0	108	99.1	97	98.0	167	100.0
The extent of plans or preparation	748	98.3	100	92.6	170	98.8	106	100.0	108	99.1	97	98.0	167	100.0
The intent to act on those plans	745	97.9	97	89.8	171	99.4	106	100.0	108	99.1	97	98.0	166	99.4
Past suicidal behavior	734	96.5	89	82.4	169	98.3	105	99.1	109	100.0	95	96.0	167	100.0
Risk factors and protective factors	738	97.0	94	87.0	169	98.3	105	99.1	108	99.1	95	96.0	167	100.0
Standard screening tools administered within 1 day of admission^b														
SAFE-T	18	2.4	0	0.0	0	0.0	0	0.0	18	16.5	0	0.0	0	0.0
SSRI	17	2.2	0	0.0	0	0.0	0	0.0	17	15.6	0	0.0	0	0.0
Any standard tool	35	4.6	0	0.0	0	0.0	0	0.0	35	32.1	0	0.0	0	0.0

SOURCE: Administrative data and medical records from 6 IPFs, corresponding to all discharges from October 1, 2013, to December 31, 2013.

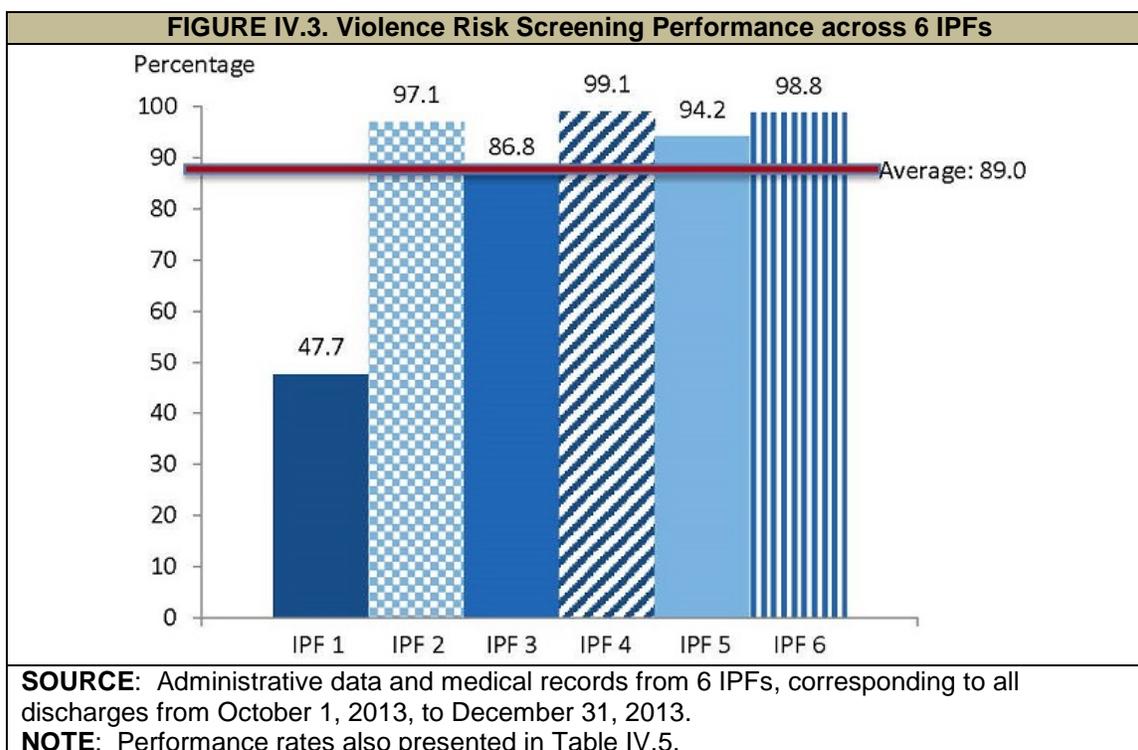
- a. Elements were met if the element was covered in a validated screening tool, a non-validated tool, or through a non-structured screening or assessment. Documented full completion of the SAFE-T constituted a full assessment, as it covered all 5 screening elements. IPFs did not change their screening or admission processes to test these measures. Skip logic was employed for the second and third screening elements: credit was automatically given for extent of plans and intent to act on plans if patients reported no ideation.
- b. Choice of screening tools was at the discretion of the clinician or IPF.

2. Violence Screening

To meet the numerator requirement for the violence screening measure, the medical record must contain documentation that both screening elements were completed within one day of admission -- either through use of a validated screening tool or a non-validated assessment protocol.³⁰ These screening elements are listed below.

Violence Risk Screening Measure Numerator Requirement	
The medical record must provide documentation that both of the following screening elements were completed within one day of admission:	
1.	Presence or absence of threats of violence.
2.	Any history of violence.

Overall Performance. Overall, 89 percent of patients in the denominator received a full violence risk screening. Performance varied across IPFs; four IPFs had performance rates ranging from 94 percent to 99 percent, one IPF had a performance rate of 87 percent, and one IPF had a performance rate of only 48 percent (Figure IV.3 and Table IV.5). In most cases of incomplete screenings, only one of the two required elements was missing in the patient record.



³⁰Documented full completion of the V-RISK-10 constituted a full assessment, as it contained both required screening elements.

TABLE IV.5. Violence Risk Screening Performance, by IPF

	All IPS		Psychiatric Units						Freestanding Hospitals					
			IPF 1 (private)		IPF 2 (private)		IPF 3 (private)		IPF 4 (public)		IPF 5 (public)		IPF 6 (private)	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Denominator (after exclusions)	765	100.0	107	100.0	173	100.0	106	100.0	109	100.0	103	100.0	167	100.0
Overall rate (met numerator requirement)	681	89.0	51	47.7	168	97.1	92	86.8	108	99.1	97	94.2	165	98.8
1 element missing	66	8.6	49	45.8	0	0.0	14	13.2	0	0.0	1	1.0	2	1.2
Both elements missing	18	2.4	7	6.5	5	2.9	0	0.0	1	0.9	5	4.9	0	0.0
Screening elements met within 1 day of admission^a														
Threats of violence	731	95.6	98	91.6	168	97.1	92	86.8	108	99.1	98	95.1	167	100.0
Any history of violent episodes	697	91.1	53	49.5	168	97.1	106	100.0	108	99.1	97	94.2	165	98.8
Standard screening tools administered within 1 day of admission^b														
V-RISK-10	27	3.5	0	0.0	0	0.0	0	0.0	27	24.8	0	0.0	0	0.0
BVC	1	0.1	0	0.0	0	0.0	0	0.0	1	0.9	0	0.0	0	0.0
Any standard tool	27	3.5	0	0.0	0	0.0	0	0.0	27	24.8	0	0.0	0	0.0
SOURCE: Administrative data and medical records from 6 IPFs, corresponding to all discharges from October 1, 2013, to December 31, 2013.														
NOTE: Totals for individual tools do not always sum to any standard tool, given that more than 1 tool could have been used for a patient.														
a. Requirements were met if the element was covered in a validated screening tool, a non-validated tool, or through a non-structured screening or assessment. Documented full completion of the V-RISK-10 constituted a full assessment, as it contained both required screening elements. IPFs did not change their screening or admission processes to test these measures.														
b. Choice of screening tools was at the discretion of the clinician or IPF.														

Performance on Individual Elements. Examining performance on individual screening elements, screening for threats of violence was common (average of 96 percent across all IPFs), with performance ranging from 87 percent to 100 percent of patients. Similarly, five IPFs screened between 94 percent and 100 percent of patients for history of violence. However, one psychiatric unit (IPF 1) screened only half of its patients for history of violence. In a debriefing session, clinical staff from this psychiatric unit stated that they do not systematically inquire into patients' history of violence, and that their measure performance likely reflects this fact. This stands in contrast to the same IPF's statement regarding suicide screening -- namely that their suicide screenings cover all required elements, but patients' responses are not necessarily reflected in the patient record.

Use of Standard Screening Tools. The use of standardized screening tools for violence screening was rare. IPF staff used standard tools to conduct violence screenings for only 4 percent of patients across all six IPFs. For these patients, IPFs generally used the V-RISK-10 to administer the screening.

3. Substance Use

In order for the IPF to meet the numerator requirement for the substance use screening measure, the medical record must contain documentation that all four screening elements were completed within one day of admission -- either through use of a validated screening tool or a non-validated assessment protocol.³¹ Because each element pertains to both alcohol and drug use, they can be interpreted as eight screening elements -- four for alcohol and four for drug use. All screening elements are listed below.

Substance Use Screening Measure Numerator Requirement
The medical record must provide documentation that all four of the following screening elements were completed within one day of admission:
1. Type, frequency, and amount of alcohol <u>and</u> drug use.
2. Adverse effects of this use (if use is reported).
3. Dependence upon these substances (if use is reported).
4. Any history of drug <u>and</u> alcohol abuse.

First, we summarize IPF performance on alcohol and drug screening as separate components; then we discuss performance on the full measure, which requires a complete alcohol and drug screening within one day of admission.

³¹ Documented full completion of the AUDIT was equivalent to completing the first three elements for alcohol use (but not substance use).

a. Alcohol Screening

Overall Performance. Average performance on the alcohol screening component of the measure was 90 percent across all six IPFs. However, there was variation among IPFs. Four IPFs had similar performance, with between 93 percent and 96 percent of patients meeting the numerator requirement. Of the remaining two IPFs, IPF 5 had a performance rate of 84 percent and IPF 1 had a performance rate of 66 percent (Table IV.6).

Performance on Individual Elements. Average IPF performance on the measure's four screening elements was uniformly high, ranging from 93 percent on any history of alcohol abuse to 96 percent on frequency and amount of alcohol use. Across all IPFs, most incomplete screenings had only one of four screening elements missing. At the IPF that had the lowest alcohol screening rate (IPF 1: 66 percent performance), element-level performance was lowest for any history of alcohol abuse (78 percent).

b. Drug Screening

Overall Performance. Average performance on the drug screening component of the measure was 90 percent across six IPFs. Four IPFs screened patients for drug use at similar rates, with between 93 percent and 98 percent of patients screened (Table IV.6). Similar to their performance on the alcohol screening component, IPFs 3 and 1 screened the lowest proportion of patients for drug use, with performance rates of 85 percent and 65 percent, respectively.

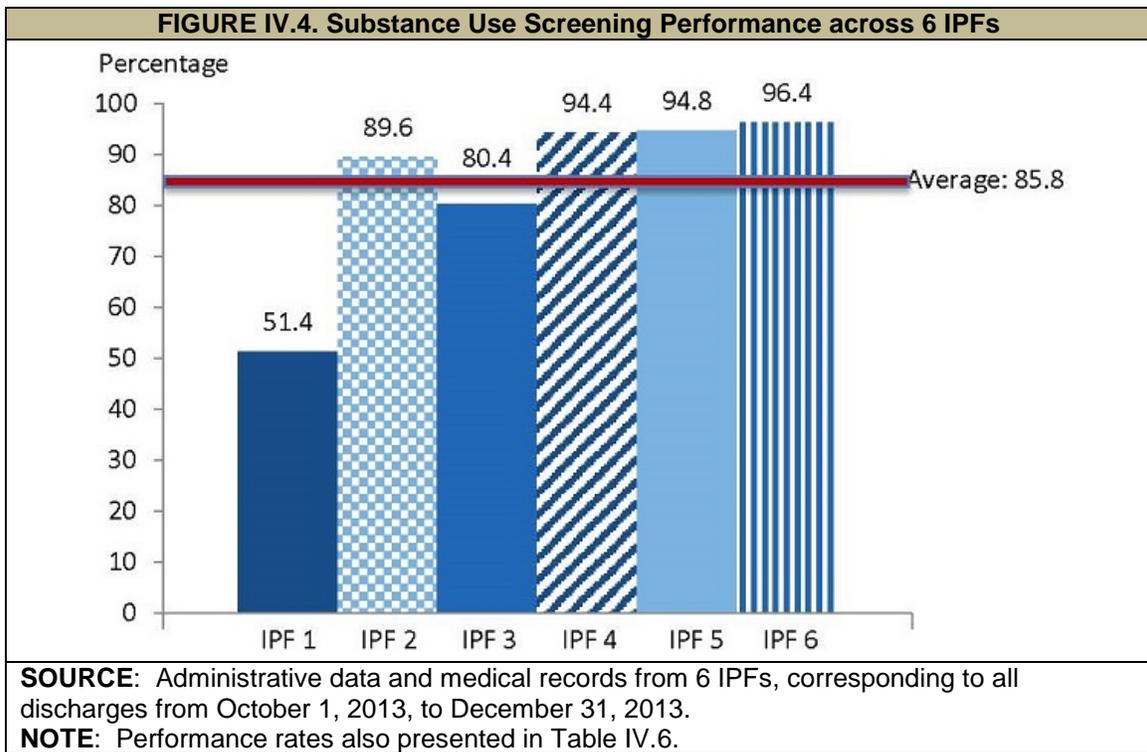
Performance on Individual Elements. Average IPF performance on the measure's four drug screening elements was somewhat uniform, ranging from 94 percent on adverse effects of reported substance use to 96 percent on type, frequency, and amount of substance use. Most incomplete screenings had only one screening element missing. At the IPF that had the lowest drug screening rate (IPF 1: 65 percent across all elements), individual-level performance was lowest for adverse effects of drug use (79 percent). As with the suicide screening measure, this IPF reported that it routinely gathers this information from patients, but that some of its records fail to reflect information provided by patients on these screening elements.

c. Combined Alcohol and Substance Use Screening

Overall, 86 percent of patients at the six IPFs received a full screening for substance use -- including all four elements for alcohol use and all four elements for drug use. The proportion of patients screened for all elements varied notably by IPF, from 51 percent at IPF 1 to 96 percent at IPF 6 (Figure IV.4 and Table IV.6).

TABLE IV.6. Substance Use Screening Performance, by IPF

	All IPS		Psychiatric Units						Freestanding Hospitals					
			IPF 1 (private)		IPF 2 (private)		IPF 3 (private)		IPF 4 (public)		IPF 5 (public)		IPF 6 (private)	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Denominator (after exclusions)	754	100.0	107	100.0	173	100.0	102	100.0	108	100.0	97	100.0	167	100.0
Overall rate: alcohol and drugs	647	85.8	55	51.4	155	89.6	82	80.4	102	94.4	92	94.8	161	96.4
Overall rate: alcohol	675	89.5	71	66.4	160	92.5	86	84.3	105	95.4	92	94.8	161	96.4
1 element missing	50	6.6	23	21.5	7	4.1	13	12.8	3	2.8	1	1.0	3	1.8
2 or more elements missing	29	3.9	13	12.1	6	3.5	3	2.9	0	0.0	4	4.1	3	1.8
Alcohol screening elements met within 1 day of admission^a														
Frequency and amount of alcohol use	723	95.9	101	94.4	164	94.8	92	90.2	107	99.1	95	97.9	164	98.2
Adverse effects of reported alcohol use	720	95.5	95	88.8	163	94.2	93	91.2	108	100.0	94	96.9	167	100.0
Dependence upon alcohol	716	95.0	90	84.1	165	95.4	92	90.2	108	100.0	94	96.9	167	100.0
Any history of alcohol abuse	706	93.6	83	77.6	169	97.7	94	92.2	106	98.1	93	95.9	161	96.4
Overall rate: drugs	678	89.9	70	65.4	160	92.5	87	85.3	103	95.4	94	96.9	164	98.2
1 element missing	46	6.1	22	20.6	6	3.5	13	12.8	5	4.6	0	0.0	0	0.0
2 or more elements missing	30	4.1	15	14.0	7	4.1	2	2.0	0	0.0	3	3.0	3	1.8
Drug screening elements met within 1 day of admission^a														
Type, frequency, and amount of substance use	725	96.2	102	95.3	166	96.0	90	88.2	108	100.0	95	97.9	164	98.2
Adverse effects of reported substance use	711	94.3	85	79.4	165	95.4	91	89.2	108	100.0	95	97.9	167	100.0
Dependence upon substances	717	95.1	89	83.2	167	96.5	91	89.2	108	100.0	95	97.9	167	100.0
Any history of substance abuse	713	94.6	87	81.3	168	97.1	97	95.1	103	95.4	94	96.9	164	98.2
Standard screening tools administered within 1 day of admission^b														
CAGE (alcohol)	18	2.4	0	0.0	0	0.0	0	0.0	18	16.7	0	0.0	0	0.0
AUDIT (alcohol)	11	1.5	0	0.0	0	0.0	0	0.0	5	4.6	0	0.0	6	3.6
Other standard tool	9	1.2	9	8.4	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
DAST-10 (drugs)	6	0.8	0	0.0	0	0.0	0	0.0	6	5.6	0	0.0	0	0.0
Any standard tool	40	5.3	9	8.4	0	0.0	0	0.0	25	23.2	0	0.0	6	3.6
SOURCE: Administrative data and medical records from 6 IPFs, corresponding to all discharges from October 1, 2013, to December 31, 2013.														
NOTE: Totals for individual tools do not always sum to any standard tool, given that more than 1 tool could have been employed for 1 patient.														
a. Elements were met if the element was covered in a validated screening tool, a non-validated tool, or through a non-structured screening or assessment. Documented full completion of the AUDIT was equivalent to completing the first 3 elements for alcohol use (but not substance use). IPFs did not change their screening or admission processes to test these measures. Skip logic was employed for the second and third screening elements: credit was automatically given for adverse effects and dependence if the patient reported no drug or alcohol use.														
b. Choice of screening tools was at the discretion of the clinician or IPF.														



Use of Standard Screening Tools. Standard screening tools were used to conduct alcohol or drug screenings for just over 5 percent of patients sampled. These tools included the AUDIT for alcohol use, and the CAGE and the DAST-10 for drug use. Three of the six IPFs did not use any standardized instruments to screen patients for alcohol or drug use.

Screening Results

Throughout the chart-abstraction process, abstractors noted patients' responses to screening questions in cases in which these responses would determine if additional screening elements were necessary. For example, information on suicide plans was required only in cases in which patients reported suicidal ideation. The following screening results reflect data collected on patients' responses to screening questions, averaged across all IPFs:

- 50% of patients reported suicidal ideation, and 76% of patients who reported suicidal ideation also reported making plans or preparations for suicide.
- 31% of patients reported drug use.
- 28% of patients reported alcohol use.

4. Comparisons with Existing Screening Measures

Table IV.7 and Figure IV.5 compare performance on these admission screening measures with performance on similar existing measures, including HBIPS-1 and SUB-1. As described earlier, the major difference between these new measures, HBIPS-1, and SUB-1 is that these screening measures require screening within one day of

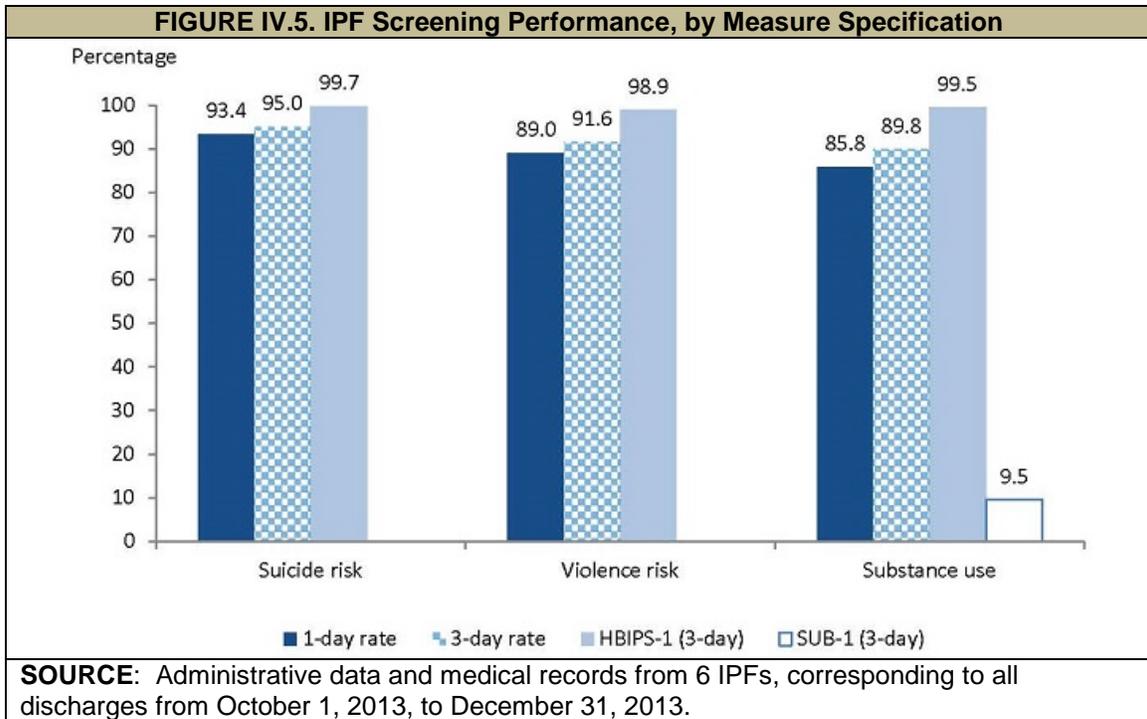
admission (whereas HBIPS-1 and SUB-1 require screening within three days of admission), and these new measures require all designated screening elements to be completed (whereas HBIPS-1 requires only that a screening be completed, and SUB-1 requires that a validated tool be used to screen for alcohol use).

TABLE IV.7. Performance Rates across Screening Measures				
	Suicide	Violence	Substance Use	Alcohol Component of Substance Use*
New IPF measure performance (screening within 1 day that meets all required elements)	93.4%	89.0%	85.8%	89.5%
HBIPS-1 PERFORMANCE (screening within 3 days with no specific element requirements)	99.7%	98.9%	99.5%	NA
SUB-1 PERFORMANCE (screening within 3 days with no specific element requirements)	NA	NA	NA	9.5%
Difference in performance rates (average across 6 IPFs)	6.3 points higher on HBIPS-1	9.9 points higher on HBIPS-1	13.7 points higher on HBIPS-1	80 points lower on SUB-1
SOURCE: Administrative data and medical records from 6 IPFs, corresponding to all discharges from October 1, 2013, to December 31, 2013. * This is the alcohol screening component of the substance use screening measure, which comprises an alcohol screening component as well as a drug screening component. This alcohol screening component is the closest analogue to SUB-1, which measures completion of an alcohol use screening.				

The six IPFs that participated in testing scored, on average, between 6 and 14 percentage points lower on the new measures compared to their analogue components in HBIPS-1, predominantly due to the new measures' stricter requirements regarding screening elements that must be documented in patient charts.³² On average, the six IPFs involved in testing had performance rates of nearly 100 percent for the Risk to Self component of HBIPS-1 (comparable to the suicide screening measure) and 99 percent for the Risk to Others component (comparable to the violence screening measure). However, their performance on the new suicide and violence risk screening measures was lower at 93 percent and 89 percent, respectively. Similarly, average IPF performance on the Substance Use screening component of HBIPS-1 was nearly 100 percent, compared to 86 percent on the new substance use screening measure.

In contrast, across the six IPFs, only 10 percent of patients met the numerator requirement for SUB-1. However, there was substantial variation across sites on SUB-1 performance. For example, IPF 2 had a 23 percent performance rate on the measure, whereas three sites had performance rates of around 0 percent (not shown). IPFs performed much higher on the alcohol screening component of the new substance use measure than on SUB-1, due primarily to SUB-1's requirement of a validated screening tool.

³² It should be noted that one psychiatric ward with relatively low performance (IPF 1) effectively lowers average performance on the new measures across the six IPFs. If this IPF's data were excluded from the calculation, performance on the new measures would be approximately 2-8 percentage points lower than performance on relevant HBIPS-1 components.



5. One-Day Versus Three-Day Performance

Another objective of measure testing was to determine the extent to which IPF performance varied using a one-day versus a three-day version of these screening measures. The three-day version of the measures tabulated the proportion of patients who were screened within three days of admission (as opposed to within one day of admission).³³

There was little change in average measure performance when screening within three days was required rather than screening within one day. Average performance on the substance use measure was four points higher when a three-day requirement was used versus a one-day requirement (89.8 percent versus 85.8 percent, respectively). This trend was also evident in the substance use measure's two components: for the alcohol component, performance on the three-day measure was around 3 percentage points higher than performance on the one-day measure (92.6 percent versus 89.5 percent, respectively). For the drug component, performance on the three-day measure was around 3 percentage points higher than performance on the one-day measure (92.6 percent versus 89.9 percent, respectively). For the suicide and violence measures, performance was 2 and 3 percentage points higher, respectively, when a three-day requirement was used compared to a one-day requirement (see Table IV.8 and Figure IV.5).

³³ The three-day version of the measures required alternate exclusions: notably, patients with stays of fewer than three days were excluded from the measures.

TABLE IV.8. 1-Day versus 3-Day Measure Performance			
	Suicide	Violence	Substance Use
1-day specification (after exclusions)	93.4	89.0	85.8
3-day specification (after exclusions)	95.0	91.6	89.8
Difference in performance rates	1.6 points higher on 3-day version	2.6 points higher on 3-day version	4.0 points higher on 3-day version
<p>SOURCE: Administrative data and medical records from 6 IPFs, corresponding to all discharges from October 1, 2013, to December 31, 2013.</p> <p>NOTE: Denominator size for the 1-day measure is 761, 765, and 754 for the suicide, violence, and substance use measures, respectively. Denominator size for the 3-day measure is 642, 646, and 637 for the suicide, violence, and substance use measures, respectively.</p>			

6. Performance on the Metabolic Screening Measure

Next, we summarize testing results for the metabolic screening measure. As illustrated below, the medical record must contain documentation that all four of the following tests were completed at least once in the 12 months preceding IPF discharge for patients discharged on antipsychotic medications, either during or prior to the index IPF stay. Either an HbA1c or glucose test meets the requirement in the third element. For the fourth element, a lipid panel must be a full panel, as no credit is given for a partial panel.

The medical record must provide documentation of the completion of all four of the following tests at least once in the 12 months preceding IPF discharge for patients discharged on antipsychotics:

1. BMI.
2. Blood pressure.
3. HbA1c or glucose.
4. Lipid panel (includes total cholesterol, triglycerides, high-density lipoprotein, and low-density lipoprotein).

Overall Performance. Across all sites, approximately 42 percent of all patients discharged on antipsychotic medications received complete metabolic screenings in the 12 months prior to their IPF discharge. Performance rates on metabolic screening differed greatly between sites; the percentage of patients screened ranged from 6 percent to 99 percent (Figure IV.6 and Table IV.9). In nearly half of all denominator-eligible patient stays (46 percent across all sites), patients received all but one screening measurement or test.

TABLE IV.9. Metabolic Screening Performance, by IPF

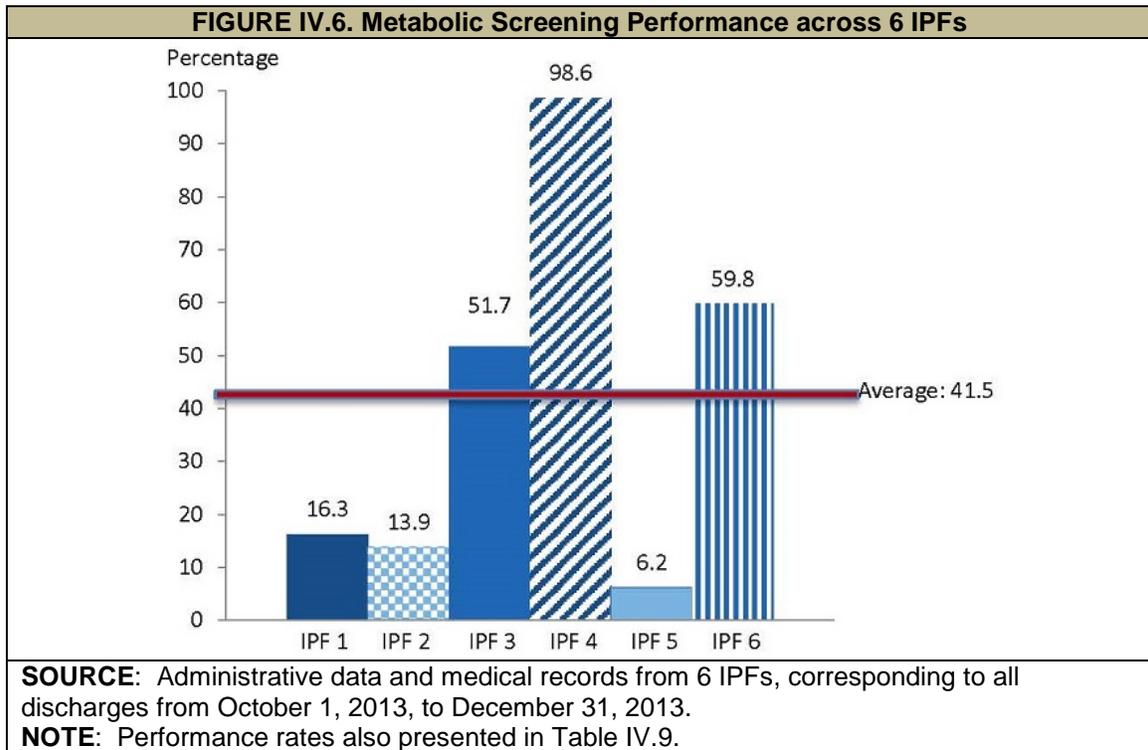
Patient Characteristics	All IPS		Psychiatric Units						Freestanding Hospitals					
			IPF 1 (private)		IPF 2 (private)		IPF 3 (private)		IPF 4 (public)		IPF 5 (public)		IPF 6 (private)	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Denominator (after exclusions)	434	100.0	43	100.0	108	100.0	58	100.0	73	100.0	65	100.0	87	100.0
Overall rate (met numerator requirement)	180	41.5	7	16.3	15	13.9	30	51.7	72	98.6	4	6.2	52	59.8
1 element missing	197	45.4	29	67.4	88	81.5	22	37.9	1	1.4	29	44.6	28	32.2
2 or more elements missing	57	13.1	7	16.3	5	4.6	6	10.3	0	0.0	32	49.2	7	8.1
Screening elements met within 1 year prior to discharge*														
BMI	345	79.5	36	83.7	104	96.3	41	70.7	73	100.0	21	32.3	70	80.5
Blood pressure	431	99.3	43	100.0	106	98.1	58	100.0	73	100.0	64	98.5	87	100.0
Glucose or HbA1C	407	93.8	40	93.0	105	97.2	58	100.0	73	100.0	51	78.5	80	92.0
Glucose	405	93.3	40	93.0	104	96.3	58	100.0	73	100.0	50	76.9	80	92.0
HbA1C	151	34.8	11	25.6	18	16.7	13	22.4	73	100.0	12	18.5	24	27.6
Full lipid panel	220	50.7	8	18.6	14	13.0	40	69.0	72	98.6	19	29.2	67	77.0

SOURCE: Administrative data and medical records from 6 IPFs, corresponding to all discharges from October 1, 2013, to December 31, 2013.

NOTE: Of the 180 patients with a completed metabolic screening in the 12 months prior to discharge, only 3 (1% of patients screened) had the screening completed prior to the index discharge, but within 12 months of discharge. Only 5 patients (or 1.2% of the denominator population) completed a partial lipid panel, as opposed to a full lipid panel.

* IPFs did not change their screening or admission processes to test these measures.

Performance on Individual Elements. Although a large majority of patients discharged on antipsychotic medications received a blood pressure measurement (99 percent), a glucose/HbA1c test (94 percent), and a measurement of their BMI (80 percent), only around half of all patients received a complete lipid panel (51 percent).³⁴ IPFs were much more likely to measure patients' glucose than HbA1C; however, as noted above, either test satisfied the element requirement.



7. Screening Measure Performance With and Without Exclusions

The methodology used to exclude patients from the final denominator for all screening measures was modeled after the HBIPS measure set, which excludes patient stays of less than three days, stays of greater than one year, and patient inability or unwillingness to complete screenings. We applied these exclusions as well, in addition to exclusions related to multiple admissions during the same hospitalization, as psychiatric units could not be expected to re-screen patients who transferred out of their units for a brief period of time, and then returned during the same hospital stay. However, we modified the HBIPS exclusion of a patient stay of less than three days to be a patient stay of less than one day, to accommodate the admission screening measures' one-day time frame.

³⁴ Slightly more than 1 percent of sampled patients received a partial lipid panel, which did not count toward the measure numerator.

TABLE IV.10. Screening Measure Exclusions								
	Suicide		Violence		Substance Use		Metabolic	
	N	%	N	%	N	%	N	%
Denominator before exclusions	825	100.0	825	100.0	825	100.0	506	100
Exclusions Identified								
Patients with a length of stay less than 1 day	41	5.0	41	5.0	41	5.0	NA	NA
Patients with a length of stay less than 3 days	NA	NA	NA	NA	NA	NA	63	12.5
Patients with a length of stay equal to or greater than 365 days	10	1.2	10	1.2	10	1.2	6	1.2
Patients who had previous admissions to psychiatric units during a single hospitalization	2	0.2	2	0.2	2	0.2	2	0.4
Patient inability or unwillingness ^b	11	1.3	7	0.8	20 ^a	2.4	1	0.2
Total exclusions	64	7.8	60	7.3	71	8.6	72	14.2
Denominator after exclusions								
Denominator for screening within 1 day of admission	761	92.2	765	92.7	754	91.4	434	85.8
<p>SOURCE: Administrative data and medical records from 6 IPFs, corresponding to all discharges from October 1, 2013, to December 31, 2013.</p> <p>NOTE: Totals for individual tools do not always sum to any standard tool, given that more than 1 tool could have been employed for 1 patient.</p> <p>a. Across the alcohol and drug components, a total of 20 patients were unable or unwilling to perform the screening within 1 day. This includes 19 individuals who could or would not perform the alcohol screening and 16 individuals who could not perform the drug screening, with large overlap between the 2 groups.</p> <p>b. Patient charts had to demonstrate evidence of patient inability or unwillingness -- including the day and time of attempted screenings -- for this exclusion to be applied.</p>								

Admission Screening Measures. As shown in Table IV.10, exclusions did not have a major impact on the size of the denominator population for the three admission screening measures. Across the six IPFs, the admission screening measures excluded 41 patients for lengths of stay of less than one day, ten patients for lengths of stay equal to or greater than 365 days, and two patients who had previous admissions to psychiatric units during a single hospitalization. The denominator used by each measure differed only by the number of patients excluded due to inability or unwillingness to complete the screening within one day admission. Notably, 20 patients were excluded from the substance use measure due to patient unwillingness or inability to complete the screening, 11 were excluded from the suicide screening measure, and seven were excluded from the violence measure for this reason. (See Appendix Table D.2 for the number of patients who met measure exclusions at each IPF, presented for each screening measure.)

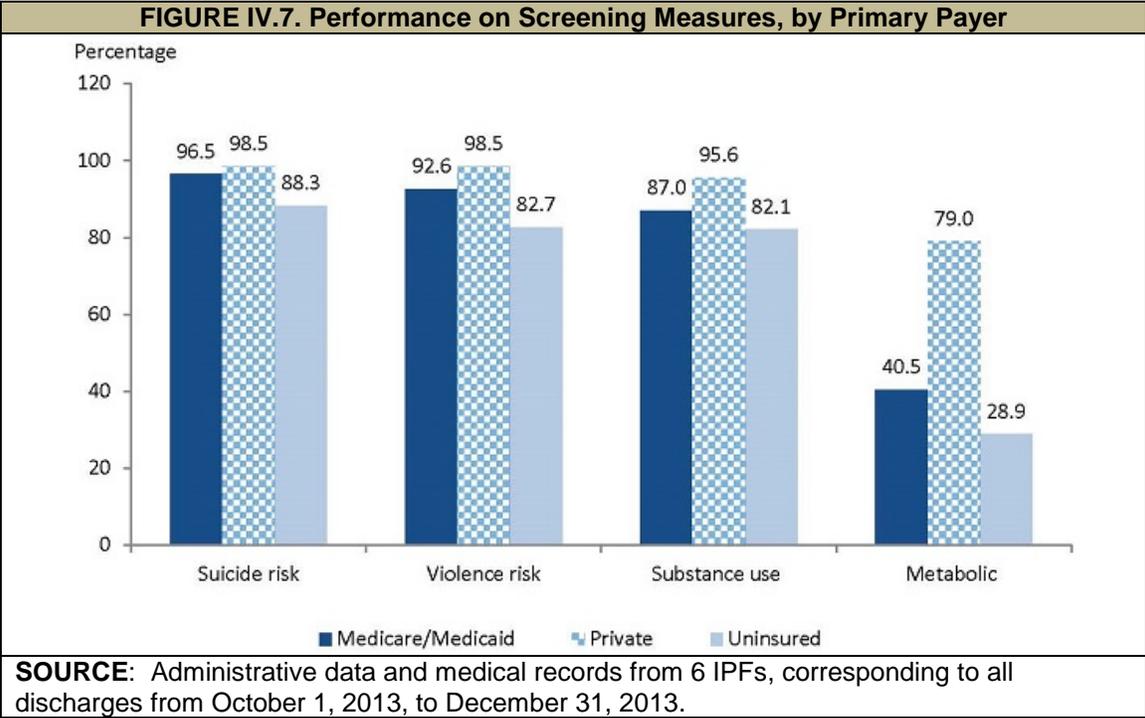
Metabolic Screening Measure. Patients with stays of fewer than three days were excluded from the metabolic screening measure based on the rationale that IPFs could not be expected to complete all metabolic screening tests (or verify that they were completed elsewhere within the previous 12 months) within that short time period. As a result, exclusions had a more substantial impact on the size of the denominator population of the metabolic screening measure relative to the admission screening measures; exclusions totaled 14 percent of discharges for the metabolic screening measure, versus less than 9 percent of discharges for the admission screening measures.

Performance rates were slightly higher among the denominator population after exclusions, compared with the population before exclusions were applied (Table IV.11). However, these differences were not large in magnitude (ranging from one point on the violence screening measure to 2 percentage points on the substance use screening measure) and were not statistically significant. We found similar results for the alcohol and drug components of the substance use measure: for the alcohol component, performance increased by 2 percentage points once exclusions were applied. For the drug component, performance increased by 1 percentage point once exclusions were applied.

	Suicide	Violence	Substance Use	Metabolic
Performance before exclusions	92.0	88.1	83.8	40.3
Performance after exclusions	93.4	89.0	85.8	41.5
Difference in performance rates*	1.4 points higher after exclusions	0.9 points higher after exclusions	2.0 points higher after exclusions	1.2 points higher after exclusions
SOURCE: Administrative data and medical records from 6 IPFs, corresponding to all discharges from October 1, 2013, to December 31, 2013.				
* A t-test of means was conducted to determine if performance before exclusions was statistically different from performance after exclusions. These tests found no statistically significant differences (all p-values > 0.10).				

8. Performance by Patient Characteristics

Admission Screening Measures. There were some statistically significant differences in measure performance across patient subgroups. When data were combined across IPFs, there were no statistically significant differences in performance on the admission screening measures (suicide, violence, and substance use) by patient gender (Table IV.12). However, overall screening rates for patients between the ages of 18 and 64 (ranging from 84 percent to 92 percent) were lower than for other age groups, and these differences were statistically significant for the suicide and substance use measures.



There were statistically significant differences in overall performance rates by race for all three admission measures. In each case, measure performance was highest among African American patients (with a minimum performance rate of 95 percent across all measures) and lowest among White patients (with a minimum performance rate of 82 percent across all measures). These differences in performance likely reflect the underlying differences in patient demographics across the six IPFs. The African American patient population was concentrated in IPFs 4, 5, and 6 -- the three IPFs with the highest performance on most of these measures.

TABLE IV.12. Performance on Screening Measures across Demographic Characteristics

	Suicide	p-Value	Violence	p-Value	Substance Use	p-Value	Metabolic	p-Value
Gender								
Female	95.0	0.07	88.1	0.40	86.4	0.67	37.8	0.11
Male	91.8		90.0		85.3		45.5	
Race								
White	91.8	0.05	85.5	<0.01	82.4	<0.01	37.3	<0.01
African American	96.5		97.5		94.9		53.7	
Other	96.6		90.0		84.7		24.1	
Age								
Under 13	100.0	0.02	95.7	0.06	82.6	0.04	22.2	0.26
13-17	100.0		100.0		100.0		41.2	
18-64	91.8		87.5		84.3		44.0	
Greater than 64	97.7		91.7		89.3		34.4	
Insurance Coverage (primary payer)								
All Medicare and Medicaid	96.5	<0.01	92.6	<0.01	87.0	<0.01	40.5	<0.01
Dually eligible	100.0		87.0		90.5		57.1	
Medicare-only	97.2		93.0		90.0		40.5	
Medicaid-only	94.3		91.7		77.9		36.0	
Private insurance	98.5		98.5		95.6		79.0	
Uninsured	88.3		82.7		82.1		28.9	
Primary Diagnosis								
Schizophrenia	95.7	0.02	91.5	0.03	88.1	0.34	51.2	<0.01
Bipolar disorder	90.1		84.0		83.0		30.8	
Depression	97.6		95.2		90.5		69.2	
Other	95.5		92.9		86.9		40.8	
<p>SOURCE: Administrative data and medical records from 6 IPFs, corresponding to all discharges from October 1, 2013, to December 31, 2013.</p> <p>NOTE: Statistically significant differences at 5% are in bold. Readers should be cautious interpreting these differences given that there are multiple comparisons; these bivariate differences may not hold when controlling for other patient or facility characteristics.</p>								

In addition, admission screening measure performance varied by patient insurance; patients with private insurance and Medicare generally had higher screening rates than uninsured patients and Medicaid-only beneficiaries (Figure IV.7). Also notable, performance was highest among patients with a depression diagnosis and lowest among patients with a bipolar disorder diagnosis; these differences were statistically significant for the suicide and violence screening measures.³⁵

Metabolic Screening Measure. Similar to the admission screening measures, performance on the metabolic screening measure was lowest among patients in the bipolar disorder diagnoses (31 percent versus over 40 percent for other diagnoses). In addition, performance varied by patient insurance; patients with private insurance had much higher screening rates than patients with alternate insurance or no insurance (Figure IV.7). As noted later in this report, several stakeholders noted that this finding may reflect that some types of private insurance provide more generous coverage of lab tests compared with public insurance programs.

Also similar to the admission screening measures, there were significant differences in metabolic screening rates in our sample by race, with highest performance among African Americans patients (54 percent compared to 37 percent among Whites and 24 percent among other ethnicities). These differences in performance likely reflect the higher concentration of African American patients in IPFs 4 and 6; these two IPFs had the highest performance on the metabolic screening measure.

9. Inter-Rater Agreement for Screening and Monitoring Measures

Inter-rater reliability assesses whether two chart abstractors independently reviewing data from the same record agreed on whether the patient met the requirements for the numerator, denominator, and/or exclusions for the measure. In order to assess inter-rater reliability, each IPF had two abstractors independently abstract the same record for a sample of charts. We used Cohen's kappa statistic, a measure of agreement adjusted for chance, to quantify agreement among these abstractors.

Inter-rater reliability was moderate to high for all measures. Percentage agreement ranged from 93 percent to 98 percent across measure elements (Table IV.13). The kappa coefficients for measure exclusions were 0.85 or above for all three admission screening measures, which is considered near-perfect agreement (Landis and Koch 1977). For the numerator, element kappa coefficients ranged from 0.49 (substance use) to 0.93 (metabolic screening). In addition, we calculated a kappa coefficient for each of the alcohol and drug screening components of the substance use measure. Kappa for

³⁵ For all admission screening measures the highest performance was among the group missing a primary diagnosis. However, we excluded this category from the analysis because it is relatively small (approximately 3 percent of the denominator).

the alcohol screening numerator was 0.99 whereas kappa for the drug screening numerator was 0.49. As such, the drug screening component is responsible for the lower overall kappa coefficient for the substance use screening measure.

TABLE IV.13. Inter-Rater Agreement on Screening Measures				
	Number of Patient Charts Double Extracted Across All IPFs	Percent Agreement	Kappa Coefficient	95% Confidence Interval
Suicide				
Exclusions	58	98.3	0.85	0.56-1.00
Numerator	58	96.6	0.65	0.20-1.00
Violence				
Exclusions	58	98.3	0.85	0.56-1.00
Numerator	58	93.1	0.63	0.31-0.96
Substance Use				
Exclusions	58	98.3	0.85	0.56-1.00
Numerator	58	96.6	0.49	-0.11-1.00
Metabolic				
Exclusions	58	98.3	0.95	0.84-1.00
Numerator	58	96.5	0.93	0.83-1.00
SOURCE: Administrative data and medical records from 6 IPFs, corresponding to all discharges from October 1, 2013, to December 31, 2013.				
NOTE: A kappa of 0.21-0.40 indicates fair agreement; a kappa of 0.41-0.60 indicates moderate agreement; a kappa of 0.61-0.80 indicates substantial agreement; a kappa of 0.81 or higher indicates almost perfect agreement.				

Interpreting these scores, the substance use numerator exhibits moderate agreement, the suicide and violence numerators exhibit substantial agreement, and the metabolic screening numerator exhibits near-perfect agreement. This near-perfect agreement on the metabolic screening measure likely reflects the non-subjective nature of the numerator, which required explicit tests to have been performed, as opposed to the admission screening measures, which required a higher level of abstractor interpretation of written notes in the patient record documenting clinical interviews. See Appendix Table D.3 for percentage agreement at each IPF, presented for each screening measure.

C. Qualitative Testing Results

Feedback from stakeholders -- including IPF staff, measure experts, policymakers, consumer and advocacy group representatives, and TEP members -- focused on the importance of the measures, the face validity of measure elements, the required time frame for completion of screening, and the measures' usability and feasibility. This feedback is summarized below.

1. Admission Screening Measures

Importance. Focus group participants and IPFs involved in measure testing acknowledged the importance of all three screening measures as the starting point for high quality care in IPFs. However, many thought CMS should also place a strong measure focus on whether patients receive appropriate treatment and demonstrate positive outcomes. The TEP underscored the importance of screening in IPF

populations and was supportive of all three screening measure concepts. Some TEP members pointed to higher screening rates among African Americans compared to White and Latino patients as evidence of disparities in care.

Validity. Overall, focus group participants and IPFs involved in testing supported the validity of all three screening measures. Notably, five of the six IPFs that participated in the testing reported that the measures accurately captured their completion of screenings within one day of patient admission.³⁶ Most participants viewed the measures' required screening elements (for example, the suicide screening measure's required documentation of ideation, plans, intent, and so on) as an improvement over HBIPS-1 in terms of mandating a high quality screening. A majority of participants preferred requiring these screening elements over requiring the use of validated instruments for all three measures. However, several participants voiced support for validated suicide screening tools, given their usefulness in treatment planning.

At least three of nine TEP members were surprised at the high average performance rates for the screening measures (performance of between 85 percent and 95 percent for all three measures), and noted that the measures might be "topped out." Overall, however, the TEP was supportive of the screening elements, and four out of five TEP members³⁷ preferred individual screening elements over the required use of validated screening instruments.

Time Frame. At least 13 of the 25 focus group participants expressed concern about IPFs' capacity to gather accurate or complete information from screenings within one day of admission, due to patients' medical status on admission, IPF staffing constraints, or the difficulty of obtaining collateral information within one day. However, three focus group participants favored a one-day time frame for the suicide and violence risk measures -- given the clinical importance of such screening immediately upon admission -- and a three-day timeline for the substance use screening measure.

Three of nine TEP members preferred a three-day time frame over a one-day time frame for all three measures, citing the importance of allowing sufficient time to complete comprehensive and accurate screenings, as well as the need to incentivize IPF staff to re-screen patients within the first three days of their stay to complete and refine initial assessments.³⁸ Although one TEP participant suggested using non-uniform time frames across screening measures (for example, one day for violence and suicide risk screening, and three days for substance use screening), two participants noted that it would be more feasible to use a standard time frame across all three screening measures. One TEP member argued for a one-day time frame for all three screening measures.

³⁶ One IPF reported that it routinely gathers information on suicide and violence risk from patients but that some of its records fail to reflect information provided by patients on these screening elements.

³⁷ Five TEP members expressed an opinion on this topic.

³⁸ Other TEP members did not articulate a preference for one time frame over another during the meeting.

Usability. The majority of the 25 focus group participants and the six IPFs involved in testing generally agreed that these screening measures would be useful to inform quality improvement activities. However, some thought that the value of the measures would be stronger in the short term as IPFs would modify their screening and tracking practices to fully comply with the measures. Consumer advocates mentioned that although consumers generally have little choice in IPFs, advocacy groups have a strong interest in examining publicly available performance rates, and calling attention to poor IPF performance if necessary. Such efforts could motivate IPFs to improve their performance on these screening measures, among others.

The majority of TEP members agreed that the screening measures could contribute to quality improvement. However, given the high performance rates on these measures, two TEP members questioned whether further improvement on these measures would be possible for a subset of high-performing IPFs. These participants suggested that future IPFQR program measures be oriented toward processes and outcomes with more sizable gaps in performance.

Feasibility and Burden. The majority of focus group participants and IPFs involved in testing agreed that the burden of abstracting and reporting these three screening measures was manageable. In particular, the six IPFs that piloted the measures noted that the full abstraction of a patient's record for all three screening measures took 10-15 minutes, on average, as required data elements were readily available in records. However, representatives of three of 11 IPFs, one consumer representative, one quality measurement expert, and one IPFQR program vendor noted that if IPFs choose to integrate new screening tools into their electronic medical record to ensure reliable data capture for the measures, the transition could take at least six months to implement.

The TEP agreed that the task of abstracting and reporting the suicide, violence, and substance use screening measures was manageable. However, three TEP members strongly recommended that the measures be harmonized with HBIPS-1 to decrease IPF burden and redundancy across psychiatric inpatient measures.

2. Metabolic Screening Measure

Importance. Representatives from IPFs, IPFQR program vendors, policy experts, and consumer groups agreed that monitoring patients' metabolic functioning is extremely important, given the psychiatric inpatient population's high risk for metabolic conditions, and expressed strong support for CMS's focus on the issue. TEP members also supported the importance of the metabolic screening measure, citing large performance gaps across IPFs that participated in testing.

Validity. Focus group participants generally supported the measure concept and its elements. However, some IPFQR program vendors, measurement experts, and consumer representatives were unsure if all of the required elements were clinically

relevant -- particularly a full lipid panel. One vendor and a clinician expressed concern that the measure might inadvertently encourage IPFs and other clinicians to conduct unnecessary tests, and that test results might not be transmitted to patients' next level of care. Three of six IPFs involved in measure testing expressed some resistance to the measure's full lipid panel requirement. According to these IPF representatives, clinicians generally used their judgment in determining whether a panel was necessary. As such, requiring a full panel for all patients on antipsychotics might "over treat" patients.

Three of nine TEP members expressed similar concern over the measure requirement of a full lipid panel. One TEP member stated that lipid panels were relatively expensive lab tests, and that some IPFs serving patients with state or county insurance could not get adequate compensation for a full lipid panel for all patients on antipsychotics. Two participants suggested excluding the lipid panel requirement, or structuring the measure such that IPFs are required to complete blood pressure, BMI, and glucose tests, and then proceed to a full lipid panel if necessary. However, one TEP member strongly disagreed, citing the strength of scientific evidence in support of a full lipid panel at least once per year for individuals on antipsychotics.

Usability. Representatives from IPFs, IPFQR program vendors, policy experts, and consumer groups generally supported the potential usability of the metabolic screening measure, reasoning that poor performance could help IPFs identify gaps in care and areas for improvement. However, one IPF representative noted that some hospitals with poor performance on the measure may face cost constraints -- particularly related to reimbursement from some payers for a full lipid panel -- that might preclude their improvement on the measure.

TEP members supported the measure's usability, noting that IPFs' low performance on the measure (around 42 percent performance across six IPFs, on average) suggests strong potential for quality improvement efforts around metabolic screening. However, two participants noted that screening could contribute to improved quality only if it informs care beyond screening (such as treatment options).

Feasibility and Burden. Focus group participants expressed concern that the metabolic screening measure would place a large burden on IPF staff, as they must search patient records to determine whether required tests were completed within the past year. IPFs involved in measure testing verified that chart-abstraction of this measure was more labor-intensive than the other screening measures, but generally did not exceed 20 minutes for any given discharge. One IPF involved in testing stated that chart abstractors often did not have access to relevant lab test results due to data sharing restrictions. However, the other five IPFs involved in testing mentioned that abstractors have easy access to relevant lab records. Only one of nine TEP members explicitly expressed concerns about the burden of abstracting or reporting the metabolic screening measure.

D. Summary of Findings and Proposed Revisions

Admission Screening Measures. Measure performance was quite high across IPFs on the suicide, violence, and substance use screening measures, with average performance on the measures ranging from 86 percent (in the case of substance use) to 93 percent (in the case of suicide). Reliability was moderate for the substance use measure and substantial for the suicide and violence measures. In addition, IPF performance on the measures was not substantively affected by exclusions or specification as one-day versus three-day measures, thus providing support for the measures' face validity. Stakeholders were generally supportive of the measures and thought their screening element requirements represented an improvement over existing screening measures used in an inpatient psychiatric setting, including HBIPS-1.

Regarding changes to measure specifications, stakeholders generally recommended that the final specification of the substance use, violence, and suicide screening measures use a three-day time frame to allow for complete screening. Obstacles to performing accurate screenings within one day of admission include staff shortages, patient uncooperativeness, and lack of patient lucidity. Citing harmonization needs, the TEP also noted that a three-day version of this measure would be consistent with HBIPS-1. Some stakeholders noted that the suicide and violence measures should be conducted within a one-day time frame, given the importance of obtaining that information quickly. Based on this feedback, the research team recommends changing the time frame for the substance use screening measure from one day to three days, and keeping the suicide and violence screening specifications at one day (as currently specified; see Table IV.14). The additional days will facilitate the capture of complete and accurate information regarding patients' alcohol and drug use, without compromising the need to capture important information on suicide and violence risk in the first day of admission.

Metabolic Screening Measure. Performance on the metabolic screening measure was low, on average, across the six IPFs. The measure's average performance rate of 42 percentage points signals strong room for improvement on the measure. The metabolic screening measure demonstrated non-trivial variation in performance among IPFs as well as by patient characteristics. In addition, it demonstrated near-perfect agreement between chart abstractors (kappa of 0.93 for the measure numerator).

Overall, stakeholders found the metabolic screening measure to be important for addressing a notable gap in psychiatric care. However, focus group participants and TEP members were divided over whether to keep the requirement of a full lipid panel, as some felt that blood pressure, BMI, and glucose/HbA1c tests were sufficient screening requirements. In particular, one of the six IPFs that participated in pilot testing and two of nine TEP members expressed concern that the measure might inadvertently encourage IPFs and other clinicians to conduct unnecessary tests -- namely a full lipid panel in instances in which there is no clinical need. However, given the preponderance of clinical evidence supporting a full lipid panel on an annual basis for patients taking

regularly prescribed antipsychotic medications, we suggest that the full lipid panel remain a screening element in the metabolic screening measure (see Table IV.14).

TABLE IV.14. Summary of Testing Results, Stakeholder Feedback, and Proposed Revisions for Screening Measures			
Measure	Quantitative Testing Results	Stakeholder Input	Revisions to Specification Following Testing
Suicide screening	<ul style="list-style-type: none"> • Credible results • Some variation among IPFs and by patient characteristics, but generally high performance 	<ul style="list-style-type: none"> • Important and useful for quality improvement • Disagreement over whether to implement as a 1-day or 3-day measure • Overall high average performance--potentially "topped out" • Measure should be harmonized with related measures to decrease redundancy and burden 	<ul style="list-style-type: none"> • No revisions
Violence screening	<ul style="list-style-type: none"> • Credible results • Some variation among IPFs and by patient characteristics, but generally high performance 	<ul style="list-style-type: none"> • Important and useful for quality improvement • Disagreement over whether to implement as a 1-day or 3-day measure • Overall high average performance--potentially "topped out" • Measure should be harmonized with related measures to decrease redundancy and burden 	<ul style="list-style-type: none"> • No revisions
Substance use screening	<ul style="list-style-type: none"> • Credible results • Some variation among IPFs and by patient characteristics, but generally high performance 	<ul style="list-style-type: none"> • Important and useful for quality improvement • Disagreement over whether to implement as a 1-day or 3-day measure • Overall high average performance--potentially "topped out" • Measure should be harmonized with related measures to decrease redundancy and burden 	<ul style="list-style-type: none"> • Change time frame from 1 day after admission to 3 days after admission
Metabolic screening	<ul style="list-style-type: none"> • Credible results • Substantial room for improvement • Strong variation among IPFs and by patient characteristics 	<ul style="list-style-type: none"> • Important and useful for quality improvement. • Disagreement on whether to include a lipid panel as a required screening element • Concern about reporting burden 	<ul style="list-style-type: none"> • No revisions--maintain full lipid panel as a screening element

V. TESTING RESULTS FOR THE FOLLOW-UP AFTER IPF HOSPITALIZATION MEASURE

We used Medicare and Medicaid claims data from calendar year 2008 to test the measure of follow-up after IPF hospitalization. This measure is closely based on the HEDIS FUH measure (NQF #0576). As part of quantitative testing, we examined measure performance across various numerator options, as well as measure reliability. In addition, we held stakeholder focus groups and debriefing sessions to gather qualitative input on the validity, credibility, and representativeness of the measure. In the next section, we present the results of quantitative measure testing.

A. Characteristics of Medicare Beneficiaries Who Used IPFs

Across all 50 states in this analysis, roughly 210,000 FFS Medicare beneficiaries had an IPF stay with a mental health diagnosis from January 1, 2008, to December 1, 2008 (Table V.1). These beneficiaries were almost equally divided between male and female, and most were over 50 years old. The majority of beneficiaries were White; approximately half received a principal diagnosis of bipolar disorder and one-third a diagnosis of schizophrenia during their IPF stay.

The denominator for the measure is based on discharges rather than individual patients. The 210,326 FFS Medicare beneficiaries in the denominator population had a total of 321,454 IPF discharges during 2008, generating an average of around 1.5 IPF stays per beneficiary over the calendar year. A total of 1,702 IPFs had at least one FFS Medicare discharge during 2008 (not shown). On average, each IPF had 189 FFS Medicare discharges during the year, but this varied significantly (range = 1-1,300 discharges; median of 139 FFS Medicare discharges).

Our primary quantitative analyses used Medicare claims data to examine measure performance among non-dual eligible beneficiaries. At the outset of the project, CMS determined that, due to feasibility issues and data limitations, the follow-up measure would be calculated using only Medicare data.³⁹ Including dual eligible beneficiaries in the denominator of a measure that relies solely on Medicare data would be problematic because these data would not capture dual eligible beneficiaries' receipt of Medicaid-financed services. For this reason, we excluded dual eligible beneficiaries from the measure denominator prior to conducting most validity and reliability analyses.

As illustrated in Table V.1, the non-dual eligible population represented 37 percent of all FFS Medicare beneficiaries who had at least one IPF hospitalization in 2008. The

³⁹ Medicare claims are available within a few months of service receipt, whereas finalized Medicaid claims generally have a multiyear time lag (depending on the state).

remainder of this chapter summarizes measure performance among this non-dual eligible population, using only FFS Medicare claims. Chapter VI presents supplemental analyses of measure performance among all FFS Medicare beneficiaries, including dual eligible beneficiaries, using a combination of Medicare and Medicaid claims.

TABLE V.1. Characteristics of FFS Medicare Beneficiaries with At Least One Mental Health IPF Hospitalization in Calendar Year 2008		
Characteristic	Number of FFS Medicare Beneficiaries	Percentage of FFS Medicare Beneficiaries (%)
All beneficiaries	210,326	100.0
Age		
Under 27	8,897	4.2
27-39	36,355	17.3
40-49	48,429	23.0
50-64	56,002	26.6
65-79	42,889	20.4
80 or older	17,745	8.4
Gender		
Male	98,186	46.7
Female	112,140	53.3
Race		
White (non-Hispanic)	161,436	76.8
African American (non-Hispanic)	36,227	17.2
Asian/Pacific Islander	2,044	1.0
Hispanic/Latino	5,910	2.8
American Indian/Alaskan Native	1,290	0.6
Other	2,870	1.4
Unknown	549	0.3
Primary Diagnosis		
Schizophrenia	69,764	33.2
Bipolar disorder	109,769	52.2
Major depressive disorder	8,448	4.0
Psychosis	12,720	6.1
Other	9,625	4.6
Insurance Coverage		
Medicare FFS only	77,928	37.1
Medicare/Medicaid dual beneficiary	132,398	62.9
SOURCE: FFS Medicare claims from calendar year 2008.		
NOTE: Counts in this table are Medicare FFS beneficiaries discharged from IPFs with a mental health diagnosis. Sample size is 321,454 discharges at 1,702 IPFs that had at least 1 FFS Medicare discharge during 2008.		

B. Performance by Numerator Options

We tested four numerator options (Table V.2) for the follow-up measure. These options define a follow-up mental health visit in four slightly different ways:

Option 1 uses a definition of outpatient follow-up care that closely follows the HEDIS FUH measure: a mental health procedure (identified by CPT/HCPCS codes) provided by a mental health practitioner, who is a psychiatrist, psychologist, social worker, psychiatric nurse or physician assistant with a psychiatric specialty.⁴⁰ This option does not use diagnosis codes to define follow-up mental health services.

Option 2 uses the same CPT/HCPCS codes used in option 1, but requires that these codes be accompanied by a principal mental health diagnosis instead of requiring that the procedure be performed by a mental health practitioner (as in option 1). Following the HEDIS FUH measure specification, diagnoses of schizophrenia, depression, bipolar disorder, stress, and personality disorders are considered mental health diagnoses, but diagnoses of alcohol and drug dependency, as well as some anxiety disorders, are not.⁴¹

It should be noted that CPT/HCPCS codes used in options 1 and 2 include services that are explicitly behavioral health care (such as a psychiatric evaluation and psychotherapy), as well as services that are not exclusively behavioral health care (such as office visits and home visits). However, procedure codes that are not explicitly for behavioral health care only count toward the numerator if they are either provided by a mental health practitioner (option 1) or provided in conjunction with a principal mental health diagnosis (option 2).

Option 3 requires a mental health diagnosis, but does not require a specific CPT/HCPCS procedure code. Rather, it requires that the service was billed as an outpatient service. This less-restrictive definition of mental health care -- any outpatient visit accompanied by a principal mental health diagnosis -- captures mental health services performed by primary care providers and other practitioners who are not mental health practitioners according to the HEDIS FUH definition.

Option 4 requires only an outpatient visit, regardless of diagnosis or provider. This is not a viable option for measuring follow-up mental health care. However, we included it in our analysis to provide context for the other three numerator options.

For each of the 1,669 IPFs that treated at least one non-dual FFS Medicare beneficiary in 2008, we calculated the proportion of patients that received follow-up care within seven days and within 30 days of their IPF discharge.⁴² For example, an IPF with

⁴⁰ The HEDIS FUH measure uses health plan records to determine whether a provider is a mental health practitioner. We use provider taxonomy and Specialty codes in Medicare and Medicaid claims to make this determination. Specifically, we use providers' NPI numbers to determine their exact Taxonomy code and specialty. Based on consultation with stakeholders, the team expanded the list of eligible providers to include psychiatric nurses and physician assistants with a psychiatric specialty.

⁴¹ These diagnosis codes are used in the HEDIS FUH measure to determine the denominator of individuals hospitalized for mental illness. In adapting and testing the FUH measure for IPFs, we use these same diagnosis codes to calculate the numerator for options 2 and 3.

⁴² This is slightly less IPFs than the 1,703 IPFs that had at least one FFS Medicare-paid stay in 2008. A total of 34 IPFs in the sample dropped out of the non-dual analysis because all of their FFS Medicare-paid stays pertained to dual beneficiaries.

ten FFS Medicare discharges in 2008, in which four of these stays were accompanied by a follow-up visit within the month, would have a performance rate of 40 percent for the year. In the tables below, we present pooled results, or the average of IPF performance rates for 2008 on all four numerator options.

TABLE V.2. Numerator Options for Follow-Up Measure	
Option 1	An outpatient visit, an intensive outpatient encounter or partial hospitalization with a mental health practitioner within 30 or 7 days of discharge. (Uses CPT/HCPCS codes and NPI/Taxonomy codes.) (This definition most closely aligns with the HEDIS FUH measure.)
Option 2	An outpatient visit, an intensive outpatient encounter or partial hospitalization (with any provider) with a principal mental health diagnosis within 30 or 7 days of discharge. (Uses CPT/HCPCS codes and mental health diagnosis codes.)
Option 3	Any outpatient visit with a principal mental health diagnosis within 30 or 7 days of discharge (excluding emergency department, ambulance, lab, and other non-ambulatory claims).
Option 4	Any outpatient visit, regardless of diagnosis or provider.

Overall IPF Performance. On average, numerator options 1, 2, and 3 yield similar estimates of follow-up care: slightly less than one-third of non-dually eligible FFS Medicare beneficiaries received follow-up mental health care within seven days, whereas slightly more than 50 percent received such follow-up care within 30 days (Table V.3). Seven-day rates are between 28 percent and 30 percent for all three options, and 30-day rates are between 52 percent and 55 percent for all three options. Among the first three numerator options, option 3 has the highest average performance on both the seven-day and 30-day measures; this is not surprising given that it has the least restrictive definition of follow-up care among the three options. Also notable, option 1 has slightly higher performance than option 2 (30-day follow-up rate of 54 percent for option 1 versus 52 percent for option 2). This difference is due to option 2's requirement that the outpatient visit have a principal mental health diagnosis, which identified slightly less numerator hits than option 1's requirement that the visit take place with a mental health practitioner.

TABLE V.3. Follow-Up within 7 and 30 Days of IPF Hospitalization, among Non-Dual FFS Medicare Beneficiaries							
	Mean	Min	25 th Percentile	Median	75 th Percentile	Max	IQR
Option 1							
7-day follow-up	28.7	0	16.7	27.8	39.5	100	22.8
30-day follow-up	53.5	0	42.3	55.0	67.3	100	25.0
Option 2							
7-day follow-up	28.2	0	16.7	27.3	38.5	100	21.8
30-day follow-up	52.0	0	40.0	53.1	65.9	100	25.9
Option 3							
7-day follow-up	30.3	0	18.8	29.3	40.9	100	22.2
30-day follow-up	54.5	0	42.9	55.9	67.7	100	24.8
Option 4							
7-day follow-up	50.3	0	40.0	50.0	60.6	100	50.3
30-day follow-up	80.4	0	73.7	83.3	90.2	100	80.4
SOURCE: FFS Medicare claims from calendar year 2008.							
NOTE: Sample size 61,871 index discharges among 1,669 facilities with at least 1 non-dually eligible FFS Medicare discharge in 2008.							

Variation in Performance. As shown below, all numerator options demonstrate wide variation in performance across IPFs, with IQRs of greater than 21 for all numerator options.⁴³ The follow-up measure's relatively large IQR -- indicative of a large distribution of performance among IPFs -- suggests widespread opportunities for improvement on the measure.

In addition, there is large dispersion of IPF performance on the measure, with a minimum of zero and a maximum of 100 percent performance for the seven-day and 30-day versions of the measures for all numerator options. However, some of the extremes in these minima and maxima reflect instances of small sample sizes -- generally IPFs that had fewer than ten FFS Medicare-paid discharges in 2008.

Stability of Performance Across Numerator Options. To determine the extent to which IPFs' performance under one numerator option is correlated with performance under other numerator options, we tabulated the portion of IPFs that score within the same decile -- or within 10 percentile points -- on any combination of options. For this analysis, we converted each IPF's score for each numerator option into a percentile score, reflecting that IPF's performance relative to other IPFs in the sample. Next, we determined the portion of IPFs whose percentile scores for two numerator options were within ten points of each other. For example, if an IPF's score on option 1 is the 87th percentile and its score for option 2 is the 93rd percentile, the IPF would qualify as remaining in the same decile in a comparison of options 1 and 2.

There is considerable consistency in IPF performance across numerator options 1, 2, and 3. As illustrated in Table V.4, at least 73 percent of IPFs score within the same decile between any two numerator options. In addition, the median percentile difference between numerator options was relatively small, ranging between 3 and 5 percentile points. These results illustrate that IPFs have similar performance relative to other IPFs under all three numerator options -- essentially that a high-performing IPF under any option would likely be high-performing under the other two options.

TABLE V.4. Performance among Numerator Options: Follow-Up within 30 Days of IPF Hospitalization, among Non-Dual FFS Medicare Beneficiaries					
Options 1 and 2		Options 1 and 3		Options 2 and 3	
Median Percentile Difference	IPFs That Remained in the Same Decile (%)	Median Percentile Difference	IPFs That Remained in the Same Decile (%)	Median Percentile Difference	IPFs That Remained in the Same Decile (%)
4	77.1	5	72.8	3	94.4
SOURCE: FFS Medicare claims from calendar year 2008.					
NOTE: Sample size 61,871 index discharges among 1,669 IPFs.					

Final Numerator Selection. In consultation with CMS, ASPE, and the TEP during the measure testing process, we determined that numerator option 1 could be accurately

⁴³ The IQR is the difference between the values at the 25th and 75th percentiles of a distribution. A larger IQR indicates greater variation in performance. Measures with a low IQR (for example, less than 10 percentage points) may be less useful for comparing entities.

calculated using only Medicare claims. As mentioned above, option 1 closely adheres to the NQF-endorsed HEDIS approach, which allows for a direct comparison with health plan performance on the FUH measure. Because option 1 has similar performance relative to the other two numerator options for non-dual beneficiaries, and has the additional advantage of closely following FUH measure specifications, stakeholders viewed it as the most appropriate numerator option. For this reason, we used numerator option 1 to conduct the remaining analyses of measure exclusions, variation in performance by IPF size and demographics, and reliability.

C. Impact of Measure Exclusions on Follow-Up Measure Performance

Measure exclusions for the follow-up measure align with those of the HEDIS FUH measure, which excludes IPF discharges followed by acute care and non-acute care during the 30-day follow-up period.⁴⁴ The rationale for exclusions related to acute and non-acute care is that an inpatient or institutional stay could interfere with the ability of the beneficiary to seek and obtain follow-up care after an IPF hospitalization. We examined the impact of these measure exclusions on the denominator size and measure performance.

TABLE V.5. Proportion of Eligible Discharges Excluded from the Follow-Up Measure Denominator, among Non-Dual FFS Medicare Beneficiaries			
Exclusion	Rationale for Exclusion	Proportion of IPF Discharges (%)	Number of Discharges
1. Death within 30 days of IPF discharge	Death within the follow-up period does not allow for follow-up care	0.6	718
2. For an IPF discharge where the patient visited an IPF in the previous 30 days, exclude the previous IPF discharge	Including these IPF discharges would influence the number of discharges in the denominator and measure performance	17.8	18,787
3. IPF discharges with a non-IPF inpatient or other residential stay during follow-up period	An inpatient or otherwise residential stay may interfere with patients' receipt of follow-up care	38.0	40,374
All exclusions (proportion excluded for any of the exclusions above)		41.7	44,268
SOURCE: FFS Medicare claims from calendar year 2008.			
NOTE: Sample size is 106,139 index discharges among 1,669 IPFs. The exclusions presented in this table are not mutually exclusive. For example, exclusions 1 and 2 may both apply to the same discharge and would be counted toward the proportion reported for each exclusion. Exclusions apply to both 7-day and 30-day versions of the measure.			

Roughly 42 percent of IPF discharges are excluded from the denominator after all exclusions are implemented (Table V.5). One percent of IPF discharges were followed

⁴⁴ Exclusions are defined using the 30-day follow-up period, but exclusions are applied uniformly across the seven-day and 30-day versions of the measure.

by death within 30 days (exclusion 1), approximately 18 percent of IPF discharges were followed by another IPF stay within 30 days (exclusion 2), and 38 percent of IPF discharges were followed by a (non-IPF) inpatient or other residential stay during the 30-day follow-up period (exclusion 3; Table V.5).

After implementing these exclusions, average performance is about 5 points higher for seven-day follow-up (29 percent versus 24 percent) and average performance was about 10 points higher for the 30-day follow-up measure (54 percent versus 44 percent; Table V.6). These differences are largely attributable to exclusion 3, in which IPF discharges are excluded if they are followed by a non-IPF inpatient or other residential stay within 30 days.

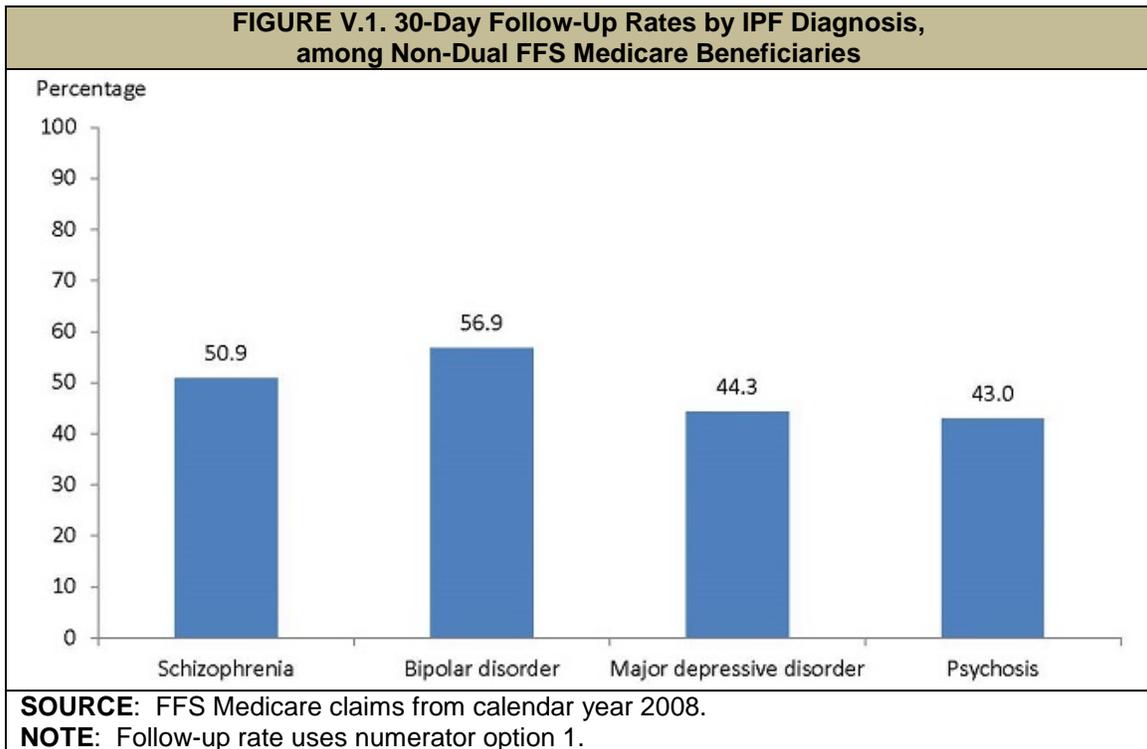
TABLE V.6. Impact of Measure Exclusions on Follow-Up Rates, among Non-Dual FFS Medicare Beneficiaries		
Exclusions	Average 7-Day Follow-Up Rate (%)	Average 30-Day Follow-Up Rate (%)
No exclusions	24.0	43.8
All exclusions	28.7	53.5
Exclusion 1: Death	24.1	44.0
Exclusion 2: Exclude readmission to IPF	25.1	47.0
Exclusion 3: Exclude admission to non-IPF inpatient care or other residential stay	28.6	52.7
SOURCE: FFS Medicare claims from calendar year 2008.		
NOTE: Follow-up rates use numerator option 1. Sample size is 106,139 index discharges among 1,669 IPFs.		

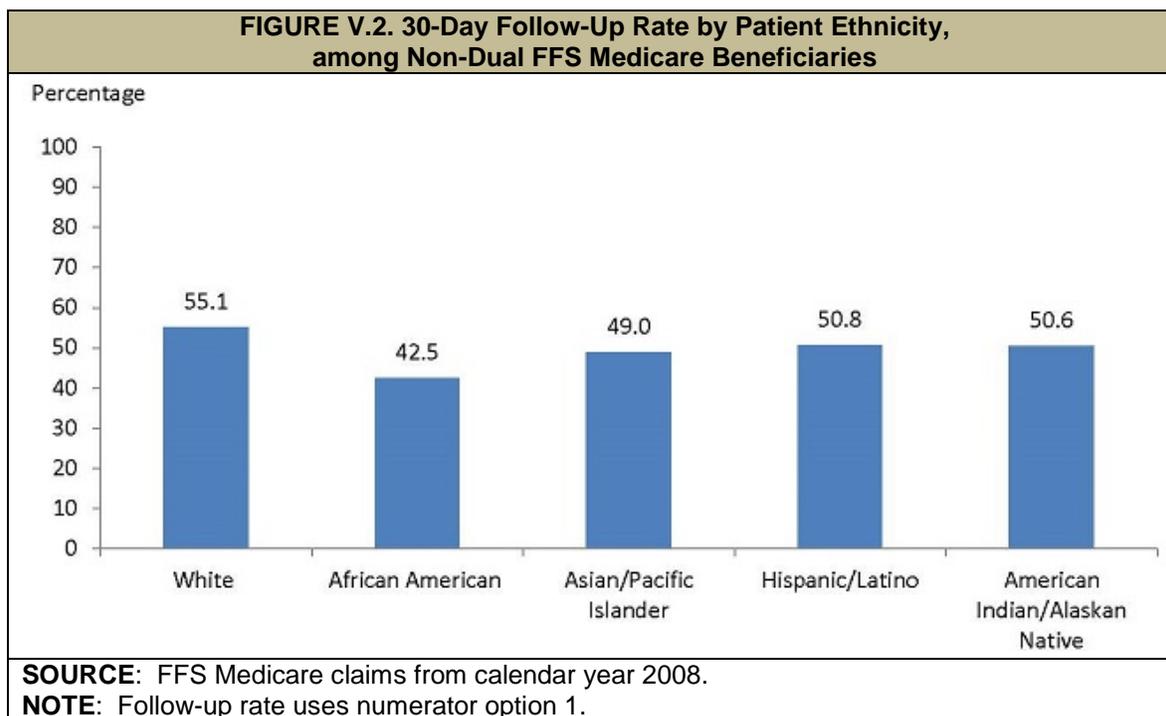
Although these exclusions result in a substantial decrease in the denominator size, the TEP considered all exclusions to be necessary for the face validity of the measure and to maintain consistency with the NQF-endorsed HEDIS measure (which may facilitate more accurate comparison between IPF and health plan performance).

D. Follow-Up after IPF Hospitalization by Beneficiary Characteristics and Geographic Location

IPF performance on the follow-up measure varies significantly according to patient characteristics. On average, seven-day and 30-day follow-up rates are lower for adults ages 18-26 relative to other age groups (Table V.7). In addition, males have lower average rates of follow-up care compared to females, as do African Americans and patients with IPF diagnoses of depression and psychosis (Figure V.1 and Figure V.2). All these differences are statistically significant (however, we are likely to detect some spurious bivariate relationships, given the large sample size).

TABLE V.7. Follow-Up after IPF Hospitalization by Patient Characteristics, among Non-Dual FFS Medicare Beneficiaries				
Characteristic	Average 7-Day Follow-Up	p-Value	Average 30-Day Follow-Up	p-Value
All beneficiaries	28.7		53.5	
Age				
18-26	24.7	<0.01	42.6	<0.01
27-39	30.9		52.9	
40-49	30.3		53.3	
50-64	30.0		53.8	
65-79	31.0		58.9	
80+	23.3		50.6	
Gender				
Male	26.3	<0.01	48.9	<0.01
Female	30.7		57.6	
Race				
White	29.5	<0.01	55.1	<0.01
African American	22.5		42.5	
Asian/Pacific Islander	28.5		49.0	
Hispanic/Latino	27.9		50.8	
American Indian/Alaskan Native	35.3		50.6	
Other	31.7		57.6	
Unknown	31.0		59.2	
Primary Diagnosis				
Schizophrenia	26.7	<0.01	50.9	<0.01
Bipolar disorder	30.7		56.9	
Major depressive disorder	22.6		44.3	
Psychosis	21.3		43.0	
Other	22.7		45.4	
SOURCE: FFS Medicare claims from calendar year 2008.				
NOTE: Follow-up rates use numerator option 1. Sample size 61,871 index discharges among 1,669 IPFs. The average rate presented here is a pooled average across patients, based on their demographic characteristics. Performance can be interpreted as IPF performance for each subpopulation. Statistically significant differences at 5% are in bold.				





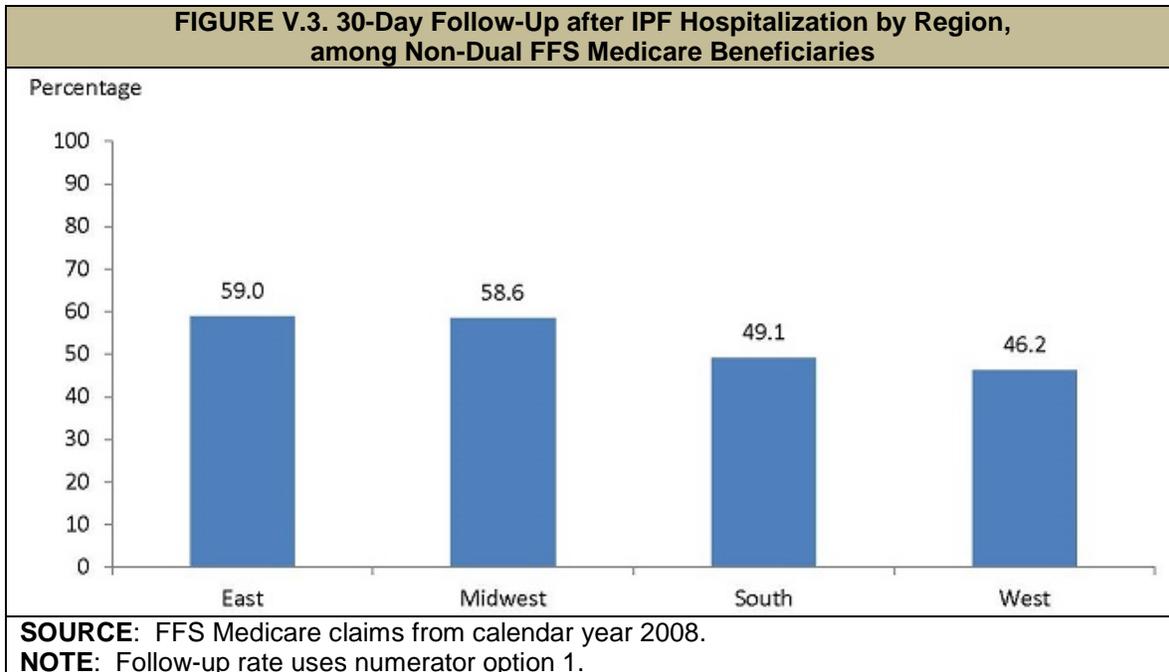
IPF performance on the follow-up measure also varies by state and region (Table V.8 and Figure V.3). On average, seven-day and 30-day follow-up rates are lower in Southern and Western states relative to Eastern and Mid-Western states. States with particularly high follow-up rates include New Hampshire and Vermont, with 74 percent and 72 percent follow-up on the 30-day measure, respectively. States with particularly low rates of follow-up care include Idaho and Alaska, with 36 percent and 37 percent follow-up on the 30-day measure, respectively. However, most of these highest and lowest-performing states had fewer than ten IPFs with FFS Medicare beneficiaries in 2008; therefore, these state-level averages should be interpreted with caution.

TABLE V.8. Follow-Up after IPF Hospitalization by State, among Non-Dual FFS Medicare Beneficiaries

Characteristics	Number of Facilities	Average Number of Discharges Per Facility in 2008	Average 7-Day Follow-Up	Average 30-Day Follow-Up
All facilities	1,669	37.1	28.7	53.5
East	398	37.5	36.0	59.0
New Hampshire	12	36.3	50.0	73.7
Vermont	4	28.3	53.8	71.5
Maine	7	21.9	41.7	65.8
Massachusetts	55	31.5	40.9	67.3
Connecticut	33	34.1	39.4	64.0
Pennsylvania	100	32.6	30.3	57.2
Rhode Island	6	57.8	30.3	55.8
New Jersey	52	46.4	33.3	55.3
New York	129	41.6	36.6	55.1
Delaware	4	80.3	32.9	53.2

TABLE V.8 (continued)				
Characteristics	Number of Facilities	Average Number of Discharges Per Facility in 2008	Average 7-Day Follow-Up	Average 30-Day Follow-Up
Mid-West	429	34.2	29.3	58.6
Michigan	62	53.8	37.2	65.1
Iowa	28	17.4	33.5	62.9
North Dakota	10	24.3	30.0	62.6
Wisconsin	32	32.2	32.5	61.4
South Dakota	4	41.0	23.9	60.1
Indiana	50	20.6	32.1	59.0
Nebraska	9	23.4	35.8	58.8
Minnesota	29	26.2	23.1	58.7
Missouri	44	42.7	26.7	57.1
Kansas	26	21.2	28.2	56.7
Illinois	64	37.9	27.0	55.4
Ohio	71	35.8	23.6	53.6
South	620	41.4	24.6	49.1
West Virginia	14	36.4	30.5	61.2
Virginia	39	44.4	28.4	58.3
North Carolina	41	48.6	27.0	54.7
South Carolina	21	59.2	23.0	54.5
Florida	61	70.3	28.4	52.7
Kentucky	27	31.3	23.5	52.2
Texas	86	53.5	26.4	51.0
Tennessee	49	34.4	20.8	50.8
Alabama	41	32.3	18.0	47.3
Louisiana	76	21.0	33.2	46.3
Georgia	37	68.3	19.2	43.6
Maryland	8	49.5	28.5	43.5
Arkansas	34	19.6	17.4	42.6
Mississippi	49	18.1	20.6	40.8
Oklahoma	33	30.9	16.3	40.0
West	210	29.4	25.6	46.2
Montana	3	26.3	30.2	64.2
Wyoming	6	cs	17.2	64.2
Utah	11	25.1	33.3	57.9
New Mexico	10	27.3	26.7	54.7
Colorado	20	28.4	31.7	52.6
Washington	20	25.8	27.3	52.3
Arizona	18	45.0	27.6	47.0
Hawaii	4	cs	35.5	45.8
Oregon	10	28.1	24.2	43.9
California	88	28.6	24.0	41.1
Nevada	10	60.9	21.0	40.0
Idaho	8	16.6	15.3	36.9
Alaska	2	cs	17.9	35.7
Other	12	39.3	25.5	44.2
District of Columbia	6	26.3	31.4	46.7
Puerto Rico	6	52.3	19.7	41.7

SOURCE: FFS Medicare claims from calendar year 2008.
NOTE: Follow-up rates use numerator option 1. Sample size 61,871 index discharges among 1,669 IPFs. Within each region, states are arranged from highest to lowest performance on the 30-day measure.
cs = Suppressed in adherence to CMS DUA governing use of Medicare data.



E. Reliability Analysis

We conducted a beta-binomial test to examine the reliability of the measure specification using numerator option 1. This test estimates the measure’s ratio of “signal to noise.” The “signal” is the proportion of the variability in measured IPF performance that can be explained by *actual* differences in performance between IPFs, whereas “noise” is any measurement error due to sampling in the time period of interest. This test produces a reliability score ranging from 0 to 1: a score of 0 implies that all the variability in the measure is attributable to measurement error; a score of 1 implies that all the variability is attributable to actual differences in performance.

The measure demonstrated high reliability: 0.94 for the seven-day measure and 0.93 for the 30-day measure. This suggests that the measure can strongly detect actual variation in IPF performance. (We note that beta-binomial statistics are generally high with large sample sizes -- in this case, a large number of facilities.)

Next, we calculated reliability estimates for subgroups of IPFs, based on their number of FFS Medicare discharges per year (Table V.9). Only IPFs with fewer than 11 discharges per year had reliability estimates that dropped below the lower limit of 0.70 that is considered necessary to distinguish the quality of care at one facility from that at another (Adams 2009). This is consistent with CMS rules regarding public reporting, which require that measures based on a sample size of fewer than 11 discharges not be publicly reported.⁴⁵

⁴⁵ This rule is largely in place to protect patient privacy.

TABLE V.9. Follow-Up after IPF Hospitalization by Number of Discharges per Facility, among Non-Dual FFS Medicare Beneficiaries					
Facility Characteristic	Average 7-Day Follow-Up Rate	Reliability	Average 30-Day Follow-Up Rate	Reliability	Number of Facilities
All facilities	28.7	0.94	53.5	0.93	1,669
By number of Medicare FFS discharges per year:					
0-10	19.5	0.72	41.3	0.65	75
11-20	24.1	0.91	49.8	0.88	122
21-30	22.4	0.93	45.5	0.91	112
31-40	26.3	0.94	50.9	0.93	113
41-50	27.5	0.95	53.1	0.94	118
51-60	29.6	0.95	54.8	0.94	108
61-70	33.2	0.93	59.4	0.93	85
71-80	28.8	0.95	54.2	0.95	98
81-90	30.4	0.96	54.4	0.96	92
91-100	31.6	0.95	56.9	0.96	86
101-125	31.4	0.98	56.6	0.98	177
126-150	31.2	0.98	58.8	0.98	123
151-200	29.4	0.98	53.8	0.98	130
201 or more	30.7	0.99	54.6	0.99	230
SOURCE: FFS Medicare claims from calendar year 2008.					
NOTE: Follow-up rates use numerator option 1. Sample size 61,871 index discharges among 1,669 IPFs. Reliability rates below the lower limit of 0.7 are shaded above.					

In addition, we examined the follow-up measure's stability across quarters -- or the extent to which IPFs maintained their performance relative to other IPFs throughout the course of the year. We found that approximately 54 percent of all IPFs remained in the same performance quartile during the first three quarters of 2008.⁴⁶ However, performance rates were more stable in IPFs with a substantial number of FFS Medicare patients during the year. Approximately 80 percent of IPFs with at least 50 FFS Medicare discharges in 2008 remained in the same performance quartile during all three quarters. These findings support the measure's reliability, in that IPF performance is generally stable throughout the calendar year, particularly among IPFs with at least 50 FFS Medicare discharges.

F. Stakeholder Feedback on the Follow-Up Measure

Stakeholder support for the follow-up measure was mixed. Three of the six IPFs involved in testing, and at least 11 of the 25 focus group participants expressed concern that the measure may inappropriately hold IPFs accountable for patient behavior and system inadequacies that are outside of facilities' control, including limited provider availability and long wait-times in some regions. However, at least five focus group participants -- primarily policymakers and measurement experts -- noted that this measure could help to drive innovative partnerships between facilities, community mental health agencies, health plans, and providers to improve follow-up care for IPF patients.

⁴⁶ The last quarter of 2008, from October 1 to December 31, could not be included in this analysis because claims data from January 2009 would have been required to calculate follow-up rates for December 2008.

Although the TEP generally agreed that follow-up for IPF patients was important to measure, two TEP members were quite vocal in expressing the concern that the measure would unfairly hold IPFs accountable for factors outside of their control -- particularly the availability of follow-up care in their community and patients' disposition to keep follow-up appointments. Two other TEP members expressed strong support for the follow-up measure, arguing that it could identify opportunities for quality improvement. One TEP member suggested that this measure should not be publicly reported, given that it could have the unintended consequence of diverting patients from IPFs that provide high quality inpatient services, but have poor performance on the measure due to factors outside their control.

Regarding the measure numerator, focus group participants and TEP members supported specifying follow-up care as services provided by licensed behavioral health practitioners (numerator option 1). They favored this numerator because it closely adheres to the NQF-endorsed HEDIS measure (and would therefore better facilitate comparisons with health plan follow-up rates) and because they felt that patients discharged from an IPF require follow-up care from a mental health professional rather than from a primary care or other provider.

Regarding the measure denominator, focus group participants and TEP members noted that limiting the denominator to non-dual Medicare beneficiaries would exclude a large proportion of IPF patients; therefore, measure performance may not be representative of all patients discharged from IPFs. However, TEP members agreed that non-dual Medicare beneficiaries still represent a sizable portion of all IPF patients and thus constitute "a good place to start" for measuring follow-up care. TEP members stated that IPF patients without a principal mental health diagnosis should be excluded from the measure, given the lack of consensus regarding what follow-up services are most appropriate for patients with dementia, Alzheimer's, or substance abuse diagnoses. This feedback is consistent with the current denominator specification.

G. Summary and Revisions to Follow-Up Measure Specification

The claims-based follow-up measure demonstrated strong quantitative performance; there was good variation in measure performance across IPFs and among patient characteristics. In addition, IPFs' relatively low average performance for all numerator options highlights room for improvement on the measure. The measure also demonstrated very good reliability. Nonetheless, many stakeholders were opposed to the implementation of this measure because they did not think IPFs should be solely accountable for the performance it measures, given the many other community-level factors that could influence patients' receipt of care.

If this measure is to be implemented with Medicare claims, we recommend using numerator option 1 -- an outpatient visit or partial hospitalization with a mental health professional -- to determine the proportion of non-dual Medicare beneficiaries who receive follow-up care (Table V.10). This numerator option most closely adheres to the

NQF-endorsed measure from which this measure was adapted and is consistent with stakeholder and TEP support for follow-up care provided by mental health specialists.

TABLE V.10. Testing Results, Stakeholder Feedback, and Proposed Revisions to the Follow-Up after IPF Hospitalization Measure		
Testing Results	Stakeholder Input	Revisions to Specifications
<ul style="list-style-type: none"> • Substantial room for improvement • Strong variation among IPFs and by patient demographics • Very high reliability 	<ul style="list-style-type: none"> • General stakeholder agreement on the numerator (option 1) and patient population (non-dual eligible beneficiaries) • Mixed stakeholder support, generally related to holding IPFs solely accountable for their patients' follow-up care 	<ul style="list-style-type: none"> • No revisions; numerator option 1 selected for non-dual eligible FFS Medicare population (using only Medicare claims)

To maintain consistency with the NQF-endorsed measure and maintain the face validity of the measure, we recommend no changes to the measure exclusions. However, in combination with the exclusion of dual eligible beneficiary stays, these exclusions substantially reduce the measure denominator to a fraction of all FFS Medicare-paid stays. We discuss this issue in more depth in Chapter VII, which presents a comparison of a chart-based and claims-based version of the follow-up after IPF hospitalization measure.

VI. FOLLOW-UP MEASURE PERFORMANCE USING MERGED MEDICARE-MEDICAID CLAIMS

Given that more than 65 percent of Medicare IPF discharges are dual eligible beneficiaries, we completed a supplemental analysis of the follow-up measure performance using both Medicare and Medicaid claims from calendar year 2008 to calculate follow-up care among all Medicare beneficiaries, including dual eligible beneficiaries. Using both Medicare and Medicaid claims allows us to more accurately calculate follow-up care among dual eligible beneficiaries.

In this chapter, we present analyses of merged Medicare and Medicaid data for dual eligible beneficiaries discharged from an IPF in 2008. In contrast to the analysis in Chapter V, which uses Medicare data from all 50 states, this analysis uses Medicare and Medicaid data from only those states that had: (1) at least 75 percent of dual eligible beneficiaries enrolled in Medicaid FFS in 2008; and (2) provider Specialty codes in 2008 MAX data that could be used to identify outpatient services provided by mental health practitioners. A total of 26 states met these criteria. (See Appendix E for details on the selection of these 26 states.)

In Table VI.1, we present 30-day follow-up rates for dual eligible beneficiaries (using merged Medicare and Medicaid claims) versus non-dual eligible beneficiaries (using only Medicare claims). As illustrated, dual eligible beneficiaries have higher average rates of follow-up care compared with non-dual eligible beneficiaries on all numerator options. However, differences in performance are most pronounced for options 2 and 3; follow-up rates among dual eligible beneficiaries are approximately six points higher than Medicare-only beneficiaries when outpatient care is defined with a principal mental health diagnosis (as opposed to a mental health practitioner in option 1). This higher rate likely reflects the fact that dual eligible beneficiaries have access to a wider range of services under their Medicaid benefits compared with Medicare-only beneficiaries.

The higher performance of options 2 and 3 relative to option 1 among dual eligible beneficiaries is likely attributable, in part, to the fact that several states do not provide sufficient information in Medicaid claims to identify all mental health providers (relevant to option 1). Similarly, option 3 likely has slightly higher average performance than options 1 and 2 due to the fact that state Medicaid programs use many procedure codes besides those in the HEDIS FUH measure to reimburse outpatient behavioral health care (relevant to options 1 and 2); outpatient visits using state-specific procedure codes that do not appear in the FUH measure could qualify as follow-up care under option 3, provided that they have a principal mental health diagnosis. Given that option 3 is not affected by the two primary data concerns related to Medicaid -- a lack of complete information on providers, as well as state-specific procedure codes -- it is likely the most

attractive numerator option for measuring follow-up care among all FFS Medicare beneficiaries with combined Medicare and Medicaid claims.

TABLE VI.1. Facility Performance by Numerator Option: Follow-Up within 30 Days of IPF Hospitalization among Dual and Non-Dual Eligible Beneficiaries			
Numerator Option	Dual Eligible (%)	Non-Dual Eligible (%)	Difference
Option 1: Mental health practitioner + CPT/HCPCS Code (HEDIS approach)	54.7	54.3	0.4
Option 2: Principal mental health diagnosis + CPT/HCPCS Code	58.7	52.9	5.8
Option 3: Principal mental health diagnosis + Outpatient visit	61.6	55.4	6.2
Option 4: Any outpatient visit	84.3	80.0	4.3
SOURCE: Medicare and Medicaid claims from calendar year 2008.			
NOTE: The sample of IPF stays is restricted to those 26 states with complete and reliable MAX data and at least 75% of dual eligible beneficiaries enrolled in FFS Medicaid.			

We also conducted an analysis of the stability of performance rates when Medicaid claims are used to supplement Medicare claims for all beneficiaries, including dual eligible and non-dual eligible beneficiaries.⁴⁷ This analysis is also based on Medicare and Medicaid data from the 26 states that had at least 75 percent of dual eligible beneficiaries enrolled in FFS Medicaid and MAX data that could be used to identify outpatient services provided by mental health practitioners. As shown in the Table VI.2, performance rates for all four numerator options were between 1 and 4 percentage points higher when the measure numerator was calculated using Medicaid claims in addition to Medicare claims. This increase in performance is due solely to additional follow-up services provided to dual eligible beneficiaries, as measured by Medicaid claims.

⁴⁷ This analysis used a fixed population of Medicare beneficiaries to ensure that measured differences in performance were the result of supplementing Medicare claims with Medicaid claims.

TABLE VI.2. Follow-Up within 30 Days of IPF Hospitalization among All Medicare Beneficiaries, by Data Source			
Numerator Option	All Beneficiaries		Difference
	Using Only Medicare Claims	Using Medicare and Medicaid Claims	
Option 1: Mental health practitioner + CPT/HCPCS Code (HEDIS approach)	54.3	55.0	0.7
Option 2: Principal mental health diagnosis + CPT/HCPCS Code	53.5	57.2	3.7
Option 3: Principal mental health diagnosis + outpatient visit	56.7	59.9	3.2
Option 4: Any outpatient visit	81.2	83.3	2.1
SOURCE: Medicare and Medicaid claims from calendar year 2008.			
NOTE: The sample of IPF stays is restricted to those 26 states with complete and reliable MAX data and at least 75% of dual eligible beneficiaries enrolled in FFS Medicaid.			

VII. COMPARISON OF A CHART VERSUS CLAIMS-BASED APPROACH TO THE FOLLOW-UP AFTER IPF HOSPITALIZATION MEASURE

Although CMS anticipates implementing the claims-based follow-up measure as part of IPFQR program in 2015, the measure may be transitioned to draw on other data sources in future years, including patient medical records or other administrative data sources (such as electronic patient tracking systems or registries). We conducted an exploratory analysis to examine the feasibility of measuring follow-up care using other data sources. Our analysis included comparing denominator sample sizes if different data sources were used, and gathering qualitative feedback on the strengths and limitations of different data sources and measurement approaches. This analysis is designed to provide CMS and other stakeholders with insights into the advantages and disadvantages of a chart-based approach to the follow-up measure, relative to a claims-based approach. In this chapter, we use the term “chart-based” to refer to a range of data sources available to IPFs, including electronic records, patient charts, and administrative data. These analyses should be interpreted with caution, given the limited number of stakeholders and IPFs included in the analysis.

A. Methods

In 2014, Mathematica and NCQA held three focus groups with administrative staff from nine IPFs, as well as debriefing sessions with each of the six IPFs that piloted the four chart-based measures (Table VII.1). During these conversations, we gathered feedback on IPFs’ efforts to encourage and track follow-up care, including factors that facilitate and constrain IPFs’ ability to capture accurate data on their patients’ follow-up care. In addition, we asked about the feasibility of a chart-based approach to the follow-up measure, including data collection and reporting burden, as well as infrastructure and resources that would be necessary to support reporting. Finally, we asked participants questions about their patients’ insurance coverage and demographics to provide additional context for quantitative analyses discussed below.

	Freestanding Facilities	Psychiatric Wards
Private facilities	8	3
Public facilities	4	0

In addition, we used administrative data provided by the six IPFs that participated in measure testing to conduct a quantitative analysis of patient characteristics and IPF discharges, with the goal of comparing and contrasting the denominator sizes and characteristics of a chart-based versus a claims-based approach to the follow-up

measure. These six IPFs included one private freestanding hospital, two public freestanding hospitals, and three private psychiatric wards (included in Table VII.1). Below we present our qualitative and quantitative findings of this analysis.

B. IPFs' Efforts to Encourage and Track Follow-Up Care

Encouraging Follow-Up Care. Many of the IPFs that participated in focus groups and debriefing sessions have basic processes in place to facilitate patients' follow-up care after they are discharged. Nearly all of the 15 participating IPFs stated that they schedule patients' follow-up appointments prior to discharge, and five IPFs noted that they coordinate with outpatient providers on a regular basis to encourage patients' continuity of care. However, IPFs vary considerably in the level of effort and resources allocated to such efforts. For example, one IPF schedules follow-up appointments for all psychiatric patients, but makes no additional efforts to encourage or track follow-up care. Other IPFs reported much more labor-intensive processes, in which IPF staff schedule aftercare appointments prior to patient discharge, arrange for patients to meet aftercare providers directly preceding or following discharge, and then contact patients by phone to remind them of upcoming appointments.

"We have very close relationships with our care providers. We will provide bus passes for patients to get to the next care provider. We buy [them] clothes. We do a lot to keep patients from being readmitted."

- Representative of a public freestanding IPF

Around half of IPFs indicated some effort to contact patients after discharge, generally by phone, to check on their status and encourage them to attend follow-up appointments. IPF staff noted that these follow-up efforts require a large investment of staff time and resources. For example, one facility is working to dedicate a full-time staff person to making calls to patients in advance of their scheduled follow-up appointments. Although a small number of IPFs attempt some follow-up with all discharged patients, other facilities' social work teams focus their limited resources more narrowly on contacting individuals most likely to miss follow-up appointments or be readmitted to the IPF.

Examples of IPFs' Efforts to Coordinate Follow-Up Care

As part of its quality improvement strategy, one IPF has entered into a partnership with a local university to develop and implement a pilot project that uses innovative technologies such as instant phone messaging to remind discharged patients about upcoming aftercare appointments. The hospital also employs certified peer specialists to conduct follow-up outreach by phone, primarily to encourage high-risk patients to keep follow-up appointments.

In addition, a private freestanding IPF recently opened a transitional outpatient clinic to promote its patients' continuity of care. The clinic uses motivational interviewing to identify and mitigate patients' barriers to outpatient care -- including transportation difficulties and motivational constraints. The IPF noted a substantial improvement in follow-up care linked to patients' engagement with clinic staff.

Tracking Follow-Up Care. Although most IPFs reported having processes in place to schedule follow-up appointments or track whether patients received reminder calls prior to their scheduled appointments, it was much less common for IPFs to track whether or not patients actually *received* follow-up care. Five out of 15 IPFs reported using simple information systems to track whether follow-up care occurred, based on information provided by patients, community partners, managed care organizations (MCOs), or some combination of these sources. One public hospital, for example, has built a form into its electronic medical record to capture who was contacted and what information was gathered regarding each patient's follow-up status. In contrast, two public IPFs mentioned that although their social workers schedule aftercare appointments for patients prior to discharge, the IPFs are not permitted to access patients' files after discharge, thereby eliminating opportunities for patient follow-up and tracking.⁴⁸

"We're trying to figure out how to get a dedicated staff member to do follow-up calls after someone is discharged. The other piece [to tracking follow-up care] is building relationships with MCOs to share data, trying to get more information on patients' follow-up care."

- Representative of a private freestanding IPF

IPFs mentioned partnerships with external providers and stakeholders as a key facilitator of tracking patient follow-up (see Table VII.2 for a summary of all facilitators mentioned by more than one IPF). In particular, public IPFs noted strong working relationships with community providers, and mentioned that such relationships enable them to follow discharged patients closely. Private IPFs noted that data sharing with partner organizations was critical to their efforts to track patients' care. Three IPFs mentioned collaborating with MCOs to receive and share data related to follow-up care. IPFs also indicated the necessity of well-established patient consent or release processes to allow IPF staff to contact patients and share information with providers.

Examples of IPFs' Efforts to Track Follow-Up Care

One IPF works closely with the largest outpatient providers and MCOs in its region to construct follow-up and readmission rates for a subset of patients. The goal of these analyses is to identify high-risk and high-utilizing IPF patients and to focus staff resources on encouraging these patients to keep follow-up appointments. Another IPF systematically tracks whether patients readmitted to the hospital kept their initial follow-up outpatient appointments as planned. The representative reported that most patients readmitted to the IPF within 30 days did not keep their follow-up appointment. In both cases, IPFs track only a subset of their full patient population -- namely those individuals for whom data on follow-up care is available.

⁴⁸ Staff's reported lack of access to patient files appears to reflect administrative procedures that block access to patient records following discharge. In at least one IPF, these procedures appear to be designed to protect patient privacy following their stay.

TABLE VII.2. Commonly Cited Facilitators to Tracking Follow-Up Care	
Comment	Frequency Mentioned
Partnerships with external providers, particularly community mental health care providers	9
Collaborations with MCOs and other payers	5
Well-established patient consent and release process	3
Information-sharing systems with in-network providers	3

C. IPF Perspectives on a Chart-Based Approach to Measuring Follow-Up Care

IPFs participating in focus groups were strongly opposed to a chart-based approach to measuring follow-up care, citing sizable obstacles related to data availability and human resources, discussed below and summarized in Table VII.3.

TABLE VII.3. Commonly Cited Constraints to Tracking Follow-Up Care	
Comment	Frequency Mentioned
Tracking and reporting would require hiring or dedicating additional staff	7
Patient consent/release process would be burdensome, or privacy would be a concern	7
Obtaining complete data would be difficult due to patient non-response	4
Monitoring patients outside of IPF networks would be challenging	4
Tracking would require changes to electronic health records or other systems	3
Tracking via telephone would impose a considerable time burden on IPF staff	2
Tracking via telephone would require significant expense	2

Data Availability. More than half of IPFs reported that they do not have a data collection infrastructure to accurately capture the actual receipt of follow-up care in the community, and that creating such an infrastructure would require substantial financial resources. In particular, staff from several IPFs stated that accessing and tracking data on patients that are outside their network would be extremely difficult. IPFs noted that changes to existing data systems and administrative procedures would be necessary to accommodate a chart-based follow-up measure, as patient records or other data tracking systems would require new data elements and modified processes to remind staff to populate these elements following patient discharge. IPF staff noted that these modifications are not trivial, and would likely be expensive to implement, both in terms of information technology investments and staff training.

“If the follow-up was with an in-network provider, it would not be difficult to determine whether follow-up occurred, but with an outside provider, I’m not sure how we would get information about whether a patient was seen.”

- Representative of a private psychiatric ward

Also related to data availability, several IPFs noted that some patients refuse to sign information release forms upon discharge. As such, IPFs cannot contact these patients or their outpatient providers to determine whether follow-up care occurred.

Human Resources. More than half of IPF respondents noted that tracking and reporting follow-up care for the full patient population would require a significant investment of staff time, particularly to determine if scheduled follow-up appointments were kept. IPFs cited the transient nature of their patient populations as a major barrier to locating them in the community, and noted that some patients do not have phones or are reluctant to maintain contact with the IPF. Three facilities mentioned that they would likely have to hire additional staff to conduct systematic patient follow-up by phone if a chart-based follow-up measure were introduced to the IPFQR program.

"We're developing relationships with community health providers. The availability of follow-up data [with those providers] is pretty much there...But for those we don't have a relationship with, we'd have to add [follow-up calls] to the social workers' responsibilities. That would be so much additional work.

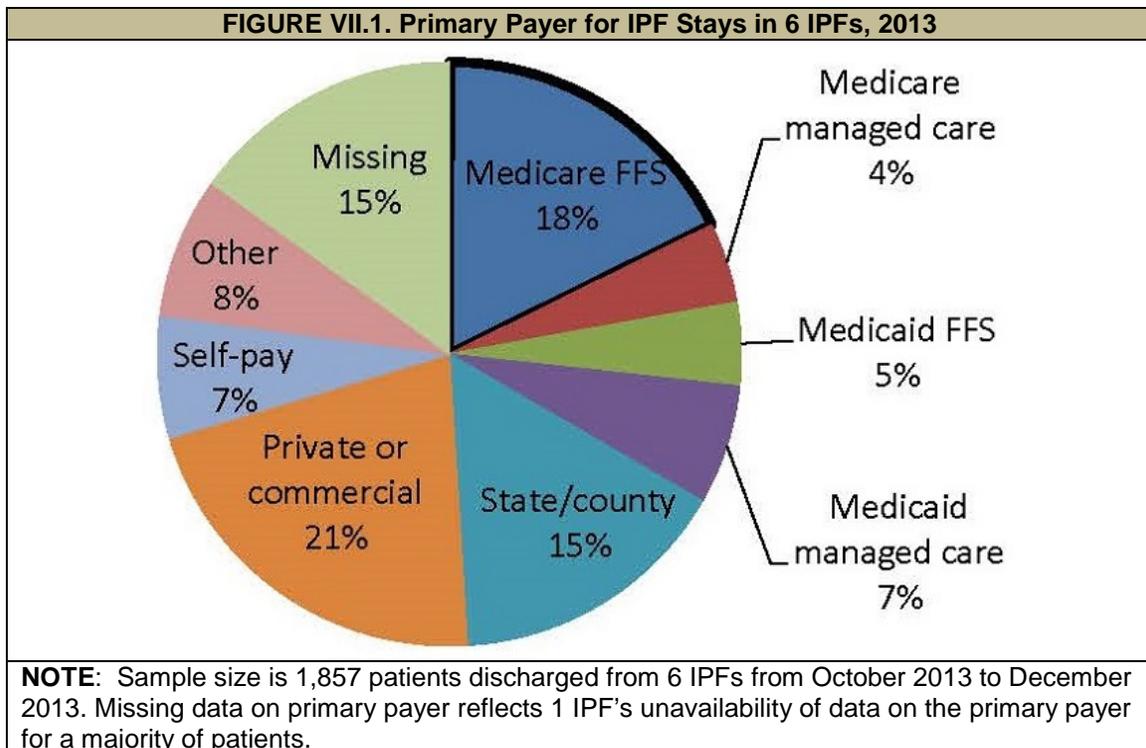
- Representative of a public freestanding hospital

The IPFs that already conduct patient and provider outreach indicated that obtaining follow-up information is labor-intensive and requires a considerable amount of staff time and effort. One IPF noted that follow-up calls to patients do not generate a strong return in terms of the amount of information gained per hour of staff time devoted to the task. Commonly, IPF staff will make multiple phone calls over the course of several days to determine if a patient attended a single follow-up appointment.

Citing these feasibility concerns, IPF staff heavily favored a claims-based approach over a chart-based approach to the follow-up measure. IPFs overwhelmingly viewed the claims-based measure's minimal burden (to IPFs) as a primary advantage over a chart-based follow-up measure, which several IPFs viewed as "impossible" to implement under current conditions. However, as described below, some IPFs had concerns about the representativeness of a claims-based approach that uses only FFS Medicare beneficiaries as the measure denominator.

D. Analysis of Insurance Coverage, Patient Demographics, and Sample Sizes

In this section, we present analyses of patient characteristics and IPF discharges, with the goal of comparing and contrasting the denominator sizes and characteristics of a chart-based versus a claims-based approach to the follow-up measure. For this analysis, we use administrative data from the six IPFs that participated in measure testing under the contract, corresponding to all discharges at these IPFs from October 1, 2013, to December 31, 2013.



Insurance Coverage. We collected information regarding patients' insurance coverage from the six IPFs that participated in chart-based measure testing. Three of these IPFs were private psychiatric wards with fewer than 50 beds, two were public freestanding facilities with over 100 beds, and one was a private freestanding facility with 400 beds.⁴⁹ As illustrated in Figure VII.1, FFS Medicare was the primary payer for only 18 percent of patient stays at the six IPFs. Because dually eligible Medicare-Medicaid beneficiaries account for over half of FFS Medicare-paid stays,⁵⁰ we can estimate that non-dual FFS Medicare beneficiaries accounted for less than 10 percent of stays at these IPFs.⁵¹ This is an important statistic, because a claims-based approach to measuring follow-up care will likely rely on claims for non-dual FFS Medicare patients -- a relatively small portion of IPFs' total patient population. In contrast, a chart-based approach to the measure would potentially draw from all patients, including patients covered by FFS Medicare, Medicare managed care, Medicaid, state or county payment sources, private insurance, and the uninsured. This stronger representativeness -- or "generalizability" of the chart-based follow-up measure to the entire IPF population -- represents a potential advantage of a chart-based specification versus a claims-based specification that uses only FFS Medicare data.

⁴⁹ Additional statistics on these IPFs are available in Table IV.2.

⁵⁰ Based on 2008 Medicare data, we found that dually eligible beneficiaries accounted for 59 percent of IPF stays paid by FFS Medicare. A recent MedPac report had similar findings: MedPac (2012). A Data Book: Health Care Spending and the Medicare Program, Section 6. <http://www.medpac.gov/chapters/Jun12DataBookSec6.pdf>.

⁵¹ IPFs were only able to report the primary payer for hospital stays. Because IPFs were not able to systematically report patients' full insurance coverage, Medicare and Medicaid dual eligible status was not directly measurable using information provided by the IPFs. As a result, we apply the basic assumption that approximately half of FFS Medicare beneficiaries that used IPFs were dually eligible, based on findings from our claims analysis.

However, if FFS Medicare beneficiaries are not systematically different from IPFs' total patient population, a claims-based approach would likely yield similar performance results to a chart-based measure that draws on a broader denominator population. Below, we examine potential differences between FFS Medicare beneficiaries and IPFs' total patient population in more depth.

TABLE VII.4. Comparison of Patient Demographics for FFS Medicare IPF Discharges versus All IPF Discharges				
Characteristics	Stay Covered by FFS Medicare		All IPF Discharges	
	N	%	N	%
Age at Discharge				
Under 18	0	0	103	5
18-26	15	5	388	21
27-44	86	26	630	34
45-64	135	41	589	32
65 and older	93	28	147	8
Primary Diagnosis				
Schizophrenia	107	33	361	19
Bipolar disorder	109	33	704	38
Delusional disorder	29	9	134	7
Major depressive disorder	4	1	90	5
Psychosis	19	6	111	6
Alcohol or drug dependency	11	3	151	8
Alzheimer's/dementia/degeneration	20	6	37	2
Other	18	5	206	11
Missing	12	4	63	3
Race/Ethnicity				
African American	65	20	389	21
Caucasian	234	71	1,120	60
Other	5	2	57	3
Missing	25	8	291	16
Gender				
Male	166	50	964	52
Female	163	50	893	48
Length of Stay				
0, 1, or 2 days	25	8	248	13
3 or 4 days	43	13	320	17
5, 6, or 7 days	77	23	454	24
8 to 14 days	82	25	466	25
15 to 21 days	37	11	129	7
22 to 30 days	20	6	70	4
>30 days	45	14	170	9
Sample Size	329	100	1,857	100

NOTE: Sample size is 1,857 patients discharged from 6 IPFs from October 2013 to December 2013.

Patient Demographics. Table VII.4 presents a comparison of beneficiary demographics for patients whose IPF stays were paid by FFS Medicare versus IPFs' full patient populations. As illustrated, patients with FFS Medicare stays were different from IPFs' total patient population in several respects. First, they are generally older than the full patient population, with nearly 70 percent of FFS Medicare patients over the age of 43, compared to 40 percent among all IPF patients. FFS Medicare patients are also more likely to have schizophrenia as a principal diagnosis than the full patient population (33 percent versus 19 percent in the full population), and more likely to have IPF stays of over two weeks (31 percent versus 20 percent of all patients). These data suggest that FFS Medicare patients are systematically different from the full IPF patient population. As such, it is possible that FFS Medicare patients could have different rates of follow-up care than non-FFS patients, related to demographic characteristics or

length of stay. (For example, it is possible that IPFs may have more opportunities to arrange follow-up care for FFS Medicare patients with longer lengths of stay if discharge planning begins upon admission). If this were the case, a claims-based follow-up measure would either underestimate or overestimate follow-up care among IPFs' full patient population.

Stakeholder Feedback on Medicare Beneficiaries

Qualitative input from IPFs involved in focus groups and debriefing sessions suggests potential systematic differences between patients covered by Medicare and patients with other forms of insurance, as well as the potential for these differences to affect rates of follow-up care -- although the direction of the potential bias is unclear. One IPF stated that Medicare patients tend to have more co-morbidities than other patients, and these co-morbidities could serve as obstacles to outpatient follow-up care. In addition, representatives from one IPF expressed concern that many outpatient providers do not accept Medicare patients, potentially generating lower follow-up rates for a claims-based measure specification (drawing only from Medicare beneficiaries) versus a chart-based specification (drawing from all IPF patients). Staff from another two IPFs could not estimate whether Medicare patients would have higher or lower rates of follow-up care compared with other patients, but they conjectured that Medicare patients would likely have more outpatient visits and thus higher rates of follow-up care than uninsured patients.

Sample Sizes. We analyzed IPF administrative data to estimate the number of IPFs for which a chart-based approach to calculating the follow-up measure would result in larger sample sizes than those generated under a claims-based approach that only uses non-dual FFS Medicare patients as the denominator. A chart-based follow-up measure would likely use a sampling method that is similar to chart-based measures in the HBIPS measure set.⁵² Under this method, IPF staff or an IPFQR program vendor would select a random sample of discharged patients as the measure denominator. Following the HBIPS algorithm, a chart-based sampling approach to the follow-up measure would require that a minimum of 20 percent of patients with a mental health diagnosis be sampled on a quarterly basis.⁵³ We compared this minimum chart-based sample to the estimated sample size for a claims-based follow-up measure corresponding to the same time period: one-half of the total number of FFS Medicare beneficiaries with a mental health diagnosis that were served by IPFs. We divide the total number of FFS Medicare beneficiaries in half because an estimated 50 percent of these patients are likely dually eligible beneficiaries and would thus be excluded from the claims-based measure denominator.⁵⁴

⁵² TJC is the steward for these measures. For more information, visit <http://www.jointcommission.org/hospital-based-inpatient-psychiatric-services/>.

⁵³ Consistent with the HEDIS FUH measure, the IPF follow-up measure's denominator is composed of patients with a principal mental health diagnosis. This includes patients with diagnoses of schizophrenia, bipolar disorder, and depression, but excludes patients with diagnoses of substance use disorders, Alzheimer's, and dementia.

⁵⁴ As many as 63 percent of FFS Medicare stays may involve dual beneficiaries, on average; however, we use a 50 percent estimate for the sake of simplicity. As such, claims-based measure sample sizes presented below can be interpreted as the maximum potential sample size under a claims-based approach that uses only FFS Medicare beneficiaries as the denominator.

TABLE VII.5. Quarterly Sample Sizes for the Follow-Up after IPF Hospitalization Measure							
	All Sites	IPF 1	IPF 2	IPF 3	IPF 4	IPF 5	IPF 6
All discharges	1,857	152	272	172	382	409	470
All discharges with a mental health diagnosis	1,400	109	203	136	347	365	240
Minimum chart-based sample size^a	280	22	41	27	69	73	48
FFS Medicare discharges	329	42	15	28	94	106	44
FFS Medicare discharges with a mental health diagnosis	268	34	14	26	75	101	18
Estimated claims-based sample size^b	134	17	7	13	38	51	9
Percentage of all patients covered by FFS Medicare	18%	28%	6%	16%	25%	26%	9%
NOTE: Sample size is 1,857 patients discharged from 6 IPFs from October 2013 to December 2013.							
a. 20% of all IPF discharges with a mental health diagnosis.							
b. 50% of all IPF FFS Medicare discharges with a mental health diagnosis, assuming that dual beneficiaries account for approximately half of all FFS Medicare discharges.							

As illustrated in Table VII.5, sample sizes for a chart-based follow-up measure are larger than projected sample sizes for a claims-based follow-up measure for all six IPFs, particularly for two IPFs that had less than 10 percent of stays covered by FFS Medicare (IPF 2 and IPF 6). For these two IPFs, the quarterly sample size for a claims-based measure is less than ten patients, whereas the sample size for a chart-based version of the measure is more than 40 patients. This illustrates a primary disadvantage of a claims-based follow-up measure, in that IPFs serving a small number of FFS patients relative to their full patient population can generate quarterly sample sizes that do not meet minimum CMS requirements for public reporting.

Stakeholder Feedback Related to Follow-Up Measure Sample Size

IPF staff who participated in focus groups -- representing freestanding hospitals and psychiatric wards, as well as public and private facilities -- confirmed that FFS Medicare beneficiaries represent a minority of their full patient population. A private hospital noted that approximately 10-12 percent of its patient population has Medicare coverage, saying, "It's only a sliver of the individuals we work with." Most IPFs involved in focus groups made similar statements regarding the proportion of their full patient population covered by Medicare.

E. Conclusion

Based on this qualitative and quantitative analysis, a chart-based approach to tracking and reporting follow-up care would be difficult for IPFs to implement in the short term, given the constraints they noted related to data availability and IPF human and financial resources. However, a chart-based approach to the follow-up measure could be more generalizable to the full IPF patient population, assuming that data on follow-up care could be collected for in-network as well as out-of-network patients. In addition, a chart-based follow-up measure would likely generate fewer instances in which small sample sizes preclude public reporting.

A claims-based approach to follow-up care would be more feasible in the short term, primarily because it imposes no burden on IPFs. In addition, a claims-based follow-up measure can precisely measure follow-up care among FFS Medicare beneficiaries, as it draws from all Medicare claims.⁵⁵ Table VII.6 provides a high-level summary of the relative advantages and disadvantages of these two approaches to the follow-up measure.

TABLE VII.6. Advantages and Disadvantages of Chart and Claims-Based Approaches to the Follow-Up after IPF Hospitalization Measure		
Approach	Advantages	Disadvantages
Chart-based	<ul style="list-style-type: none"> • Strong generalizability across all patients, assuming data availability • Potentially fewer inherent sample size constraints 	<ul style="list-style-type: none"> • Substantial reporting burden for IPF staff • Incomplete data for out-of-network and hard-to-reach patients
Claims-based	<ul style="list-style-type: none"> • No reporting burden for IPFs • Complete data for Medicare reimbursed mental health services for FFS Medicare population 	<ul style="list-style-type: none"> • Weak generalizability across all IPF patients • Small sample sizes that could preclude public reporting

⁵⁵ However, Medicare beneficiaries could still obtain mental health services outside of Medicare reimbursement, such as from community mental health centers or physicians that do not accept Medicare. These follow-up services would not be captured in a claims-based measure.

VIII. CONCLUSIONS AND LESSONS

The measures developed in this project are an attempt to fill gaps and improve upon existing screening measures for an inpatient psychiatric setting. At the time of this report, CMS has not made final decisions about these measures' inclusion in IPFQR program, and none of the measures has been submitted to NQF.

Of the screening measures, the one with the most potential for improving the quality of IPF care appears to be the metabolic screening measure, which has average performance of 42 percent across IPFs. Although the suicide, violence, and substance use screening measures had generally high performance among the six IPFs in which they were piloted, they also exhibit potential to standardize screening elements across IPFs. CMS may collaborate with TJC to harmonize these screening measures with HBIPS-1. Such harmonization efforts would ensure that new measures do not place undue burden on IPFs.

This analysis found substantial room for improvement with respect to follow-up after IPF hospitalization, with 30-day follow-up rates below 55 percent. In its current specification as a Medicare claims-based measure, the follow-up measure poses some concerns regarding its generalizability to the full IPF population. However the alternative of a chart-based specification does not appear feasible to implement in the short term, due to data availability and human resource constraints.

Together with existing IPFQR program measures, these new measures provide a strong foundation for monitoring and improving the quality of inpatient behavioral health care; however, some gaps persist, particularly between the care patients experience and the outcomes they report. Unlike screening measures (that assess a specific process of care within the IPF), some of these measurement concepts (like the follow-up measure) may require some type of shared accountability between actors because they deal with system-level issues of access to care, insurance coverage, and the care continuum spanning inpatient and outpatient care. There is a tension associated with developing new measures that are premised upon this shared responsibility, particularly because measures are often confined to one unit of analysis.

Future measure-development efforts for the IPFQR program should attempt to involve as many IPFs and different types of stakeholders as possible, to understand variation in performance across different types of facilities, patient populations, and in different community contexts. Measure-development should also draw on the latest studies and clinical guidelines, as substantive advancements in the field occur on a regular basis.

REFERENCES

- Adams, John (2009). *The Reliability of Provider Profiling*. RAND.
- American Diabetes Association (2006). *Antipsychotic Medications and the Risk of Diabetes and Cardiovascular Disease*. Available at: [http://professional.diabetes.org/admin/UserFiles/file/CE/AntiPsych%20Meds/Professional%20Tool%20%231\(1\).pdf](http://professional.diabetes.org/admin/UserFiles/file/CE/AntiPsych%20Meds/Professional%20Tool%20%231(1).pdf).
- Casey, D.E. "Dyslipidemia and atypical antipsychotic drugs." *J Clin Psychiatry*, 2004; 65 Suppl(18): 27-35.
- American Diabetes Association and American Psychiatric Association. "Consensus development conference on antipsychotic drugs and obesity and diabetes." *Diabetes Care*, February 2014, 27(2).
- Institute of Medicine, Committee on Crossing the Quality Chasm: Adaptation to Mental Health and Addictive Disorders (2006). *Improving the Quality of Health Care for Mental and Substance-Use Conditions*. Washington, DC: National Academies Press.
- Landis J.R., and G.G. Koch. "The measurement of observer agreement for categorical data." *Biometrics*, 1977, 33: 159-174.
- Marder. S.R., S.M. Essock, A.L. Miller, R.W. Buchanan, D.E. Casey, J.M. Davis, J.M. Kane, et al. "Physical health monitoring of patients with schizophrenia." *American Journal of Psychiatry*, August 2004, 161(8): 1334-1349.
- MedPac (2012). *A Data Book: Health Care Spending and the Medicare Program*, Section 6. Available at: <http://www.medpac.gov/chapters/Jun12DataBookSec6.pdf>.
- NRI's Behavioral Healthcare Performance Measurement System (2012). *History of HBIPS Core Measure Set*. 2012. Available at: <http://www.nri-inc.org/projects/bhpms/docs/HBIPSHistory.pdf>. Accessed on July 1, 2013.
- Prela, C., G. Baumgardner, G. Reiber, L. McFarland, C. Maynard, N. Anderson, and M. Maciejewski. "Challenges in merging Medicaid and Medicare databases to obtain healthcare costs for dual-eligible beneficiaries." *PharmacoEconomics*, 2009, 27(2): 167-177.
- Roohafsa, H, A. Khani, H. Afshar, A. Garakyaraghi, and B. Ghodsi. "Lipid profile in antipsychotic drug users: A comparative study." *ARYA Atheroscler*. May 2013; 9(3): 198-202.
- TJC (2012). *Specifications Manual for Joint Commission National Quality Measures*. Available at: <https://manual.jointcommission.org/releases/TJC2013A/HospitalBasedInpatientPsychiatricServices.html>. Accessed on January 1, 2015.

APPENDIX A. IPF TECHNICAL EXPERT PANEL MEMBERS

TABLE A.1. IPF Technical Expert Panel (TEP) Members

Name	Title	Organization/Agency
Frank Ghinassi	Vice President, Quality and Performance Improvement	Western Psychiatric Institute
Eric Goplerud	Senior Vice President and Director of Substance Abuse, Mental Health, and Criminal Justice Studies	NORC
Richard Hermann	Director, Center for Quality Assessment and Improvement in Mental Health	Institute for Clinical Research and Health Policy Studies, Tufts
Mary E. Johnson	Professor	Rush University/American Psychiatric Nurses Association
Kathleen McCann	Director, Quality and Regulatory Affairs	National Association of Psychiatric Health Systems
Lucille Schacht	Senior Director of Performance and Quality Improvement	National Association of State Mental Health Program Directors Research Institute
Elizabeth Stallings	Former Chief Operating Officer	John Muir Behavioral Health Center
Ann Watt	Associate Director, Department of Quality Measurement, Division of Quality Measurement and Research	The Joint Commission
Richard Wohl	President, Princeton House Behavioral Health	Princeton HealthCare System
Joel Streim*	Professor	Department of Psychiatry, University of Pennsylvania School of Medicine
Alice Lind*	Senior Clinical Officer	Center for Health Care Strategies

* Did not attend final TEP meeting.

APPENDIX B. SCREENING MEASURE SPECIFICATIONS

TABLE B.1. Measure Specifications: Screening for Risk of Suicide

Measure Dimension	Description
Description	Percentage of discharges from an IPF for which a structured suicide screening for 5 elements was completed.
Denominator	Psychiatric inpatient discharges during the measurement period.
Numerator	Suicide risk screening completed within the first day of admission.
Numerator Details	<p>Screening Content The medical record must provide documentation that information was obtained--either from the patient or from a collateral source--regarding the following 5 topic areas: (1) suicidal ideation; (2) the extent of plans or preparation (if ideation is reported); (3) the intent to act on those plans (if plans are reported); (4) past suicidal behavior; and (5) risk factors and protective factors.</p> <p>Timing of Screen All screening content must be obtained within the first day of admission (any time on the day of admission or the following day).</p> <p>Screening Administration Screening must be completed by a qualified psychiatric practitioner. The titles of qualified psychiatric practitioners may vary from state to state. Written and electronic collateral information and information provided by the patient in intake forms is acceptable if it has been reviewed by a qualified psychiatric practitioner.</p>
Exclusions	<p>Patient stays for which a screening could not be completed within the first day of admission due to the patient's enduring unstable medical or psychological condition.</p> <p>Patient stays with a length of stay equal to or greater than 365 days, or less than 1 day.</p> <p>Patient stays with multiple admissions to psychiatric units during a single hospitalization.</p>
Stratification and Risk Adjustment	The measure is currently stratified by age in to 4 categories: children (age 1-12), adolescents (age 13-17), adults (age 18-64) and older adults (age 65+). No risk adjustment is planned.
Sampling	These measures will rely on a sampling methodology, by which cases are sampled quarterly or monthly. Facilities will be required to sample at least 20% of each stratum population for the quarter or month, with a minimum of 15 cases in each stratum per month. (Sampling will follow HBIPS-1 sampling guidelines.)

TABLE B.2. Measure Specifications: Screening for Risk of Violence	
Measure Dimension	Description
Description	Percentage of discharges from an IPF for which a structured violence screening for 2 elements was completed.
Denominator	Psychiatric inpatient discharges during the measurement period.
Numerator	Violence risk screening completed within the first day of admission.
Numerator Details	<p>Screening Content The medical record must provide documentation that information was obtained--either from the patient or from a reliable source--regarding the following 2 topic areas: (1) threats of violence; and (2) any history of violent episodes.</p> <p>Timing of Screen All screening content must be obtained within the first day of admission (any time on the day of admission or the following day).</p> <p>Screening Administration Screening must be completed by a qualified psychiatric practitioner. The titles of qualified psychiatric practitioners may vary from state to state. Written and electronic collateral information and information provided by the patient in intake forms is acceptable if it has been reviewed by a qualified psychiatric practitioner.</p>
Exclusions	<p>Patients stays for which a screening could not be completed within the first 3 days of admission due to the patient's enduring unstable medical or psychological condition.</p> <p>Patient stays with a length of stay equal to or greater than 365 days, or less than 1 day.</p> <p>Patient stays with multiple admissions to psychiatric units during a single hospitalization.</p>
Stratification and Risk Adjustment	The measure is currently stratified by age in to 4 categories: children (age 1-12), adolescents (age 13-17), adults (age 18-64) and older adults (age 65+). No risk adjustment is planned.
Sampling	These measures will rely on a sampling methodology, by which cases are sampled quarterly or monthly. Facilities will be required to sample at least 20% of each stratum population for the quarter or month, with a minimum of 15 cases in each stratum per month. (Sampling will follow HBIPS-1 sampling guidelines.)

TABLE B.3. Measure Specifications: Screening for Substance Use	
Measure Dimension	Description
Description	Percentage of discharges from an IPF for which a structured substance use screening for 4 elements was completed.
Denominator	Psychiatric inpatient discharges during the measurement period.
Numerator	Alcohol AND drug use screening completed within the first day of admission.
Numerator Details	<p>Screening Content The medical record must provide documentation that information was obtained--either from the patient or from a collateral source--regarding the 4 key topic areas: (1) type, frequency, and amount of alcohol AND substance use; (2) adverse effects of this use (if use is reported); (3) dependence upon these substances (if use is reported); and (4) any history of drug/alcohol abuse.</p> <p>Timing of Screen All screening content must be obtained within the first day of admission (any time on the day of admission or the following day).</p> <p>Screening Administration Screening must be completed by a qualified psychiatric practitioner. The titles of qualified psychiatric practitioners may vary from state to state. Written and electronic collateral information and information provided by the patient in intake forms is acceptable if it has been reviewed by a qualified psychiatric practitioner.</p>
Exclusions	<p>Patients stays for which a screening could not be completed within the first 3 days of admission due to the patient's enduring unstable medical or psychological condition.</p> <p>Patient stays with a length of stay equal to or greater than 365 days, or less than 1 day.</p> <p>Patient stays with multiple admissions to psychiatric units during a single hospitalization.</p>
Stratification and Risk Adjustment	The measure is currently stratified by age in to 4 categories: children (age 1-12), adolescents (age 13-17), adults (age 18-64) and older adults (age 65+). No risk adjustment is planned.
Sampling	These measures will rely on a sampling methodology, by which cases are sampled quarterly or monthly. Facilities will be required to sample at least 20% of each stratum population for the quarter or month, with a minimum of 15 cases in each stratum per month. (Sampling will follow HBIPS-1 sampling guidelines.)

TABLE B.4. Measure Specifications: Metabolic Screening	
Measure Dimension	Description
Description	Percentage of discharges from an IPF for which a structured metabolic screening for 4 elements was completed in the past year.
Denominator	Psychiatric inpatient discharges with 1 or more routinely scheduled antipsychotic medications during the measurement period.*
Denominator Details	PRN antipsychotic medications or short-acting intramuscular antipsychotic medications do not count towards the denominator of this measure.
Numerator	The total number of patients who received a metabolic screening in the 12 months prior to discharge--either prior to, or during the index IPF stay. (If no record of a complete metabolic screening prior to the stay is found, the IPF must conduct the screening.)
Numerator Details	<p>Screening Content The medical record must provide documentation of the completion of all 4 of the following tests/measurements: (1) BMI; (2) blood pressure; (3) glucose or HbA1c; and (4) a lipid panel (which includes total cholesterol, triglycerides, high density lipoprotein, and low density lipoprotein).</p> <p>Timing of Screen Screenings must have been completed at least once in the 12 months prior to the patient's date of discharge. Screenings can be conducted either at the reporting facility or another facility for which records are available to the reporting facility.</p>
Exclusions	<p>Patients stays for which a screening could not be completed due to the patient's enduring unstable medical or psychological condition.</p> <p>Patient stays with a length of stay equal to or greater than 365 days, or less than 3 days.</p> <p>Patient stays with multiple admissions to psychiatric units during a single hospitalization.</p>
Stratification and Risk Adjustment	The measure is currently stratified by age in to 4 categories: children (age 1-12), adolescents (age 13-17), adults (age 18-64) and older adults (age 65+). No risk adjustment is planned.
Sampling	These measures will rely on a sampling methodology, by which cases are sampled quarterly or monthly. Facilities will be required to sample at least 20% of each stratum population for the quarter or month, with a minimum of 15 cases in each stratum per month.
* Medications that fall under this classification are identical to medications identified in HBIPS-4 "Multiple Antipsychotic Medications at Discharge."	

APPENDIX C. FOLLOW-UP AFTER IPF HOSPITALIZATION MEASURE SPECIFICATIONS

Measure Specifications: Follow-Up after IPF Hospitalization

Eligible Population	
Data Sources	<ul style="list-style-type: none"> • Denominator is populated using Medicare claims to include all FFS IPF discharges (Table C.1) with a principal mental health diagnosis at the IPF (Table C.2). Denominator exclusions are also defined using Medicare claims. The denominator is restricted to Medicare FFS-eligible beneficiaries, defined as individuals with continuous Part A and B coverage, and no HMO membership (HMO=0) during the month of the index discharge and the following month. (See continuous enrollment details below.) • Numerator is populated using Medicare FFS claims. Medicaid FFS claims (in addition to Medicare FFS claims) could be used to populate numerator for dual eligible beneficiaries (with a considerable time lag). Dual eligible beneficiaries can be defined as beneficiaries with values of 1 for TPBBEG or TPBEND for any month in the year/quarter of interest (found in the EDB).
Ages	No Restriction on Age
Continuous Enrollment	<ul style="list-style-type: none"> • Date of discharge through 30 days after discharge. • Continuous enrollment is defined as a value of 3 (Part A and B) or C (Part A and B, state buy-in) in monthly Medicare entitlement/buy-in indicator variables (bi_ind) in Medicare denominator file for the month of the index discharge and the following month.
Event/Diagnosis	<ul style="list-style-type: none"> • IPF discharges from January 1 through December 1 of the measurement year. • The denominator for this measure is based on discharges, not individuals. If individuals have more than 1 discharge in the covered period of time, then include all discharges.
Exclusions	
IPF Readmission or Direct Transfer	<ul style="list-style-type: none"> • A 30-day follow-up period is used for all exclusions related to the 7-day and 30-day measures. If the discharge is followed by readmission or direct transfer to an IPF within the 30-day follow-up period (with a principal mental health diagnosis), count only the readmission discharge or the discharge from the IPF to which the individual was transferred. • Exclude both the initial discharge and the readmission/direct transfer discharge if the IPF readmission/direct transfer discharge occurs after December 1 of the measurement year.

Exclusions (continued)	
Non-IPF Admission or Direct Transfer	<ul style="list-style-type: none"> • Exclude discharges followed by admission or direct transfer to an acute (non-IPF) facility--essentially a non-IPF hospital--within the 30-day follow-up period. These discharges are excluded from the measure because the admission or transfer to the hospital may prevent an outpatient follow-up visit from taking place. Refer to Table C.3 for rules to identify acute facilities. • Exclude discharges followed by admission or direct transfer to a non-acute (non-IPF) facility within the 30-day follow-up period. These discharges are excluded from the measure because the admission or transfer to a non-acute facility may prevent an outpatient follow-up visit from taking place. Refer to Table C.4 for rules to identify admission to non-acute care facilities.
Exclusions Related to Discharge Status or Death	<ul style="list-style-type: none"> • Exclude discharges followed by discharge/transfer to other institutions--including direct transfer to a prison--within the 30-day follow-up period. Refer to Table C.5 for rules to identify these transfers. • Exclude individuals who died during the stay and discharges followed by patient death within the 30-day follow-up period. Refer to Table C.5 for relevant codes related to patient death.

Description

The percentage of discharges from an IPF that had a follow-up outpatient visit, intensive outpatient encounter or partial hospitalization for their mental disorder after discharge. Two rates are reported:

- The percentage of discharges that received follow-up within 30 days of discharge.
- The percentage of discharges that received follow-up within seven days of discharge.

There are two approaches to calculating the measure numerator: (A) Using only Medicare claims to calculate follow-up care for non-dual eligible beneficiaries; and (B) using Medicare and Medicaid claims to calculate follow-up care for all Medicare beneficiaries, including dual eligibles.

TABLE C.1. Codes to Identify IPF Discharges
Inpatient claim lists the following facility codes (Medicare inpatient file)
<ul style="list-style-type: none"> • Last 4 digits of 4000-4499 (Psychiatric Hospital excluded from PPS) • 3rd digit of "S" (distinct part Psychiatric Unit in an acute care hospital) • 3rd digit of "M" (Psychiatric Unit in a CAH)
NOTE: A stay in any facility that meets 1 of the 3 criteria above constitutes an IPF stay.

TABLE C.2. Codes to Identify Principle Mental Health Diagnosis
ICD-9-CM diagnosis (Medicare carrier/outpatient and Medicaid OT files)
295-299, 300.3, 300.4, 301, 308, 309, 311-314
NOTE: Diagnosis codes are uniform across Medicare and Medicaid. Additional codes related to substance abuse diagnoses, particularly 303-305, can be included if the measure is intended to capture related services.

TABLE C.3. Codes to Identify Acute Care Facilities	
Description (Medicare inpatient files)	Provider Number
Provider number corresponding to acute care facilities	5 th thru 8 th digit (0001-0899)
NOTE: Medicare files are used to identify all exclusions. Any acute care stay during the 30-day period following IPF discharge constitutes an exclusion.	

TABLE C.4. Codes to Identify Admission to Non-Acute Care				
Description	HCPCS (Medicare carrier file)	UB Revenue (Medicare SNF, hospice, outpatient, and HHA files)	UB Type of Bill (Medicare SNF, hospice, outpatient, and HHA files)	POS (Medicare carrier files)
Hospice		0115, 0125, 0135, 0145, 0155, 0650, 0656, 0658, 0659	81x, 82x	34
SNF		019x	21x, 22x, 28x	31, 32
Hospital transitional care, swing bed or rehabilitation			18x	
Comprehensive inpatient rehabilitation facility		0118, 0128, 0138, 0148, 0158		61
Respite		0655		
ICF				54
Residential substance abuse treatment facility		1002		55
Psychiatric residential treatment center	T2048, H0017- H0019	1001		56
NOTE: Medicare files are used to identify all exclusions. Any code corresponding to the 30-day period following IPF discharge constitutes an exclusion.				

TABLE C.5. Codes to Identify Patient Deaths and Transfer/Discharge to Another Institution	
Description	Discharge Code (Medicare inpatient file)
Discharged/transferred to other short-term general hospital for inpatient care.	02
Discharged/transferred to SNF with Medicare certification in anticipation of covered skilled care--for hospitals with an approved swing bed arrangement, use Code 61--swing bed. For reporting discharges/transfers to a non-certified SNF, the hospital must use Code 04--ICF.	03
Discharged/transferred to ICF	04
Discharged/transferred to another type of institution for inpatient care	05
Expired	20
Discharged/transferred to a federal hospital	43
Hospice--home	50
Hospice--medical facility	51
Discharged/transferred within this institution to a hospital-based Medicare approved swing bed	61
Discharged/transferred to an inpatient rehabilitation facility including distinct parts units of a hospital	62
Discharged/transferred to a long-term care hospitals	63
Discharged/transferred to a nursing facility certified under Medicaid but not under Medicare	64
Discharged/transferred to a psychiatric hospital or psychiatric distinct unit of a hospital	65
Discharged/transferred to a CAH	66
Discharged/transferred to another type of health care institution not defined elsewhere in code list	70
Died during the 30-day follow-up period	Death_DT from Medicare Enrollment File
NOTE: Medicare files are used to identify all exclusions. Any code corresponding to the 30-day period following IPF discharge constitutes an exclusion.	

Administrative Specification

Denominator	<p>Medicare discharges from IPFs that are not excluded due to the rules above.</p> <p><i>Below we outline 4 potential options for identifying follow-up care: option 1 uses a combination of provider codes and procedural codes; options 2 and 3 use a combination of diagnosis codes and place/type of service and procedural codes; and option 4 uses only general place/type of service and procedural codes to identify follow-up care.</i></p>
Numerator Option 1	An outpatient visit, an intensive outpatient encounter or partial hospitalization with a mental health practitioner within 30 or 7 days of discharge. (Uses CPT/HCPCS/POS/Revenue/FACTYP/TOS codes in Table C.6 and NPI/Taxonomy codes in Table C.7; this definition most closely aligns with the HEDIS follow-up after mental health hospitalization measure.) Consult Table C.8 for a list of Medicaid Specialty codes that can be used to determine whether Medicaid-paid services were provided by a mental health practitioner--only applicable if using Medicaid claims to determine follow-up among dual eligible beneficiaries.
Numerator Option 2	An outpatient visit, an intensive outpatient encounter or partial hospitalization (with any provider) with a principal mental health diagnosis within 30 or 7 days of discharge (Uses CPT/HCPCS/POS/Revenue/FACTYP/TOS codes in Table C.6 and diagnosis codes in Table C.2)
Numerator Option 3	Any outpatient visit with a principal mental health diagnosis within 30 or 7 days of discharge. (Uses general outpatient codes in Table C.9: POS and FACTYP/TYPSVC in Medicare claims and TOS and PLC_OF_SRVC codes in Medicaid claims, and diagnosis codes in Table C.2.)
Numerator Option 4	Any outpatient visit with any principal diagnosis within 30 or 7 days of discharge. (Uses general outpatient codes in Table C.9: POS and FACTYP/TYPSVC in Medicare claims and TOS and PLC_OF_SRVC codes in Medicaid claims.)
<p>NOTE: It is important to note that Numerator options 1 and 2 use a narrow definition of outpatient care --generally assessment, management, and therapy provided by mental health specialists--whereas options 3 and 4 use a more general definition of outpatient care that can be provided in a variety of settings and by a larger range of providers.</p>	

TABLE C.6. Codes to Identify Outpatient Visits, Intensive Outpatient Encounters, and Partial Hospitalizations (numerator options 1 and 2)		
CPT (Medicare carrier/outpatient and Medicaid OT files)		HCPCS (Medicare carrier/outpatient and Medicaid OT files)
90804-90815, 98960-98962, 99078, 99201-99205, 99211-99215, 99217-99220, 99241-99245, 99341-99345, 99347-99350, 99383-99387, 99393-99397, 99401-99404, 99411, 99412, 99510		G0155, G0176, G0177, G0409-G0411, H0002, H0004, H0031, H0034-H0037, H0039, H0040, H2000, H2001, H2010-H2020, M0064, S0201, S9480, S9484, S9485
CPT with POS (Medicare carrier and Medicaid OT files)		
90801, 90802, 90816-90819, 90821-90824, 90826-90829, 90845, 90847, 90849, 90853, 90857, 90862, 90870, 90875, 90876	<i>with</i>	03, 05, 07, 09, 11, 12, 13, 14, 15, 20, 22, 24, 33, 49, 50, 52, 53, 71, 72
99221-99223, 99231-99233, 99238, 99239, 99251-99255	<i>with</i>	52, 53
CPT with TYP SVC/FACTYP (Medicare outpatient files)		
90801, 90802, 90816-90819, 90821-90824, 90826-90829, 90845, 90847, 90849, 90853, 90857, 90862, 90870, 90875, 90876	<i>with</i>	TYP SVC of 2 or 3 if FACTYP = 1-6 or 9
99221-99223, 99231-99233, 99238, 99239, 99251-99255		FACTYP = 7 FACTYP = 8
UB Revenue (Medicare outpatient and Medicaid OT files)		
0513, 0900-0905, 0907, 0911-0917, 0919--does not have to be with mental health practitioner for option 1		
0510, 0515-0517, 0519-0523, 0526-0529, 077x, 0982, 0983--must be with either a principal mental health diagnosis or practitioner for option 1		
ER visits cannot count toward numerator (Medicare carrier/outpatient and Medicaid OT files)		
No line-level claim information related to ER services can count as a numerator hit. These are defined as [revenue center code values of 0450-0459 (ER) or 0981 (Professional fees-ER) or HCPCS codes associated with ER use (99281, 99282, 99283, 99284, 99285) in the OP file] OR [Place of Service code (23=ER-hospital) or BETOS code (M3 = ER visit) in the carrier file]		
NOTE: A claim meeting any of the requirements above constitutes an outpatient visit.		

TABLE C.7. Codes to Identify Mental Health Practitioners in Medicare		
HEDIS Definition of Mental Health Practitioner	Specialty Code	Taxonomy (linked to NPI)
An MD or DO who is certified as a psychiatrist or child psychiatrist	26	2084P0800X; 2084P0804X
Neurologist (not in original HEDIS specification)	13	2084V0102X; 2084N0400X; 2084N0402X
An MD or DO who successfully completed an accredited program of graduate medical or osteopathic education in psychiatry or child psychiatry and is licensed to practice patient care psychiatry or child psychiatry	86	2084A0401X; 2084P0802X; 2084B0002X; 2084N0600X; 2084D0003X; 2084F0202X; 2084P0805X; 2084H0002X; 2084P0005X; 2084N0008X; 2084P2900X; 2084P0015X; 2084S0012X; 2084S0010X
Licensed Psychologist	62	103T00000X; 103TA0400X; 103TA0700X; 103TC0700X; 103TC2200X; 103TB0200X; 103TC1900X; 103TE1000X; 103TE1100X; 103TF0000X; 103TF0200X; 103TP2701X; 103TH0004X; 103TH0100X; 103TM1700X; 103TM1800X; 103TP0016X; 103TP0814X; 103TP2700X; 103TR0400X; 103TS0200X; 103TW0100X
Certified in Clinical Social Work	80	1041C0700X
Psychiatric Nurse, Physician Assistant, or Occupational Therapist	not applicable	364SP0808X; 364SP0809X; 364SP0807X; 364SP0810X; 364SP0811X; 364SP0812X; 364SN0800X; 364SP0813X; 363LP0808X; 225XM0800X
<p>NOTE: All codes are found in Medicare outpatient/carrier files. Either a Medicare Specialty code OR Taxonomy code qualifies as a numerator hit.</p> <p>Specialty codes and Taxonomy codes are the best match with mental health practitioners defined in HEDIS specifications (see additional resource table below). NPI codes and Specialty codes in Medicare claims can be used to determine provider taxonomy and type, respectively. Specialty codes in Table C.9 should be used to identify mental health providers in Medicaid claims.</p>		

**TABLE C.8. Codes to Identify Mental Health Practitioners in Medicaid
(numerator option 1)**

State	Physician/ Neurology	MD or DO Certified as Psychiatrist or Child Psychiatrist	MD or DO Completed Accredited Program in Psychiatry or Child Psychiatry	Licensed Psychologist	Certified in Clinical Social Work	Psych Nurse, Psych Physician Assistant, or Occupational Therapist
Alaska	13	26, 27		69		
Arkansas	13	26	P5	62		
California	13, 79	26	27, 36			
Connecticut		26		92		26
Delaware	N040	P080			M080	
Florida	22	23	42, 43	44		
Illinois	CHN, N	CHP, P		PYA		
Indiana	326	339		112	113	117
Louisiana	13	26	27	62		
Maryland	050	051, 052		196		
Massachusetts	144	153, 154		186		
Michigan	0708		0290-0295			
Missouri	17	27		45	42, 43, A2, A3	
Nebraska	13	26		62	80	
New Hampshire	13	26,27			69	
New Jersey	130	260	270	610		262
New Mexico	013	026, 027		062		
New York	193, 194	191,192, 195, 931, 945, 946, 964		780	781	
North Carolina	13	26		109	110	112
North Dakota	13	26		62		
Oklahoma	326	339		112	116	
South Carolina	22	48, 49		82		
Texas	13	26	27	62	A7	
Virginia	71			77		
West Virginia	N1	R5		W8		
Wisconsin	13	26		62	78	

NOTE: Any of these codes qualifies as a numerator hit. State-specific Specialty codes and Taxonomy codes are the best match with mental health practitioners defined in HEDIS specifications (see Table C.10 below). Data source is states' initial MSIS applications. Any Specialty code qualifies as a numerator hit for the claim in question. All codes can be found in Medicaid OT file. Codes are subject to change over time.

TABLE C.9. Codes to Identify Outpatient Visits (numerator options 3 and 4)	
Codes	Files
No line-level claim information related to ER services can count as a numerator hit. These are defined as [revenue center code values of 0450-0459 (ER) or 0981 (Professional fees-ER) or HCPCS codes associated with ER use (99281, 99282, 99283, 99284, 99285) in the OP file] OR [Place of Service code (23=ER-hospital) or BETOS code (M3 = ER visit) in the carrier file]	Medicare Outpatient and Carrier; Medicaid OT
The following POS codes count as a numerator hit if they appear in any line-level claim: 03, 05, 07, 09, 11, 12, 13, 14, 15, 20, 22, 24, 33, 49, 50, 52, 53, 71, 72	Medicare Carrier
The following codes count as a numerator hit: TYP SVC of 2 or 3 if FACTYP = 1-6 or 9 OR FACTYP = 7 or 8	Medicare Outpatient
Numerator hit in Medicaid if 1 of 2 conditions are present as defined by TOS and/or POS codes: 1. MSIS_TOS = 11 or 12 OR 2. MSIS TOS = 08, 10, 13, 19, 33, 37, or 99 AND POS = 03, 05, 07, 09, 11, 12, 13, 14, 15, 20, 22, 24, 33, 49, 50, 52, 53, 71, 72, 99, or unassigned--but cannot have TOS of 99 and POS of 99 or unassigned.	Medicaid OT
NOTE: A claim meeting any of the requirements above constitutes an outpatient visit.	

TABLE C.10. Additional Resource: HEDIS Definition of Mental Health Practitioner
Use HEDIS Definition:
A mental health practitioner is a practitioner who provides mental health services and meets any of the following criteria:
<ul style="list-style-type: none"> • An MD or DO who is certified as a psychiatrist or child psychiatrist by the American Medical Specialties Board of Psychiatry and Neurology or by the American Osteopathic Board of Neurology and Psychiatry; or, if not certified, who successfully completed an accredited program of graduate medical or osteopathic education in psychiatry or child psychiatry and is licensed to practice patient care psychiatry or child psychiatry, if required by the state of practice. • An individual who is licensed as a psychologist in his/her state of practice. • An individual who is certified in clinical social work by the American Board of Examiners; who is listed on the National Association of Social Worker's Clinical Register; or who has a master's degree in social work and is licensed or certified to practice as a social worker, if required by the state of practice.

APPENDIX D. SUPPLEMENTAL TABLES FOR SCREENING MEASURES

TABLE D.1. Average Performance on Screening Measures, Including and Excluding IPF		
Screening Measure	Average Among All 6 IPFs (including IPF 1)	Average Among 5 IPFs (excluding IPF 1)
	%	%
Suicide risk	93.4	97.7
Violence risk	89.0	95.7
Substance use	85.8	91.5
Metabolic	41.5	44.3
SOURCE: Administrative data and medical records from 6 IPFs, corresponding to all discharges from October 1, 2013, to December 31, 2013.		

TABLE D.2. Number and Proportion of Patients Excluded from Screening Measures, by IPF

	All IPS		Psychiatric Units						Freestanding Hospitals					
			IPF 1 (private)		IPF 2 (private)		IPF 3 (private)		IPF 4 (public)		IPF 5 (public)		IPF 6 (private)	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Suicide denominator before exclusions	825	100.0	120	100.0	176	100.0	120	100.0	120	100.0	118	100.0	171	100.0
Suicide screening exclusions														
Patients with a length of stay less than 1 day	41	5.0	12	10.0	3	1.7	13	10.8	2	1.7	8	6.8	3	1.8
Patients with a length of stay equal to or greater than 365 days	10	1.2	0	0.0	0	0.0	0	0.0	7	5.8	3	2.5	0	0.0
Patients who had previous admissions to psychiatric units during a single hospitalization	2	0.2	0	0.0	0	0.0	0	0.0	1	0.8	0	0.0	1	0.6
Patient unwillingness or inability ^a	11	1.3	0	0.0	1	0.6	1	0.8	1	0.8	8	6.8	0	0.0
Suicide denominator after exclusions	761	92.2	108	90.0	172	97.7	106	88.3	109	90.8	99	83.9	167	97.7
Violence denominator before exclusions	825	100.0	120	100.0	176	100.0	120	100.0	120	100.0	118	100.0	171	100.0
Violence screening exclusions														
Patients with a length of stay less than 1 day	41	5.0	12	10.0	3	1.7	13	10.8	2	1.7	8	6.8	3	1.8
Patients with a length of stay equal to or greater than 365 days	10	1.2	0	0.0	0	0.0	0	0.0	7	5.8	3	2.5	0	0.0
Patients who had previous admissions to psychiatric units during a single hospitalization	2	0.2	0	0.0	0	0.0	0	0.0	1	0.8	0	0.0	1	0.6
Patient unwillingness or inability ^a	7	0.8	1	0.8	0	0.0	1	0.8	1	0.8	4	3.4	0	0.0
Violence denominator after exclusions	765	92.7	107	89.2	173	98.3	106	88.3	109	90.8	103	87.3	167	97.7
Substance use denominator before exclusions	825	100.0	120	100.0	176	100.0	120	100.0	120	100.0	118	100.0	171	100.0
Substance use screening exclusions														
Patients with a length of stay less than 1 day	41	5.0	12	10.0	3	1.7	13	10.8	2	1.7	8	6.8	3	1.8
Patients with a length of stay equal to or greater than 365 days	10	1.2	0	0.0	0	0.0	0	0.0	7	5.8	3	2.5	0	0.0
Patients who had previous admissions to psychiatric units during a single hospitalization	2	0.2	0	0.0	0	0.0	0	0.0	1	0.8	0	0.0	1	0.6
Patient unwillingness or inability ^{a,b}	20	2.4	1	0.8	0	0.0	5	4.2	3	2.5	11	9.3	0	0.0
Substance use denominator after exclusions	754	91.4	107	89.2	173	98.3	102	85.0	108	90.0	97	82.2	167	97.7

TABLE D.2 (continued)														
	All IPS		Psychiatric Units						Freestanding Hospitals					
			IPF 1 (private)		IPF 2 (private)		IPF 3 (private)		IPF 4 (public)		IPF 5 (public)		IPF 6 (private)	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Metabolic screening denominator before exclusions	506	100.0	56	100.0	118	100.0	85	100.0	79	100.0	78	100.0	90	100.0
Metabolic screening exclusions														
Patients with a length of stay less than 1 day	63	12.5	13	23.2	10	8.5	27	31.8	2	2.5	9	11.5	2	2.2
Patients with a length of stay equal to or greater than 365 days	6	1.2	0	0.0	0	0.0	0	0.0	3	3.8	3	3.8	0	0.0
Patients who had previous admissions to psychiatric units during a single hospitalization	2	0.4	0	0.0	0	0.0	0	0.0	1	1.3	0	0.0	1	1.1
Patient unwillingness or inability ^a	1	0.2	0	0.0	0	0.0	0	0.0	0	0.0	1	1.3	0	0.0
Metabolic screening denominator after exclusions	434	85.8	43	76.8	108	91.5	58	68.2	73	92.4	65	83.3	87	96.7
SOURCE: Administrative data and medical records from 6 IPFs, corresponding to all discharges from October 1, 2013, to December 31, 2013.														
NOTE: Totals for individual tools do not always sum to any standard tool, given that more than 1 tool could have been employed for 1 patient.														
a. Patient charts had to demonstrate evidence of patient inability or unwillingness--including the day and time of attempted screenings--for this exclusion to be applied.														
b. Across the alcohol and drug components, a total of 20 patients were unable or unwilling to perform the screening within 1 day. This includes 19 individuals who could or would not perform the alcohol screening and 16 individuals who could not perform the drug screening, with large overlap between the 2 groups.														

TABLE D.3. Percent Agreement on Screening Measures, by IPF														
Screening Measure	All IPS		Psychiatric Units						Freestanding Hospitals					
			IPF 1 (private)		IPF 2 (private)		IPF 3 (private)		IPF 4 (public)		IPF 5 (public)		IPF 6 (private)	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Metabolic														
Exclusions	57	98	10	100	10	90	10	100	7	100	9	100	11	100
Numerator	57	96	10	100	10	100	10	100	7	100	9	100	11	82
Suicide														
Exclusions	58	98	10	100	10	90	10	100	7	100	10	100	11	100
Numerator	58	97	10	100	10	100	10	100	7	86	10	90	11	100
Violence														
Exclusions	58	98	10	100	10	90	10	100	7	100	10	100	11	100
Numerator	58	93	10	100	10	100	10	100	7	43	10	100	11	100
Substance use														
Exclusions	58	98	10	100	10	90	10	100	7	100	10	100	11	100
Numerator	58	97	10	100	10	100	10	100	7	71	10	100	11	100
SOURCE: Administrative data and medical records from 6 IPFs, corresponding to all discharges from October 1, 2013, to December 31, 2013.														
NOTE: Kappas are not presented due to small sample sizes and lack of variation between abstractors in some IPFs.														

APPENDIX E. SUMMARY OF STATE SELECTION PROCESS FOR FOLLOW-UP MEASURE ANALYSIS WITH MEDICAID CLAIMS

In Chapter VI, we present an analysis of dual eligible Medicare-Medicaid beneficiaries. To conduct this analysis, we merged Medicare and MAX data for dual eligible Medicare beneficiaries who visited IPFs in 2008 (the latest year for which MAX data were available at the beginning of this project). MAX data are created from eligibility and claims files submitted by states to CMS and then standardized into variables that can be used to create comparable measures of service use across states. These variables include information such as demographic characteristics (for example, race, ethnicity, gender, and age), diagnoses, and procedures performed for each beneficiary enrolled in Medicaid at any point during the year. A DUA with CMS governed our use of these MAX data.

We limited our analysis of MAX data to FFS claims because although MAX includes some encounters submitted by health maintenance organizations (HMOs) and managed behavioral healthcare organizations, Medicaid managed care encounter data do not undergo the data validation process applied to MAX FFS data, and are generally considered to be of lower quality.

Starting with all 50 states and the District of Columbia, we used a two-step process to select states that would be included in this analysis. We describe these steps below:

Step 1. As noted above, Medicaid managed care encounters are not reliably captured in MAX data. Thus, if a substantive proportion of dual eligible beneficiaries in a state are enrolled in Medicaid managed care, these analyses would underestimate their receipt of care. To avoid this potential bias, states with more than 25 percent of dual eligible beneficiaries enrolled in Medicaid managed care were excluded from the analyses during the first step of state selection. A total of 14 states were excluded in this first step: Arizona, Colorado, Hawaii, Iowa, Kansas, Maine, Minnesota, Montana, Oregon, Pennsylvania, Rhode Island, Tennessee, Utah, and Vermont.

Step 2. The completeness and reliability of MAX data varies across states. After excluding 14 states due to their relatively high proportion of dual eligible beneficiaries in Medicaid managed care (step 1), we excluded additional states that did not use provider Specialty codes in 2008 MAX data in a manner that could facilitate the identification of outpatient services provided by a mental health practitioner. States were excluded if they met any of the following three exclusion criteria: (1) the state did not use provider Specialty codes on any claims; (2) provider Specialty codes did not identify mental

health practitioners; or (3) nearly all claims were missing provider Specialty codes.⁵⁶ A total of 11 states were excluded in this second step: Alabama, District of Columbia, Georgia, Idaho, Kentucky, Mississippi, Nevada, Ohio, South Dakota, Washington, and Wyoming.

This two-step selection process resulted in a total of 26 states included in the analysis of dual eligible beneficiaries presented in Chapter VI, after excluding a total of 25 states (see Table E.1 for a full listing of included and excluded states).

⁵⁶ The primary concern with MAX data was that a lack of uniform practitioner codes would underestimate IPF performance on the measure using numerator option 1, which defines follow-up care as a procedure (as specified by a CPT/HCPCS code) performed by a mental health practitioner (as specified by a practitioner Specialty code). For this reason, our data quality checks focused on identifying instances in which a lack of information on the practitioner specialty would result in an underestimation of follow-up care under option 1.

TABLE E.1. Results of State Selection Process for Merged Medicare-Medicaid Analysis					
State	Step 1: Excluded Because More Than 25% of Dual Eligible Beneficiaries Enrolled in Medicaid Managed Care	Step 2: Excluded Due to Data Quality Concerns			Included in the Analysis
		States do not Use Provider Specialty Codes	Provider Specialty Codes do not Identify Mental Health Practitioners	A Substantial Portion of Claims are Missing Provider Specialty Codes	
Alaska					X
Alabama			X		
Arkansas					X
Arizona	X				
California					X
Colorado	X				
Connecticut					X
District of Columbia				X	
Delaware					X
Florida					X
Georgia		X			
Hawaii	X				
Idaho				X	
Illinois					X
Indiana					X
Iowa	X				
Kansas	X				
Kentucky			X		
Louisiana					X
Maine	X				
Massachusetts					X
Maryland					X
Michigan					X
Minnesota	X				
Missouri					X
Mississippi		X			
Montana	X				
North Carolina					X
North Dakota					X
Nebraska					X
New Hampshire					X
New Jersey					X
New Mexico					X
Nevada	X				
New York					X
Ohio		X			
Oklahoma					X
Oregon	X				
Pennsylvania	X				
Rhode Island	X				
South Carolina					X
South Dakota	X				
Tennessee	X				
Texas					X
Utah	X				
Vermont	X				
Virginia					X
Washington				X	
Wisconsin					X
West Virginia					X
Wyoming		X			

DEVELOPMENT AND TESTING OF BEHAVIORAL HEALTH QUALITY MEASURES

Reports Available

Development of Quality Measures for Inpatient Psychiatric Facilities: Final Report

Executive Summary <http://aspe.hhs.gov/daltcp/reports/2015/IPFes.cfm>

HTML <http://aspe.hhs.gov/daltcp/reports/2015/IPF.cfm>

PDF <http://aspe.hhs.gov/daltcp/reports/2015/IPF.pdf>

Strategies for Measuring the Quality of Psychotherapy: A White Paper to Inform
Measure Development and Implementation

Executive Summary <http://aspe.hhs.gov/daltcp/reports/2014/QualPsyres.cfm>

HTML <http://aspe.hhs.gov/daltcp/reports/2014/QualPsy.cfm>

PDF <http://aspe.hhs.gov/daltcp/reports/2014/QualPsy.pdf>

To obtain a printed copy of this report, send the full report title and your mailing information to:

U.S. Department of Health and Human Services
Office of Disability, Aging and Long-Term Care Policy
Room 424E, H.H. Humphrey Building
200 Independence Avenue, S.W.
Washington, D.C. 20201
FAX: 202-401-7733
Email: webmaster.DALTCP@hhs.gov

NOTE: All requests must be in writing.

RETURN TO:

Office of Disability, Aging and Long-Term Care Policy (DALTCP) Home
http://aspe.hhs.gov/office_specific/daltcp.cfm

Assistant Secretary for Planning and Evaluation (ASPE) Home
<http://aspe.hhs.gov>

U.S. Department of Health and Human Services (HHS) Home
<http://www.hhs.gov>