# Contents

_Foreword_ ................................................................. 2

_Acknowledgments_ ....................................................... 4

_Population Health Overview_ ........................................... 6

  _Industry Update_ ........................................................ 6

  _Outcomes Initiative Update_ ......................................... 8

  _Evolution of the Industry and the Guidelines_ ................... 11

_Population Health Management_ ........................................ 16

  _Overview_ ................................................................. 16

  _Methodological Considerations_ .................................... 26

  _Measure Sets_ ........................................................... 35

_Wellness_ ........................................................................ 36

  _Overview_ ................................................................. 36

  _Methodological Considerations_ .................................... 38

  _Measure Sets_ ........................................................... 51

_Chronic Care Management – Five Core Chronics_ ................... 55

  _Overview_ ................................................................. 55

  _Methodological Considerations_ .................................... 57

  _Measure Sets_ ........................................................... 92

_References_ ..................................................................... 119

_Appendix_ ....................................................................... 121
Tables, Figures and Charts

Table I, Guidelines Locator ................................................................. 9
Table II, Alignment – Program Checklist .............................................. 12
Table III, PHM Capabilities ................................................................. 18
Table IV, Methodology Framework: Study Design ................................. 28
Table V, Comparison of Current State of Chronic Care Management/Wellness Programs .......................... 38
Table VI, Control/Comparison Groups Methods for Wellness Program Evaluation ................................................. 45
Table VII, Study Design .................................................................... 59
Table VIII, Methods to Define Outliers .................................................. 64
Table IX, Prevalence Rates .................................................................. 68
Table X, Clinical Measures .................................................................. 69
Table XI, Utilization Measures ............................................................. 69
Table XII, Financial Measures .............................................................. 70
Table XIII, Trend Recommendation Results ......................................... 72
Table XIV, Risk Adjustment Examples .................................................... 75
Table XV, Small Populations, Part I ...................................................... 83
Table XVI, Comparison of Various Utilization Measures ......................... 95
Table XVII, Medication Adherence Self-report Instruments ..................... 113
Tables, Figures and Charts, cont.

Click on a heading to jump directly to the item

Figures

Figure 1, Alignment Algorithm ................................................................. 14
Figure 2, Conceptual PHM Framework .................................................. 19
Figure 3, PHM Program Process Framework ......................................... 20
Figure 4, HIT Framework ....................................................................... 24
Figure 5, Impacts Model ........................................................................ 47
Figure 6, Predictive Risk Score Calculation ........................................... 80
Figure 7, Program Impact Over Time .................................................... 87
Figure 8, Achieving Optimal Balance in Study Design ............................. 92
Figure 9, Organizational Best Practice Framework ................................. 104
Figure 10, Diabetes Persistence Curve ................................................. 109
Figure 11, Operational Measures Flow Diagram .................................... 115

Charts

Chart I, Attribution Adequacy Illustration of Framework ................................ 31
Chart II, Attribution Adequacy Examples ............................................. 32
Chart III, Small Populations, Part II ..................................................... 84
Chart IV, Diabetes Persistence Example .............................................. 109
Foreword

The Care Continuum Alliance (CCA) Outcomes Guidelines, now in a fifth iteration, have expanded and evolved in step with the widening scope of industry programs to improve and manage population health. The Guidelines serve as a critical primer on key issues in measurement of population health management programs and offer a set of standards against which purchasers and providers of these services may consistently and reliably assess program impact. This volume provides additional guidance on assessing care management programs that go beyond single condition management, standardizing measurement of medication adherence and setting clear direction on critical measures of program activity.

Creation of these Guidelines reflects tremendous effort from many volunteers and staff. Workgroup members from a diverse group of organizations collaborated to analyze, strategize and build consensus around increasingly complex measurement issues. It is a remarkable process, surprisingly free of dissonance, that produces guidelines able to withstand intense scrutiny. The acknowledgments section of this report includes members of the Outcomes Steering Committee — workgroup members, leaders and contributing authors for this year’s work. These individuals and the organizations that support them deserve our heartfelt gratitude, as do the CCA staff who work tirelessly to facilitate these efforts. In addition, our work has been enhanced this year through collaborations with several organizations. These collaborations have helped us to maximize our resources, enhance our knowledge base and ensure that our work is truly representative and applicable to all stakeholders.

As health care consumers, providers, insurers, purchasers and service organizations continue to adjust to the implications of the health reform law, the Patient Protection and Affordable Care Act (PPACA), the consistent and credible measurement of population health management program outcomes will become increasingly important. The CCA Outcomes Guidelines will, as a result, need to adapt to the new models of support for consumers and providers that result from reform. This report begins to address these new models by offering evaluation guidelines for programs that now encompass a full spectrum of population health needs through a wide array of program offerings. For example, the Population Health Management Workgroup refined an essential framework for population health and provided critical guidance on the range of methodologies used to measure program effect. In addition, the Medication Adherence Workgroup continued to build on the important methodologies and recommendations in its area of focus by articulating the medications that can be billed on either the medical or pharmacy benefit, providing organizational adherence best practices and compiling self-reported adherence measurement tools. The definition of initial engagement, especially at the rapid pace of technology innovation and new models of care delivery, has been carefully considered as well, and makes its debut in this volume, along with a new operational measures flow diagram.
As health care continues to dominate the political agenda, we hope that the work of the Outcomes Steering Committee will be a source of information for policymakers. Of the approximately $150 billion included for health care in the 2009 American Recovery and Reinvestment Act (ARRA), the top four investments were tiered for Medicaid, health information technology (HIT), COBRA subsidies and National Institutes of Health research and facilities. The majority has been allocated to assist states with their Medicaid programs — an ongoing area of interest of the Outcomes Steering Committee as it begins to deliberate, study and adapt the current Guidelines to this unique and fast-growing population. This growth trend is expected to continue as a result of the PPACA legislation and the expected expansion of Medicaid eligibility. Also of note are PPACA’s creation of a not-for-profit Patient-Centered Outcomes Research Institute and the eventual development of a Center for Medicare and Medicaid Innovation by the Centers for Medicare and Medicaid Services. These entities represent the important responsibility of conducting rigorous, comparative effectiveness research and studies, and the Care Continuum Alliance looks forward to continued collaboration with both the federal government and states in their ongoing efforts. We have seen the renewed interest of wellness and health promotion in recent legislative discussions and hope that the guidelines for wellness program evaluation can be a resource to stakeholders, as well.

The CCA Outcomes Steering Committee has an ambitious agenda for the coming year, including continuing work to define more explicit guidance on how to measure population health management programs. The committee is likely to prepare interim guidance documents in addition to a sixth annual Guidelines document in the coming year. Each section of this report offers topic-specific suggestions for next steps. We encourage both members and stakeholders to review the work and potential next steps and to offer comment and feedback.

Jason Cooper, MS  
Co-Chair, CCA Outcomes Steering Committee  
Vice President, Clinical Analytics  
CIGNA HealthCare

David Veroff, MPP  
Co-Chair, CCA Outcomes Steering Committee  
Vice President, Evaluation  
Health Dialog Inc.
Acknowledgments

Outcomes Steering Committee Members

Committee Co-Chairs:
Jason G. Cooper, MS
CIGNA HealthCare *
David Veroff, MPP
Health Dialog Inc. *

Workgroups

PHM ALIGNMENT
Leader, Craig Nelson, MS, DC, American Specialty Health *
Chris Behling, Hooper Holmes
Helene Forte, RN, MS, PAHM, Aetna
Charles Nagurka, OptumHealth
Sarah Sampsel, MPH, WellPoint Inc.
Dexter Shurney, MD, MBA, MPH, Vanderbilt Medical Center

PHM FRAMEWORK
Leader, Jennifer Pitts, PhD, Edington Associates, LLC *
Ted Dacko
Richard Feiser, MD, MPH, FACP, Medco Health Solutions Inc.
Jane Hamerle, Blue Cross Blue Shield of Kansas City
Jeremy Margolis, PhD, Caremark Inc.
Shirley Roberts, Iowa Chronic Care Consortium

*Contributing author

PHM EVALUATION METHODOLOGY
Leader, Michael J. Connor, DrPH, Alere *
Leader, Barry Zajac, MHSA, Blue Cross Blue Shield of Louisiana *
Osman Ahmed, OptumHealth
Avery Ashby, MS, Blue Cross Blue Shield of Tennessee
Daniel Checkman, Blue Cross Blue Shield of Louisiana
Chin-Yu Chen, PhD, University of Michigan
Jen Clair, MPH, Health Management Corporation
Michael S. Cousins, PhD, CIGNA HealthCare and CareAllies
Tracy Durrah, APS Healthcare Inc.
R. Allen Frommelt, PhD, Nurtur Health Inc.
Joanne Garcia, RN, MHSM, FACHE, CVS Caremark
Neil F. Gordon, MD, PhD, MPH, Nationwide Better Health
Christine Groenier, APS Healthcare Inc.
Barbara Haddock, Deloitte Services LLP Life Sciences & Health Care
Benjamin N. Hamlin, MPH, National Committee for Quality Assurance
Adam Kaufman, DPS Health
Annemarie Kline, Coventry Health Care Inc.
Thomas P. Kotsos, Health Management Corporation
Andrew Krueger, MD, CVS Caremark
Erik Lesneski, AllOne Health
Soeren Mattke, MD, MPH, DSc, RAND Corporation
Kejian Niu, MS, Health Management Corporation
Gary Persinger, National Pharmaceutical Council Inc.
Annette Ruby, RN, SummaCare
Elizabeth Rula, PhD, Healthways
Tatiana Shnaiden, MD, ActiveHealth Management Inc.
Kelly Shreve, Capital Blue Cross
Alison B. Soucy, BS, CVS Caremark
Suzanne Stearns, AmeriChoice
Matt Stiefel, MPA, Kaiser Permanente Health Plan Inc.
Pei Tang, MS, MA, Alere
Yijin Wert, Capital Blue Cross
Steven Wright, ASA, MAAA, StayWell Health Management
Acknowledgments cont.

SELECTION CRITERIA

**Leader**, Iver Juster, MD, ActiveHealth Management Inc. *
**Leader**, Steele Stewart, FSA, MAAA, Blue Cross Blue Shield of Kansas City *
Christian Birkmeyer, MS, Solucia Consulting *
Rob Blundo, Solucia Consulting
Jen Clair, MPH, Health Management Corporation
Helene Forte, RN, MS, PAHM, Aetna
Carl Goff, RN, PhD, Blue Cross Blue Shield Association
Benjamin N. Hamlin, MPH, National Committee for Quality Assurance *
Jad Hayes, McKesson Health Solutions
Mark Heller, StayWell Health Management
Sue Jennings, PhD, Independent Healthcare Consultant *
Ying Liu, MSC, Health Management Corporation
Pei Tang, MS, MA, Aere
Min Tao, Blue Cross Blue Shield of Michigan
Tony Wang, Aere
Aaron R. Wells, PhD, Healthways

MEDICATION ADHERENCE

**Leader**, Richard Feifer, MD, MPH, FACP, Medco Health Solutions Inc. *
**Leader**, Elizabeth Skinner, PharmD, GlaxoSmithKline *
Chin-Yu Chen, PhD, University of Michigan
D.W. Edington, PhD, University of Michigan
Feride Frech-Tamas, Novartis Pharmaceuticals Corporation
Patricia A. Gladowski, RN, MSN, Highmark Blue Cross *
Garry Goddette, RPh, MBA, Aere
Christine Groenier, APS Healthcare Inc.
Kristina Hahn, Capital Blue Cross
Ron Harris, MD, FACE, Geisinger Health Plan
Diane M. Hedler, Kaiser Permanente
Jeremy A. Margolis, PhD, Caremark Inc.
Mary Jane Osmick, MD, StayWell Health Management *
Urvashi Patel, PhD, Care Management Technologies
Gary Persinger, National Pharmaceutical Council Inc.
Michelle Williams, RN, MSN, Kaiser Permanente
Richard Zabinski, OptumHealth

OPERATIONAL MEASURES

**Leader**, Cindy Hochart, RN, MBA, Blue Cross Blue Shield of Kansas City *
**Leader**, Susan Weber, RN, CCM, MHP, StayWell Health Management *
Daniel Checkman, Blue Cross Blue Shield of Louisiana
Jay Chyung, Healthways
Christopher L. Cook, PharmD, PhD, GlaxoSmithKline
Joanne Garcia, RN, MHSM, FACHE, CVS Caremark
Jane Hamerle, Blue Cross Blue Shield of Kansas City
Benjamin N. Hamlin, MPH, National Committee for Quality Assurance
Debra Hicks, Aetna
Annemarie Kline, Coventry Health Care Inc.
Rhonda Lash, RN, BSN, Capital Blue Cross
Erik Lesneski, AllOne Health
Mark Norton, MPA, PAHM, Health Management Corporation
Robert Palmer, Aere
Ann Ritchhart, StayWell Health Management
Erin Seaverson, MPH, StayWell Health Management
Travis Tucker, Aere

*Contributing author
Population Health Overview

Industry Update

The population health industry continues to meet the needs of diverse populations through partnerships and collaborations focused on helping individuals achieve and maintain their health goals. Organizing care, information and services around the health needs and desires of individuals is a concept that all health care stakeholders have embraced, as experience with these programs repeatedly demonstrates positive impact over time. As always, the devil is in the details for optimizing these efforts, but progress is being made in several key areas. These include:

Scope of Health Improvement Offerings in the Marketplace

Despite recent industry consolidation, numerous firms now compete to provide a wide variety and scope of health improvement programs, from wellness and prevention to chronic care support to end-of-life care. Those who arrange and pay for health care benefits and services never have had more options for designing flexible programs to support prevention, healthful lifestyles and chronic care support. Increasingly, many are seeking more-integrated programs that consolidate outreach, communication, intervention, transitions, reporting and assessment across the entire continuum of care to leverage economies of scope and scale.

Flexibility of Intervention/Program Design

As the population health industry continues to move toward collaborative models of care delivery, it is clear that no single plug-and-play model for all settings is appropriate. Rather, flexible models developed and aligned with the needs of the target population and existing services have enjoyed success. While resisting the temptation to over-generalize among differing pilots and delivery models, research must identify success factors common to programs that produce the greatest impacts on health and seek solutions to barriers in programs with disappointing outcomes. Iterative refinement of a variety of models is the most likely pathway to sustainable and scalable success, versus a search for the ideal “magic bullet” model for all settings and populations.

Collaborative Physician-Led Models for Patient-Centered Care

Many providers desire to transform their practice infrastructure, workflows, information technologies and partnerships to better meet the diverse needs and desires of various populations in the pursuit of improved health. This has led to innovative collaborations among health care providers who recognize that, while physicians must lead these efforts, they can benefit from additional staff and capabilities, both within and beyond their practice walls, to provide health support to patients. As new models — such as patient-centered medical homes and accountable care organizations — evolve to focus more attention on outcomes of care, population health management will continue to influence these collaborative models by offering a suite of services that can complement the core patient-centered direct care model. The population health management frameworks on pages 18-25 of this report display this collaborative model for all care settings.
Care Transitions for At-Risk Populations

Transition of care for vulnerable populations has been identified as an inviting opportunity to achieve the increased quality and efficiency reform demands of the health care system. Population health management has played a vital role by successfully managing transitions from acute care to home for many patients. Working with hospitals, providers and ancillary care organizations, population health management will lessen avoidable morbidity associated with transitions across different sites and levels of care by leveraging its experience employing the most advanced technology.

Convergence of Devices and Diagnostics with Health Support

As care models evolve to become more consumer-oriented, tools available to individuals for health support in the home are proliferating. Ideally, these devices – whether biometric monitoring, diagnostic devices or smartphone apps – will connect to the care system to populate providers’ and care managers’ electronic health records (EHRs) and individuals’ personal health records (PHRs) so that all who care for and support the patient have access to the same health data and care plan. This is a necessary evolution if patients are to take greater personal responsibility for their own health improvement, as signified by the consumer slogan, “Nothing about me, without me.” Empowering individuals to take better care of their health is an important element of health reform that gets scant attention, but can pay off with big results.

Expansion of HIT and Increased Use of Modular Applications to Promote Health Information Exchange

Electronic medical records (EMRs) and personal health records hold great promise for enhancing care coordination, eliminating waste and duplication and providing individuals with greater resources for improving their own health than ever before. Regrettably, adoption of these tools by providers has been hampered by a variety of factors, including cost and usability issues. Newer EHR offerings that are delivered on a Web-based modular platform, require little or no capital expenditure (with the “software as a service” subscription model), and allowing for piecemeal implementation at users’ preferred pace may help to accelerate deployment and usage of these important tools for improving health care. This will require not just a proliferation of stand-alone EMRs across the health care landscape, but, rather, interoperable systems providing “meaningful use” through a variety of forms of health information exchange. Only by networking these systems, while safeguarding security and privacy, will health care move to the connected “medical neighborhoods” and “medical villages” necessary for scalable health improvement. The HIT Framework on page 24 of the report visually depicts the HIT components necessary for population health.

Participant Engagement, Incentives, Personalization

As CCA reported first in 2008 and again in 2010, in its “Population Health Improvement: A Market Research Survey,” the top concern of many purchasers of population health services remains improving rates of participation by those who can benefit from these programs. This is an understandable focus for those seeking to maximize the value of health improvement investments, and population health management has responded with new modes of recruitment and engagement; communication personalized to individuals’ preferences; novel incentive structures that leverage behavioral economics principles; and contact-level reporting to show all interactions with participating individuals leading to reported outcomes. Challenges remain however – motivating, engaging and empowering individuals to become better stewards of their own health has never been easy work, but meaningful progress is apparent.
Outcomes Initiative Update

CCA released its first Outcomes Guidelines Report in 2006. That report resulted from an 18-month effort to begin development of consensus-based outcomes measurement guidelines. The guidelines and recommendations in this fifth volume of work continue the efforts to bring education and standardization to methodology and measure sets that can be used to show the true value and impact of wellness, chronic care management and broader population health management programs.

What Our Recommendations Are

- The result of a consensus effort to create a standardized method for determining population health management outcomes that meets suitability and acceptability requirements across a wide variety of populations and circumstances;

- A standardized method based on current best practices;

- An effort to better manage some of the most prevalent challenges now encountered in determining population health management outcomes in non-experimental settings such as routine reporting; and

- An intermediate step in evolving practical and reliable methods to aid comparisons of different programs’ performance.

What Our Recommendations Are Not

- A prescriptive method intended to replace all other methods for determining population health management outcomes;

- A formulaic recipe for “plug and play” outcomes determinations by unsophisticated population health management program reviewers;

- An ideal method for all populations under all circumstances; and

- The last word in evolving standardized methods that facilitate inter-program and intra-program comparisons of performance.

This year’s Volume 5 report includes the recommendations from Volumes 1, 2, 3 and 4, in addition to the recommendations developed and approved this year. The Guidelines Locator on pages 9 and 10 identifies the development stage—beginning, intermediate, or advanced—of each of the recommendations. In addition, all of the new or revised recommendations from this year are shown in bold. While much of the 2010 work focused on PHM program process and evaluation, the medication adherence organizational best practice model as well as the selection criteria and operational measures sections are noteworthy additions to the Chronic Care Management — Five Core Chronics section.
### TABLE I – GUIDELINES LOCATOR

#### ALIGNMENT

<table>
<thead>
<tr>
<th>GUIDELINE</th>
<th>PAGE #</th>
<th>BEGINNING</th>
<th>INTERMEDIATE</th>
<th>ADVANCED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alignment – Program Checklist</td>
<td>12</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alignment Algorithm</td>
<td>13</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### POPULATION HEALTH MANAGEMENT

<table>
<thead>
<tr>
<th>GUIDELINE</th>
<th>PAGE #</th>
<th>BEGINNING</th>
<th>INTERMEDIATE</th>
<th>ADVANCED</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHM Program Definition</td>
<td>16</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Program Guiding Principles</td>
<td>17</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Program Capabilities</td>
<td>18</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>PHM Program Frameworks</td>
<td>18</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Conceptual PHM Framework</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• PHM Program Process Framework (Refined 2010)</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Health Information Technology Framework for PHM</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methodological Considerations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methodology Framework: Study Design</td>
<td>27</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methodology Framework: Adjustment</td>
<td>30</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attribution Adequacy – Illustration &amp; Examples</td>
<td>30</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Measure Sets

<table>
<thead>
<tr>
<th>GUIDELINE</th>
<th>PAGE #</th>
<th>BEGINNING</th>
<th>INTERMEDIATE</th>
<th>ADVANCED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program Evaluation Core Measures</td>
<td>35</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### WELLNESS

<table>
<thead>
<tr>
<th>GUIDELINE</th>
<th>PAGE #</th>
<th>BEGINNING</th>
<th>INTERMEDIATE</th>
<th>ADVANCED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wellness Program Definition</td>
<td>37</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Program Comparison</td>
<td>37</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Methodological Considerations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of Control or Comparison Group</td>
<td>43</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Use of Claims Data for Wellness Program Evaluation</td>
<td>46</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Wellness Program Model of Impacts</td>
<td>48</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Measure Sets</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary Outcomes Measures</td>
<td>51</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Timeline</td>
<td>51</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other Outcomes Measures</td>
<td>52</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>• Quality of Life</td>
<td>52</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>• Presenteeism/Absenteeism</td>
<td>52</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Productivity Measure</td>
<td>53</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GUIDELINE</td>
<td>PAGE #</td>
<td>BEGINNING</td>
<td>INTERMEDIATE</td>
<td>ADVANCED</td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
<td>--------</td>
<td>-----------</td>
<td>--------------</td>
<td>----------</td>
</tr>
<tr>
<td>Disease Management Definition</td>
<td>55</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Methodological Considerations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluation Design</td>
<td>57</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Population Identification Methodology</td>
<td>60</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Defining the Population</td>
<td>61</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>♦ Inclusion Criteria</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>♦ Look-Back Period</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>♦ Claims Run Out</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>♦ Measurement Period</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>♦ Member Month</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methods to Define Outliers/Exclusions</td>
<td>63</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Selection Criteria</td>
<td>65</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>The Use of Trend</td>
<td>71</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>The Use of Risk Adjustment</td>
<td>73</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Evaluation Considerations for Small Populations</td>
<td>82</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Evaluation Considerations for Mature Chronic Care Management Programs</td>
<td>87</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Measure Sets</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Financial Measures</td>
<td>92</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Utilization Measures</td>
<td>94</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Clinical Measures</td>
<td>96</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>♦ Group 1</td>
<td>97</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>♦ Group 2</td>
<td>98</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>♦ Self Management</td>
<td>99</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>♦ Medication Adherence</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Organizational Approach to Best Practice</td>
<td>101</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pharmacy vs. Medical Benefit</td>
<td>102</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measure of MPR (Refined as of 2009)</td>
<td>106</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measure of Medication Persistence</td>
<td>107</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measures of Self-reported Medication Adherence</td>
<td>112</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operational Measures: Operational Flow Diagram (Refined 2010)</td>
<td>114</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Operational Measures</td>
<td>116</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>♦ Health Status</td>
<td>117</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>♦ Quality Measure – Satisfaction</td>
<td>117</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
Evolution of the Industry and the Guidelines

Over the past decade we have seen an evolution in both the science of chronic care management and the market for these services. At first and at its most basic, the industry targeted individuals who suffered from one (or possibly more than one) of a specified set of chronic illnesses. Most commonly, this set of chronic illnesses included heart disease, diabetes, COPD, asthma and congestive heart failure. For each of these chronic illnesses there is a well-established standard of care, and compliance with this standard is known to improve clinical outcomes. Thus, the focus of most early programs was to identify and eliminate the gaps in these standards of care.

Also in the past 10 years, the industry has expanded its services to include wellness and health promotion. The premise behind wellness programs is that there are multiple modifiable risk factors that lead to many of the diseases targeted by chronic care management programs. While by no means a new idea (Johnson & Johnson launched its first corporate wellness program in 1978), there was a dramatic uptick in awareness and appreciation of the underlying logic of wellness programs and of their presence in the market, as well as purchaser interest. More often than not, wellness programs and chronic condition management programs existed as discrete entities in the market. The program provider community typically consisted of companies who delivered either care management or wellness programs.

The CCA Outcomes Guidelines process followed this evolution and began with the development of a set of guidelines focused on single-condition chronic care management programs and added a set of guidelines specific to wellness programs in year three of the process. Even as this year’s work neared completion, the market has continued to evolve. This time, the direction of this evolution is toward what is known as population health management (PHM). The PHM section of this report, on pages 16-35, offers definition and guidance on this topic.

This evolution does not suggest that every program can or should follow this pattern. Certainly, there will continue to be a continuum of program types and levels of comprehensiveness in the industry based on the specific needs of a population. Just as CCA refers to a “continuum of care” that exists in the management of a population, there is a program continuum, as well, for varying populations.

Although this evolution has dramatically changed the industry in the past decade, this does not negate earlier efforts to provide program evaluation guidelines for chronic care management and wellness programs. It has simply created the need to integrate PHM programs into the existing program evaluation framework. The goal of the Alignment Workgroup this year was to align both the chronic care management and wellness guidelines as much as possible with this industry evolution and ensure their compatibility with PHM program evaluation methods. The first part of this effort is to identify exactly what type of program or programs are in place for any particular population, followed by the development of a simple decision tree that stakeholders may utilize to determine which guidelines would be most appropriate for the programs in place.
ALIGNMENT – PROGRAM CHECKLIST

The purpose of the program checklist (Table II) is two-fold. First, this checklist can serve as an auditing tool to identify possible program gaps. It is not assumed that the optimal program approach is to address every population with every mode of intervention. At the very least, both program providers and purchasers should be fully aware of program gaps so the ultimate program design represents the needs of the purchaser and the population. Second, the checklist can provide a basis for determining which program evaluation methodology to pursue.

<table>
<thead>
<tr>
<th>TABLE II – ALIGNMENT – PROGRAM CHECKLIST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Check all that apply.</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>I. GENERAL PROGRAM(S) SERVICES</td>
</tr>
<tr>
<td>A. Management of chronic conditions</td>
</tr>
<tr>
<td>B. Improvement in modifiable risk factors (e.g., diet, tobacco use, exercise)</td>
</tr>
<tr>
<td>C. Health/wellness information</td>
</tr>
<tr>
<td>II. TARGET OF PROGRAM</td>
</tr>
<tr>
<td>A. Healthy, low risk</td>
</tr>
<tr>
<td>B. Healthy, high risk</td>
</tr>
<tr>
<td>C. Healthy, high risk (targeted obesity or smoking cessation)</td>
</tr>
<tr>
<td>D. Disease management eligible/chronically ill</td>
</tr>
<tr>
<td>III. MODES OF PROGRAM INTERVENTION</td>
</tr>
<tr>
<td>A. Face-to-face</td>
</tr>
<tr>
<td>B. Telephonic coaching</td>
</tr>
<tr>
<td>C. Telephonic case management</td>
</tr>
<tr>
<td>D. On-site resources</td>
</tr>
<tr>
<td>E. Web tools</td>
</tr>
<tr>
<td>F. Written materials</td>
</tr>
<tr>
<td>IV. MODES OF INTEGRATION OF DM/WELLNESS PROGRAM ELEMENTS</td>
</tr>
<tr>
<td>A. Fully integrated from single vendor</td>
</tr>
<tr>
<td>B. Broker integrated from multiple vendors</td>
</tr>
<tr>
<td>C. Loosely coordinated among multiple vendors</td>
</tr>
<tr>
<td>D. Stand-alone program from multiple vendors</td>
</tr>
<tr>
<td>V. PROGRAM(S) FOCUS</td>
</tr>
<tr>
<td>A. Identifying gaps in care of chronic patients</td>
</tr>
<tr>
<td>B. Changing modifiable risk factors</td>
</tr>
<tr>
<td>C. Addressing illness behavior/quality of life</td>
</tr>
<tr>
<td>D. Education/information</td>
</tr>
<tr>
<td>CHECKLIST ITEM</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td><strong>VI. DATA ELEMENTS/OUTCOMES</strong></td>
</tr>
<tr>
<td>A. Claims (costs)</td>
</tr>
<tr>
<td>B. Claims (utilization)</td>
</tr>
<tr>
<td>C. HRA/risk modification</td>
</tr>
<tr>
<td>D. Coaching record</td>
</tr>
<tr>
<td>E. PBM records</td>
</tr>
<tr>
<td>F. Biometric measurements</td>
</tr>
<tr>
<td>G. Medical record</td>
</tr>
<tr>
<td>H. Productivity measures (presenteeism, absenteeism)</td>
</tr>
<tr>
<td>I. Quality of life</td>
</tr>
<tr>
<td>J. Psychosocial measures</td>
</tr>
<tr>
<td>K. Process measures</td>
</tr>
<tr>
<td><strong>VII. DATA INTEGRATION</strong></td>
</tr>
<tr>
<td>A. All data elements from all program components fully integrated</td>
</tr>
<tr>
<td>B. Some integration among data elements</td>
</tr>
<tr>
<td>C. No integration among disparate data elements</td>
</tr>
<tr>
<td><strong>VIII. PROGRAM DURATION</strong></td>
</tr>
<tr>
<td>A. Year to year</td>
</tr>
<tr>
<td>B. One to three years</td>
</tr>
<tr>
<td>C. Three + years</td>
</tr>
<tr>
<td><strong>XI. EVALUATION OBJECTIVES</strong></td>
</tr>
<tr>
<td>A. Net evaluation of all program effects</td>
</tr>
<tr>
<td>B. Attribution of individual program effects</td>
</tr>
<tr>
<td>C. Both A &amp; B</td>
</tr>
</tbody>
</table>

**ALIGNMENT ALGORITHM**

Several key aspects of the Alignment-Program Checklist have been incorporated into an Alignment Algorithm (Figure 1) that provides for four possible outcomes with respect to the evaluation methodologies included in this report.

1. NA—The program is not covered by CCA program evaluation guidelines;
2. Disease Management Evaluation—The program is best evaluated using the chronic care management guidelines (pages 55-118);
3. Wellness Evaluation—The program is best evaluated using the wellness program evaluation guidelines (pages 36-54); and
4. PHM—The program is best evaluated using the population health management program evaluation guidelines (pages 16-35).
It is not intended that any single question or answer on this checklist will drive the decision nodes on the algorithm. Rather, having completed this checklist, it should be possible to navigate the algorithm in an informed fashion and arrive at a sensible and workable end point.

The targeted Population Segment for a program is represented in the blue and green boxes in the left half of the Alignment Algorithm. The General Program Services are also represented within these boxes, along a continuum from Wellness/Prevention to Clinical Management. To determine which evaluation method to use for a particular program, identify which box most closely represents the program of interest, and follow the corresponding arrow to the right. The top three boxes on the left represent programs that meet the criteria for population health management. (See page 18 of this report for an outline of the basic enhanced capabilities of a population health management program.)

The middle column of the algorithm concerns Modes of Program and Data Integration capabilities. Depending on whether programs are integrated and whether data can be integrated, the Evaluation Methodology chosen will then depend on the evaluation question. If the desire is to understand the net outcomes associated with a program across a single population, CCA's PHM Evaluation Methodology (pages 26-35) is recommended. If the intention is to understand the impact of individual programs, the recommendation is to conduct disease management and wellness evaluations separately. However, an important caveat: There likely will be overlap in participation among population members, so it is important to understand and represent that the individual effects of these evaluations are not additive.
Population Health Management

Overview

As health care purchasers pursue new solutions to clinical, cost, wellness and other issues, a new choice of support services has emerged: “population health management” (PHM), which encompasses a much broader approach than typical chronic care management, wellness or case management programs. The scope of these services presents compelling new challenges in outcomes measurement. Typical evaluation models may be inappropriate; for example, methods that CCA earlier recommended for measuring chronic condition management outcomes likely will not work well when much broader populations are targeted with a broader set of interventions. Clarifying the expectations of population health management is a critical step in developing appropriate outcomes measurement approaches.

This section includes guidelines that focus on the evaluation approaches for population health management programs, as well as guidance on key components and capabilities of these types of programs. In addition, there are several framework models that visually display program components, key stakeholders and operational program flow, as well as the essential health information technology components. These guidelines will continue to evolve with the industry and are designed to complement any and all health care practice models.

Section guidelines include the following:

1. PHM Program Definition
2. Program Guiding Principles
3. Program Capabilities
4. PHM Program Frameworks
5. Methodology Framework: Study Design
6. Methodology Framework: Adjustment
7. Attribution Adequacy – Illustration & Examples
8. Program Evaluation Core Measures

PHM PROGRAM DEFINITION

CCA recommends the following definition of a population health management program:

- A population health management program strives to address health needs at all points along the continuum of health and well-being through participation of, engagement with and targeted interventions for the population. The goal of a population health management program is to maintain or improve the physical and psychosocial well-being of individuals through cost-effective and tailored health solutions.
**PROGRAM GUIDING PRINCIPLES**

CCA recommends the following guiding principles for evaluation of population health management programs:

**GUIDING PRINCIPLE I: UNDERSTANDING CLIENT/CUSTOMER EXPECTATIONS**

It has become increasingly clear that inclusion of client and customer expectations into the overall analysis and reporting process is a critically important ingredient in outcomes evaluation. Not only do service providers need to receive input from clients regarding expectations for the programs, clients and customers themselves need considerable education to formulate expectations around what programs can and cannot deliver. Clients, for example, are increasingly interested in total population measures that address support for the entire spectrum of health plan or employee groups, rather than isolated single- or multi-condition programs. Many programs already have incorporated a total population approach.

**GUIDING PRINCIPLE II: METHODOLOGY TRANSPARENCY**

CCA continues to recommend methodology transparency as an important guideline for any evaluation, regardless of the population targeted. Results should be clear, compelling and easily explainable, a task that is often difficult for the complex scenarios created in standard practice.

**GUIDING PRINCIPLE III: THE APPROPRIATE USE OF ADJUSTMENT TO ACHIEVE COMPARISON GROUP EQUIVALENCE**

CCA recognizes that the randomized controlled trial (RCT) is a highly regarded study design for both clinical and medical research, but also recognizes that, for a variety of reasons, RCTs may be impractical or impossible in real-time population programs at either the employer or health plan level. Absent the ability or desire to do formal RCTs, evidence from the field strongly suggests that efforts to evaluate programs be accompanied by an equal comparison group if a true randomized control is unavailable. The use of a comparison group offers a comparative view as to what might be possible directionally, if not absolutely, from the outcomes of medical management programs. These methods, often termed “quasi experimental” approaches, have their own literature and rules that must be observed.

**GUIDING PRINCIPLE IV: USE OF UTILIZATION MEASURES**

The inclusion of utilization measures in outcomes analyses is increasingly important to those individuals who believe that utilization measures represent a real proxy for program effectiveness. Given that multiple variables exist in clinical program delivery and that many are changing over time (such as unit price), it stands to reason that utilization measures are an important indicator and supporting metric for downstream claims of financial improvement. The argument here is that it is difficult to justify savings that might be methodology-driven if utilization measures do not reflect a significant impact on major variables of cost, such as hospitalization, surgeries, emergency department visits, etc. Connection of results presented to utilization changes occupies the focus of most “plausibility” type statistics.
PROGRAM CAPABILITIES

CCA recommends the following be considered as basic and enhanced capabilities of a population health management program:

<table>
<thead>
<tr>
<th>CAPABILITIES</th>
<th>BASIC CAPABILITIES</th>
<th>ENHANCED CAPABILITIES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engagement strategies</td>
<td>Active or passive</td>
<td>Combination of active and passive</td>
</tr>
<tr>
<td>Health risk assessment</td>
<td>Single data source</td>
<td>Multiple data sources</td>
</tr>
<tr>
<td>Risk stratification</td>
<td>Single dimension</td>
<td>Multiple dimensions</td>
</tr>
<tr>
<td>Tailored interventions</td>
<td>Population-oriented</td>
<td>Member-oriented</td>
</tr>
<tr>
<td>Program components</td>
<td>Multiple components along the continuum that address needs specific to healthy, at risk, and/or chronically ill</td>
<td>Multiple components integrated and coordinated</td>
</tr>
<tr>
<td>Program delivery modalities</td>
<td>Single modality</td>
<td>Multiple modalities</td>
</tr>
<tr>
<td>Information exchange with providers</td>
<td>One way – outbound</td>
<td>Two way – inbound and outbound</td>
</tr>
<tr>
<td>Outcomes measurement</td>
<td>Evaluate effect on outcome domains (separately)</td>
<td>Assess root cause impact on outcomes (such as individual behavior changes)</td>
</tr>
</tbody>
</table>

PHM PROGRAM FRAMEWORKS

The Care Continuum Alliance has developed frameworks to illustrate, both conceptually and operationally, population health management programs. These frameworks have as their origins the previously outlined definition of population health management and the program capabilities defined above. These frameworks have been developed as a guide for care delivery models seeking to integrate and implement population health management programs and processes. For example, care delivery models such as integrated delivery systems and accountable care organizations as well as primary care providers in patient-centered medical home practices can adopt the processes and key components outlined in these frameworks to assess their own capabilities and to guide the development of expanded and integrated care delivery models.

Work to develop the conceptual and operational PHM frameworks repeatedly drew attention to the expanded use of varied health information technologies as a requirement for fully operationalized PHM programs. As such, the Care Continuum Alliance developed a complementary health information technology framework. The HIT Framework identifies key information and technology components and can help all delivery models understand the components necessary to deliver a comprehensive program focused on population health.
In addition to a revised version of the 2009 released PHM Framework, a conceptual population health management framework is included in this report, (Figure 2, above). The intent of this model is to complement the detailed PHM Program Process Framework by identifying the general components and stakeholders of population health. Like the detailed PHM Program Process Framework, the Conceptual PHM Model depicts the identification, assessment and stratification of program participants. The core of the model (central blue box) includes the continuum of care, as well as health management interventions. The person is central in the model, and is surrounded by various overlapping sources of influence on the management of his or her health. This can include, but is not limited to, organizational interventions, provider interventions and family and community resources. Operational measures are represented under program outcomes, as are the core outcome domains from the PHM Program Process Framework. Finally, the cycle of program improvement based on process learnings and outcomes is prominently depicted by the large curved green arrows.

**Conceptual PHM Framework**

In addition to a revised version of the 2009 released PHM Framework, a conceptual population health management framework is included in this report, (Figure 2, above). The intent of this model is to complement the detailed PHM Program Process Framework by identifying the general components and stakeholders of population health. Like the detailed PHM Program Process Framework, the Conceptual PHM Model depicts the identification, assessment and stratification of program participants. The core of the model (central blue box) includes the continuum of care, as well as health management interventions. The person is central in the model, and is surrounded by various overlapping sources of influence on the management of his or her health. This can include, but is not limited to, organizational interventions, provider interventions and family and community resources. Operational measures are represented under program outcomes, as are the core outcome domains from the PHM Program Process Framework. Finally, the cycle of program improvement based on process learnings and outcomes is prominently depicted by the large curved green arrows.
FIGURE 3 – PHM PROGRAM PROCESS FRAMEWORK

Population Monitoring/Identification

Health Assessment
- HRA
- Medical Claims
- Lab Data
- Other

Risk Stratification
- Healthy
- Health/Emotional Risk
- Chronic Illness
- End Of Life

Enrollment/Engagement Strategies

Communication and Intervention Delivery Modalities
- Mail
- E-mail
- Telephone
- Internet/Intranet
- Social Media
- Face-to-Face Visits

Health Management Interventions

Organizational Interventions
- Culture/Environment
  - Health Promotion, Wellness, Preventive Services
  - Health Risk Management
- Care Coordination/Advocacy
  - Disease/Case Management
- Program Referrals (External/Internal)
- Integrated/Coordinated Components

Tailored Interventions

Operational Measures
- Psychosocial Drivers
  - Self-Management
  - Screening/Preventive Services

Program Outcomes

Health Status and Clinical Outcomes
- Quality of Life
- Productivity
- Satisfaction

Service Utilization

Financial Outcomes

Incentives & Rewards
- Incentive Enrollment/Engagement
- Reward Participation Outcomes

Time frame for Impact

³³³³

³ For a more detailed discussion of monitoring and identification flow please refer to the work of the operational measures workgroup.
² Represents example components for each essential element. Does not necessarily reflect the universe of components.
³ Communication may utilize one or more touch points within the delivery system.
**PHM Program Process Framework**

In 2009, the Care Continuum Alliance published the Population Health Management (PHM) Framework. In 2010, to reflect the evolving views on population health management, the workgroup made several adjustments to the original Framework. Much like its predecessor, the updated PHM Program Process Framework (see Figure 3, page 20) is intended to help improve our understanding of the essential elements of a PHM program and how they relate to one another. It outlines the process flow associated with operating a PHM program, beginning with monitoring the population and identifying individuals who are eligible for the program. It also includes an Assessment stage, followed by Risk Stratification, the application of Enrollment and Engagement Strategies, the availability of multiple Communication and Delivery Modalities, Health Management Interventions across the care continuum and Program Outcomes in multiple domains. Finally, it includes a feedback loop that reflects the need to incorporate process and program improvements based on learnings from operational measures and program outcomes. The sections below provide an updated description of the components of the PHM Framework and highlight what has changed and what has remained the same in the current version of the PHM Program Process Framework. It is clear that the population health management industry and its supporting technology will continue to evolve, so the PHM Framework also will evolve to reflect those changes in upcoming years.

**HEALTH ASSESSMENT**

While the assessment stage of the 2009 PHM Framework focused on “Health Risk Assessment,” the updated PHM Framework has been renamed “Health Assessment” to reflect the possibility that assessment of health can include additional factors other than those we more traditionally include in the health risk domain. Such factors can include, but are not limited to, environment, financial issues, psychosocial influences and outcomes, such as self-efficacy, resilience and optimism.

Health often is assessed using questionnaires to gather respondents’ self-reported information about current health behaviors, status regarding recommended screening and preventive services, safety precautions and other potential health risks. Other sources of health risk information include medical claims and pharmacy data and, if available, data on laboratory results for recommended tests. While these methods are among those commonly used, this is by no means a comprehensive list of possible health assessment approaches.

**RISK STRATIFICATION**

As outlined in the updated PHM Framework, the next step in the PHM program process is to stratify individuals into meaningful categories for personalized intervention targeting, using information collected in the health assessments. This component of the framework did not change from the previous version presented in Volume 4 of the Outcomes Guidelines Report.

Stratification should include categories that represent the continuum of care in the population. While some organizations use complicated and proprietary mathematical algorithms to predict risk, others use a simple count of risks to classify individuals. It is not our intent to prescribe how risk stratification should be conducted, but to emphasize the importance of having some type of stratification in place to help align individuals with appropriate intervention approaches and maximize the impact of the program.

**ENROLLMENT/ENGAGEMENT STRATEGIES**

Another area that remained the same from 2009 to 2010 is Enrollment and Engagement Strategies. Once individuals in a population are identified and stratified, population health management programs should utilize proactive strategies to...
enroll and engage people in the program. It is becoming increasingly evident that effective enrollment and engagement is key to impacting the health of a population. If the participation rate for a program is low, there is little chance the program will have a measurable impact on the population.

**COMMUNICATION/INTERVENTION MODALITIES**

Whenever possible, the components of a population health management program should be offered through a variety of communication or intervention modalities for efficient program implementation and/or to accommodate the preferences and technological abilities of program participants. Some individuals may prefer to receive everything through the mail, while others might want to participate through an online program. Some services are best delivered in direct communication by telephone or in face-to-face encounters. Others may want to make use of a combination of intervention modalities. The updated PHM Framework includes Social Media as a delivery modality to reflect the increasing popularity and promise of this means of health education and support.

Matching intervention modalities to the communication preferences of individuals likely will lead to an increased level of program participation and engagement, and ultimately to improved program outcomes.

**HEALTH MANAGEMENT INTERVENTIONS**

The most substantial change to the PHM Program Process Framework is the representation of health management interventions. In the updated framework, the Health Management Interventions section includes the Participant Health Continuum along which many program components can be placed. The Organizational Interventions box in the updated framework highlights the culture and environment within which many health management programs are delivered. To maximize the impact of a program, it is important to consider the environment of participants and, whenever possible, employ interventions designed to create a supportive environmental and organizational culture. The framework also reflects the partial overlap between Organizational Interventions and the Tailored Interventions box to represent that these interventions often form an integral part of the culture and environment of organizations, yet may be delivered in other ways, as well.

While the list of program types in the Tailored Interventions box is not exhaustive, Health Promotion, Wellness and Preventive Services are designed to help healthy individuals stay healthy. Health Risk Management programs are intended to help people manage any health risks they might have, and Care Coordination/Advocacy represents efforts to help people understand, navigate, manage and coordinate the health and health care resources available to them. Finally, disease management programs are designed to help people with chronic illnesses better manage their condition.

**PROGRAM OUTCOMES**

The final essential element of the PHM Framework is a conceptual outcomes framework. The outcomes domains represented in the updated PHM Framework are the same as those in the previous version with only a few minor modifications to the relationships among those outcomes. The depiction of Program Outcomes begins by representing the program processes as an early program “outcome.” As previously mentioned, a program can only be successful if it effectively touches a significant number of people in the population, and it is most likely to succeed if it is operating efficiently. Tracking these process-related “outcomes” is critical to a successful program.
The next link in the outcomes framework represents the implicit hypothesis that the population health management program will impact psychosocial variables that will then drive changes in health behaviors, including self-management and use of screening and preventive services. Improvements in these behaviors will, in turn, have a positive impact on health and clinical outcomes. For a fuller list of specific psychosocial, health behavior and health-related outcomes, please refer to the Impacts Model on page 47. For health-related outcomes specific to chronic conditions, refer to the measures sets in the Chronic Care Management - Five Core Chronics section of this report, pages 55-118.

As outlined in the outcomes section of the PHM framework, quality of life, productivity (for definition, see page 53) and satisfaction (including life satisfaction, satisfaction with care, satisfaction with the program, etc.) are overlapping constructs, all of which will be positively impacted by, and may have a reciprocal positive impact on, participants’ behavioral and health-related outcomes. While these three constructs overlapped in the earlier framework, in the updated framework their relationships with other outcome domains are represented as a group. Finally, the outcomes section of the PHM framework represents that improvements in health behaviors, health and clinical outcomes and productivity will ultimately impact service utilization and financial outcomes.

Outlining a framework for a program’s outcomes can have several practical applications. It can help systematize the program design and implementation and shape both the evaluation processes and outcomes reporting strategy. Whether the outcomes framework is created before or concurrent with the development of the program, it can help with the conceptualization of the overall program strategy and specific intervention approaches. Careful consideration of the chain of effects that will eventually lead to the ultimate program goal or outcome, and inclusion of those outcomes in the outcomes framework, can identify needed program components designed to impact those outcomes. Additionally, because there are many things that contribute to the financial impact of a program, explicitly outlining the predicted short-, intermediate- and long-term outcomes can help stakeholders understand the full range of impacts and the expected time frame for ultimately generating cost savings. Finally, a well-constructed conceptual outcomes framework can help with interpretation of program outcomes and shed light on the practical implications of evaluation findings. Demonstrating to stakeholders that short- and moderate-term program outcomes are occurring as expected can provide early evidence that a program is on track to deliver a longer-term impact. Conversely, if early outcomes are contrary to expectations, early reporting allows for midcourse corrections to the program.

INCENTIVES AND REWARDS

Many providers of population health management programs are looking to incentives and rewards to increase participation and engagement. As outlined in the framework, eligible individuals can be incentivized and/or rewarded for enrolling and/or engaging in the program, as well as for making progress. There are many considerations related to incentives and rewards, including type and amount of reward, regulatory issues associated with the Genetic Information Nondiscrimination Act (GINA), the Health Insurance Portability and Accountability Act (HIPAA) and the Americans with Disabilities Act (ADA), as well as effectively developing intrinsic motivation to help individuals maintain changes in behaviors over the long-term.
**QUALITY IMPROVEMENT PROCESS**

The final addition to the original PHM Program Process Framework is in the representation of the *Quality Improvement Process*. The quality improvement cycle depicted in the original framework primarily focused on intervention changes based on outcomes. The cycle of quality improvement in the updated framework includes changes to both interventions and program processes (including assessment, stratification and engagement/enrollment strategies) based on process learnings from operational measures, as well as program outcomes.

**Health Information Technology Framework for PHM**

Population health relies increasingly on data-driven, HIT-supported interventions. Numerous federal government initiatives are underway to expand the adoption and implementation of health information technologies (HIT). Appropriately, these initiatives focus on establishing standards, defining effective use, and measuring quality improvement achieved via the role of HIT. As technology continues to evolve, so will its ability to enhance the health consumer’s experience and enable communications and information sharing across all health care settings.

Care Continuum Alliance members develop, utilize and support numerous health technologies and components. The HIT Framework, developed jointly by the HIT Committee and the PHM Outcomes Workgroup, attempts to identify the key components of both health information and health technology necessary to fully operationalize population health management programs. These technologies extend beyond electronic medical records to encompass a wide spectrum of innovative technology devices and applications. As such, the HIT Framework is intended to complement and expand upon

![FIGURE 4 - HIT FRAMEWORK](image-url)
the Conceptual PHM Framework by first identifying the upper-most levels of data exchange, envisioned by the Health Information Technology for Economic and Clinical Health (HITECH) Act (a part of the American Recovery and Reinvestment Act, Public Law 111-5) as “health information exchanges,” through broad systems and information hubs, such as person-level databases, continuing to identify components through to end-user, or consumer-enable, devices.

**REGIONAL DATA LIQUIDITY**

This component of the HIT Framework circles the entire framework and refers to regional data hubs that compile and distribute data for a population that exists within a specific geographic region. Data can be collected from a variety of health delivery sources, including hospitals and providers, as well as non-traditional sources, such as community centers and public health agencies. Examples of these regional or high-level data hubs are envisioned through accountable care organizations, health information exchanges and regional health information organizations.

**SYSTEMS- AND PERSON-LEVEL DATABASES**

This component of the HIT Framework refers to the databases and systems used to identify, assess, stratify, and enroll a population. Both systems and databases enable the assessment, stratification and engagement processes identified in the PHM Program Process Framework, as well as ability to measure the program outcomes. This section includes both the data and technology needed to perform these functions. Examples of these databases and systems include electronic health records, as well as lab and claims processing systems.

**INFRASTRUCTURE AND SERVICES**

The third component of the HIT Framework is focused on the information and technology needed to support health providers in their efforts to enhance the services they provide to patients. This infrastructure combined with the complementary services enables the enrollment and engagement process as well as the process of communication and intervention delivery. Examples of infrastructure tools and the services included within them are rules engines, decision support tools and intervention-level databases.

**COMMUNICATION AND ENABLING DEVICES**

This component is focused on devices that allow and enhance communication between and among health care providers and health care consumers. These devices also enhance the ability for providers and consumers to exchange and share information and contribute to most of the processes outlined in the PHM Program Process Framework, including the process of enrolling and engaging, program delivery and measurement of outcomes. Examples of these devices include home health hubs, personal health records and monitoring devices.

**END-USER MEDICAL DEVICES**

This component of the HIT Framework includes devices used by health care consumers to communicate and exchange information with health care providers, including, but not limited to physicians. These devices contribute to the process of successfully communicating and delivering program components. Advancements in technology have produced a variety of such devices, including personal computers, cell phones and iPads.
Methodological Considerations

When conceiving of guidelines for PHM program evaluation, the workgroup started with some general goals and objectives. It was recognized that the environment in which PHM measurement occurs is sufficiently different, less well-developed and more varied from chronic care management; and that it would be unreasonable to imagine that there would be consensus around a given methodology or sufficient evidence to support the preference of one well-designed and executed approach over another.

The objectives are to move the industry forward in the following areas:

- **Establish minimum requirements for attribution.** As will be discussed later, evaluating programs designed and intended to affect an entire population challenges many of the assumptions embedded in the disease management evaluation guidelines presented in Volume 4. For example, the trend methodology, along with risk adjustment, helps establish estimates of what would have occurred if the disease management program had not intervened, all things being equal (ceteris paribus), which helped justify attribution — the ability to claim that changes in costs were caused by the program’s interventions. In the PHM world, where there is a potential for all strata of a population to be receiving interventions of some kind, there is no reasonable way to estimate what the “underlying” trend is and therefore no good way to establish a similar ceteris paribus assumption. Setting standards for identifying what is required to make a reasonable claim of causality, without conducting a randomized controlled trial, is an important step.

- **Set minimum standards for adjustment.** The populations targeted for PHM are more varied and opportunities for selection bias and non-comparability of study and comparison groups are expanded. There is a need to match adjustment approaches to the study design to minimize the impact of bias and confounding.

- **Establish framework to judge relative strengths of evaluation approaches.** The health services research and medical care literature are replete with examples of study designs that utilize various sources of data to evaluate the impacts of different interventions on populations. This diversity is useful to meet the needs of organizations that might have limited access to certain types of data but plentiful access to other types. In this context, and because there has not been a convergence of methodologies in PHM evaluation as in chronic care management, it is useful to remain open-minded and have a means to assess each methodology in an objective and systematic way. In addition, the opportunity for thought leaders to differentiate their products based on rigor of evaluation design is likely to raise the quality of these evaluations for the industry.

- **Define a research agenda to encourage studies to benchmark methodology and establish empirical standards for PHM evaluations.** Before identifying even minimal standards and eventually establishing a rigorous benchmark for PHM program evaluations, significant research is needed on methodology. A “comparative effectiveness” approach to various experimental and quasi-experimental study designs is needed to assess their validity and potential weaknesses. Since behavioral change is a key focus of PHM, social intervention evaluation approaches should be further studied and applied. These guidelines can help frame the research questions that will best advance the PHM industry.
**Client/Customer Expectations**

A PHM strategy includes coordinated program initiatives for the continuum of needs across a population and creates an expectation for outcomes measurement with equivalent span. Unlike stand-alone chronic care management, there is no intent to leave significant portions of the eligible population unaffected by the programs.

This comprehensiveness in intervention approach creates two specific evaluation needs: a consolidated methodology assessing performance across all program components; and a component-level assessment to differentiate the effectiveness of the individual programs and interventions within the overall approach.

For the purpose of focusing on a definable and common level of comprehensiveness and an achievable scope, these initial PHM evaluation guidelines focus on the combination of common chronic disease and wellness programs.

**METHODOLOGY FRAMEWORK: STUDY DESIGN**

The objective of measurement is to render a quantitative or qualitative description. To describe a program outcome, one can describe a state or a change that, to be valuable, must be valid. Validity has many flavors. For this purpose, one can imagine a hierarchy of validity types:

1. **Accuracy:** measurements correctly capture data and compute metrics. This definition is independent of the meaning of the results.

2. **Internal validity:** metrics and study design are constructed and executed in a way consistent with the intent of the measurement under specific circumstances. The measurement can be reasonably expected to represent what it is intended to measure — for example, a difference in the utilization of services from what would otherwise have been expected for this population.

3. **External validity:** the results of the evaluation can be generalized beyond the specific analysis, i.e., the measurement can be reasonably expected to represent the impact of the intervention across instances and settings. As such, measures and study outcomes are comparable broadly across programs and/or organizations — a difference in the utilization of services from what would otherwise be expected for any similar population, for example.

The following framework outlines the prevailing approaches that are being or could be used to assess program effectiveness. As previous Outcomes Guidelines Reports have noted, the more rigorous a method’s ability to produce conclusive or valid results, the more impractical it is likely to be for routine, periodic reporting. These are noted in both the complexity and key issues rows.
<table>
<thead>
<tr>
<th>TABLE IV – METHODOLOGY FRAMEWORK: STUDY DESIGN</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DESIGN</strong></td>
</tr>
<tr>
<td><strong>Source of Comparison Group</strong></td>
</tr>
<tr>
<td><strong>Threats to Validity</strong></td>
</tr>
<tr>
<td><strong>Complexity</strong></td>
</tr>
<tr>
<td><strong>Key Issues</strong></td>
</tr>
</tbody>
</table>

**Randomized Controlled Trial (RCT):** A randomized controlled trial is a scientific experiment to evaluate the effectiveness of an intervention (or a product). Subjects are randomly allocated to the treatment group that receives the intervention and the control group that does not. The random assignment is meant to maximize the probability that both groups are fully comparable at baseline so that later differences can be attributed to the intervention. Thus, the RCT is considered the gold standard of study design. Its disadvantages are the complexity of organizing such a trial, concerns about the limited generalizability of the findings because of the tight control over the intervention under trial conditions, and the reluctance of assigning subjects to a control group that may not receive treatment.

**Study designs with non-experimental comparison groups:** A strong non-experimental design requires the comparison group to be drawn from comparable populations or to use individuals matched to the intervention group members. Comparability can also be achieved with statistical adjustment. Below are some examples of control group opportunities:

- **Non-contracted group:** This can be referred to as natural experiment. In this design, researchers try to find a naturally occurring control group that is equivalent to the treatment group in all other aspects but exposure to the intervention. Thus, the design is similar to an RCT but without the random assignment of individuals. For example, if a health plan offers a care management intervention to the fully insured but not to the self-insured business lines in the same region, the members in the self-insured products could be investigated as an equivalent, non-randomly assigned comparison group.
• **Non-qualified individuals**: The comparison group consists of individuals who were not identified for the chronic care management intervention. This group can include individuals with no chronic conditions, with chronic conditions that are not part of the chronic care management contract, or with chronic conditions that are part of the chronic care management contract but fail to meet program inclusion criteria.

• **Non-contacted individuals**: The comparison group consists of individuals who were eligible for the intervention but not contacted by the chronic care management organization. Examples are individuals with incomplete addresses or missing phone numbers.

• **Non-participants**: The comparison group consists of individuals who were identified for a chronic care management intervention but chose not to participate.

**Concurrent Industry Trend**: This trend would be estimated by combining longitudinal data from a large number of care management organizations that reflect the overall industry reasonably well. The trend would then become the equivalent of the S&P 500 for the chronic care management industry, i.e., a national benchmark against which individual programs would be compared. For a fair comparison, the trend would have to be adjusted to account for differences in region, case-mix, etc., between the national peer group data and the individual program. Of note, a program’s effect would not reflect the impact compared to no treatment but compared to the rest of the industry.

**Historic control**: This design compares the trend in the sub-population equivalent to the intervention population prior to start of the intervention and afterward to an expected trend based on analysis of the rest of the population. Any difference in observed findings to the expected results is credited to the intervention.

**Pre-post comparison**: This design compares baseline to post-intervention data. The difference is counted as program effect.

Most industry activity has been focused in the historic control method. Comparison group methods, which are derived from within the same population in which the program was implemented, such as non-participants, are currently emerging both in terms of population and individual (matched control) comparisons. Limited opportunities arise to conduct studies, in which the comparison group is identified from outside the program population, such as a different employer group and region of the health plan. Even rarer is the situation to conduct a prospective randomized control design which is best suited for special studies and research.
METHODOLOGY FRAMEWORK: ADJUSTMENT

Measurement adjustment is another framework dimension, in addition to study design or attribution. Measurement comparability requires appropriate adjustments for factors that influence the measure but are not related to the program intervention. For instance, adjustments for patient demographics (age and gender) are commonly applied to utilization rates. Adjustment can be considered along a continuum similar to study design approach related to complexity and validity, except that the adjustment methodology continuum is effectively cumulative; each successively “higher” level of adjustment is intended to actually or effectively include the methodologies below it. Outcomes Guidelines Report, Vol. 4 (see Volume 4, page 60), documents whether and how to approach risk adjustment for a particular program, which is dependent upon the outcome being measured and whether that measure is:

- impacted only by exogenous confounders and not the interventions, where there is no concern that program impacts will be altered by the adjustment (e.g., outlier exclusions, non-chronic trend), or
- impacted by exogenous confounders as well as by program interventions that potentially may be inappropriately distorted or discounted by the adjustment (e.g., condition prevalence or severity, case mix).

Control group approaches may require statistical adjustments, in addition to stratification and outlier adjustments. Linear regression modeling simultaneously adjusts and assesses the association of multiple variables to the outcome measure. When using a matched-control design to assess outcome effect, propensity score weights can be used to determine the appropriateness of the match between the individual participant and non-participants based on the appropriate covariates. These scores are a result of regression modeling. This approach requires careful attention due to the various parameter estimates involved. Otherwise, the appearance of rigor may mask invalid matches.

ATTRIBUTION ADEQUACY — ILLUSTRATION & EXAMPLES

Attribution Adequacy – Illustration of Framework

In an effort to illustrate adequacy of various approaches for valid attribution to the interventions or program provided, the following schematic combines study design and risk adjustment dimensions along a “strength” continuum of low to high. The two dimensions create an area that has been further delineated as “adequate” and “not adequate” for routine reporting of valid program outcomes to purchasers of population health management programs.

Approaches with lower strengths of attribution, such as pre-post and historic control, may be adequate when certain levels of adjustment are included. Though the adequacy concept is new, the schematic and the shaded areas are in line with previous CCA outcomes guideline recommendations.
Attribution Adequacy Examples

Examples have been provided to clarify how this framework can be used as a guideline in assessing validity among various study designs and measurement adjustments. An example with insufficient adjustment is provided. Adequate examples include previously endorsed approaches, as well as emerging methods.

EXAMPLE A – A historical control (pre/post) comparison of program participants excluding for high-cost outliers. Factors that are considered inadequate include: participants as their own control group vs. eligible population; using their own experience for historic trend; not stratifying by chronic condition.

EXAMPLE B - A historical control (pre/post) comparison of program eligible members. This example outlines a CCA recommended approach (see Trend, pp. 71-73) incorporating non-chronic trend adjustments, condition stratification and outlier exclusions.

EXAMPLE C – A participant/non-participant comparison at the population level. The two comparison groups are stratified by population demographics and condition. To further control self-selection bias, each of the sub-stratums is normalized/adjusted for risk differences.
**EXAMPLE D** — A program eligible/non-eligible member-level comparison. Covariates in a linear regression model adjust for variables associated with the outcome measure but not directly attributed to the program intervention.

**EXAMPLE E** — A program eligible/non-eligible member-level comparison. Eligible individuals are matched to non-eligible members based on similar propensity scores. Outcome changes between baseline and program time periods are then assessed. This is commonly referred to as “difference in differences” analysis.

---

**CHART II - ATTRIBUTION ADEQUACY EXAMPLES**

**Example A: Pre-Post Comparison**

**INTERVENTION GROUP:** Members actively participating in chronic care management program

**COMPARISON STRATEGY:** Historic cost trend based on 3 years of pre-intervention data

**LEVEL OF ANALYSIS:** Aggregated data

**ADJUSTMENT:** Truncation of outlier cost at $100K/PMPY

**Example B: Pre-Post CCA Recommended Method**

**INTERVENTION GROUP:** Members eligible for chronic care management program and identified with chronic condition

**COMPARISON STRATEGY:** Historic cost trend based on 2 years of pre-intervention data for the non-chronic population

**LEVEL OF ANALYSIS:** Aggregated data

**ADJUSTMENT:** Truncation of outlier cost at $100K/PMPY; pre-intervention trend relativity between chronic and non-chronic; stratification by condition and co-morbidity
Example C: Non-Participating Group as Comparison Population

INTRODUCTION GROUP: Members participating in chronic care management program

COMPARISON STRATEGY: Non-participating members, 2 years of baseline and 2 years of intervention data

LEVEL OF ANALYSIS: Aggregated data

ADJUSTMENT: Truncation of outlier cost at $100K/PMPY; stratification by disease, sex and age; risk score normalization

Example D: Non-Contracted Group as Comparison Population

INTRODUCTION GROUP: Members eligible for chronic care management program

COMPARISON STRATEGY: Members not meeting program eligibility criteria, 2 years of baseline and 2 years of intervention data

LEVEL OF ANALYSIS: Member-level data

ADJUSTMENT: Truncation of outlier cost at $100K/PMPY; multivariate regression

Example E: Natural Experiment

INTRODUCTION GROUP: Fully insured members for health plan, eligible for chronic care management program

COMPARISON STRATEGY: Self-insured members to whom program was not available, 2 years of baseline and 2 years of intervention data

LEVEL OF ANALYSIS: Member-level data

ADJUSTMENT: Truncation of outlier cost at $100K/PMPY; Propensity score matching; Difference in differences analysis
Methodology Caveats

Selecting an approach within the acceptable continuum of attribution and adjustment does not guarantee valid results. The following key considerations should be kept in mind.

1. Population Size — Previous CCA work (see Small Populations, p. 82) has illustrated the impact of population size on the variability of financial outcome measures. Measurements of populations less than 15,000 may have confidence intervals that span zero, even when using more advanced methodologies.

2. Acceptance — When using sophisticated statistical methods and population comparisons, the presentation of results must account for limited customer understanding. Also, these approaches are more challenging to replicate for audit purposes.

3. Implementation — As was previously pointed out, methods which improve measurement validity result in higher costs as the reporting process requires assessment by a sophisticated evaluator and can’t be fully automated.

4. Control Group — As programs increase engagement rates, the availability of non-participants as the comparison group source will be limited.

5. Specific Application — The methodologies described may be more suitable to chronic care, wellness and other true “population”-based programs than to case management and other low-prevalence programs.

6. Limited Empirical Evidence — There is general recognition that evaluations using comparison groups, such as match control, are more rigorous than population-based analyses previously recommended. However, empirical comparisons with randomized control designs have yet to be published.

Future Considerations

Addressing outcome evaluation methods is an ongoing process. Improvement in data access will create opportunities for better measurement. The industry also requires more robust demonstrations of value, which consultants are responding to with different methodologies. With this in mind, consideration should be given to:

1. Encouraging benchmarking studies comparing the results of evaluations using non-experimental studies to randomized control designs.

2. Promoting more familiarity with the application of social intervention evaluation methods previously referenced. The methods contained in this current framework focus on conclusiveness of program effect. In an effort to ensure valid conclusions, these methods may conservatively measure effect and not provide insight for process improvement on why interventions are effective. Social intervention methods can address why programs work by examining both the intervention and context in terms of population characteristics and the environmental circumstances.
3. Establishing industry-accepted, empirically based standards on:

- Engagement/participation
- Behavior change
- Clinical adherence
- Service utilization
- Direct health care cost
- Indirect cost

Measure Sets

PROGRAM EVALUATION CORE MEASURES

Program outcomes are identified as a basic capability and are a key program component in the PHM Program Process Framework. In a population health management program, there is a need for cross-cutting measures that could be used to assess impact at a population level, regardless of the differences in service offerings. It is clear that there is still a need for service- and disease-specific measures as well, but cross-cutting measures could be useful for assessing overall impact and program comparison. With this understanding in mind, a core list of measures was developed to support the overall evaluation of a PHM program.

CCA recommends that the core set of measures listed below be included in an evaluation of a population health management program.

- Medical costs
- Health care utilization — appropriate use
- Health risks/behaviors
- Quality of life
- Health status
- Productivity
- Psychosocial drivers
- Program satisfaction
- Process/operations measures
Wellness

Overview

Over the past decade, a fundamentally new benefit has arisen alongside, or from within, traditional chronic care management and population health management programs: the wellness program. Chronic care management focuses on optimizing medical care for individuals with specific chronic conditions, while wellness programs seek to prevent such illness, minimize risk and improve general health. In its 2010 analysis, “Population Health Improvement: A Market Research Survey,” CCA reported that 89 percent of surveyed health plans and health systems offer wellness programs. Further, 61 percent of current purchasers view wellness programs as a “must have” within their organizations, and 89 percent of purchasers predict a trend toward more wellness programs within their organizations.

CCA considers wellness programs and chronic care management programs to be subsets of population health management programs. As such, CCA will, where appropriate, adopt the standards developed by the methods, clinical outcomes, financial outcomes and other workgroups of its Outcomes Steering Committee. Relative to traditional chronic care management programs, wellness program outcomes measures and standards are less well-developed.

Sections guidelines include the following:

1. Wellness Program Definition
2. Program Comparison
3. Use of Control or Comparison Group
4. Use of Claims Data for Wellness Program Evaluation
5. Wellness Program Model of Impacts
6. Primary Outcomes Measures
7. Timeline
8. Other Outcomes Measures — Quality of Life, Presenteeism/Absenteeism
9. Productivity Measure
WELLNESS PROGRAM DEFINITION

CCA recommends the following definition of a wellness program:

Wellness programs are designed to:

- help individuals maintain and improve their level of health and well-being by identifying health risks and educating them about ways to mitigate these risks;
- increase awareness of factors that can affect health and longevity;
- enable individuals to take greater responsibility for their health behaviors;
- prevent or delay the onset of disease; and
- promote healthful lifestyles and general well-being.

Effective wellness programs employ a variety of behavior change techniques and lifestyle management strategies.

The following are examples of wellness program components (note that this list is not exhaustive):

- Health risk appraisal
- Biometric screening (e.g., blood pressure, cholesterol)
- Smoking cessation
- Weight loss
- Diet and nutrition
- Stress reduction
- Exercise and fitness programs
- Ergonomic programs
- Safety (both at the workplace and home)
- Sleep hygiene
- Health advocacy
- Disease screening
- Immunization

PROGRAM COMPARISON

Wellness programs target the total population and participation is not primarily driven by disease state. This approach differs from a total population chronic care management approach, which could offer programs across the full health spectrum, including both wellness and disease-specific components. Table V (next page) compares these differences.
Methodological Considerations

The logic of wellness programs, as expressed graphically in the Wellness Program Impacts Model (page 47), is straightforward. That logic can be expressed in narrative form as follows:

1. Modifiable risk factors and behaviors are known to have effects on clinically important biometric variables, such as blood pressure, body mass index (BMI), serum lipids/cholesterol, serum glucose, etc.

2a. These biometric variables are, in turn, associated with the development or exacerbation of specific disease states, such as heart disease, cancer, stroke and diabetes. Utilization of health care services to address these conditions results in costs to the health care purchaser.

2b. As well, productivity and quality of life can be adversely affected by these risk factors directly and indirectly by the associated diminished health status.

| TABLE V – COMPARISON OF CURRENT STATE OF CHRONIC CARE MANAGEMENT/WELLNESS PROGRAMS |
|-----------------------------------------------|-----------------|-----------------------------------------------|
| **Health status of target population**       | Always have chronic illness | Total population which varies from optimal health to chronically ill |
| **Primary focus of intervention**            | Optimization of chronic condition management | Modification of health risk behaviors |
| **Primary outcome metrics of interest**       | Health care costs and utilization rates (hospitalizations, ED visits and use of specific diagnostic and therapeutic procedures) | Specific health risk behaviors, indirect health-related costs (productivity, etc.) |
| **Unit cost of program**                     | Moderate         | Low to moderate |
| **Claims data availability (for purposes of evaluating program impacts)** | Usually available | Variable availability, less likely than with disease management program |
| **Time frame for impact on participant behavior** | Near term | Near term |
| **Time frame for impact on health status**   | Near to medium term | Near to long term depending upon program elements and aims |
| **Time frame for impact on health care costs** | Near to medium term | Medium to long term |
| **Use of health risk assessment for baseline assessment** | Variable, less likely than with wellness program | Nearly always |
| **Use of biometric screening data for baseline assessment** | Variable | Variable |
| **Status/availability of narrative coaching/encounter notes for analysis** | Variable | Usual (depending upon program type) |
3. As a result, a program that targets the modifiable risk factors should result in a healthier and more productive population, and reduced health care costs. (For the purposes of this discussion, modifiable risk factors will be designated as “surrogate outcomes,” and health status/health care costs as “primary outcomes.”)

The following guidelines should be considered when developing an evaluation design for a wellness program:

**CCA recommends the appropriate level of rigor to be applied to a wellness program evaluation should be predicated upon the factors listed below.**

Upon consideration of these factors, the determination of evaluation design should balance feasibility, cost of evaluation and the marginal value of the incrementally better data that will be produced by more rigorous designs.

- Program costs versus evaluation costs
- Program scope
- Program duration
- Availability of appropriate control group
- Availability of claims data
- Number of participants in program (sample size)

**CCA recommends that when the factors above mitigate against any experimental or quasi-experimental design, the limitations of the subsequent evaluation should be transparent and fully communicated to the purchaser.**

Foremost among these limitations are:

- Regression to the mean
- Selection biases
- Reporting biases

A purchaser of a wellness program is most interested in the question: “Did this program result in better health, lower health care costs and increased productivity and quality of life among those participating in the program?” A valid and reliable study design for answering this question would be the randomized controlled trial (RCT), in which potential participants were randomly assigned either to an intervention group receiving the program or to a control group with no intervention. The primary outcomes measures might include: total health care costs (from claims data), incidence/prevalence of particular disease states (from claims data), rates of specified diagnostic/therapeutic procedures (from claims data) and productivity and quality of life measures (e.g., instruments measuring absenteeism or presenteeism). In addition, sufficient time would need to be allowed for the study to observe the impact (if any) of the program. Such a study would provide the best evidence of causality, the best estimate of effect sizes and ultimately the best evidence as to whether meaningful outcomes are positively affected by the program. It is unfortunate such a study is rarely
undertaken for wellness program evaluations. The cost of such a study, the elapsed time to see credible results and the perception of denying benefits to individuals in the control group make implementing evaluations of this nature difficult and often impossible in a real-world setting.

A typical evaluation design would be a pre-post, within-group, cohort analysis of surrogate outcomes measures. That is, program participants are evaluated at baseline for specific behaviors (e.g., diet, exercise, smoking habit) and/or biometric variables (e.g., lipid profile, BMI, blood pressure) and these variables are re-measured at a later time (e.g., at the conclusion of the program or at some later follow-up point). Changes in the rates of behaviors and/or values of biometric variables at these two time points are then compared.

There are any number of obvious methodological shortcomings and threats to internal validity with such an evaluation. Notably, these shortcomings and threats include:

- **REGRESSION TO THE MEAN.** Wellness programs may require that members be in certain high-risk categories to be eligible to participate in the program. These risks could be behavioral, clinical or a combination of the two. The same risk factors that define program eligibility also may be used as outcomes measures. This scenario creates a high probability of regression to the mean affecting those outcomes; utilizing the evaluation design described above will not control for these effects.

- **SELECTION EFFECTS.** Program participants are likely to be self-selecting and already inclined to behave in a manner that would lead to improvements in the outcomes measures, irrespective of program implementation.

- **REPORTING BIASES.** By definition, the follow-up outcomes data used for such an evaluation would only include data from participants inclined to make the effort to provide the data — people who are more likely to have performed well in the program. As well, there may be concerns regarding the tendency of participants to provide acceptable (to the evaluators) data, rather than accurate data.

All these factors (as well as others that might be noted) lead to uncertainty regarding the causal relationship between the wellness program and any improvements in these surrogate outcomes measures. In addition to this uncertainty, there is the question of whether these outcomes measures are reliable surrogates or proxies for those outcomes that are of principle interest: health status and health care costs. As noted above, wellness programs are predicated upon this relationship between risk factors and health status, the exact nature and magnitude of this relationship is unknown and improvements in risk factors do not necessarily or inevitably translate into improvements in health and lowered health care costs.

Given this set of methodological shortcomings, this typical wellness evaluation design might not produce a high level of confidence that reported results are an accurate reflection of program effects, particularly in terms of the primary outcomes measures (health status, costs). Why then, perform such an evaluation rather than employ more rigorous methods? The reasons are manifold, mostly having to do with practicality and expediency. Among the more pressing reasons are the following:

- **COSTS.** There is a clear and strong relationship between design rigor and cost of evaluation – a better and more rigorous program evaluation will be a higher-cost program evaluation. The cost of the program evaluation must be commensurate with program costs and, in the case of wellness programs, those costs, per person, are often quite modest. There is no logic to performing program evaluations that cost a significant percentage of the program itself.
- **PROGRAM DURATION.** The wellness program impacts model does specify that health care costs, disability and other measures of health status will be positively affected. It also specifies that the period over which these changes take place would be measured in “years or decades.” Thus, any undertaking to measure these effects must also be in place over this extended period.

- **PROGRAM SCOPE/EFFECT SIZES.** Wellness programs vary in their content from very modest (e.g., access to information and resources) to more aggressive coaching programs targeted at high-risk individuals. At the lower end of this spectrum, effect sizes (in terms of the effects on health care costs and health status) are likely to be modest and difficult to measure. Even in the more aggressive programs, the effects of the surrogate outcomes on the primary outcomes may not be immediate or definitive.

- **AVAILABILITY OF APPROPRIATE CONTROL GROUP.** Much of the uncertainty around the within-group evaluation described above derives from the absence of a control group and can be corrected by the inclusion of a control in program evaluation. In turn, the inclusion of such should be predicated upon the availability of a control group that is sufficiently similar to the intervention group to warrant a comparison. The use of statistical methods may also be used (e.g., propensity matching) to create a comparison group or to adequately control for differences. The addition of an inappropriate control group to an evaluation process many times is worse than no control group, because it replaces the knowledge of uncertainty with the illusion of rigor.

- **SAMPLE SIZE.** It is always important to consider sample size in determining whether evaluation or study findings are valid. This is particularly so in the case of claims data. The variance and distribution of these data are such that extremely large samples may be required to create confidence intervals sufficiently narrow to be of value. (Please refer to the small population topic for a more detailed discussion, page 82.)

The above discussion would apply equally to a traditional chronic care management program as to a wellness program. However, the specific characteristics and attributes of wellness programs make the program design challenges more acute and more difficult to mitigate. As seen in the comparison table on page 38, there are distinct differences between the two programs, and these differences should be considered in the design of program evaluation.

Establishing a standard for wellness program evaluation should not create a situation in which the ideal is the enemy of the good. The fact that the RCT is not an easy option in wellness program evaluation should not deter the offering of programs that do not permit a rigorous evaluation. There is a considerable gulf between the default, pre-post, within-group design and an RCT, and within this gulf there are a number of design options that might be considered. The study design table on page 45 of this report reviews several useful design options. Ultimately, the question of whether or how to fill this gulf should be determined by answering the question:

> “Given a wellness program’s cost, scope, duration and number of participants, and given the availability (or non-availability) of claims data and of an appropriate control group, what program evaluation design is indicated, such that the additional cost associated with the additional rigor is warranted and can be justified by higher quality of data that results?”
The previous discussion describes the trade-offs between rigor and practicality that must be thought through when considering wellness program evaluation. As well, this discussion highlights the fact that there is not a “one-size-fits-all” model for program evaluation. How each organization balances those trade-offs will depend upon the specific characteristics of the program itself, as well as the extent to which practical limitations restrict evaluation options. The methodological recommendations that follow reflect those trade-offs and present a selection of options within which an appropriate program evaluation methodology can be identified that is both practical and meaningful.

A broad variety of programs can be categorized under the definitional umbrella of wellness programs. These may involve many different interventions, including telephonic coaching, worksite health seminars and demographically driven reminders for preventive screenings. But only a subset of wellness programs can be evaluated by using the recommended evaluation framework included in the CCA guidelines. The five criteria listed below include the minimum threshold a wellness program needs to meet for the recommended outcome methodology to be applied to evaluate its outcome.

**Minimum threshold for application of CCA recommendations in evaluating wellness programs:**

1. The program is designed to address modifiable health risk factors;
2. A tool is used to assess health risks;
3. Targeted intervention is used to support healthful behavior;
4. Individual member-level information is collected to measure outcomes; and
5. The outcome is measured against a comparison group or a baseline measure.

The first criterion is that the wellness program should address modifiable risk factors. A list of potential modifiable risk factors that may be used in wellness program evaluations is included in these guidelines, on page 51. The number of risk factors that should be addressed is not specified. It is entirely possible that a single-purpose wellness program might still qualify for evaluation based on the rest of the criteria if it only impacts one risk factor, such as smoking or obesity.

The second criterion is that a risk assessment tool should be used to assess an individual’s health risk. This tool may be used to identify the at-risk individuals for intervention and to assign different levels of intervention intensity based on the different risk levels. Health risk assessment (HRA) tools available in the marketplace fit these criteria extremely well, but this criterion does not exclude the use of home-grown questions or tools. There is no specific guidance on utilizing a health risk assessment tool that is validated for a particular use.

The third criterion is that there should be a targeted intervention designed to help members to establish healthful behaviors. A targeted intervention is not simply a “one-size-fits-all” approach. Different interventions should be applied for different risk factors and different intervention intensities should be designed for members at different risk levels.

The fourth criterion is that the information at the member level should be collected to measure program outcomes. For example, it is not only important to measure how many people participated in a program, it is also important to know who participated and who did not, and consequently who successfully achieved a specific health behavior change. The member-level outcome information provides an insightful picture on how the program achieved the desired outcome and proffers the ability to provide additional important lessons learned for improvement of the program.
The fifth criterion is that the outcomes of interest are measured against a comparison group or a baseline measure. It is important to compare the results from the intervention group with a comparison group so the effectiveness of the program can be measured by the incremental value. The use of comparison groups also lends itself to a more scientifically valid method of establishing the cause-and-effect relationship between the intervention and outcome. There are multiple types of comparison groups that can be used and each has its advantages and disadvantages. A discussion of different types of comparison groups is included later in this recommendation.

USE OF CONTROL OR COMPARISON GROUP

CCA recommends that the use of an appropriate control/comparison group should be determined by the size and scope of the program being evaluated and by purchaser preferences and should reflect the understanding of the strengths and limitations of the various evaluation methodologies shown in Table VI.

The fundamental question to be addressed by this recommendation is: Against what are the program’s effects being compared? And depending upon how that question is answered, the answers to other questions are provided: What effects can be attributed to the wellness program? How do these results compare with other groups that have participated in the program? How do program participants’ health and risk status compare before and after the implementation of the program?

The accompanying table (page 45) is a modification of Table II: Study Design, on page 16 of the CCA Outcomes Guideline Report, Volume 3 (2008). This modification reflects the differences between chronic care management programs and population health and wellness programs and populations. As one moves through the options from left to right, the level of rigor decreases and the level of practicality increases.

The most rigorous method of control is of an RCT – the first column in the accompanying table. As is true of any evaluation of a treatment or intervention, a randomized control is the only way of controlling for both known and unknown, measured and unmeasured confounding variables, and thus the surest method of establishing causality. However its practical limitations are considerable (cost, time frame, institutional review board (IRB) considerations) and a randomized control can be considered only in the context of scientific inquiry, rather than of program evaluation.

The next level of rigor is the non-randomized control groups (matched, and unmatched). A matched control would consist of a set of non-participants who have been identified as resembling the set of participants in a relevant manner. This might be done through sophisticated propensity matching techniques or through more basic matching of age, gender and health status. The goal of such matching is to attempt to control for important confounding variables that might skew evaluation results. The important distinction between this type of non-randomized control group and a randomized control is that the randomized group can control for confounding variables, both known and unknown, while the non-randomized group can only control (or attempt to control) for known and measured confounding variables.

An unmatched control is simply a set of non-participants selected without regard to any resemblance to the intervention group. Obviously, the degree to which an unmatched control does or does not resemble the intervention group will determine how effective a control it represents.
These three controls (randomized, matched and un-matched non-randomized) all have as their primary goal the establishment of a causal relationship between the program and the measured results. The term *Comparison Groups* used in the next two columns reflects the fact that this evaluation approach makes no attempt to control for any confounding variables or to make any inferences about causality. It is merely a comparison of a client's results to some other set of results. In the case of the *Comparable Employee Population*, the comparison group might be from a similar industry type, from the same geographic region or from a similar socioeconomic profile. The *Book of Business* comparison group is just that: a comparison of one client's results to the overall average book of business results. These comparison group evaluations have the virtue of being practical and of being relatively easy to perform, but, as noted, they cannot inform the question of causality.

Finally, the rightmost column, *Own Control*, describes the pre-post change that takes place in program participants without reference to any control or comparison group. This represents the most minimal and basic level of wellness program evaluation.
<table>
<thead>
<tr>
<th>METHOD TO DEVELOP A COMPARISON GROUP</th>
<th>Randomized Control</th>
<th>Non-Randomized Control</th>
<th>Comparison Group</th>
<th>Own Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>General description of method</td>
<td>Intervened population compared with individuals randomly selected to have services withheld</td>
<td>Intervened population compared with non-participants</td>
<td>Intervened population compared with similar employer group(s) in same program</td>
<td>Pre/post comparison of intervened population.</td>
</tr>
<tr>
<td>Comparison time frame</td>
<td>Concurrent to intervention</td>
<td>Concurrent to intervention</td>
<td>Prior period</td>
<td>Concurrent to intervention</td>
</tr>
<tr>
<td>Primary outcome measures</td>
<td>Modifiable risk factors, biometric variables</td>
<td>Modifiable risk factors, biometric variables</td>
<td>Modifiable risk factors, biometric variables</td>
<td>Modifiable risk factors, biometric variables</td>
</tr>
<tr>
<td>Population selection bias</td>
<td>None</td>
<td>Somewhat significant</td>
<td>Significant</td>
<td>None</td>
</tr>
<tr>
<td>Source of comparison group</td>
<td>Population for whom program was implemented, randomly selected group withheld from program</td>
<td>Population for whom program was implemented, purchase decision to not participate</td>
<td>Vendor data</td>
<td>Intervened population</td>
</tr>
<tr>
<td>Credibility of causal statements</td>
<td>Extremely strong</td>
<td>Moderate</td>
<td>Poor</td>
<td>Very poor</td>
</tr>
<tr>
<td>Control of confounding variables</td>
<td>Controls known and unknown, measured and unmeasured confounding variables.</td>
<td>Does not control confounding variables</td>
<td>Does not control confounding variables</td>
<td>Does not control confounding variables</td>
</tr>
<tr>
<td>Program sponsor resistance to approach</td>
<td>High</td>
<td>Moderate/High</td>
<td>Moderate</td>
<td>None</td>
</tr>
<tr>
<td>Ease of implementation</td>
<td>Very difficult</td>
<td>Difficult</td>
<td>Somewhat difficult</td>
<td>Easy</td>
</tr>
<tr>
<td>Clarity of method to lay audience</td>
<td>Very clear</td>
<td>Very unclear</td>
<td>Somewhat unclear</td>
<td>Clear</td>
</tr>
<tr>
<td>Multiyear application vs single year application</td>
<td>Much harder</td>
<td>Much harder</td>
<td>Somewhat harder</td>
<td>Same</td>
</tr>
<tr>
<td>Method availability</td>
<td>Rarely possible</td>
<td>Occasionally possible</td>
<td>Usually possible</td>
<td>Always possible</td>
</tr>
<tr>
<td>Bleed of interventions to comparison population</td>
<td>Control group may get provider-based interventions, other vendor interventions, secular interventions and self-care</td>
<td>Control group may get provider-based interventions, other vendor interventions, secular interventions and self-care</td>
<td>Comparison population already intervened</td>
<td>NA</td>
</tr>
<tr>
<td>Key strengths</td>
<td>Gold standard evaluation method</td>
<td>Possible to infer reasonable level of causality without experimental design</td>
<td>Least costly and easiest method of compared intervened to non-intervened population</td>
<td>Low cost, availability of data, ability to answer question asked: (How does my group compare to other groups that have been in the program?)</td>
</tr>
<tr>
<td>Key problems/biases</td>
<td>Cost, sponsor resistance, IRB imperatives, low generalizability</td>
<td>Cost, availability of control group HRA data</td>
<td>Limited ability to make causal inferences,</td>
<td>No inferences on causality are possible</td>
</tr>
</tbody>
</table>

**TABLE VI – CONTROL/COMPARISON GROUPS METHODS FOR WELLNESS PROGRAM EVALUATION**
USE OF CLAIMS DATA FOR WELLNESS PROGRAM EVALUATION

CCA recommends that when claims data are to be used to establish either a financial baseline or a utilization baseline, the following criteria should be met:

- Program scope and cost are commensurate with such an analysis.
- The claims data are routinely available.
- There is a minimum of three years of program exposure.
- The sample size is adequate.
- There is a means of attributing effects.
- Appropriate claims analysis methodology is established.

Although changing modifiable risk factors is considered the primary outcomes measure of wellness programs, the ultimate goal of wellness is to improve health and reduce health care costs. Regarding the latter, the most obvious and intuitively appealing way of evaluating costs would be through the direct measurement of health care expenditures, i.e., through the use of claims data. However, the opportunities to do so in a practical, valid manner are likely to be limited. When these limitations can be overcome, it may indeed be worthwhile and appropriate to conduct a claims-based analysis. This recommendation identifies what conditions should be met before undertaking a claims data analysis.

Herewith, a brief discussion of these conditions:

1. **Program scope and cost are commensurate with such an analysis.** A claims analysis will be a costly undertaking. The cost of this analysis must be considered in the context of the cost of the actual program. Only when the cost of a claims analysis represents a small fraction of program costs is such an analysis a reasonable option.

2. **The claims data are routinely available.** The implementation of wellness programs does not routinely require that claims data be in the hands of program vendors. Unless the vendor has routine access to client claims data, a claims analysis is likely to be impractical.

3. **There is a minimum of three years of program exposure.** As discussed previously in Outcomes Guidelines Report, Volume 2 (2007), and shown in the Impacts Model, (Figure 5, next page), the time frame of the anticipated cascade of events (behavior change > biometric change > health status change > health cost change) that will culminate in a reduction of health care costs is measured in years or even decades. The three-year minimum time frame of this recommendation is indeed a minimum, and depending upon the specifics of program scope and intensity, an even longer time frame might be necessary to measure cost changes in the claims data.

4. **The sample size is adequate.** The adequacy of the sample size should be established through a power analysis prior to any attempt at claims data analysis. The appropriate sample size will vary depending upon the variance and distribution of the specific data set and upon the hypothesized effect size. However, based on the sample size analysis shown in Volume 3 (Table XIV, page 71), it is possible to generalize and note that an adequate sample size (of program participants) will certainly, at a minimum, number in the thousands.
5. **There is a means of attributing effects.** It is commonplace for an individual employer to have in place multiple health management programs simultaneously. These might include chronic care management programs, population health management programs and focused wellness programs. Any effort to isolate the effects (specially, the effects on claims data) of an individual program must be predicated upon an evaluation methodology that succeeds in controlling for these variables.

6. **Appropriate claims analysis methodology is established.** Finally, any valid analysis of wellness program effects on claims costs will ultimately be dependent on the details of the actual analysis. These details and variables will include the length of baseline period, trending assumptions and specific statistical methods employed.
WELLNESS PROGRAM MODEL OF IMPACTS

The Wellness Program Model of Impacts is designed to outline the range of relevant outcomes for the evaluation of wellness programs (see Figure 5, page 47). The framework of this model also is designed to convey two important points. First, outcomes associated with wellness programs (and chronic care management programs, for that matter) are multidimensional and interdependent and, for the most part, follow a logical chain of effects. Second, given the temporal dimension implicit in this chain of impact, the expected time frames for effecting different types of outcomes may vary considerably. Proposing a comprehensive framework for the different domains of wellness program outcomes might lead to the expansion of current thinking about the range of relevant measurement areas and help set realistic expectations about reasonable time frames for demonstrating results for each.

Domains of Impact

Four primary domains of impact are outlined in the Model of Wellness Program Impacts:

- Process Measures
- Behavior Change and/or Modifiable Risk Factors
- Productivity/Quality of Life
- Utilization and Medical Costs

There also are several subcategories of impact within the four primary domains. While this framework is not meant to be an exhaustive list of all possible measures of interest for the domains, it is intended to be a relatively comprehensive representation of relevant wellness outcomes. Each domain of impact will be discussed in the remainder of this section, as will the hypothesized interrelationships among and within the domains.

PROCESS MEASURES

Process Measures include indicators of intervention contact or intensity (i.e., contact frequency, duration, type) that are assumed to contribute to the impact of the program. For example, effective outreach and engagement methods and high participation rates likely will contribute to the success of a program. It also is arguable that the number of appropriate contacts may be positively associated with improvements in program participants and that certain types of contacts may be more effective than others.

BEHAVIOR CHANGE/MODIFIABLE RISK FACTORS

As previously stated, the Wellness Program Model of Impacts in Figure 5 (page 47) reflects the general hypothesis that improvements in various relevant outcome domains occur sequentially. Thus, wellness interventions (the components of which are represented by the Process Measures) lead to improvements in the second domain of impact, Behavior Change and Modifiable Risk Factors. There are several subcategories under this domain, including psychosocial drivers, health behaviors and health and clinical outcomes. Among the psychosocial drivers listed are motivational concepts common in popular theories of health behavior change. For example, self-efficacy is a major tenet of social cognitive theory and represents an individual’s belief about his or her ability to produce desired effects. Also, readiness to change, a primary component of the transtheoretical model, assumes that people move through discrete stages in the process of fully
adopting behaviors and that a better understanding of this process can help individually tailor self-management support in wellness interventions. Other psychosocial drivers that might impact change include, but are not limited to, social isolation, depression and perceptions of health. According to the Care Continuum Alliance (CCA) model, improvements in psychosocial drivers can positively influence health behaviors\textsuperscript{14,15}. Two categories of health behaviors are proposed: self-management behaviors; and preventive and screening services. Self-management behaviors are actions that people can usually undertake on their own to positively impact their health. Examples include eating a proper diet, exercising regularly, getting enough sleep and adhering properly to prescribed medication. Preventive and screening services generally require some type of contact with a health care provider and include obtaining recommended laboratory and cancer screenings, having blood pressure and BMI measured regularly and getting appropriate immunizations on a regular basis.

The final subcategory under \textit{Behavior Change/Modifiable Risk Factors} is health and clinical outcomes. According to the CCA model, improvements in health behaviors can positively impact clinical indicators of health\textsuperscript{16,17}. These include measured physical indicators of health, such as laboratory values (cholesterol, triglycerides, blood glucose), blood pressure and BMI, as well as self-reported health status as measured by reliable and validated measures.

While the primary directional impact among these subcategories of \textit{Behavior Change/Modifiable Risk Factors} is from left to right (represented by dark, solid arrows), the model reflects the possibility of reciprocal relationships among the subcategories (represented by lighter, broken arrows). For example, success in changing health behaviors can have a positive impact on self-efficacy for behavior change, which may help sustain the behaviors in the longer-term. There is also increasing evidence that exercise can positively impact depression\textsuperscript{18}, which could then improve motivation to maintain the behavior. Finally, experiencing clinical improvements may positively impact the maintenance of health behaviors and also strengthen efficacy, perceptions of health and other psychosocial drivers of health behaviors.

\textbf{PRODUCTIVITY/QUALITY OF LIFE}

The third major domain of impact represented in the model is quality of life and productivity. Beyond risk factor reduction and the associated or inferred financial benefit, there are additional benefits from a wellness program, such as improvement of quality of life (QOL) and improvements in productivity. These additional benefits should also be measured as part of the wellness program outcomes.

Quality of life also is included under this domain to capture the ability to improve and/or maintain important activities that contribute to a productive life for both employed and non-employed populations. The most common QOL and productivity measures are member self-reported metrics potentially collected through telephonic or written surveys.

A major benefit of wellness programs, especially in employer settings, is improved worker productivity. This includes both absenteeism, or the number of days work missed due to illness; and presenteeism, which is the capacity of an employee to work at his/her optimal level of productivity (i.e., to be fully present while on the job).

Productivity measures, especially absenteeism measures, through other data sources could be used when available. Scientifically validated instruments for measuring productivity and QOL are readily available and may be utilized; conversely, proprietary questionnaires designed to measure productivity and quality of life also are acceptable.
The fourth domain outlined in the model of wellness program impacts is behavior change and modifiable risk factor improvements that positively impact utilization and medical costs\textsuperscript{19,20}. This domain encompasses appropriate utilization to maintain wellness, including regular physician visits for recommended physical examinations, laboratory tests and preventive services. Also included are utilization and costs for avoidable emergency department visits and hospitalizations, and costs associated with workers’ compensation and short- and long-term disability.

**Time Frame for Impact**

As outlined in the model, there is a considerable range in reasonable time frames to expect an impact for the four major domains. Process Measures are obviously among the first available metrics of program success, followed by impacts in the Behavior Change and/or Modifiable Risk Factors domain. But even within this domain, there is quite a range of expected impact time frames. Psychosocial measures would be among the first outcomes measures expected to show a change following the initiation of a program, as it is possible to see improvements in a matter of weeks or months, depending on the metric (changes in self-efficacy may occur sooner than improvements in depression or perceptions of health). Next, changes in behavior also would be expected relatively early (weeks or months), but clinical impact would not be expected until quite a bit further along in a wellness population (months or years). Improvements in Productivity/Quality of Life might also occur in the moderate-term, depending on the population.

Program impact on some types of Utilization and Medical Costs may not occur until much longer-term. However, as with the Modifiable Risk Factors domain, there is a range of time frames for expected impact in the Utilization and Medical Costs domain. In the shorter term (weeks or months), physician visits and laboratory tests might increase due to more appropriate adherence to recommendations for screening and preventive services. In the moderate term (months or years), reductions may be seen in emergency department visits or outpatient procedures, but it might be years (or even decades) before significant impact is shown in medical costs, workers’ compensation offsets or long-term disability.
Measure Sets

PRIMARY OUTCOMES MEASURES

CCA recommends that the primary outcomes measures of wellness program evaluation should be modifiable behavioral risk factors and related biometric variables. The minimum set of these outcome measures may include:

- Diet
- Exercise/physical activity
- Tobacco use
- Alcohol/drug use
- BMI
- Blood pressure (BP)
- Blood glucose
- Stress levels
- Lipid profile

TIMELINE

CCA recommends that the measurement of primary outcomes measures should be made at baseline. In addition, re-measurement of primary measures should take place at 12-month intervals.

To the extent possible, every program participant’s risk factor profile, as defined above, should be measured at baseline, as closely as possible in time to that individual’s entry into the program. The choice of a 12-month follow-up period is arbitrary, but it does reflect the following considerations: that there should be sufficient time for program effects to manifest and that sufficient follow up time should have elapsed to measure the durability of program effects.

This recommendation should not be interpreted as precluding any shorter-term interim measures that might be relevant, such as engagement rates, levels of participation or even interim behavior change. Rather, the 12-month time frame is intended only to identify the time frame for primary program evaluation.
OTHER OUTCOMES MEASURES — QUALITY OF LIFE, PRESENTEEISM/ABSENTEEISM

CCA recommends that quality of life should be measured at baseline. Such measures can be made via single questions or by validated QOL instruments. In addition, both presenteeism and absenteeism should be measured at baseline.

Re-measurement of these outcomes measures should be made at 12-month intervals.

It is recommended that QOL and/or productivity should be measured at baseline when a member starts participating in the program. At the completion of the program, a re-measure of QOL and/or productivity should also be made so the QOL or productivity change could be captured. The re-measurement of these outcomes should be made at a minimum 12-month interval because wellness health behavior change typically occurs over a longer period of time on the order of many months to years.

CCA recommends program evaluations of worksite wellness and chronic care management programs consider adding assessments of productivity, especially presenteeism, as outcome measures used to evaluate the program.

While the ability of wellness and chronic care management programs to impact presenteeism is not well-established in the literature, the strong association between presenteeism and health/illness suggests that this is a potentially important outcomes measure.

CCA recommends careful consideration be given when selecting from the instruments available to ensure compatibility with the specific needs of the program, the setting and the target population.

CCA recommends that, given the lack of consensus within population health management on methodology and few results in the peer-reviewed literature, the conversion of presenteeism measures into financial outcomes warrants continued caution, especially when using these outcomes to establish or support far-ranging fiscal policies (e.g., benefit design or other corporate policies).

The monetization method selected should be agreed upon at the beginning of the implementation, and the benefits and limitations of that method should be made clear. Evaluators might also consider use of multiple methods.
PRODUCTIVITY MEASURE

Health-related, on-the-job productivity losses typically arise from two sources: absenteeism and presenteeism. Work absenteeism and its related costs have been measured using self-reporting of days missed for health-related reasons, as well as data from administrative records of absences, short- or long-term disability, workers’ compensation and Family Medical Leave. Challenges arise in using administrative data to assess the impact of wellness or chronic care management programs on workplace productivity, as it might be difficult to differentiate days lost to illness from absenteeism associated with non-health-related causes, such as child care, personal days and vacations. In addition, policy differences in disability and workers’ compensation may vary by state or by employer, further complicating comparisons using these data sources across multiple program sites.

Presenteeism – a Definition

Variations in the definition of presenteeism result from differences in focus – i.e., health-related, work-related or more general/life-related. Moreover, presenteeism has been used in both a positive sense (being fully present and productive while at work) and in a negative sense (being unproductive while being present at work) in more recent studies. For the purposes of our work, the definition of presenteeism will reflect the prevailing attitude of researchers and employers and therefore will retain the more pejorative connotation and the focus on both health and the workplace.

Presenteeism definition: Presenteeism is decreased on-the-job productivity associated with health concerns or problems.

Measuring Presenteeism

Presenteeism losses are challenging to both quantify and measure. They do not show up in claims data or appear on time cards. Productivity losses are also intrinsically difficult to measure, since much work output is not quantifiable and many measures must rely solely on self-report. Several different approaches have been developed as a result. These approaches include: assessment of perceived impairment; comparison of one’s own performance and productivity to that of others without health related-problems; and an estimation of unproductive time spent at work. All methods attempt to measure the same construct, loss of productivity, by asking the respondent to evaluate his or her own work performance as a function of time not on task and the quality and quantity of work produced as a result. Numerous survey instruments have been developed and assessed for validity, consistency and reproducibility.

No matter which questionnaire is used, the following factors should be considered in selecting an instrument that is appropriate for the setting, the population and the program:

- Instrument reliability and validity.
- Applicability across industries and occupations (appropriate for the target population).
- Applicability across the health care continuum (appropriate for the target population and program).
- Representativeness of the behavior recall period used by the survey to the study period.
• User-friendliness, available languages and reading level (appropriate for the target population).
• Mechanism of administration.
• Length of survey.
• Ability to integrate into other program processes.
• Licensing and cost requirements.

**Monetization**

An additional consideration in selecting a presenteeism instrument is whether the findings can be converted to some measure of economic impact. The most common method is to use salary information about the study population or the industry to estimate costs. Given that there are many presenteeism instruments, as well as several methods for monetization, it is not surprising that there is no clear consensus on how best to quantify presenteeism-related productivity costs. As one reviewer noted, “the greatest impediment to estimating the cost of productivity lost to illness is the lack of established and validated methods for monetization.”

It is not necessary to monetize presenteeism losses to evaluate program effectiveness. Relative changes in presenteeism are in themselves meaningful outcomes. Careful consideration of the general and disease-specific presenteeism measures and judicious choice of the most appropriate instrument for the study population and program will help to ensure the validity and reliability of the resulting productivity data.
Chronic Care Management — Five Core Chronics

Overview

Traditionally, chronic care management/disease management has focused on the “big five” chronic diseases: coronary artery disease (CAD), diabetes, chronic obstructive pulmonary disease (COPD), asthma and congestive heart failure (CHF). Chronic care management programs are generally offered telephonically, involving interaction with a trained nursing professional, and require an extended series of interactions, including a strong educational element. Patients are expected to play an active role in managing the disease.

Because of the presence of comorbidities or multiple conditions in most high-risk patients, this approach may become operationally difficult to execute, with patients being cared for by more than one program. Over time, the industry has moved more toward a whole person model in which all the diseases a patient has are managed by a single chronic care management program.

The work in this section applies to single condition traditional disease management programs. Although application to other chronic care management programs and other specific conditions is feasible, the recommendations were designed to apply to the evaluation of programs targeting diabetes, asthma, chronic obstructive pulmonary disease (COPD), coronary artery disease (CAD) and heart failure. The Care Continuum Alliance definition for disease management is as follows:

Disease management is a system of coordinated health care interventions and communications for populations with conditions in which patient self-care efforts are significant.

Disease management:

- Supports the physician or practitioner/patient relationship and plan of care;
- Emphasizes prevention of exacerbations and complications utilizing evidence-based practice guidelines and patient empowerment strategies; and
- Evaluates clinical, humanistic, and economic outcomes on an ongoing basis with the goal of improving overall health.

Disease management components include:

- Population identification processes;
- Evidence-based practice guidelines;
- Collaborative practice models to include physician and support-service providers;
- Patient self-management education (may include primary prevention, behavior modification programs, compliance/surveillance);
• Process and outcomes measurement, evaluation, and management; and
• Routine reporting/feedback loop (may include communication with patient, physician, health plan and ancillary providers, and practice profiling).

Full-service disease management programs must include all six components. Programs consisting of fewer components are disease management support services.

The focus of disease management is on chronic conditions with certain characteristics that make them suitable for clinical intervention:
• Once contracted, the disease remains with the patient for the rest of the patient’s life;
• The disease is often manageable with a combination of pharmaceutical therapy and lifestyle change; and
• The average cost to some chronic patients is sufficiently high to warrant the expenditure of resources by the health plan or employer to manage the condition.

Section guidelines include the following:

1. Evaluation Design
2. Population Identification Methodology
3. Defining the Population
4. Methods to Define Outliers
5. Selection Criteria
6. The Use of Trend
7. The Use of Risk Adjustment
8. Evaluation Considerations for Small Populations
9. Evaluation Considerations for Mature Chronic Care Management Programs
10. Financial Measures
11. Utilization Measures
12. Clinical Measures
13. Self Management
14. Medication Adherence
15. Operational Measures
16. Additional Measures
Methodological Considerations

EVALUATION DESIGN

The goal of an evaluation design is to measure the impact of the intervention and to determine if the effects found were due to the intervention. In general, there are three types of evaluation design that have been applied to chronic care management program evaluation. These include true experimental designs (e.g., randomized controlled trials), quasi-experimental designs (e.g., pre-post with some form of comparison group) and pre-experimental (e.g., pre-post with no comparison group).

The randomized controlled trial (RCT), in which subjects are randomly assigned into concurrent control and intervention groups, is a highly regarded study design for scientific evaluation of outcomes because it allows the evaluation to rule out many competing explanations for changes observed. The RCT design may be difficult to implement routinely in a real-world setting, where it is often not possible to restrict access to the chronic care management program by assigning some individuals to a “usual treatment” group. At the other extreme, a pre-experimental design is the least rigorous category of evaluation designs and generally easy to implement.

The most common of these pre-experimental designs is the pre-post with no comparison group. In this design, a group is measured on metrics of interest at baseline, receives the intervention and is measured on the metrics at the end of the measurement period. While this design provides information on the changes that occur between the baseline and the post-intervention periods, it is difficult to rule out competing explanations for the changes that occur. For example, there might be a substantial decrease in the percentage of participants who are smokers, but it might not be possible to attribute that change to the chronic care management intervention if members also were exposed to other programs designed to reduce smoking.

Quasi-experimental designs, while still subject to potential confounding, are intended to reduce threats to internal validity and thereby increase the confidence with which one can attribute changes to the chronic care management interventions. Campbell and Stanley’s “Experimental and Quasi-Experimental Designs for Research” describes and assesses several quasi-experimental designs.

Experimental models will control for bias and confounders better than the quasi-experimental design noted above. CCA acknowledges the desirability and value of the randomized controlled trial model to reach conclusions about chronic care management value, but also recognizes the impracticality of expecting all business purchasers of chronic care management services to set up this model for evaluation purposes. Therefore, CCA proposes that the goal of program evaluation be practical and not necessarily be held to the rigors and complexity of a true experimental study design.
CCA recommends the use of a pre-post study design with an internal or external comparison group that is equivalent to and assessed over the same time period as the group receiving the intervention.

A comparison group that is both equivalent and concurrent may not always be available in applied settings. Accordingly, CCA recommends that evaluations using a pre-post study design without a comparison group make explicit efforts to control potential biases and error introduced by the design and that the potential impact of the design on the interpretation of the findings be made clear.

There are other study designs population health management uses to evaluate chronic care management program outcomes. With this in mind, a matrix (Table VII) that compares several of the most commonly used study designs is included on page 59. Both chronic care management service providers and purchasers are encouraged to review the information included in the matrix when selecting a study design that differs from the CCA-recommended design. CCA recognizes the challenges of conducting evaluations in real-world settings, but encourages programs to select the most rigorous study design possible within existing constraints and to understand the limitations on interpretation imposed by the design selected.
<table>
<thead>
<tr>
<th><strong>METHOD TO DEVELOP A COMPARISON GROUP</strong></th>
<th><strong>Randomized concurrent control</strong></th>
<th><strong>Matched control</strong></th>
<th><strong>Non-participating individuals comparison</strong></th>
<th><strong>Historical control (as defined by CCA guidelines)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>General description of method</td>
<td>Intervened population compared with individuals randomly selected to have services withheld</td>
<td>Intervened population compared with individuals from non-participating groups matched to have similar characteristics as intervened individuals</td>
<td>Intervened participating individuals compared with intervened non-participating individuals</td>
<td>Intervened population compared with similarly identified population in a baseline period (with costs trended forward)</td>
</tr>
<tr>
<td>Comparison time frame</td>
<td>Concurrent to intervention</td>
<td>Concurrent to intervention</td>
<td>Concurrent to intervention</td>
<td>Prior period</td>
</tr>
<tr>
<td>Population selection bias</td>
<td>None</td>
<td>Somewhat significant</td>
<td>Significant</td>
<td>None</td>
</tr>
<tr>
<td>Source of comparison group</td>
<td>Population for whom program was implemented, randomly selected group withheld from program</td>
<td>Population for whom program was implemented, purchaser decision to not participate</td>
<td>Population for whom program was implemented, individual decision to not participate</td>
<td>Population for whom program was implemented, in prior period</td>
</tr>
<tr>
<td>Trend factor</td>
<td>Not required</td>
<td>Not required</td>
<td>Not required</td>
<td>Individuals without measured conditions in population for whom program was implemented</td>
</tr>
<tr>
<td>&quot;Regression to mean&quot; issues</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>Low</td>
</tr>
<tr>
<td>Credibility of causal statements</td>
<td>Extremely strong (gold standard)</td>
<td>Strong (with proper design)</td>
<td>Poor</td>
<td>Moderate</td>
</tr>
<tr>
<td>Applicability to all types of programs/program designs</td>
<td>Strong</td>
<td>Strong</td>
<td>Strong</td>
<td>Uncertain</td>
</tr>
<tr>
<td>Program sponsor resistance to approach</td>
<td>Strong</td>
<td>Moderate (confidentiality of non-participating groups can be an issue)</td>
<td>None</td>
<td>Low</td>
</tr>
<tr>
<td>Ease of implementation</td>
<td>Very difficult</td>
<td>Difficult</td>
<td>Easy</td>
<td>Somewhat difficult</td>
</tr>
<tr>
<td>Clarity of method to lay audiences</td>
<td>Very clear</td>
<td>Very unclear</td>
<td>Clear</td>
<td>Somewhat unclear</td>
</tr>
<tr>
<td>Multiyear application vs. single-year application</td>
<td>Much harder</td>
<td>About the same</td>
<td>About the same</td>
<td>Somewhat harder</td>
</tr>
<tr>
<td>Method availability</td>
<td>Rarely possible</td>
<td>Occasionally possible</td>
<td>Frequently possible</td>
<td>Usually possible</td>
</tr>
<tr>
<td>Bleed of interventions to comparison population</td>
<td>Control group likely to get provider-based interventions</td>
<td>Control group likely to get provider-based interventions</td>
<td>Control group likely to get provider-based interventions and softer member-based interventions</td>
<td>Control group unlikely to get any interventions</td>
</tr>
<tr>
<td>Key strengths</td>
<td>Generally seen as gold standard evaluation method</td>
<td>Potential for a concurrent comparison population with many of the same characteristics as the intervened population</td>
<td>Easy to implement</td>
<td>Relatively universally available method</td>
</tr>
<tr>
<td></td>
<td>Ease/strength of interpretation</td>
<td>Credibility of causal statements can be very strong</td>
<td>Method is easy to understand</td>
<td>Application of well-accepted actuarial processes</td>
</tr>
<tr>
<td>Key problems/biases</td>
<td>Sponsor resistance to implementation</td>
<td>Difficult to get access to non-participating groups of sufficient size in same geographic area, with similar benefit structure</td>
<td>Very significant bias associated with differences in motivation between participants and non-participants</td>
<td>Difficult to ensure equivalence between baseline and intervention year (particularly in populations with many shifts in size, composition, and/or benefit structure)</td>
</tr>
<tr>
<td></td>
<td>Difficulty of multiyear assessment</td>
<td>“Black box” approach difficult to understand</td>
<td>Likely not possible in “opt-out” program models</td>
<td>Difficulty in deriving credible trend factor</td>
</tr>
</tbody>
</table>
POPULATION IDENTIFICATION METHODOLOGY

Developing a claims-based methodology that identifies appropriate patients for inclusion in a chronic care management program evaluation can be challenging, largely because of the inherent complexity and diversity of claims data. The first step is to decide which codes will be used to select the conditions of interest (see selection criteria, page 65). Claims-identification codes might be International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnostic codes; Current Procedural Terminology (CPT®) codes; Healthcare Common Procedure Coding System (HCPCS) Level II codes; or National Drug Code (NDC) codes.

Once the appropriate codes are determined, the selection algorithm must be defined. One critical issue is that of false positives (i.e., patients identified by the search algorithm who do not have the condition of interest). These may occur because current diagnostic coding convention makes no distinction between diagnoses and “rule-out” diagnoses. A person with shortness of breath may have claims for office visits and diagnostic tests with diagnoses of angina pectoris that could be used to identify the person as a potential coronary heart disease program participant when, in fact, the person has an esophageal problem. Chronic care management programs might differ on the relative importance of false positives (being selective) versus false negatives (being inclusive), such that different selection algorithms might generate different levels of false positives. The effect on return on investment (ROI) calculations of using different algorithms is that different definitions of populations using different algorithms will define different groups, different prevalence of the illness in question, different average per-member-per-month (PMPM) costs, and, ultimately, different estimated levels of impact for a given population.

The period over which the algorithm applies also is important. Identification of individuals within a given year selects only those who have been ill enough to generate a claim. Individuals with the disease who did not have a claim (and likely are less ill) are not counted. Program evaluations often seek to use a “look-back” year to identify these individuals for a period, counting individuals as having a disease if they have triggered the algorithm in either period. This issue can impact ROI calculations if different methods for identification are used in the baseline and the program. A number of strategies have been suggested to work through this problem.

The guidelines initially offered two methods of applying identification criteria for defining the population for a given measurement year.

Method I – Annual Qualification

- Each measurement period population (e.g., pre-program baseline or any post-implementation year) is defined uniquely, based on application of identification criteria specific to the measurement period.
- Identical identification criteria are used to define the population, and also are applied to each measurement period, baseline or post-implementation, in the same manner.
- As a result, no members automatically qualify for inclusion in later periods, or are automatically carried forward into later periods, simply because they were identified in an earlier period.
Method II – Prospective ID (once chronic, always chronic)

- In contrast to the “annual qualification” process, those identified for the baseline, or initial year evaluation periods, are automatically carried forward into the post-implementation measurement period populations, as long as they remain eligible with the chronic care management program purchaser. That is, they are assumed to still have the previously identified condition irrespective of claims evidence for that condition in the post-implementation period.

- The same identification criteria are used to define the population (e.g., the logic used to define someone with a specific condition) for each year. However, those criteria are applied differently between periods, as people are prospectively “carried forward” into post-baseline years, while no members are typically carried forward into the baseline year population.

- Thus, the post-implementation measurement period population includes all those meeting the identification criteria applied for the current period, as well as those who met criteria applied to define the prior periods.

While both methods were considered acceptable for identifying groups for the purpose of program evaluation, CCA recommends the adoption of an annual qualification process for defining a population, due to its closer correlation to the principle of equivalence between measurement period populations.

The key points for this process are:

- Apply the same criteria for defining the presence of a condition across all measurement periods. For example, use the same algorithms to define a diabetic in all measurement periods. Do not change from one year to the next.

- Apply the criteria in the same manner for each measurement period. In whatever way claims data are used to define one year's population, ensure the criteria are applied in the same manner for each year. If the populations are not defined in an identical manner, it is less likely they will be as equivalent, and more likely that biases will be introduced into the results.

- Use at least 24 months of claims to define the population for each year, and apply identical criteria in terms of number of months of claims used to define the population in each year. For example, do not use 24 months of claims to define one population and 36 to define a comparative population.

DEFINING THE POPULATION

Defining the population is important to an effective and accurate measurement of the effects of a chronic care management program. Guidelines included in this report to help define the population include the length of time for measurement periods, baseline, run-out and look-back periods, as well as the definition of a member month, exclusion criteria and selection criteria.
For the purpose of chronic care management program evaluation,

CCA recommends that one year’s worth of information be included in the baseline year, as well as in the look-back period of the analysis and all subsequent years used for the measurement.

In addition, when claims are involved in the evaluation,

CCA recommends that paid-through run out for each measurement period be three months with a completion factor and six months without.

This run-out period would increase the probability that all claims incurred in the measurement period were included in the claims available for analysis.

CCA recommends that the measurement period be at least six months for the purchaser of the commercial program and at least one month for Medicaid TANF participants.

Many metrics in evaluations use member-months as the denominator. Although not all chronic care management programs have information on enrollment dates,

CCA recommends that, when this information is available, members be counted only in those months in which they were enrolled on or before the 15th of the month.

This guideline applies to both commercial and Medicare populations, but not to Medicaid programs. CCA concluded that there should be three types of exclusion criteria that allow a member’s experience to be excluded from the evaluation. These criteria need to be made clear to all stakeholders before the evaluation.

**Criteria I** - excludes all data from the evaluation for patients who have comorbidities that would make it difficult for the patient to gain benefit from the chronic care management program. Examples of these conditions include:

- ESRD
- HIV/AIDS
- transplants
- non-skin cancers with evidence in claims of active treatment
- hemophilia

**Note that patients with these conditions may or may not participate in the program, but would not be included in evaluations of the program.**

**Criteria II** - excludes claims from the evaluation for events and diagnoses that are potentially costly but clearly unrelated to the chronic care management program—for example, trauma with hospitalization or skin cancers.

Note that this recommendation excludes specific claims, but not the individual from the evaluation.

**Criteria III** - excludes outlier costs from the evaluation by using a stop-loss approach at the member level, such as removing claims greater than $100,000 annually, indexed to grow at future years concurrent with an appropriate Trend.
Methods to Define Outliers

Patients can incur extraordinarily high costs for numerous reasons. These costs often are for events randomly distributed over a population and unrelated to a chronic care management program—accidental trauma, for example. High costs create substantial volatility to claims cost trends and can distort financial savings calculations, particularly for smaller populations.

A stop-loss approach excludes, for the purpose of measurement, claims costs for individual members in excess of the stop-loss threshold during the year. This approach is preferable to excluding the participant’s entire experience because it enables the inclusion of a greater proportion of the managed population in measurement. Moreover, it does not create a distortion if the program is involved in shifting a member above or below the stop-loss threshold. There are several methods commonly used across population health management to identify and mitigate these outlier costs. These methods are outlined in Table VIII, Methods to Define Outliers, page 64. CCA recommends a review of the information in the table prior to selection of a method.
<table>
<thead>
<tr>
<th>METHOD</th>
<th>Stop-Loss Method</th>
<th>Percentile Distribution Method</th>
<th>Standard Deviation (SD) Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>General description of method</td>
<td>A threshold value is determined (e.g., $100K) and costs above threshold are excluded and are indexed to grow in future years concurrent with an appropriate trend.</td>
<td>A threshold value is determined based on the X percentile of claims costs (e.g., 99.5%); costs above this threshold are excluded.</td>
<td>A threshold value is determined based on X (e.g., 3SD) standard deviations from population mean. Costs above this threshold are excluded.</td>
</tr>
<tr>
<td>Trend factor</td>
<td>Should be used to account for medical cost trends year to year.</td>
<td>Not required</td>
<td>Not required</td>
</tr>
<tr>
<td>Applicability to all types of programs/program designs</td>
<td>May require lower threshold to offset variability in small populations.</td>
<td>N/A</td>
<td>May not be appropriate for populations with non-normal distributions.</td>
</tr>
<tr>
<td>Ease of implementation</td>
<td>Very Easy</td>
<td>Easy</td>
<td>More difficult</td>
</tr>
<tr>
<td>Clarity of method to lay audiences</td>
<td>Very Clear</td>
<td>Clear</td>
<td>Requires knowledge of simple statistics.</td>
</tr>
<tr>
<td>Multiyear application vs. single-year application</td>
<td>Trend adjustment should be considered for multiyear assessment.</td>
<td>Easy to apply.</td>
<td>Should recalculate SD for each year.</td>
</tr>
<tr>
<td>Key strengths</td>
<td>Better for evaluations where two groups are expected to have different frequency cost distributions rather than unit cost variance. Ease of application and most sensitive to varying rates of catastrophic claims.</td>
<td>Better for evaluations where two groups are expected to have unit cost variance rather than different frequency cost distributions. Can be used without adjustment on small populations.</td>
<td>Better for evaluations where two groups are expected to have unit cost variance rather than different frequency cost distributions. Can be used without adjustment on small populations.</td>
</tr>
<tr>
<td>Key limitations</td>
<td>Stop-loss threshold is arbitrary and not data-sensitive. Threshold selection for smaller populations is by nature more subjective than other methods listed. May not appropriately filter random drivers of variance in small groups without lowering the stop-loss amount. Lack of threshold trend adjustment will deflate overall trended results proportional to the presence of catastrophic cases.</td>
<td>Threshold is also arbitrary but data-sensitive to varying percentages of shock-loss claims. Handles variant cost distributions with variant thresholds, which may not be desirable with variant rates of catastrophic claims.</td>
<td>Threshold is also arbitrary but data-sensitive, and is only somewhat sensitive to a varied percentage of claims categorized as shock-loss. The calculations are the most involved, but involve standard simple statistics. Using standard concepts of significance for selecting a standard deviation threshold may not apply to the proper selection of shock-loss claims. Handles variant cost distributions with variant thresholds, which may not be desirable with variant rates of catastrophic claims.</td>
</tr>
</tbody>
</table>
**SELECTION CRITERIA**

**Background and Prior Work**

The work to develop standardized selection criteria began in 2008 with the recognition that achieving consensus on how to select populations for the evaluation of chronic care management programs would contribute to improving standardized program evaluation, as well as help facilitate rigorous performance comparisons and increase transparency for purchasers.

The term “selection criteria” refers to standardized characteristics (observed in data sets) used to identify people for inclusion in the measurement pools (denominators) of outcomes metrics. The project’s goal was to develop a fundamental approach to guide denominator specifications for comparison of chronic care management programs for the five chronic conditions. Specification of selection criteria requires: identifying data sources to be used; specifying the algorithm to be used to query the data; and selecting the diagnostic, procedural and other codes for use in the algorithm.

It’s important to emphasize that selection criteria represent an identification algorithm’s intent to accurately identify people who have a disease while not (falsely) identifying those who do not have the disease. It is well-recognized that administrative data cannot completely succeed at both tasks—that algorithms that identify all (or nearly all) people who have a disease will include some false-positive identifications, and vice versa. In previous volumes of the Outcomes Guidelines, we discussed the nature of this “sensitivity-specificity balance.” An additional issue with using selection criteria—discussed in Volume 4—is that standardized selection criteria will produce denominators that may not completely overlap with those produced by programs. This is important to understand for those who use standardized program evaluation reports to compare programs.

Volumes 2 and 3 of the CCA Outcome Guidelines Report developed a philosophical framework to allow construction and evaluation of identification algorithms for the five core chronic conditions: diabetes, asthma, chronic obstructive pulmonary disease, coronary artery disease and congestive heart failure. The resulting selection criteria are the standardized characteristics used to identify people for inclusion in the measurement pools (denominators) of outcomes metrics.

In 2009, the work focused on testing the denominators specified in Volume 3 using varying identification time frames and minimum eligibility time frames for their suitability for program comparison. Organizations used their own data to test whether CCAs denominator specifications produced: prevalence rates that were consistent with their own experience; reasonable specificity without unduly sacrificing sensitivity when tested over time; and acceptable overlap between individuals identified using CCA selection criteria and those using proprietary criteria.

**2010 Scope of Work**

Building on the work in 2009’s Outcomes Guidelines Report Volume 4, the Selection Criteria Workgroup tested selection criteria (denominators) for the five common chronic conditions for their **suitability** for comparing the performance of chronic care management programs on clinical, utilization and financial outcomes measures.

Consistent with our prior work, we defined “**suitability**” to mean that measures derived from the selection criteria are fair and representative of program experience and are acceptable to most programs for the purpose of outcomes comparisons.

While these criteria may also be appropriately used for internal program evaluation and improvement, it is not our expectation that they will be; however, we recognize that programs will not be willing to be compared unless the CCA selection criteria identify a relevant set of individuals and is capable of showing the impact of their programs.

*These denominators are appropriate for measures in any population.*
To assess suitability, the workgroup tested the five chronic condition denominators’ performance with a basket of clinical, utilization and financial measures. Good performance was defined as yielding stable results over time and across vendors (testers). Last year—as reported in Volume 4—the first criterion was tested (overlap); this year the focus was on the second and third. The fourth and fifth criteria were not explicitly tested (measure overlap and overlap stability over time), but the expectation is that programs will do so using the published criteria:

1. (Reported in Outcomes Guidelines Report, Volume 4) The identification overlap between the tester (vendor) and CCA selection criteria is adequate (i.e., most of the people identified by the tester are also found by CCA, and CCA does not require tester to report on too many people who don’t fit their identification criteria).

2. In a given measurement year, the measurements using the CCA criteria correlate sufficiently well across testers (i.e., the measure gives stable results across testers).

3. The results from using CCA selection criteria over multiple, consecutive measurement years is plausibly stable (i.e., doesn’t vary more than expected—the measure gives stable results across time).

4. (Not explicitly tested) In a given measurement year, the measurements using the CCA and tester’s own selection criteria correlate sufficiently well.

5. (Not explicitly tested) In a series of measurement years, the results from the CCA selection criteria are, on average, directionally consistent with that found by using the testers’ (vendors’) own criteria.

We devised clinical, utilization, and financial metrics to address the second and third questions. It is important to bear in mind that the purpose of constructing these metrics was not to define actual detailed measures but to test the denominator suitability over a variety of conditions and outcomes types.

**General Measure Criteria**

Each measure embodies the following elements:

- **Identification frame** (incurred begin/end dates for claims): 24 months with a 3 months’ paid-through run out past the end of the identification frame. This is the timeframe based on date of service, used to determine whether a member qualifies as “having” the denominator’s disease.

- **Measurement frame** (incurred begin/end dates for claims): last 12 months of the 24 month identification frame

- **Minimum eligibility** (during the measurement frame)
  - Clinical measures: “HEDIS continuous” - eligible for the entire measurement frame with a single allowable gap of up to 45 days
  - Utilization and financial measures: Any six months, not necessarily contiguous. While testing was done with this criterion (to conform to that specified in previous editions of the Guidelines), it is recognized that some organizations may wish to use the same denominators for all outcomes measures. It is strongly recommended that reporting organizations specify which eligibility criterion was used.
**Clinical Test Measures**

The test measures were simplified from formal comparison measures in that clinical exclusions were not allowed, and simplified, calendar-based identification timeframes were used. The denominator definitions may be found in the Appendix, page 121. The following measures were tested (numerator evidence as per 2010 HEDIS Technical Specifications):

- Diabetes: A1c test during the measurement year.
- Diabetes: Nephropathy screen or at least one fill for ACE inhibitor or ARB during the measurement year.
- Asthma (ages 5-17): At least one fill for an asthma controller medication during the measurement year.
- Asthma (ages 18-56): At least one fill for an asthma controller medication during the measurement year.
- COPD: At least one fill for a (short- or long-acting) beta-agonist.
- CHF: At least one fill for an ACE inhibitor or ARB during the measurement year.
- CHF: At least one fill for a beta-blocker during the measurement year.
- CAD: LDL-cholesterol testing during the measurement year.

**Utilization Measures**

The following general measure specification criteria were used:

- Outcome: All-cause (except maternity and perinatal) emergency department (ED) visits and hospitalizations per 1,000 chronic (by disease) members per year, reported separately for members with each of CAD, CHF, diabetes, asthma, and COPD. Also report for all diseases combined (eliminating double-counting).
- Denominators: One for each population with CAD, CHF, diabetes, asthma (separate for population age 5-17 and 18-56), and COPD who were eligible at least six months in the measurement year.
- Numerators: Number of all-cause (except maternity and prenatal) ED visits or hospitalizations for members in each denominator.
- A hospitalization was counted as occurring on its date of admission (not its date of discharge).
- Hospital transfers were not counted as separate hospitalizations.

*Note: For clinical test measures and utilization measures, asthma was tested through age 56, to align with the original specifications that were consistent with HEDIS at that time. HEDIS now goes through age 50 for asthma.*
Financial Measures

The following general measure specification criteria were used:

- Outcomes: Year 1 $PMPM (paid) and Year 1 – Year 2 trend for the chronic populations with CAD, CHF, diabetes, asthma (separate for population age 5-17 and 18-56), and COPD.
- Minimum eligibility: at least six months in the measurement year.

Results and Commentary

Several CCA members participated in the testing process. The following results are reported for the four populations on whom all tests were performed. These populations are large, geographically diverse, commercial populations with chronic care management programs.

It was concluded that the criterion of suitability is met because the test results represent both plausible results and a reasonable range considering that the four populations measured might have different demographics, coinsurance and deductible levels, physician practice patterns, population risk and years of disease and population health management.

Table IX displays the average condition-specific prevalence rates across the four populations, as well as the average prevalence rates reported in Volume 4. For the 2010 data, both the average prevalence and the +/- 95th percentiles for Year 1 and Year 2 populations are presented. Measured prevalence rates for the five chronic conditions (with asthma split into two age groups) for Year 1 and Year 2 data met expectations: They were consistent with rates expected by the testers and the workgroup members, based on their experience; they were consistent across testers; and, with the exception of CAD, in line with prevalence rates found in the 2009 tests. Note that in the 2009 tests, it was not possible to present separate averages for juvenile and adult asthma. Finally, prevalence rates were found to be consistent between Years 1 and 2.

<table>
<thead>
<tr>
<th>TABLE IX – PREVALENCE RATES</th>
</tr>
</thead>
<tbody>
<tr>
<td>PERCENT OF POPULATION WITH</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>DIABETES</td>
</tr>
<tr>
<td>Rate</td>
</tr>
<tr>
<td>V4 CCA Outcome Report (DOR)</td>
</tr>
<tr>
<td>Rate</td>
</tr>
<tr>
<td>95% CI (+/-)</td>
</tr>
<tr>
<td>V5 Testers: Year 1</td>
</tr>
<tr>
<td>Rate</td>
</tr>
<tr>
<td>95% CI (+/-)</td>
</tr>
<tr>
<td>V5 Testers: Year 2</td>
</tr>
</tbody>
</table>
Table X presents condition-specific average rates and ± 95th percentile for Year 1 clinical outcomes across the four populations. Measured clinical outcome rates for Year 1 data for the chronic conditions likewise met expectations: They were consistent across testers.

<table>
<thead>
<tr>
<th>TABLE X – CLINICAL MEASURES</th>
<th>PERCENTAGE WITH SPECIFIED TEST OR TREATMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>DIABETES</strong></td>
</tr>
<tr>
<td>Annual A1c Test</td>
<td>Rate</td>
</tr>
<tr>
<td>Nephropathy screen</td>
<td>95% CI (+/-)</td>
</tr>
<tr>
<td>&gt;= 1 fill for controller</td>
<td></td>
</tr>
<tr>
<td>LDL Screen</td>
<td></td>
</tr>
<tr>
<td>&gt;= 1 fill for beta-blocker</td>
<td></td>
</tr>
</tbody>
</table>

Table XI presents condition-specific rates and the ± 95 percentiles for Year 1 utilization measures. Measured utilization (ED and hospitalization) rates for Year 1 data for the chronic conditions were plausible according to testers’ expectations based on their experience, though showed fairly wide inter-tester variation (not unexpected given the composition differences among populations and the lack of risk adjustment in our results).

| TABLE XI – UTILIZATION MEASURES | **DIABETES** | **ASTHMA (5-17)** | **ASTHMA (18-56)** | **COPD** | **CAD** | **CHF** |
|-------------------------------|--------------|------------------|--------------------|==========|--------|--------|
| Admits/1,000                  | 183          | 38               | 92                 | 400       | 530    | 860    |
| 95% CI (+/-)                  | 20           | 5                | 24                 | 69        | 165    | 158    |
| ED Visits/1,000               | 272          | 376              | 396                | 438       | 444    | 558    |
| 95% CI (+/-)                  | 65           | 80               | 115                | 148       | 161    | 131    |
Finally, Table XII presents the average, condition-specific, incurred per diseased member per month (PDMPM) paid amounts and ± 95 percentiles for the four populations. Also included is the ± 95 percentile for the Year 1/Year 2 trend. Measured financial (Year 1 SPDMPM and trend) rates for the chronic conditions were plausible according to testers’ expectations based on their experience, though as with utilization showed fairly wide inter-tester variation.

<table>
<thead>
<tr>
<th>TABLE XII – FINANCIAL MEASURES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PMPM (for members</strong>&lt;br&gt;<strong>with identified conditions)</strong></td>
</tr>
<tr>
<td>$907</td>
</tr>
<tr>
<td>$158</td>
</tr>
<tr>
<td>4%</td>
</tr>
</tbody>
</table>

It was concluded that the results of all tests are consistent with meeting the criterion of suitability. Based on this conclusion, **CCA recommends that the five condition selection criteria be used as denominators for clinical, utilization and financial measures specifically for program comparison.**

**Further Commentary and Next Steps**

CCA invites member and public commentary on the test results, in light of the fact that organizations may now choose to utilize its selection criteria not only for comparison with other programs, but also, internally, for program evaluation and improvement.

Updates for new ICD-9, Revenue, CPT4, NDC and LOINC codes may be made annually by CCA. Where possible, these codes will be consistent with those published by the National Committee for Quality Assurance (NCQA) for use in performance measures; where CCA does not specify a denominator, CCA will annually publish updates as appropriate.

CCA recommends that HEDIS-continuous eligibility be applied to clinical measures and that utilization and financial measures require at least six months’ (not necessarily contiguous) eligibility in the measurement year. However, organizations may choose to use HEDIS-continuous eligibility for all measures. Reporting organizations should specify the eligibility criteria used.

**2011 Provisional Scope of Work**

- Explore replacing the disease CAD with Ischemic Vascular Disease (IVD) as defined by the NCQA in its performance measures. IVD comprises CAD, peripheral arterial disease and cerebrovascular disease. Rationale: These atherosclerotic or atherothrombotic conditions are highly clinically-related and share many important clinical practice guidelines. Clinical outcomes measures are generally identical for the IVD conditions. As with the NOQA, programs that treat one or two IVD conditions may report them as “IVD” and specify which conditions are included.

- Consider testing a six-month vs. HEDIS-continuous eligibility criterion for utilization and financial measures.

- Consider further testing of the **relevance criterion** for financial measures to provide guidance on the feasibility of program comparison.
The major thrust of the Selection Criteria Workgroup—to define outcomes denominators suitable for program comparison—is complete. Therefore, in 2011, this workgroup may be combined with the Population Health Management Workgroup to provide a deeper synergy among initiatives that assess the impact of identification, engagement and outcomes measurement methodologies.

THE USE OF TREND

The most difficult part about evaluating the financial performance of chronic care management programs is that the comparison of costs are made with what the costs “might have been” without the program in place. Imagine being asked to evaluate the performance of a country’s president by comparing it against what “might have been” under a different president during those same years. Each citizen would have a different response based on constructing different scenarios of how the macro-economy would have evolved independent or consequent to that president’s actions and what global events might have occurred independently or consequent, etc., and no individual’s response could be tossed aside as wrong. Similarly, to calculate a trend that would represent what “might have been” without a chronic care management program, efforts might be undertaken to explore the depths of several macro and micro health economic factors. Or, the program evaluators might acknowledge the difficulty of the very proposition and accept something simple up front. It is standard practice that, in the absence of an equivalent control group, program evaluations relying on pre-post comparisons should be adjusted for the trend that would be expected to occur in the absence of the program interventions.

CCA recommends the use of a non-chronic population to calculate this trend. For this purpose, the non-chronic population is defined as those members not identified as having any of the following “common chronic” conditions: diabetes, CAD, heart failure, asthma and COPD.

It has been empirically demonstrated that chronic trend (for any of the five common chronic conditions) can differ significantly from non-chronic trend. Accordingly, it is desirable to have a method for adjusting the non-chronic trend to represent a more accurate surrogate for what chronic trend would have been in the absence of the chronic care management program.

CCA recommends the use of the average difference between historical chronic and non-chronic trends to adjust current year non-chronic trend.

- To adjust the current intervention year’s non-chronic trend to estimate the expected chronic trend in the absence of intervention, use the historical average difference in chronic and non-chronic trend. Use two to three years of data (pre-intervention) to compare chronic trend with non-chronic trend for the same population as being intervened currently (this assumes the differences in these rates are relatively stable from year to year and that trend is calculated consistently for both groups in the historical years and current program year).

- While this can be done for any individual chronic care management program with accessible historical claims, it may be desirable to develop a national reference database of chronic and non-chronic trend as part of a future CCA research project.
This would allow chronic care management programs to utilize an empirically derived national or regional trend adjustment factor, permitting a more standardized calculation of financial outcomes and facilitating comparison of financial outcomes for different programs, which is difficult now due to the use of differing trends and trend adjustments for the evaluation of individual chronic care management programs.

**Step-by-Step Approach**

**Step 1** – calculate risk-adjusted non-chronic and chronic trends in historical time periods using identification methodologies identical to those in the program period; measure the relationship between these two trends.

**Step 2** – calculate the risk-adjusted non-chronic trend for the program period, and then modify it to represent a “what might have been” chronic trend based on the relationship measured in Step 1. Thereafter, as in the preceding method, proceed to calculate the savings by first ensuring that any risk profile changes in the chronic population are accounted for.

<table>
<thead>
<tr>
<th>Client #1, Annual Re-qualification</th>
<th>1 Year Trend</th>
<th>2 Year Trend</th>
<th>3 Year Trend</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>MM</td>
<td>PMPM Medical</td>
<td>PMPM RX</td>
</tr>
<tr>
<td>Non-Chronic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td>2,345,429</td>
<td>$ 114.65</td>
<td>$ 38.14</td>
</tr>
<tr>
<td>2005</td>
<td>2,933,300</td>
<td>$ 123.29</td>
<td>$ 40.61</td>
</tr>
<tr>
<td>2006</td>
<td>3,950,584</td>
<td>$ 135.39</td>
<td>$ 43.10</td>
</tr>
<tr>
<td>2007</td>
<td>5,536,021</td>
<td>$ 153.05</td>
<td>$ 44.35</td>
</tr>
<tr>
<td>Chronic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td>181,316</td>
<td>$ 454.61</td>
<td>$ 178.61</td>
</tr>
<tr>
<td>2005</td>
<td>246,631</td>
<td>$ 498.19</td>
<td>$ 188.72</td>
</tr>
<tr>
<td>2006</td>
<td>329,180</td>
<td>$ 566.01</td>
<td>$ 207.41</td>
</tr>
<tr>
<td>2007</td>
<td>469,303</td>
<td>$ 664.90</td>
<td>$ 217.06</td>
</tr>
<tr>
<td>All</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td>2,526,745</td>
<td>$ 139.04</td>
<td>$ 48.22</td>
</tr>
<tr>
<td>2005</td>
<td>3,179,931</td>
<td>$ 152.36</td>
<td>$ 52.10</td>
</tr>
<tr>
<td>2006</td>
<td>4,279,764</td>
<td>$ 168.51</td>
<td>$ 55.74</td>
</tr>
<tr>
<td>2007</td>
<td>6,005,324</td>
<td>$ 193.05</td>
<td>$ 57.84</td>
</tr>
</tbody>
</table>
In many cases, the relationship between historical chronic trend and non-chronic trend may be quite stable, allowing the use of the difference between these trends for adjustment of non-chronic trend.

Empirical testing was conducted to determine if the historical relationship between chronic and non-chronic trend is stable. A large claims database representing three years worth of data on a population that did not have a rigorous chronic care management program in place was used for testing. Testing was conducted on the five common chronic diseases to which the trend recommendation was designed to apply, and the CCA-recommended annual qualification method was used to identify the population for testing.

Table XIII (page 72) summarizes the results for the non-chronic, chronic and combined samples. The trends to the right of the table for Year 1 through Year 3 show there is stability between the two samples for the three-year historical trend.

**THE USE OF RISK ADJUSTMENT**

Risk adjustment, as applied to evaluation of chronic care management financial outcomes, consists of a series of techniques that account for the individual characteristics of patients within a defined population when retrospectively evaluating the impact of a chronic care management intervention on the financial outcomes for that population.

Performing chronic care management program evaluation of financial outcomes when an equivalent comparison group is available eliminates the need for risk adjustment of the outcome—ideally, the comparison group differs from the intervention group only by the impact of the chronic care management program, as all other relevant factors that contribute to the financial outcome are equivalent.

General considerations for using risk adjustment for chronic care management program pre-post evaluations:

- In measuring the changes in an outcome over time resulting from an intervention, it is often the case that the outcome may be influenced by both the intervention, as well as other factors external to the intervention (e.g., demographic or case mix shifts over time that would alter the population characteristics to a material degree).

- Risk adjustment serves to adjust for changes in an outcome of interest that result from those factors that are “exogenous” or external to the intervention being evaluated.

- The goal of using risk adjustment in the evaluation of chronic care management programs is to adjust for these exogenous confounders to the greatest extent possible, while not altering or distorting chronic care management program impacts.

- Risk adjustment methods should be transparent, simple, reliable, affordable and suitable for the data available.

- Risk adjustment methods also should be validated; this is most likely the case when the method is simple (age, gender) or a commercial tool or published non-proprietary method is used. Examples of such tools include, but are not limited to, ACGs, DCGs, CRGs, ERGs or the CDPS grouper system.

Cautions in applying risk adjustment to chronic care management program evaluations:

- Many approaches to adjusting for exogenous variables have some risk of inadvertently adjusting for variables that are positively impacted by chronic care management programs; this can result in combined adjustment for
confounding factors, as well as target factors at the same time, potentially discounting the desired chronic care management impacts while attempting to adjust for the exogenous variables beyond the influence of the program.

- All risk adjustment tools are imperfect and the goal of risk adjustment can never be achieved completely; even academic research studies often are stymied by how, exactly, to adjust for risk.

- Risk adjustment methods are not “general purpose”; they must be individualized to the outcomes of interest, the populations involved and the data available.

- Performing risk adjustment is neither simple nor formulaic; no single approach to risk adjusting chronic care management outcomes can be universally applied to all program evaluations to achieve the goal of risk adjustment without possible unintended consequences.

**Recommendation**

In deciding (1) whether and (2) how to approach risk adjustment for a particular chronic care management program for a specific population, it is useful to categorize outcomes of interest into one of the following two categories:

- **Category 1**: Those believed to be impacted only by exogenous confounders and not the chronic care management interventions, where there is no concern that program impacts will be altered by risk adjustment (e.g., non-chronic trend).
  - For this category of variables, one should utilize an appropriate risk adjustment method, ideally a commercially available risk adjustment tool or other non-proprietary validated method.

- **Category 2**: Those believed to be impacted by exogenous confounders, as well as by program interventions that potentially may be inappropriately distorted or discounted by risk adjustment (e.g., condition prevalence or severity, case mix).
  - For this category of variables, the next step is to examine the potential magnitude and importance of the potential exogenous confounder(s). If the potential magnitude is large and/or highly important, then one must consider which available risk adjustment methods permit a reasonable job of adjusting for the offending confounders without seriously distorting or discounting program impacts.

If more than one method is available, then the one with the least likelihood of distorting program impacts while reasonably adjusting for confounding factors is preferred. In some cases, using a “minimalist” approach, such as age-gender or simple prevalence adjustment, may be more suitable than more complex risk adjustment tools. That is because the more comprehensive or explanatory the risk adjustment method, the greater the likelihood that some of the input variables for that method are factors that chronic care management programs positively impact.

The application of risk adjustment to chronic care management program evaluation is a complex issue—even when desirable, it can be quite difficult. The decision to utilize risk adjustment and the choice of which method to use necessarily involves thoughtful trade-offs of the associated benefits and risks of whether to do, and exactly how to do, risk adjustment of financial outcomes. Table XIV offers examples of how risk adjustment could be used for various situations.
To provide more detail and to highlight some advantages and disadvantages of specific techniques, CCA offers two case study examples of risk adjustment below. In both examples, the methods described were applied using actual health claims experiences. CCA hopes these case studies can help Outcomes Guidelines users understand the importance and complexity of risk adjustment. CCA is not, at this point, advocating either technique as a single appropriate methodology.

**Case Study Example I**

**Use of Risk Adjustment Approaches in Chronic Care Management Evaluation**

**Introduction:** A widely used method of performing chronic care management program savings evaluations is the CCA recommended adjusted historical control method. This method uses the concurrent non-chronic population’s trend as the benchmark trend to approximate the chronic population trend in the absence of the chronic care management program. One enhancement to this method has been to nullify any impact on the non-chronic benchmark trend of changes in risk profile of the non-chronic population. Another enhancement has been to adjust the non-chronic benchmark trend by the pre-disease management relativity between chronic and non-chronic trends. Both of these enhancements, while well-intentioned, can still result in chronic care management program savings evaluations that are faulty, due to non-recognition of changes in risk profile of the chronic population.

The case study being presented here recognizes the importance of changes in risk profile of the chronic population.
**Methodology:** The overall methodology can be summarized in a simple manner as follows:

- Stratify the chronic population in the base period and program evaluation period into high-level homogenous risk strata.
- Calculate the PMPM costs for each risk stratum in the chronic population in the baseline period.
- Calculate the member months distribution across the risk strata of the chronic population in the program evaluation period.
- Calculate the weighted average baseline period chronic PMPM cost using the baseline period’s risk strata specific PMPM costs and the program evaluation period’s member months distribution across the risk strata.
- Then, proceed as in the standard adjusted historical control method (that is, trend the above weighted average baseline period chronic PMPM cost by the benchmark trend to the program evaluation period and subtract the actual PMPM cost experienced by the chronic population in the program evaluation period to calculate gross savings).

**Benefits:** The above methodology adjusts the starting point (baseline chronic cost) of the savings calculation by making it mimic the program evaluation period’s chronic population risk profile. This represents a significant improvement in the savings evaluation methodology, because it ensures that any changes in the chronic population’s risk profile from the baseline period to the program evaluation period does not remain embedded within the savings calculation. If the risk profile improved, the savings would be lowered from an otherwise unduly optimistic conclusion; if the risk profile worsened, the savings would be increased from an otherwise unduly unfavorable conclusion. In other words, savings conclusions would not be affected by such a factor (chronic population’s risk profile change) that is other than the chronic care management program’s interventions’ effectiveness.

**Case Study Example:** Below is a table summarizing the chronic population’s PMPM costs and member months distribution across three relatively homogenous strata: the segment of chronics who persist or continue to remain in the study population, those who newly enter the population and those who terminate.

<table>
<thead>
<tr>
<th>GROUP</th>
<th>AVERAGE COST BASELINE PMPM</th>
<th>NUMBER OF MEMBER MONTHS BASELINE</th>
<th>PERCENT OF MEMBER MONTHS</th>
<th>AVERAGE COST YEAR 1 PMPM</th>
<th>NUMBER OF MEMBER MONTHS YEAR 1</th>
<th>PERCENT OF MEMBER MONTHS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terminating</td>
<td>$929.75</td>
<td>31,407</td>
<td>11%</td>
<td>$721.73</td>
<td>53,938</td>
<td>18%</td>
</tr>
<tr>
<td>Continuing</td>
<td>$706.53</td>
<td>186,918</td>
<td>65%</td>
<td>$623.80</td>
<td>180,522</td>
<td>60%</td>
</tr>
<tr>
<td>Newly Identified</td>
<td>$601.71</td>
<td>70,571</td>
<td>24%</td>
<td>$528.11</td>
<td>64,582</td>
<td>22%</td>
</tr>
<tr>
<td>Total</td>
<td>$705.19</td>
<td>288,896</td>
<td>100%</td>
<td>$620.80</td>
<td>299,042</td>
<td>100%</td>
</tr>
</tbody>
</table>
If the change in risk profile of the chronic population were not recognized, then the analysis would focus only on the bottom row and compare the actual $620.80 to the baseline $705.19 trended at a benchmark trend. This is shown below.

**Basic Savings Calculation (unadjusted):**

<table>
<thead>
<tr>
<th>Action</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Chronic PMPM</td>
<td>$705.19</td>
</tr>
<tr>
<td>x Benchmark Trend</td>
<td>$1.05</td>
</tr>
<tr>
<td>= Expected PMPM</td>
<td>$740.45</td>
</tr>
<tr>
<td>- Actual Intervention PMPM</td>
<td>$620.80</td>
</tr>
<tr>
<td>= Estimated Savings PMPM</td>
<td>$119.65</td>
</tr>
</tbody>
</table>

If, however, the change in risk profile of the chronic population (as captured by the change in distribution of member months across the risk strata) is to be recognized, then the baseline period cost $705.19 should be replaced by a number that re-weights each risk stratum’s baseline PMPM costs using the program evaluation period’s (Year 1’s) member months distribution. This is shown below.

<table>
<thead>
<tr>
<th>GROUP</th>
<th>AVERAGE COST BASELINE PMPM</th>
<th>YEAR 1 MEMBER MONTH PERCENT</th>
<th>AVERAGE COST YEAR 1 PMPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terminating</td>
<td>$929.75</td>
<td>18%</td>
<td>$721.73</td>
</tr>
<tr>
<td>Continuing</td>
<td>$706.53</td>
<td>60%</td>
<td>$623.80</td>
</tr>
<tr>
<td>Newly Identified</td>
<td>$601.71</td>
<td>22%</td>
<td>$528.11</td>
</tr>
<tr>
<td>Total</td>
<td>$705.19</td>
<td>100%</td>
<td>$620.80</td>
</tr>
<tr>
<td>Re-weighted</td>
<td>$724.15</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Accordingly, the savings concluded changes, as shown below.

**Savings Calculation (adjusted):**

<table>
<thead>
<tr>
<th>Action</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Chronic PMPM</td>
<td>$724.15</td>
</tr>
<tr>
<td>x Benchmark Trend</td>
<td>$1.05</td>
</tr>
<tr>
<td>= Expected PMPM</td>
<td>$760.36</td>
</tr>
<tr>
<td>- Actual Intervention PMPM</td>
<td>$620.80</td>
</tr>
<tr>
<td>= Estimated Savings PMPM</td>
<td>$139.56</td>
</tr>
</tbody>
</table>

Thus, the worsening of the risk profile does not allow the savings to be underestimated.
Another way of dividing a chronic population into homogenous risk strata is by condition and comorbidity categories:

<table>
<thead>
<tr>
<th>CHRONIC CONDITION</th>
<th>AVERAGE COST BASELINE PMPM</th>
<th>NUMBER OF MEMBERS BASELINE</th>
<th>AVERAGE COST YEAR 1 PMPM</th>
<th>NUMBER OF MEMBERS YEAR 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asthma</td>
<td>$587.43</td>
<td>5,426</td>
<td>$606.75</td>
<td>6,073</td>
</tr>
<tr>
<td>CAD</td>
<td>$521.51</td>
<td>2,151</td>
<td>$525.47</td>
<td>1,857</td>
</tr>
<tr>
<td>CHF</td>
<td>$574.26</td>
<td>533</td>
<td>$555.47</td>
<td>480</td>
</tr>
<tr>
<td>COPD</td>
<td>$595.08</td>
<td>1,696</td>
<td>$525.82</td>
<td>1,618</td>
</tr>
<tr>
<td>Diabetes</td>
<td>$495.50</td>
<td>7,554</td>
<td>$521.55</td>
<td>6,995</td>
</tr>
<tr>
<td>Asthma &amp; CAD</td>
<td>$698.99</td>
<td>237</td>
<td>$688.11</td>
<td>196</td>
</tr>
<tr>
<td>CAD &amp; CHF &amp; COPD &amp; Diabetes</td>
<td>$1,640.84</td>
<td>584</td>
<td>$1,440.18</td>
<td>472</td>
</tr>
<tr>
<td>Asthma &amp; CAD &amp; CHF &amp; COPD &amp; Diabetes</td>
<td>$1,799.27</td>
<td>382</td>
<td>$1,716.11</td>
<td>269</td>
</tr>
<tr>
<td>Total</td>
<td>$725.99</td>
<td>30,513</td>
<td>$697.04</td>
<td>28,590</td>
</tr>
<tr>
<td>Re-weighted</td>
<td>$709.94</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Basic Savings Calculation (unadjusted):**

Baseline Chronic PMPM ........................................ $725.99
x Benchmark Trend ........................................... $1.05
= Expected PMPM ........................................... $762.29
- Actual Intervention PMPM ................................ $697.04
= Estimated Savings PMPM ................................ $65.25

**Savings Calculation (adjusted):**

Baseline Chronic PMPM ........................................ $709.94
x Benchmark Trend ........................................... $1.05
= Expected PMPM ........................................... $745.43
- Actual Intervention PMPM ................................ $697.04
= Estimated Savings PMPM ................................ $48.39

Thus, the improvement of the risk profile does not allow the savings to be overestimated.
A typical question is: Why not use standard risk adjustment models to evaluate the change in risk of the chronic population? The answer is that because the chronic population is subject to chronic care management interventions, the application of risk-adjustment to this population would potentially neutralize the effect of the outcome that the evaluation is attempting to capture. This case study illustrates alternative approaches to assessing the change in risk of the chronic population without nullifying the chronic care management interventions’ impact. The end result is a savings estimate that avoids confounding from changes in the chronic population risk profile, a factor that is extrinsic to the chronic care management program.

**Case Study Example II**

**Lagged Prospective Approach**

**Introduction:** To evaluate the financial outcome of a chronic care management program, it is important to take into consideration the difference in risk profiles during the baseline period and the program period.

Potential differences in the risk profile can be detected by comparing the predictive risk scores at the beginning of each period. For the purpose of risk adjustment, any of the commercially available risk scores could be used. Proprietary predictive models could also be used, provided the model is validated as a good risk predictor. If the risk profiles in the two periods are significantly different, then the risk levels can be adjusted using the average predictive risk scores under certain circumstances.

**Methodology:**

- Identification of members with chronic diseases in the baseline period. Scoring the identified members using a predictive model at the beginning of the baseline period. Calculation of average risk score for the baseline period - RS0.

- Identification of members with chronic diseases in the program period. Scoring the identified members using a predictive model at the beginning of the program period. Calculation of average risk score for the program period - RS1.

- Calculation of the per diagnosed member per month (PDMPM) costs for the chronic population in the baseline period.

- Adjustment of PDMPM costs for the risk difference using the average risk scores - [Adjusted Baseline PDMPM Costs] = [Baseline PDMPM Costs] * RS1/RS0.

- Trending of adjusted baseline PDMPM costs using the benchmark trend and comparing it to the observed PDMPM cost in the program evaluation period to calculate the cost savings.
Benefits and Limitations: There have been concerns over using the risk scores to directly adjust for different risk profiles. Since the chronic care management program is designed to intervene in the way identified members manage their chronic conditions, this program intervention could change the risk profile of the chronic population in the program period. Directly adjusting for the risk could confound the program effect.

In the proposed approach, however, the predicted risk score in the beginning of each period is used for risk adjustment. Typically, information from a 12- to 18-month period before the time of scoring is used to derive the risk score. The effect of the intervention during the program year is not factored in the calculation of the predictive risk scores. The graph below illustrates the period of calculating the predictive risk scores.

This risk adjustment method can be used to evaluate the effect of chronic care management programs in the following situations:

• to compare the program effect in program Year 1 when the program was first implemented to the baseline year;
• to measure the cumulative program impact over several years, starting from the program implementation in Year 1, using the baseline year outcome as an adjusted control; or
• to measure the incremental year-over-year improvement beyond Year 1 in cases in which it is the main objective.

However, this risk adjustment approach is not appropriate for comparing the costs in any given program period beyond Year 1 to those in the baseline period. This is because the effect of the intervention in early program years would be used to calculate the predictive risk scores.
**Case Study Example:** The table below summarizes the chronic population’s PDMPM costs and the average predictive risk scores at the beginning of each period.

<table>
<thead>
<tr>
<th></th>
<th>PDMPM Costs</th>
<th>Average Risk Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Period</td>
<td>$686.84</td>
<td>2.24</td>
</tr>
<tr>
<td>Program Period</td>
<td>$773.54</td>
<td>2.35</td>
</tr>
</tbody>
</table>

The average predictive risk score for the baseline period is 2.24 and the average predictive risk score for the program period is 2.35. This indicates a difference in the risk profile between the two time periods.

If the risk adjustment were not done to account for the difference in risk profile, the analysis would directly compare the observed $686.84 in the baseline period with the $773.54 in the program period with a benchmark trend.

Following, is the analysis that shows the program to have a cost increase of $18.02 per identified member using a benchmark trend of 10 percent:

**Basic Cost Savings Calculation (Unadjusted):**

- Baseline Chronic PDMPM ................. $686.84
- x Benchmark Trend ............... $1.10
  = Expected PDMPM ................. $755.52
- - Actual Intervention PDMPM ..... $773.54
  = Estimated Savings PDMPM ........ $18.02

To account for the change in risk profile, the average predictive risk scores can be applied to adjust for the difference in the two time periods. The risk-adjusted PDMPM in the baseline period is calculated this way:

Risk-adjusted PDMPM in baseline = $686.84*2.35/2.24 = $720.57
Then, this risk-adjusted PDPM costs for the baseline period can be applied in the normal, pre-post calculation for cost savings. Assuming the same 10 percent benchmark trend, the cost savings is $19.09. Following, is the adjusted cost savings calculation:

\[
\begin{align*}
\text{Baseline Chronic PDPM} & \quad \$720.57 \\
\times \text{Benchmark Trend} & \quad \$1.10 \\
= \text{Expected PDPM} & \quad \$792.63 \\
- \text{Actual Intervention PDPM} & \quad \$773.54 \\
= \text{Estimated Savings PDPM} & \quad \$19.09
\end{align*}
\]

It appears that the initial cost increase in the basic savings calculation was due mostly to the higher predicted risk in the program period, which skewed the results. Once the risk profile was adjusted, it showed a positive cost savings in the program period.

**EVALUATION CONSIDERATIONS FOR SMALL POPULATIONS**

The value and risks of applying the methods recommendations articulated in these Guidelines to small populations is an important concern. The ultimate end users of outcomes results, typically, are organizations representing groups of people (such as employers, state and federal agencies, health plans and provider groups). Given the importance of employer-based health care management, the number of potential users of outcomes information through our employer groups alone is extremely large. These groups can vary widely in size. Employer groups ranging in size from 50 people to more than 250,000 people are actively engaged in chronic care management and have an increasingly active interest in understanding outcomes from these services. This section is intended to provide important contextual information for understanding outcomes measures for groups of individuals aggregated into relatively small numbers. In fact, the information below provides relative information for groups that have a wide range of sizes, up to 50,000 individuals.

Medical cost data are highly variable in small populations. Average costs for individuals with many of the common chronic care management conditions can show severe variation with high standard deviations. Even a few participants with high costs can have an impact on averages calculated for PMPM costs and result in wide confidence intervals around estimates of these measures. This is, in fact, one reason why medical management professionals were attracted to these conditions in the first place: The elimination of unnecessary variation was considered one of the key goals of chronic care management interventions. In larger populations, the impact of a few cost outliers on the measure variance does not have as significant an impact.

Table XV, Small Populations (page 83), demonstrates the significance of the impact of this variability on medical cost assessments. The information presented in the table is meant to be an example of the range of differences that can be seen in a sample population and is not intended to represent results that would be seen for all populations of this size. The table was created by starting with a large population of individuals who participate in a chronic care management program, then repeatedly taking various sample sizes and computing the economic impact of the intervention for each sample.
The table shows values for the upper and lower confidence intervals on first year PMPM medical cost savings estimates in a population with a robust chronic condition management program in place.

The savings estimates were derived from methods compliant with the CCA Outcomes Guidelines Report.

The variance was estimated using repeated samples from a large commercial plan population for which a chronic condition management program was implemented.

For example, in the first line, 500 samples of members were pulled from the entire population, generating, in each case, 30 members with chronic conditions. The medical cost savings algorithms applied to each of these samples and the variation in these results are measured to produce the confidence interval cited.

This process was then repeated for different size samples, up to 3,000 members, and with chronic conditions, a procedure called “bootstrapping” in mathematical programming circles.

### TABLE XV – SMALL POPULATIONS, PART I

<table>
<thead>
<tr>
<th><strong>POPULATION WITH CHRONIC CONDITIONS</strong></th>
<th><strong>TOTAL POPULATION</strong></th>
<th><strong>TOTAL POPULATION PMPM MEDICAL COST SAVINGS</strong>&lt;sup&gt;*&lt;/sup&gt; (95% Confidence Intervals)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>500</td>
<td>-$17.29 to $29.20</td>
</tr>
<tr>
<td>30</td>
<td>1,000</td>
<td>-$21.79 to $35.50</td>
</tr>
<tr>
<td>60</td>
<td>2,000</td>
<td>-$11.07 to $24.02</td>
</tr>
<tr>
<td>120</td>
<td>3,000</td>
<td>-$7.02 to $17.77</td>
</tr>
<tr>
<td>180</td>
<td>4,000</td>
<td>-$3.91 to $15.24</td>
</tr>
<tr>
<td>240</td>
<td>5,000</td>
<td>-$3.74 to $14.71</td>
</tr>
<tr>
<td>300</td>
<td>10,000</td>
<td>-$2.88 to $13.26</td>
</tr>
<tr>
<td>600</td>
<td>15,000</td>
<td>$0.24 to $10.24</td>
</tr>
<tr>
<td>900</td>
<td>20,000</td>
<td>$0.74 to $10.00</td>
</tr>
<tr>
<td>1,200</td>
<td>25,000</td>
<td>$1.43 to $9.04</td>
</tr>
<tr>
<td>1,500</td>
<td>30,000</td>
<td>$1.98 to $8.86</td>
</tr>
<tr>
<td>1,800</td>
<td>40,000</td>
<td>$2.06 to $8.36</td>
</tr>
<tr>
<td>2,400</td>
<td>50,000</td>
<td>$2.53 to $8.09</td>
</tr>
<tr>
<td>3,000</td>
<td>10,000</td>
<td>$2.88 to $13.26</td>
</tr>
</tbody>
</table>

*Savings for the population with chronic conditions is divided by the total population member months*
Clearly, caution is advised in producing medical cost savings measures in subpopulations with small numbers of members receiving management for chronic medical conditions. High variability will frequently result in conflicting, misleading and/or grossly inaccurate indications of program-related impact.

Looking at this table slightly differently, the upper and lower confidence interval limits were graphed vs. total sample size to get an appreciation of the direction and magnitude of this effect.

Of interest, several items should be noted.

- As the sample size increases, the upper and lower confidence intervals converge on the PMPM savings that a large population likely would recognize from this chronic care management program.

- For small numbers of participants, the range of measured PMPM savings can be striking—from as high as $35 PMPM to as low as -$20 PMPM for population sample sizes in the 1,000 to 2,000 range. In other words, a small company or group being serviced by this program could show PMPM impacts ranging from -$20 to $35 just on chance alone.

- Also of interest, the lower curve rises gradually, crossing $0 savings at a total population of 15,000 members/900 with the condition. One might argue that to avoid misleading clients or coming to the incorrect conclusion that there were no cost savings or even a loss, a minimum number of 15,000 population size should be included in a calculation before PMPM savings calculations are made for this program.
Typically, this large number of participants is not commonly available in modest size or small companies. Consider that in diabetes, where the prevalence rate is approximately 5 percent, you would need an employer with 50,000 employees to identify 2,500 diabetics—a number that, on this chart, still is in a range characterized by wide variation.

Ideally, the owner of a chronic care management program would calculate the number of individuals necessary in a program to ensure statistical significance of results, a process called a “power calculation” by statisticians, before embarking on a program. This would prepare them for the level of certainty they would have later in estimating outcomes.

CCA considered several alternative recommendations as possible solutions to evaluating program outcomes for small companies or institutions.

**Alternative I - Blend results (using standard medical cost savings methods) for the small population with results from “book of business” or larger reference population.**

Proponents of this approach note that it blends customer-specific results with results from a more stable, larger “book of business” reference population that is assumed to be comparable, typically done without severity, age, sex or similar adjustments.

Also, this approach seems in line with standard actuarial processes for premium rating, which blend the computed premiums calculated by book of business and the premium calculated by small group experience rating in a ratio that more heavily weighs the book of business at smaller sample sizes. The “credibility ratio” describes the relative percentage of the client’s own data that is mixed with a book of business or “manual” rate.

Finally, this approach enables the results of the small population to still be factored into the ultimate result, providing some sense of contribution from the small populations actually represented.

**Alternative II - Using a “book of business” where results from a larger reference population derive a factor that estimates the percentage of total medical costs saved per member who receives a significant level of support.**

Presumably, this support level will be defined and agreed to by all parties. The multiplier so calculated will then be multiplied by the number of participants to estimate the PMPM medical cost savings for individuals receiving the standardized level of support used to develop the statistic. This result will provide an estimate of savings for the group being managed. In the example above, the total population in the chronic care management program appears to save approximately $5 per member per month. If another program using the identical system enrolled only 50 people, then the total savings would still be $5 PMPM times the number of enrolled participants.

Proponents of this approach would note that it enables group level activity and, perhaps, cost data to be utilized in deriving savings. It also utilizes a more stable “book of business” or larger reference population results which has enhanced statistical validity.
Alternative III - Using “book of business” or results from a larger reference population to build a statistical model that assesses all the factors that drive savings.

An example here might be a linear regression model that develops termed weights around each of the various program interventions. Applying the coefficients from the model to a smaller group yields data to calculate savings estimates.

Proponents of this approach would point out that it enables group level information to be utilized in deriving savings. This approach also utilizes the more stable book of business or large reference population results and uses a mathematical weighting based on measured impact on drivers for medical cost savings. This method also permits the use of other studies, such as experiments connected in the medical literature that are more diffuse and more removed than actual empiric association.

All three alternatives should be clearly noted to provide only estimates of savings projected by the programs used to calculate them. For full transparency, it is necessary to counsel purchasers or developers of chronic care management programs in small populations that precise accounting of impact is not knowable for statistical reasons, but that reasonable estimates can be made and assist in the projection of program impact.

In summary, the principal reason for discussing the problems with small numbers in these guidelines is to provide a formally recognized concern that programs with small numbers of participants have serious, substantial issues with creation of inaccurate results if simple averaging techniques are used alone. While the issues are not insurmountable and some of the above alternative recommendations can be used to estimate savings in a way that represents mathematical validity, computation of savings to several decimal places has little meaning in small populations, and statistical consultation should be solicited whenever there is uncertainty around course of action.
EVALUATION CONSIDERATIONS FOR MATURE CHRONIC CARE MANAGEMENT PROGRAMS

The CCA methodological considerations have, until this point, addressed evaluation principles for a program for a previously unmanaged population or for a population whose unmanaged time frame was very recent and, therefore, would allow for comparison to a time period before management. More mature program populations – those being managed for three or more years, either by a single program provider or multiple program providers – may present unique evaluation challenges for a variety of reasons.

To help outline and expand on these reasons, the CCA Long-Term Evaluation Workgroup developed a frequently asked questions (FAQ) section that reviews the topic in depth (next page). In addition, the workgroup included a graphic to illustrate the considerations reviewed in the FAQ section. The graphic is a simple example, based on general experience rather than data analysis, and is intended to illustrate that a specific intervention does not continually bend the trend year over year. While the incremental impact of an intervention levels out, there is a continued enduring benefit realized, as the trend is lower than it would have been without the intervention. Different interventions can impact the expected trend in different ways; the slope changes in relation to the aggressiveness of the cost control intervention.

**FIGURE 7 – PROGRAM IMPACT OVER TIME**

As effective program interventions are applied, cost trajectory (trend) is impacted. Impact levels out, and to make further incremental impact, additional interventions are needed.

Each effective intervention impacts the trend, though the impact levels off. Continued savings are realized as the ongoing “impacted” trend is lower than it would have been, absent the intervention. Stopping the intervention results in an eventual return to the baseline.
Frequently Asked Questions Regarding the Outcome Evaluation of Mature Program Populations

Q: My condition management program has been in place for several years. How many years should a baseline per member per month be trended forward? Is there a limit?

A: While it is technically possible to trend a baseline PMPM dollar figure forward many years, there is a practical limit to the number of years for which this makes sense. Remember that the intent with pre-post analysis is to project what the cost for these members would have been, absent the program. In an ideal world, the evaluation would be done in a concurrent time period, with no trending necessary.

With each passing year, more confounds are brought into play that can have an impact on the efficacy of applying a trend derived from multiple years prior. Three years is a common limit and is a reasonable cutoff.

Q: When a program reaches its third year or longer, does it make sense to measure impact year-over-year? What should the expectation be, if the program is delivered in the same manner?

A: Once a program has been established and in place for a lengthy period for which it no longer makes sense to compare with an unmanaged baseline, it may make sense to perform a year-over-year analysis to ensure the program is still maintaining prior performance levels.

If the program has been deemed a success, with a verified financial return, one may simply measure year-over-year from that point forward. The goal is to maintain sustained impact. If the year-over-year analysis is flat, that can be interpreted to mean prior savings achieved are still in place. Costs may no longer be incrementally reduced year-over-year, but as long as identifiable member PMPMs do not creep back up, the prior year gains still hold. Put differently, while the trend is no longer being "bent," the effect of the prior years’ bending of the trend is still in place, so the observed trends are lower than what they would be without the program.

It may be assumed that a mature (e.g., in place for three or more years) program has been delivered to a population, and continues to be delivered to the same general population in a similar manner. When comparing one managed year with the next managed year, in this scenario, the program would not be expected to deliver incremental, year-over-year savings. To achieve incremental gross savings, something additional has to be done. Examples include:

- more conditions managed;
- higher level of outreach/engagement within conditions managed; and
- additional interventions applied.
Q: What if I replace my vendor? Can’t we start from scratch and treat the new program provider’s first year as “Year 1” for the program and have substantial ROI expectations?

A: Replacing a program provider with the intent of having “substantial ROI expectations” can be a slippery slope. Replacing a vendor makes sense if:

- the current program provider significantly fails basic operational functions;
- the current program provider does not appear to be having any impact in any year – i.e. is ineffective across key domains evaluated; and
- the replacement program provider will have more of an impact than the current vendor, adjusted for the amount spent.

If the program provider is ineffective in the basic task it sought to contract to perform, it indeed makes sense to change program providers. In the case of significant operational failure, such as member complaints, promise failures, corporate culture incompatibility, etc., changing business partners makes sense from a variety of perspectives but may not necessarily improve return on investment. Note, however, that there might be a difference between ineffective program function and the sustained improvement seen in a population when the program is operating optimally to extract all available benefit from the program strategy. In this case, changing the savings paradigm with the same program provider may be less costly and continue to improve effectiveness. Examples of techniques include benefit design changes, addition of biometrics, increased participation by the corporate sponsor through increased communication, etc.

Simply replacing a program provider with the hope that return on investment will increase, that the program will “reset” and again show return on investment, is likely to occur only if the new program provider represents existing capabilities plus substantial paradigmatic change. If the first program provider was effective over several years and the second program provider uses a similar model, there may be very little change that can be anticipated. Switching costs, corporate culture change and the risk of changing business strategies midstream all add to the cost of switching and should be considered. If a program is discontinued or changed, some mechanism for monitoring the decay of effect absent the program should be continued to confirm that a correct choice was made.

Q: If a program’s financial results are flat, does that mean the program is not doing what it is intended to do?

A: The answer to this question depends on how financial results are defined and at what time in the program evolution they are being considered. Financial results on a year-over-year change basis may show little change, yet represent optimal functioning and sustained improvement from what would have occurred in the absence of a program. Estimated cumulative results may be quite pronounced if the annual savings remain at a constant or unchanging level.

Time series analyses of financial program performance should be included with simple year-over-year calculations. Additionally, examining a program from a multidimensional perspective often will provide additional clues that it is not performing or has ceased being effective. Closely evaluating operational measures, clinical results, utilization management and intangible factors, such as quality of life indicators, are as equally important in evaluating program effectiveness as are simple accounting calculations of impact.
Q: If improvements are made to my program, does it make sense to measure year-over-year? What type of ROI should be expected?

A: If improvements are across-the-board programmatic changes and the elapsed time from the un-intervened baseline period is still relatively short, the program could still be measured relative to the baseline. This presumes that the nature of the improvement is such that the same population is managed, as was true pre-implementation. The challenge will be separating the incremental increase to savings.

- One approach would be to adjust the ROI benchmark for the period to reflect both the expected increase in savings, as well as increased program costs.
- An alternative is that performance for certain types of improvements may be measured by non-financial metrics, such as clinical compliance, medication adherence or other process or intermediate-term outcomes.

If the improvement affects only a certain segment of the population, it might be possible to determine the incremental savings by comparing current year savings between population segments with the prior year’s results. Any incremental savings could then be compared with costs associated with the improved program.

It may also be possible to stage the implementation of program enhancements so that the affected segment of the managed population can be directly compared with another unaffected segment. Again, this might allow the incremental impact of the enhancement to be calculated as the difference in PMPM savings between the two segments.

Q: How should we evaluate a chronic care management program that has been in place so long that trending forward the unmanaged baseline no longer makes sense?

A: The first question to address: At what point does the baseline year no longer make sense to use as a historical control?

- The pre-post historical control methodology CCA endorsed assumes the baseline and measurement year populations are comparable in term of key risk characteristics, such as age, gender, disease prevalence and health risk. To the extent that there is a significant amount of churn in the population, the risk characteristics of the population could change such that the baseline may be an inappropriate comparison population after a relatively short period of time.

- The fact is that almost any baseline year will no longer represent a reasonable comparison group after three or four years, whether due to the cumulative impact of turnover in membership or other confounding factors.

An obvious alternative is to “reset” the baseline year to a more recent intervened period. In this scenario it is reasonable to expect that any ROI performance guarantees would be lowered to reflect previously attained savings in the new baseline year. The extent to which the guarantees are lowered would depend upon the program, covered population and historical level of savings. In practice, this situation is no different than full replacement of an existing care management program, but the impact may often be disregarded.
Another option may be to move to alternative performance measures, such as sustaining levels of member activation, clinical compliance or medication adherence for long-term, continuing population segments.

If the managed population is large enough, it may be possible to estimate year-over-year savings on defined population segments using statistical methods, such as multiple linear regression or matched controls (each requiring a sizable population). In theory, comparison groups could be obtained from un-intervened populations, passively managed populations or even from historical program experience on comparable groups. Selection bias is an obvious issue in these approaches.

Q: What about results outside of the financial area? Should those keep improving, or do they flatten out as well? Do we need to measure them differently in later years?

A: Measures that are markers for the drivers of financial change – utilization measures, for example – will follow the same trajectory as the results for financial change. Thus, the decrease in admissions, emergency department visits, etc., for a particular condition will flatten out over time, following the same pattern as financial change. For clinical measures, the change in adherence will similarly flatten, but the time frame can vary depending on the rate of adherence at the starting point. Measures which start with high levels of adherence will level out more quickly than those that start with moderate or low levels of adherence. There is no need to change the measurement approach for clinical measures, but there is more flexibility with these over time. As a measure reaches the flat part of the curve, it can be replaced with other measures that have lower levels of adherence or that represent more advanced types of interventions. For example, when measuring diabetes program impact, an initial measure might be achieving two instances of a hemoglobin A1c during a 12-month period. Over time, this measure could be replaced by a measure of the level of control of hemoglobin A1c values.

Q: Is this concern limited to just chronic care management programs? Or does it apply to other programs as well?

A: Case management, with its focus on intensively managing a small subset of those with complex multisystem disease and a resultant continually refurbished “pool” due to catastrophic events, is unlikely to experience similar flattening of the curve of financial savings. The same confounders previously noted concerning time-series issues and need to establish a new baseline will exist in any population-based program impact measurement.
Measure Sets

FINANCIAL MEASURES

CCA recognizes that stakeholders utilize a wide variety of methods and measures for assessing financial outcomes. Some of these methods have their basis in experimental science and statistics, some are empirical or quasi-experimental and some are just convenient measures chosen for simplicity. Finding a mutually agreeable method for determining financial outcomes can be challenging for purchasers and providers of programs and services, as each party might have different views on the trade-offs between suitability and acceptability for different methods of measuring these outcomes. For this purpose, these terms are used as follows:

- **Suitability**: method achieves a threshold level of precision and reliability; consistent with the most rigorous standards of experimental, biostatistical and epidemiological sciences in dealing with random variation, confounders, regression to the mean, bias and equivalence.

- **Acceptability**: adequate transparency, ease, simplicity, practicality and utility of the method for the purpose for which outcomes are desired; methods are able to be understood and implemented by smaller purchasers, without external consultants or academics.

Any standardized method for financial outcomes that CCA endorses must meet this dual test of suitability and acceptability to achieve widespread acceptance and adoption. Is there a single method that will invariably meet these tests for all possible stakeholders, all populations and all circumstances? Can we define measures and methods that are equally satisfying to actuaries, consultants, statisticians, CFOs, benefits managers and medical directors?

---

**FIGURE 8 – ACHIEVING OPTIMAL BALANCE IN STUDY DESIGN**

<table>
<thead>
<tr>
<th>“Suitability”</th>
<th>“Acceptability”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rigor</td>
<td>Cost</td>
</tr>
<tr>
<td>Precision</td>
<td>Time</td>
</tr>
<tr>
<td>Replicability</td>
<td>Ease</td>
</tr>
<tr>
<td>Evidence-based</td>
<td>Simplicity</td>
</tr>
<tr>
<td>Bias, Confounders</td>
<td>Accessibility</td>
</tr>
<tr>
<td>Causal Association</td>
<td>Transparency</td>
</tr>
<tr>
<td>Experimental Design</td>
<td>Diverse Users</td>
</tr>
</tbody>
</table>
The CCA recommendation for a financial outcomes measure will, in most cases, satisfy these dual needs for the majority of stakeholders, populations and circumstances. Where it does not, CCA understands that parties will need to mutually agree on some alternative method better suited to the particular situation in question.

To find a financial outcomes approach that combines suitability and acceptability in balanced proportions and that also is generalizable to the majority of chronic care management settings, CCA’s goal has been to recommend a “middle of the road” approach that can be accepted by the majority, if not all, stakeholders. As such,

CCA recommends health care cost outcomes for the financial measure in program evaluation.

- Both per-member-per-month costs (applied over all covered lives) and per-diagnosed-member-per-month costs (applied to those members who are eligible for the program per predefined eligibility criteria) should be used to represent gross and net health care cost savings.

- CCA recognizes that return on investment will inevitably be computed by decisionmakers, but that this should not be the primary financial metric for program evaluation.

  --- Why cost savings as opposed to return on investment (ROI)?

- Return on investment describes the size of a return relative to the investment, not in absolute terms that facilitate comparisons of financial outcomes:

  - Projects with the same return on investment can have very different total savings.
  - For two projects, the one with the lower return on investment may have the higher total savings.

- Analogous to using net present value (NPV) versus internal rate of return (IRR) for capital investment decisions:

  - Return on investment suffers the same problems for comparing alternative chronic care management choices as IRR for capital investment decisions.
  - Neither IRR nor ROI can be relied on to select the option that maximizes value in all cases.
  - Total cost savings, like NPV, incorporates all cash flows and produces a result that can be compared with other options.
  - With savings, one can compute “net present savings” to incorporate time value of cost (as in NPV) if desired.

- Simple algebra illustrates the problem:

  - $1 million cost, $3 million gross savings = $2 million net savings, 3:1 ROI.
  - $10 million cost, $20 million gross savings = $10 million net savings, 2:1 ROI.
  - Would you rather have 3:1 ROI or the additional $8 million savings?
  - Savings, you can deposit at the bank; ROI, you can’t.
How To Measure Health Care Cost Outcomes

CCA recommends using actual and/or paid dollars to calculate savings reported in program evaluation.

Advocates of using allowed dollars prefer this approach to neutralize benefit differentials across different time periods; those preferring to see financial outcomes expressed in paid dollars are less concerned with benefit distortions and more sensitive to using the same paid cost that determines their premium trend. Note that one can determine cost in allowed dollars and then multiply by the paid/allowed ratio to approximate actual paid dollars.

Utilization Measures

Within population health management, there is widespread use of utilization measures as complements to the primary financial outcomes measure for understanding and validating program savings. Multiple utilization measures are currently used for this purpose, with different utilization measures revealing different information about chronic care management program performance.

Hospital admission measures (typically, admission rate expressed as the number of hospitalizations per thousand members per year) and emergency department utilization measures (typically, ED visit rate expressed as the number of visits per thousand members per year) are the utilization measures chronic care management programs most directly impact. These measures, derived from medical claims, are suitable for pre-post comparison, as well as year-on-year tracking for programs beyond their baseline year, to complement and corroborate the primary financial measures.

Two of the most commonly used admissions measures are:

- **All-cause admission rate for the diseased or eligible population.**
  - Relatively sensitive for major cost driver impacted by chronic care management programs.
  - Measures impact on comorbidities, as well as primary conditions of interest.
  - Does not test accuracy of identification algorithm for diseased population.

- **Condition-specific admission rate for the entire insured or covered population (using principal diagnosis only).**
  - Specific for one expected impact of chronic care management programs, but insensitive to other possible program impacts.
  - Does not measure chronic care management impact on comorbidities.
  - Sensitive to condition prevalence changes, so prevalence adjustment required.
  - May serve as “end to end” test of identification, outreach, enrollment, engagement, impact, retention, etc.

Other utilization measures that might be of interest to purchasers and suppliers of population health management programs may be collected and reported at the option and agreement of the parties. Table XVI, page 95, “Comparison of Various Utilization Measures,” elaborates on several of these measures.
<table>
<thead>
<tr>
<th>Measure</th>
<th>Numerator</th>
<th>Denominator</th>
<th>Utility/Comments</th>
</tr>
</thead>
</table>
| All-cause utilization of diseased population                           | All admissions for any cause for the period                                | Member mos. for the defined chronic care management-eligible or diseased population | • Relatively sensitive for major cost driver impacted by chronic care management programs  
  • Measures impact on comorbidities as well as primary conditions of interest  
  • Does not test accuracy of identification of diseased population                                                                                     |
| Condition-specific utilization of diseased population                  | Admissions for the period for which principal diagnosis is condition in question | Member mos. for the defined chronic care management-eligible or diseased population | • More sensitive for expected key impact of chronic care management programs but does not measure chronic care management impact on comorbidities  
  • Does not test accuracy of identification of diseased population                                                                                           |
| Condition-specific utilization of entire population – known as “plausibility indicator” by Disease Management Purchasing Consortium International (DMPC) (Variant 1 – one measure) | Admissions and ED visits (separately) for the period for which principal diagnosis is condition in question | Member mos. for the entire insured or covered population | • Specific for one expected impact of chronic care management programs but insensitive to other possible program impacts  
  • Does not measure chronic care management impact on comorbidities  
  • Sensitive to condition prevalence changes, so prevalence adjustment required  
  • May serve as “end to end” test of chronic care management identification, outreach, enrollment, engagement, impact, retention, etc.                                                                                      |
| Condition-specific utilization of entire population – (Variant 3 – one combined measure) | Combined Admissions and ED visits (aggregated on cost-weighted basis) for the period for which principal diagnosis is condition in question | Member mos. for the entire insured or covered population | • Same as for Variant 1  
  • Adds ED visits, which reflect additional chronic care management impact beyond admissions reductions                                                                                                                |
| All-cause utilization of entire population                              | All admissions for any cause for the period                                | Member mos. for the entire insured or covered population | • The most inclusive measure of utilization but with attribution problems  
  • May be insensitive to chronic care management program impact, as many other influences impact total utilization changes over time                                                                                                 |
| Utilization x unit cost to build up total cost estimate (surrogate for measured cost approach) | Utilization by service type x standard unit cost by service type, added together | Member mos. for the entire insured or covered population (for PMPM cost estimate) | • Relatively sensitive for major cost driver impacted by chronic care management programs  
  • Measures impact on comorbidities as well as primary conditions of interest  
  • Does not test accuracy of identification of diseased population                                                                                           |
CLINICAL MEASURES

The clinical measures recommended and discussed in this report were developed in collaboration with the National Committee for Quality Assurance (NCQA). Measures are categorized into Group I and Group II measure sets. Metrics were developed for five chronic conditions: diabetes, heart failure, coronary artery disease (CAD), asthma and chronic obstructive pulmonary disease (COPD). Existing clinical measures for the Group I recommendations were reviewed against the following criteria:

— The measure addressed an important gap in care.
— Programs could impact the measure.
— The measure had previously been endorsed or was in wide use (e.g., National Quality Forum or HEDIS measures).

The Group I measure development focused on adapting measures for which a national consensus on the measure and its specification was available (e.g., LDL testing and control for coronary artery disease). Although efforts were made to use existing measures “as is,” changes were made to adapt language or to define data sources that would be widely acceptable across population health management. Group II includes measures that are in wide use in chronic care management programs, but for which national consensus on measure specification is not yet available (e.g., sodium intake self-monitoring by heart failure patients).
**Group I Measures**

**Asthma**
- Flu vaccination
- Pneumococcal vaccination
- Smoking cessation identification and advice
- Current medication use
  - Controller medication

**COPD**
- Flu vaccination
- Pneumococcal vaccination
- Smoking cessation identification and advice
- Spirometry evaluation
- Medication use
  - Bronchodilator

**Heart Failure**
- Flu vaccination
- Pneumococcal vaccination
- Smoking cessation identification and advice
- Medication persistence
  - beta blockers
  - ACE/ARB
  - anticoagulants (with chronic or paroxysmal AF)

**Coronary Artery Disease**
- Flu vaccination
- Pneumococcal vaccination
- Smoking cessation identification and advice
- LDL testing and control
  - annual test
    - LDL < 100; < 130
- Blood pressure
  - Blood pressure < 140/90
- Medication persistence
  - beta blockers
  - ACE/ARB
  - aspirin

**Diabetes**
- Flu vaccination
- Pneumococcal vaccination
- Smoking cessation identification and advice
- Daily aspirin use
- LDL testing and control
  - annual test
    - LDL < 100; < 130
- HbA1c testing and control
  - annual test
    - HbA1c < 7.0; > 9.0
- Blood pressure
  - blood pressure < 130/80; < 140/90
- Eye exam
- Nephropathy testing
Group II Measures

These measures include:

Heart Failure
- Screening for depression
- Knowledge/self-efficacy
- Diet/weight management
- Physical activity
- Self-management/activation
- Care coordination
- Sodium intake monitoring
- Volume overload monitoring
- Alcohol use

Coronary Artery Disease
- Screening for depression
- Knowledge/self-efficacy
- Diet/weight management
- Physical activity
- Self-management/activation
- Care coordination

Asthma
- Screening for depression
- Knowledge/self-efficacy
- Self-management/activation
- Presenteeism/productivity
- Medication use (persistence)
- Action plan

COPD
- Screening for depression
- Knowledge/self-efficacy
- Self-management/activation
- Presenteeism/productivity
- Inhaler technique sufficient

Diabetes
- Screening for depression
- Knowledge/self-efficacy
- Diet/weight management
- Physical activity
- Self-management/activation
- Care coordination
Measures identified for the Group II measure set do not have nationally accepted specifications. As such, two measures from the identified group were chosen to develop for the Volume 3 report. These measures include self-management and medication adherence.

SELF MANAGEMENT

CCA recognizes the critical role individuals play in managing their health on a daily basis. Within the context of managing a chronic disease—diabetes, for example—this role becomes more complex with the daily need to self-administer and manage multiple medications, self-monitor and manage blood sugar levels and respond appropriately, implement and follow dietary recommendations and incorporate healthful lifestyle behaviors, such as daily exercise. Successful self-management of a chronic disease can slow disease progression and improve overall quality of life.

Chronic care management programs, therefore, should incorporate self-management assessment and education to increase awareness of and compliance with treatment guidelines, facilitate problem solving skills, support and motivate individuals in making healthful behavioral changes and promote open communications with providers.

This section includes:

- Definition of self-management.
- Criteria for selecting and prioritizing the development of (self-management) metrics.
- Specification of metrics to assess self-management both on the individual and program level.

CCA defines self-management as:

**Self management consists of the ongoing processes and actions taken to manage/control one's own condition, with the goal of improving clinical outcomes, health status and quality of life.**

- Core components of the self-management process include incorporating the needs, goals and life experiences of the individual, in addition to being guided by evidence-based standards.
- The objectives of self-management interventions are to support informed decisionmaking, improve and promote use of self-care skills and behaviors and encourage problem solving and active collaboration among participants, family/caregivers and others on the health care team.
- Assessment of an individual’s self-management capabilities relies on behavioral measures that include self-efficacy, health beliefs and readiness to change, knowledge of the condition and its treatment and self-care skills required to manage the condition.
Criteria for Self Management Metrics

- Metric can be influenced by a chronic care management program.
- Metric assesses an issue or problem that has a substantive impact on the participant cohort over time.
- There is an evidence base for designing or selecting the metric.
- The metric is routinely measured or is measurable by use of a validated tool or method.
- The resulting information is useable to assess and refine the intervention to lead to improved patient outcomes.

Self Management Metrics

Eight possible metrics applicable to the development and implementation of disease self management education programs have been identified.

- The knowledge of condition/health literacy.
  — Of condition/issue.
  — Of solution/intervention.
- Readiness to change on applicable behaviors (both generic and condition-specific). The stages in the process that individuals may go through to engage in and fully adopt behaviors.
- Self-efficacy (both generic and condition-specific).
  — Individual’s belief about his/her ability to produce desired effects.
  — Related constructs.
  • Confidence.
  • Perceived control.
- The use of devices and tools designed to support self management.
- The presence of collaborative goal-setting activities.
- The use and content of assessments of participant self management skills.
- The presence of individual action plans designed to guide self management.
- The presence and frequency of use of specific self-monitoring activities (both generic and condition-specific).
MEDICATION ADHERENCE

Introduction

Each day, every individual who takes medication — either for the purpose of treatment or prevention of a condition — is presented with a personal moment of truth that results in either taking the medication as prescribed, or not. The individual may or may not be aware of that fleeting moment of truth, but in that second, U.S. health outcomes are catalogued. Annually, medication non-adherence has a negative impact on health outcomes and contributes to increased health care costs. This includes $177 billion in direct and indirect costs and, more specifically, $100 billion in hospitalization costs alone. This annual impact to the U.S. health care system represents a challenge for all health care. Individual patients experience the cost of medication non-adherence through preventable hospitalizations and emergency department visits, poorer clinical outcomes, loss of quality of life and lower productivity at work.

Causes of medication non-adherence represent a definable set of attributes that cross personal, psychosocial, cognitive-behavioral and cultural boundaries. “Reasons (for non-adherence) may include a lack of (individual) knowledge about the medication; side effects or adverse events; forgetfulness; lack of social support; cultural, health and/or religious beliefs; denial of conditions; financial challenges; poor relationships with clinicians; and lack of health literacy.”

Given the multiple issues that drive non-adherence, turning the tide on the problem requires developing and carrying out a systematic, strategic, multi-layered approach to the dilemma.

Medication non-adherence represents a critical care system failure — not unlike that of inadequate pain management. Like pain management, medication adherence is a fundamental tenet on which Western medicine outcomes rely. In the last decade, when inadequate pain control was recognized as a critical driver of poor health care outcomes, national medical associations and societies declared pain the “fifth vital sign” — raising provider and institutional awareness and setting national care standards to improve patient outcomes. Following the declaration, standards for pain control were written and subsequently adopted by any organization delivering or supporting patient care, for example, health care systems, hospitals, nursing homes, ambulatory care centers, home care agencies.

The systematic approach taken to improve pain management across the United States mirrors the need for a similar approach to improve medication adherence. It might be that a highly visible national focus is required to improve this critical aspect of health care, and it also might be that medication adherence will become a “sixth vital sign” in patient care.

Despite the topic’s magnitude, health management was once absent a standardized methodology to measure and compare medication adherence across managed populations. This report offers the following guidelines for the purpose of implementing and measuring programs design to improve the medication adherence of a population.

- Organizational Approach to Medication Adherence Best Practice
- Measure of Medication Possession Ratio
- Measure of Medication Persistence
- Measures of Self-reported Medication Adherence
Organizational Approach to Medication Adherence Best Practice

The workgroup goal of developing a best practice framework for medication adherence was achieved through the creation of the Organizational Best Practice Framework (page 104). The Framework includes a step-wise set of systematic interventions that will enable organizations to promote best practice in supporting patients to achieve high rates of medication adherence. Through review of the current literature and discussions, the workgroup identified a variety of approaches used by providers and organizations aimed at improving medication adherence. A discussion that identifies surveys and tools used in self-reported medication adherence, based on the literature review, is included in this section (pages 112-114). Less information, however, was available on how organizations can systematically change medication adherence rates in their patient populations; more research appears to be needed.

While no two organizational approaches to the problem of medication adherence are alike, there are common areas that define a potentially generalizable medication adherence improvement continuum. The best practice continuum, as currently defined, is based upon available literature, as well as expert input from the workgroup, representing multiple areas of the health delivery system. The Framework is not meant to be comprehensive in scope, but rather offers a starting point from which organizations may begin to assess themselves and their focus on improving medication adherence. Theoretically, as more organizations work to improve medication adherence systematically, the current Framework will evolve to include new research and additional key interventions that support best patient and population outcomes. The purpose of defining the level of best practice, therefore, is to begin to assist organizations to: 1) evaluate their current state of function; and 2) work in a step-wise manner to systematically improve their approach to medication adherence support over time.

To define the level of the Framework, the workgroup first envisioned a fully functional system and infrastructure that promoted medication adherence in every person with whom it interacted. Various common elements appeared to be required, and include:

- surveys/tools and training to raise individual practitioner awareness of potential patient medication non-adherence;
- organization-practitioner alignment that medication non-adherence is a critical element to achieve positive patient outcomes;
- a defined organization-specific suite of practitioner/organization interventions to lessen the risk of individual patient non-adherence;
- a method to aggregate adherence/non-adherence data across populations;
- goal-setting among practitioners and the organization to achieve high rates of adherence across populations;
- incentives/reimbursement to practitioners/providers and organizations based on best practice outcomes;
- real-time medication adherence data and decision support for providers at point of care;
- organizational value-chain alignment in support of medication adherence; and
- overall organization success linked to high rates of population medication adherence.
These elements were gauged as more or less systematic in approach, and arrayed across a grid (left to right). Of particular note, the grid takes into account changes and interventions required from both individual providers and organizations to best support medication adherence. On the grid, the term “HCP” or “Health Care Provider” refers to any provider of service, including physicians, pharmacists, case or disease managers and ancillary providers. Also, the term “Organization” refers to any organization that plays a role in health care, including those that directly deliver care to patients (e.g., health delivery systems, individual or group providers, pharmacies, home care agencies, long term care facilities), as well as those removed from direct patient care (e.g., health plans, population health vendors, other health care-related organizations).

The elements are ordered into a stepped continuum, arraying the less systematic interventions to the left of the continuum and the more systematic approaches to the right. Related HCP and organization interventions are “stacked.” The model defines five levels of systematic approach — assigned Levels I through V — with organizations progressing along the best practices continuum from Level I to Level V. The five levels include:

- Level I – Person assessment;
- Level II – Identification of at-risk individuals;
- Level III – Basic intervention to improve medication adherence;
- Level IV – Advanced intervention — goals-based management to improve medication adherence; and
- Level V – Fully functional system in support of medication adherence for the individual and the population.

This visual representation of the five levels defines the steps to completion of each level, based on progression from lower to higher “Systematic Approach.” See Figure 9, page 104.
### FIGURE 9 – ORGANIZATIONAL BEST PRACTICE FRAMEWORK

Building a Best-practice Approach to Improve Medication Adherence in Organizations*

<table>
<thead>
<tr>
<th>Level I – Person Assessment</th>
<th>Level II – Identify At-Risk Individuals</th>
<th>Level III – Basic Intervention</th>
<th>Level IV – Advanced Intervention (Goals-Based Management)</th>
<th>Level V – Fully Functional System for Medication Adherence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validated tools available to assess individual patient medication non-adherence</td>
<td>HCPs** review patient adherence responses at point of care: at-risk individuals are identified</td>
<td>HCPs assess at critical points of care to identify and overcome barriers</td>
<td>HCPs: a) receive/access data on their patients’ adherence; b) training occurs; c) Goals are set</td>
<td>Organization links medication adherence to overall system success</td>
</tr>
<tr>
<td>Cross-organizational alignment to assess patient risk of non-adherence</td>
<td>Organization aggregates data to identify medication adherence trends</td>
<td>Organization builds formal database to track and report trends</td>
<td>Organization offers suite of targeted assessments and interventions to lessen risk</td>
<td>Organization aligns value-chain to support ongoing adherence</td>
</tr>
<tr>
<td>Organization collectively sets targets and goals to improve organizational performance</td>
<td>Organization integrates data base into daily provider function</td>
<td>Organization challenges low HCP performers to improve outcomes</td>
<td>Organization builds formal database to track and report trends</td>
<td>Organization offers suite of targeted assessments and interventions to lessen risk</td>
</tr>
</tbody>
</table>

* Definition of “Organization” is “any organization” in the health care space – could be health care delivery systems, pharmacies, analytics companies, population health vendors, etc.

**HCPs are any provider of any type, including physicians, case managers, pharmacists, ancillary staff, etc.
How to use the best practice model in your practice or organization

The five-level organizational best-practice medication adherence model allows providers and organizations to evaluate their current level of medication adherence progress and to envision and implement sequential systematic process improvements over time. It is important to note that the model links the various steps or related interventions in each Level, and requires the providers and their organization to take a systematic approach to improvement. Specifically, the model reminds organizations about the importance of:

- consistent provider practice (through training and expectation setting);
- identification of key areas of system failure (and ways to overcome failure);
- measurement and reporting on population-based medication adherence outcomes; and
- the link between medication adherence in populations and positive health outcomes.

Any organization can follow these steps to improve medication adherence:

1. Perform an organizational self-assessment. Identify practices currently in place or needed to improve medication adherence. Use the model to determine your organization’s level. Identify processes that require improvement to move up a level. Set a goal to achieve a new level by a specific point in time. Identify challenges and barriers to improvement.

2. Identify medication adherence “champions” in your organization. Showcase the opportunity in your organization.

3. Identify medication adherence self-report tools already in use in your organization to identify at-risk individuals.

4. Pilot the use of self-report surveys and tools in several sites in your organization. Strive for cultural acceptance and regular use of tools. Report on the outcomes at the pilot sites based on use of the tools, noting the prevalence of medication non-adherence in your populations.

5. Identify a few patient interventions to support improved medication adherence. Identify the “right” practitioners to carry out the interventions. Identify how providers and other practitioners can support significant medication adherence improvement.

6. Implement a simple mechanism to track performance over time in the sites in your organization.

7. Showcase your findings (e.g., accrediting agencies, quality improvement group settings). Add new pilot sites, identifying at each site the similarities/differences, available resources and the setting, based on the population. Capture success stories from the pilot sites and celebrate organizational success.

8. Consider making medication adherence a “key success indicator” in your organization.
**Measuring Medication Adherence**

**PHARMACY VS. MEDICAL BENEFIT**

To accurately evaluate adherence and identify non-adherence using claims data, it is important to ensure that all pertinent claims are available. For drugs covered primarily through the pharmacy benefit, this is fairly straightforward. However, some medications—often injectable biologic medications—can be covered under either the pharmacy or the medical benefit. A few examples include erythropoietin-stimulating agents, tumor necrosis factor (TNF) blockers and interferons, among others. These drugs are covered by most pharmacy benefit programs and may also be provided and billed by the health care provider using a HCPCS billing code.

Over time, some patients might change from physician-administered setting (and billed) medication to home administration under the pharmacy benefit, or vice versa. One reason involves plan design differences, since one benefit might be more favorable than the other. A potential reason for switching in the reverse direction, from self-administered to physician-administered, could involve reduced ability to self-administer due to disability or lack of fine motor agility.

In cases where individual patients might have their medications covered under either benefit, it is important to ensure that claims data from both sources are available and integrated. This also would involve a crosswalk between HCPCS and NDC codes.

**MEASURE OF MEDICATION POSSESSION RATIO**

CCA recommends the measures of Medication Possession Ratio and Persistence when assessing the outcomes of a program.

Medication possession ratio (MPR) is an operational definition of adherence: a retrospective assessment of days supplied over an evaluation period. A recommended method for calculating MPR, one aspect of the total adherence equation, is described in the section below.

**Methodology**

MPR is a population-based measure reported as a percentage:

- **Data sources**: Administrative pharmacy claims and eligibility data.
- **Evaluation Period**: A fixed calendar length: 12 months (annual). One month run-out will be allowed for claims lag; therefore, the measure can be calculated at the end of month 13.
- **Enrollment Criteria**: A continuous evaluation period with no more than a 45-day gap in pharmacy benefits coverage.
- **Denominator**: The duration from first (index) prescription to the end of the evaluation period.
- **Numerator**: The days supplied over the same period.
- **Numerator Limit**: To control for potential confounders of concomitant therapies and overlap (due to medication switches within drug class, lost medications, etc.), prior to disease-specific or drug class-specific, population-based MPR roll-ups (see “Whom to Report,” below), each individual’s MPR numerator/denominator set will be limited
such that the numerator is less than or equal to the denominator (i.e., days supply ≤ days; individual MPR cannot exceed 100 percent).

- **What to Report**: MPR as a percentage and by quartiles (e.g., box plot, box-and-whisker diagram).

- **Whom to Report**: Reported by condition and by drug classes applicable to that condition. Individuals with multiple conditions (e.g., CAD and diabetes) will be counted for all conditions and for all appropriate drug classes. See a representative drug class list, per condition, on the CCA Web site, www.carecontinuum.org.

- **How to Report**: Each appropriate condition/drug class combination, per the representative drug classes list, would have an MPR reported. For example, for beta blockers in CAD, this measure would include the sum of all days (for all members with CAD taking beta blockers) in the denominator and the sum of all days supplied (for all members with CAD taking beta blockers) in the numerator. To calculate disease-specific (columns in drug class table) or drug class-specific (rows in drug class table) MPR totals, one must sum all days and days supply for appropriate column/row and then perform the division to calculate percentage.

**Inclusion/Exclusion Criteria**

- Intended for more prevalent common chronic conditions (persistent asthma, CAD, CHF, diabetes, hypertension and hyperlipidemia).

- Intended for oral medications only (CAD, CHF, diabetes, hypertension and hyperlipidemia).

- Intended for oral medications and inhaled medications only (persistent asthma).

- Excludes liquid-form medications.

- Index prescription must occur within the first six (6) months of the evaluation period.

- A minimum of two claims, per member, for a specific drug class must be incurred to include the member in the calculation.

- Excludes “carry in” from a prior evaluation period. For example, a 30-day supply filled on Dec. 15, 2006, would not be considered for the 2007 evaluation period.

- Excludes “carry out,” when medication supply goes beyond the evaluation period. Using the above example, the same 30-day supply is filled on Dec. 15, 2006. For the 2006 evaluation period, only the days supply from Dec. 15 to Dec. 31 (17 days) would count in the numerator.

- Potential confounders of contraindications, samples and clinical utilization (inpatient admissions) may impact MPR.

**MEASURE OF MEDICATION PERSISTENCE**

Medication persistence can be defined as the “amount of time that an individual remains on chronic drug therapy.” A recommended method for calculating persistence, another aspect of the total adherence equation, is described in the section below.
**Methodology**

Persistence is a population-based measure reported as a percentage over time:

- **Data sources:** Administrative pharmacy claims data.
- **Permissible Refill Gap:** 60 days.
- **Annual Evaluation Period:** A fixed calendar length: 12 months (annual). One month run-out will be allowed for claims lag.
- **Evaluation Time Periods (ETP):** The time from last fill’s run-out date (i.e., date of fill + days supply) + permissible refill gap (not to exceed the end of the calendar year).
- **Denominators:** The number of eligible members, in each ETP period, from first (therapy initiation) prescription to the end of the evaluation period.
- **Numerators:** The number of eligible members, in each ETP period, from first (therapy initiation) prescription to the end of the evaluation period who do not exceed the permissible refill gap.
- **What to Report:** Persistence as a percentage per ETP period.
- **Whom to Report:** Reported by condition and by drug classes applicable to that condition. Individuals with multiple conditions (e.g., CAD and diabetes) will be counted for all conditions and for all appropriate drug classes. See a representative drug class list, per condition, on the CGA Web site, www.carecontinuum.org.
- **How to Report:** Each appropriate condition/drug class combination, per the representative drug classes list, would have a time series persistence reported. This can be visualized as a survivability graph (much like a Kaplan-Meier curve). The x-axis consists of the ETP periods and the y-axis consists of the persistence percentage at each PET period.

**Inclusion/Exclusion Criteria**

- Intended for more prevalent common chronic conditions (persistent asthma, CAD, CHF, diabetes, hypertension and hyperlipidemia).
- Intended for oral medications only (CAD, CHF, diabetes, hypertension and hyperlipidemia).
- Intended for oral medications and inhaled medications only (persistent asthma).
- Excludes liquid-form medications.
- Excludes “carry in” from prior evaluation period. For example, a 30-day supply filled on Dec. 15, 2006, would not be considered for the 2007 evaluation period.
- Excludes “carry out,” when medication supply goes beyond the evaluation period. Using the above example, the same 30-day supply is filled on Dec. 15, 2006. For the 2006 evaluation period, only the days supply from Dec. 15 to Dec. 31 (17 days) would count in the numerator.
- Potential confounders of contraindications, lost medications, samples and clinical utilization (inpatient admissions) may impact persistence.
**Examples**

**CHART IV – DIABETES PERSISTENCE EXAMPLE**

<table>
<thead>
<tr>
<th></th>
<th>Time 0</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
<th>Time 4</th>
<th>Time 5</th>
<th>Time 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mbr1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Mbr2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mbr3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Mbr4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

- **P%**
  - Time 0 = Initiation of Therapy, regardless of calendar date @ start
  - – = “Null” entry (i.e. non-measurable data point due to start date)

**FIGURE 10 – DIABETES PERSISTENCE CURVE**

- % Persistence
  - Biguanides
Frequently Asked Questions Regarding MPR and Measures of Persistence

Q: What are medication possession ratio (MPR) and persistence meant to measure?
A: MPR and persistence measurements are complementary. MPR attempts to highlight the proportion of filled doses while on a therapy regimen. Persistence generally captures those individuals who have halted (dropped off) therapy. The intervention strategies to close the gaps for such heterogeneous groups can differ greatly.

Q: What are commonly recurring barriers to optimal MPR and persistence?
A: Forgetfulness, dose/schedule alteration and/or cost are often cited as reasons for poor MPR outcomes. Whereas side effects, cost and/or administration issues might be linked to persistence failure. These are examples taken from published literature; however, the reasons for each can significantly overlap. Since both MPR and persistence are affected by differential response, based on an individual’s diagnosis and the therapeutic drug class utilized, multiple elements should be considered when determining effective clinical strategies for such populations.

Q: Why is measurement more variable with inhaled medications (e.g., asthma)?
A: An example of differential response, and a challenge of administrative pharmacy claims data, lies in the domain of pharmacotherapy measurements for persistent asthma. This form of asthma is chronic with acute exacerbations. Days supply for inhaled medications is also more difficult to perfect – individual lung volume and inhalation habits are inconsistent, and may change with age, circumstance or other variable. In addition, the accuracy of the days supply field is variable, depending on factors such as prescriber, prescribing system, filling adjudication system, etc. For these reasons, it is not uncommon to see MPR rates for asthmatics well below that of other chronic conditions. And, for many people, asthma severity can be influenced by seasonal variance – so, that might explain periods of adherence versus non-adherence (when a person is “off-season”).

Q: When might be considered a “triggering point” for interventions regarding MPR and persistence?
A: Many of the effectiveness studies performed by the pharmaceutical industries have focused on MPR ≥ 80 percent. Therefore, many presume MPR rates below that to be suboptimal. An optimal MPR threshold has not been causally proved, via RCTs originally designed for such, and experience reasons that such optimal MPR thresholds might differ based on multiple factors (e.g., population demographics, socioeconomic factors, side effect profiles, etc.). However, a recent study by Hansen et al. in the March 2009 issue of The Annals of Pharmacotherapy suggests that 80 percent might represent a point to determine adherent/non-adherent cohorts, while retaining parity between sensitivity and specificity. For persistence, once a person has missed an allowable refill gap, they are considered to be non-adherent. Triggering points will also depend on how one views the data: by book of business, by individual, by lines of business/markets or other dimensions. The MPR/persistence methods are meant to standardize outcomes reporting in this domain. In either case, differential response and each individual’s motivation and behavioral skills will play significant roles in determining appropriate intervention triggers and strategies.
Q: Although the recommendations call for annual measurement with appropriate claims run-out, could my organization calculate on an ongoing basis or quarterly?

A: A method for routine measurement that minimally impacts inclusion/exclusion criteria and remains statistically less variant would be a rolling 12 months view. One would merely adjust the “book ends” of the analysis time period to be 12 months. Regardless of measurement frequency (monthly, quarterly, annually), 12 months of claims data will minimize impact of some known confounders, such as plan design changes, mail versus retail fills and medication guidelines changes.

Q: Why did you decide to exclude liquid form medications?

A: Liquid medications have similar challenges as inhalers, with days supply fields being subject to considerable variation and data collection issues. For example, blood glucose control in insulin-dependent diabetics can be driven by several factors — current weight or prior blood glucose readings, for example. The adherence workgroup, from a complexity reduction standpoint, decided to approach the more straightforward forms to vet these standardized methods.

Q: How do you define drug class?

A: The recommended therapeutic drug classes were chosen with clinical practice in mind (i.e., the way the medications are employed to treat specific conditions). A drug class encompasses medications with similar modes of action that would not, under the majority of circumstances, be used as concomitant therapy.

Q: How did you arrive at the drug classes for each condition?

A: The list is recommended as a general starting point, but it is by no means all-inclusive. The conditions were chosen due to their higher overall prevalence rates. The medications had to be commonly used and relevant in the medication management of the designated condition (guidelines-based), be a chronic medication (not just taken at acute phases or used with tapering regimens) and have relatively consistent claims data availability.

Q: How do discount drug programs impact these measurements?

A: The growth of discounted prescription drug programs might have an effect on measurement accuracy. As companies begin to offer and expand low-cost discount drug lists, prescriptions associated with such programs might not appear in claims data. This occurs primarily due to cash purchases in which a benefit card is not presented and an administrative claim is not generated.
Q: How does the drug class-level MPR methodology account for concomitant therapy and medication switches?

A: If an individual has overlapping days supply for two medications within the same drug class, such as when a patient is switched from one statin to another, then a drug class-level MPR calculation would result in that person’s days supply exceeding the days in the evaluation period. As such, those days supply (numerator) would be capped so that they could be equal to, but not greater than, the days (denominator). If a person switches medications between two drug classes (i.e., discontinues therapy in drug class x and begins new therapy in drug class y), the MPR for drug class x would be deflated, whereas drug class y would remain a truer reflection of that individual’s ongoing adherence.

Q: How does the condition-level MPR methodology account for concomitant therapy and medication switches?

A: If an individual has claims for two drug classes for a condition, the calculation of condition-level MPR must make an assumption regarding whether the situation involves concomitant therapy or a switch in therapy. If the methodology assumes concomitant therapy, it will accumulate each drug’s days supply in the numerator, and each drug’s potential days of therapy in the denominator. If the methodology assumes switching, it will accumulate each drug’s days supply in the numerator, but will only accumulate potential days once in the denominator. Thus, the first approach can deflate MPR, and the second approach can inflate MPR. Neither is necessarily right or wrong, since there is currently no way for a claims-based analysis to ever truly know whether these situations involve concomitant therapy or switches. Therefore, it is essential for condition-level MPR reports to clearly explain which assumption and methodology is chosen.

Measures of Self-reported Medication Adherence

Self-report surveys for measuring medication adherence vary in structure and purpose. To identify empirically validated self-report medication adherence surveys, a literature search was conducted from 1980 to present. This literature search yielded instruments that were general as well as disease-specific. Specifically, adherence surveys have been developed for therapeutic areas such as hypertension, HIV and depression, among others. These surveys may include different elements, such as general self report, attitudes, compliance or identification of barriers to taking medicines. Other instruments focused on the predictability of those individuals that would be non-adherent. Typically, the surveys consisted of various questions grouped together (common factors) to help discuss known factors that impact the beliefs of taking medicines.

From an item development perspective, there is no standard methodology available. The survey instruments were typically developed using standard methodologies, such as those used to develop quality of life instruments. At minimum, items for the development and validation process may include the question development process, statistical analysis demonstrating the instrument is psychometrically sound (i.e., reliable and valid) and interpretation (statistical and clinical significance).
Additional measures may provide comparisons to self-reported medication-taking behaviors. Ogedegbe, et al. in the *Journal of Clinical Epidemiology* also outlines a process to develop a medication self efficacy scale\(^4\). Other known methods are to apply validation through the use of pharmacy claims comparisons or comparisons to other known survey instruments.

The primary goal for the self-reported instruments table was to provide those that could be used for multiple chronic conditions. Disease-specific instruments may be the subject of future work. Key words for inclusion in the literature search were: medication adherence, self-report, validation and risk assessment. References from the development and validation papers of each of the instruments were also useful in locating additional published instruments. Table XVII offers self-reported instruments that have been used to evaluate various aspects of medication adherence. The types of instruments include scales to assess low medication adherence, barriers to taking medicines, treatment satisfaction and risk for non-adherence.

| TABLE XVII – MEDICATION ADHERENCE SELF-REPORT INSTRUMENTS |
|-----------------------------------------------|------------------|------------------|------------------|
| **SURVEY OR SCALE** | **PURPOSE** | **NUMBER OF ITEMS** | **CITATION AND PERMISSIONS** |
| Brief Adherence Rating Scale (BARS) | Assessment of medication adherence of outpatients | 3 questions and visual analogue scale | Byerly M et al. Schizophrenia Research 100 (1):60-69; no special permission |
| Adherence to Refills and Medications Scale (ARMS) | Assessment of taking medicines as prescribed and refilling medicines on schedule | 12 | Kripalani S, et al. Value Health 2009; awaiting permission information |
| Adherence Estimator | Score serves as a low, medium and high risk for non-adherence | 3 | McHorney, CA. Curr Med Res Opin. 2009; 25(1). 215-238. Contact Merck for user agreement |
| Medication Adherence Report Scale (MARS) | Items to determine the willingness to take medications everyday | 3 | Thompson K et al. Schizophrenia Research 2000; 42:241-7. Copyright Robert Horne, 1998 for conditions of use – rhorne@bton.ac.uk |
Use of the instruments

The type of self-report medication instrument selected for a program or study may depend on the desired outcome. Considerations include use in a clinical setting where an immediate response can take place, or use in a pre-post study that might evaluate behavior change. Other instruments may be used in the prediction of adherence and, based on what they measure, will require different interventions (or actions). Last, the length of the survey and the ability to interpret the results have an impact on the selection and use of the desired survey instrument.

Next Steps for Medication Adherence Workgroup

The CCA Medication Adherence Workgroup acknowledges the need for continuing work in the areas of organizational best practice and self-report instruments. Improvements to the current organizational best practice continuum will include research-based evidence to support or change the areas within the model. Further, the workgroup will work toward designing and potentially pilot-testing an “Organizational Medication Adherence Best Practice Scorecard” — a self-assessment tool to assist organizations in medication adherence. CCA will establish a registry for self-report instruments on the organization Web site (www.carecontinuum.org) and update the table of instruments, as needed. Also for 2011, work will begin to address another barrier: first fill non-adherence, which describes when patients appropriately receive a prescription, but fail to fill it, thereby never generating a pharmacy claim record or an opportunity to utilize pharmacy claims to identify and measure non-adherence.

OPERATIONAL MEASURES

Purchasers of chronic care management programs are interested in both clinical and financial outcomes, but often these measures do not show short-term program impact. Both program providers and program purchasers turn to measures of program engagement and program participation as key short-term indicators of a program’s potential for successful outcomes. To evaluate a program’s ability to engage individuals, there first must be an understanding of how the program operates, including its identification, outreach, enrollment and interventions for the diseased population. Establishing operational measures guidelines that reflect each crucial step in the operational process is essential for program evaluation and a goal of this year’s work.

This section of the Guidelines includes two specific recommendations:


2. Definition of Initial Engagement.

CCA has continued to refine the operational flow diagram originally released in Outcomes Guidelines Report, Volume 3, and further in Outcomes Guidelines Report, Volume 4. The Volume 5 Diagram (page 115) maintains its focus on single condition chronic care management, but adds clarity to the process of targeting high-risk participants. CCA recognizes that, based on the population, program goal and other variables, programs might follow a slightly different operational process.
FIGURE 11 – OPERATIONAL MEASURES FLOW DIAGRAM

Total Population

- Ineligible

Identified Population

- Not Identified

Targeted for Intervention
- Not Targeted for Intervention
  - Possible Exclusions: False Positives, Ineligible
- Targeted for Intervention
  - That Do Not Include 1:1 Health Coaching
    - Process to be Determined

Targeted for 1:1 Coaching/Health Education
- Bidirectional Interaction, Risk Assessment, Plan of Care

Not Enrolled
- Enrolled through Opt-In Program
- Enrolled through Opt-Out Program

Unable to Reach
- Declined to Participate
- Referred to Another Level of Intervention

Initially Engaged
- Referred to Another Level of Intervention
- Dormant
- Graduate to Lower Risk
- Ongoing Engagement
The definition of initial engagement below focuses on individuals who typically have the highest risk and/or require one-on-one coaching intervention. This definition is only one of many that will be developed. The workgroup recognizes that there are a variety of modalities that can be used to interact with a program participant and will begin developing definitions that include these modalities in the next year’s work. In addition, further refinement of the flow diagram, recommended measures and focus on participation persistency will be ongoing CCA work efforts for 2011.

**CCA recommends the following definition of initial engagement:**

- The Initially Engaged Population is a subset of enrolled members who are working or have worked directly with a nurse or health coach in a chronic care management or health improvement program within the reporting period.
  - Members are interacting with the health professional in reference to their health improvement plan with “bidirectional interaction” meaning an exchange between the health professional and the member in both directions.
  - A participant is considered initially engaged if she has completed a clinical and lifestyle behavioral risk assessment that includes a mutually agreed upon plan of care with goals and at least one follow-up coaching discussion within 90 days.
  - Only real-time human interaction is included in this definition of initial engagement, regardless of the venue used for discussion.

**ADDITIONAL MEASURES**

As noted previously, chronic care management program purchasers may have multiple goals for introducing a program. This section includes measures that assess aspects of the chronic care management program that fall outside the financial, utilization or clinical domains more frequently included in chronic care management program evaluations.

Additional metrics in this report fall into the following categories:

- Overall health status.
- Satisfaction with the program.

These measures generally require self-reported data collected through mailed, online or telephone-based questionnaires/surveys. In addition, these measures:

- provide information on the participant’s experiences with the program;
- measure changes in participants that may precede and/or predict clinical, utilization and financial changes; and
- provide information that could be used to improve chronic care management programs.
Health Status

CCA recommends use of one of the SF health surveys for program evaluation.

These surveys include the SF-8, SF-12 and SF-36, and are used to measure general mental and physical health status. The SF surveys assess quality of life and have been validated in a large number of settings and populations, including patients with chronic illnesses, such as chronic heart failure and diabetes. The surveys can be easily administered using a variety of media, such as print and online. Among the commonly used are the SF-36 and SF-12, which measure:

- physical functioning;
- role limitations due to physical health (role-physical);
- bodily pain;
- general health perceptions;
- vitality;
- social functioning;
- role limitations due to emotional problems (role-emotional); and
- mental health

Disadvantages include the need to have the survey shielded from interpretation or “help” from people administering the survey; the relative resistance of SF quality of life measures to specific clinical interventions; and, despite availability of copies of the instrument and scoring algorithms in the public domain, the need for some users to license the instrument.

Participant Satisfaction

CCA recommends that participant satisfaction measures be incorporated into the evaluation of chronic care management programs.

Participant satisfaction is a critically important domain of health care. An extensive review of the literature and a survey of population health management found no validated survey for reliably measuring all domains of participant satisfaction in populations with chronic illness who are enrolled in chronic care management programs. As a result, CCA has developed a validated survey to measure a participant’s satisfaction with a variety of program aspects. These aspects include:

- General Module
  - Access to care
  - Coordination of care
  - Improvement in quality of care
  - Ability to self-manage condition
• Segmentation Module
  – Type of chronic care management program
  – Nature of condition
  – Respondent demographics

• Program-Specific Module (optional)
  – Biometric monitoring
  – Other program-specific assessments

There are several versions of the CCA Participant Satisfaction Survey available. More information on the CCA Participant Satisfaction Survey can be found on the CCA Web site, www.carecontinuum.org.
References


6Norman G. Disease management outcomes: are we asking the right questions yet? Population Health Management. 2008;11(4):185-188.


22For example, see Schultz AB and Edington DW. Employee health and presenteeism: a systematic review. Journal of Occupational Rehabilitation. 2007;17(3):547-579.


Appendix

CONDITION-SPECIFIC DENOMINATORS

All Conditions

Identification timeframe (incurred begin/end dates for claims)
- 24 months of incurred claims with three months’ of claim run-out.

Measurement timeframe (incurred begin/end dates for claims)
- Last 12 months of the 24 month identification frame with three months of run-out.

Minimum eligibility (during the measurement timeframe)
- Clinical measures: “HEDIS continuous” - eligible for the entire measurement frame with a single allowable gap of up to 45 days
- Utilization and financial measures
  - Any 6 months during the measurement timeframe, not necessarily contiguous, or HEDIS continuous eligibility. While we tested with the any 6 month criterion to conform to that specified in previous editions of the Guidelines, we recognize that some organizations may wish to use the same denominators (e.g. HEDIS continuous) for all outcomes measures.
  - Reporting organizations are expected to disclose the eligibility criterion used.

Codes used to identify admissions, outpatient and emergency department visits
- Use current-year HEDIS ASM-B table: Codes to Identify Visit Type

Codes used to identify specified medications:
- Use current-year NDC codes from HEDIS tables for medications listed in condition-specific denominator definitions
Denominator Specification for Diabetes

Benefits

All participants must have medical and pharmacy benefits.

Claims – Must have access to the following for all participants:

• Professional claims or encounters
• Facility claims (inpatient and outpatient)
• Pharmacy claims

Codes

• Professional claims/encounters: CPT-4, ICD9
• Facility claims: ICD9, UB revenue codes
• Pharmacy: Use current-year NDC codes from HEDIS tables for medications listed below.

Demographics

• Age is as of the last day of the measurement period.
• 18 years and older

Type and Number of Codes in ID Frame

• 1 OR MORE ACUTE INPATIENT DISCHARGE WITH
  – ICD9 codes in any position: 250.xx, 357.2, 362.0, 366.41, 648.0
• OR 2 OR MORE OFFICE/OUTPATIENT VISITS OR ENCOUNTERS OR EMERGENCY DEPARTMENT VISITS AT LEAST 14 DAYS APART WITH
  – ICD9 codes in any position: 250.xx, 357.2, 362.0, 366.41,648.0
• OR 1 OR MORE MEDICATION DISPENSING EVENTS FOR THE FOLLOWING
  – Alpha-glucosidase inhibitors
  – Anti-diabetic combinations
  – Insulin
  – Meglitinides
  – Miscellaneous antidiabetic agents
  – Sufonylureas
  – Thiazolidinediones
  – (Note: Glucophage/metformin is not included because it is used to treat conditions other than diabetes.)

Condition-Specific Exclusions

• Gestational diabetes: Exclude anyone with 1 or more claims with a code of 648.8.
**Denominator Specification for Heart Failure**

**Benefits**
- All participants must have medical benefits. May need pharmacy benefits for some of the metric numerators but not for denominator criteria described here.

**Claims – Must have access to all of the following for all participants:**
- Professional claims or encounters
- Facility claims (inpatient and outpatient)
- Note: May need pharmacy claims for some of the metric numerators but not for denominator criteria described here.

**Codes**
- Professional claims/encounters: CPT-4, ICD9
- Facility claims: ICD9, UB revenue codes
- Pharmacy: NDC. Note: May need pharmacy claims for some of the metric numerators but not for denominator criteria described here.

**Demographics**
- Age is as of the last day of the measurement period.
- 18 years and older

**Type and Number of Codes in ID Frame**
- 1 OR MORE ACUTE INPATIENT DISCHARGE WITH
  - ICD9 codes in any position: 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 428.0, 428.1, 428.20, 428.21, 428.22, 428.23, 428.30, 428.31, 428.32, 428.33, 428.40, 428.41, 428.42, 428.43, 428.9
- OR 2 OR MORE OFFICE/OUTPATIENT VISITS OR ENCOUNTERS OR EMERGENCY DEPARTMENT VISITS AT LEAST 14 DAYS APART WITH
  - ICD9 codes in any position: 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 428.0, 428.1, 428.20, 428.21, 428.22, 428.23, 428.30, 428.31, 428.32, 428.33, 428.40, 428.41, 428.42, 428.43, 428.9

**Condition-Specific Exclusions**
- None
Denominator Specification for CAD

Benefits
- All participants must have medical and pharmacy benefits.

Claims – Must have access to all of the following for all participants:
- Professional claims or encounters
- Facility claims (inpatient and outpatient)
- Pharmacy claims

Codes
- Professional claims/encounters: CPT-4, ICD9
- Facility claims: ICD9, UB revenue codes, HCPCS
- Pharmacy: Use current-year NDC codes from HEDIS tables for medications listed below.

Demographics
- Age is as of the last day of the measurement period.
- 18 years and older

Type and Number of Codes in ID Frame
- 1 OR MORE ACUTE INPATIENT DISCHARGE WITH
  – ICD9 codes in any position: 410.xx, 411.xx, 412.xx, 413.xx (except 413.1)
- OR 1 OR MORE ENCOUNTERS/CLAIMS WITH PROCEDURE CODE for CAD REVASCULARIZATION PTCA/PCI or CABG
  – Procedure codes:
    – CPT – 33140, 92980-92982, 92984, 92985, 92986, 92995, 92996, 33510-33514, 33516-33519, 33521-33523, 33533-33536, 33572, 35600
    – ICD9 Procedure codes - 00.66, 36.01, 36.02, 36.05, 36.06, 36.07, 36.09, 36.1x, 36.2x
    – HCPCS – S2205, S2206, S2207, S2208, S2209
- OR 2 OR MORE OFFICE/OUTPATIENT VISITS OR ENCOUNTERS OR EMERGENCY DEPARTMENT VISITS AT LEAST 14 DAYS APART WITH
  – ICD9 codes in any position: 410.xx, 411.xx, 412.xx, 413.xx (except 413.1)
- OR 1 OR MORE OFFICE/OUTPATIENT VISITS OR ENCOUNTERS OR EMERGENCY DEPARTMENT VISITS
  – ICD9 codes in any position: 410.xx, 411.xx, 412.xx, 413.xx (except 413.1)

AND 1 OR MORE MEDICATION DISPENSING EVENTS FOR THE FOLLOWING
- Nitrate
- Lipid-lowering drug

Condition-Specific Exclusions
- None
Denominator Specification for Persistent Asthma

Benefits
• All participants must have medical and pharmacy benefits.

Claims – Must have access to the following for all participants:
• Professional claims or encounters
• Facility claims (inpatient and outpatient)
• Pharmacy claims

Codes
• Professional claims/encounters: CPT-4, ICD9
• Facility claims: ICD9, UB revenue codes
• Pharmacy: Use current-year NDC codes from HEDIS tables for medications listed below.

Demographics
• Age is as of the last day of the measurement period.
• 5-56 years and older
• Results should be reported in two age categories:
  – 5-17
  – 18-56

Type and Number of Codes in ID Frame
• 1 OR MORE ACUTE INPATIENT DISCHARGE OR EMERGENCY DEPARTMENT VISITS WITH
  – ICD9 codes as primary diagnosis: 493.xx
• OR 4 OR MORE OFFICE/OUTPATIENT VISITS OR ENCOUNTERS OR EMERGENCY DEPARTMENT VISITS WITH
  – ICD9 code in any position: 493.xx

AND 2 OR MORE ASTHMA MEDICATION DISPENSING EVENTS FOR ANY OF THE FOLLOWING
  – Antiasthmatic combinations
  – Inhaled steroid combinations
  – Inhaled corticosteroids
  – Leukotriene modifiers
  – Long-acting, inhaled beta-2 agonists
  – Mast cell stabilizers
  – Methylxanthines
  – Short-acting, inhaled beta-2 agonists
• OR 4 OR MORE ASTHMA MEDICATION DISPENSING EVENTS FOR ANY OF THE FOLLOWING:
  – Antiasthmatic combinations
  – Inhaled steroid combinations
  – Inhaled corticosteroids
  – Leukotriene modifiers
  – Long-acting, inhaled beta-2 agonists
  – Mast cell stabilizers
  – Methylxanthines
  – Short-acting, inhaled beta-2 agonists
  – If leukotriene modifiers are the only medications used then must have either one of the prior qualifying events or have at least one diagnosis of asthma in any setting in the same ID frame as the leukotriene modifiers.

**Condition-Specific Exclusions**

• COPD as defined in next section
Denominator Specification for COPD

Benefits

- All participants must have medical and pharmacy benefits.

Claims – Must have access to all of the following for all participants:

- Professional claims or encounters
- Facility claims (inpatient and outpatient)
- Pharmacy claims

Codes

- Professional claims/encounters: CPT-4, ICD9
- Facility claims: ICD9, UB revenue codes, HCPCS
- Pharmacy: Use current-year NDC codes from HEDIS tables for medications listed below.

Demographics

- Age is as of the last day of the measurement period.
- 40 years and older

Type and Number of Codes in ID Frame

- 1 OR MORE ACUTE INPATIENT DISCHARGE WITH
  - ICD9 codes as primary diagnosis: 491.xx (excluding 491.0), 492.xx, 496.xx

- 1 OR MORE ENCOUNTERS/CLAIMS FOR LUNG VOLUME REDUCTION SURGERY/SERVICES
  - CPT code: 32491
  - HCPCS codes: G0302-G0305

- OR 2 OR MORE OFFICE/OUTPATIENT VISITS, ENCOUNTERS OR EMERGENCY DEPARTMENT VISITS AT LEAST 14 DAYS APART WITH
  - ICD9 codes in any position: 491.xx, 492.xx, 496.xx

- OR 1 OR MORE OFFICE/OUTPATIENT VISITS, ENCOUNTERS OR EMERGENCY DEPARTMENT VISITS
  - ICD9 codes in any position: 491.xx, 492.xx, 496.xx

AND AT LEAST 1 MEDICATION PRESCRIBING EVENT FOR

- Tiotropium

Condition-Specific Exclusions

- None