



701 Pennsylvania Ave. N.W., Suite 700
Washington, D.C. 20004-2694
(202) 737-5980 • (202) 478-5113 (fax)

dmaa@dmaa.org • www.dmaa.org

May 14, 2010

The Honorable Kathleen Sebelius
Secretary Health and Human Services
Hubert H. Humphrey Building
Room 445-G
200 Independence Avenue, SW
Washington, DC 20201

Attention: DHHS-2010-MLR

Dear Secretary Sebelius:

On behalf of the more than 200 members of DMAA: The Care Continuum Alliance, I respectfully offer the following comments for your consideration in response to the Request for Information on medical loss ratios issued April 14, 2010.

DMAA: The Care Continuum Alliance represents organizations providing services along the continuum of care to more than 160 million Americans through wellness, chronic care management and complex case management. DMAA: The Care Continuum Alliance members include wellness, disease management and population health management organizations; health plans; labor unions; employer organizations; pharmaceutical manufacturers; pharmacy benefit managers; health information technology innovators and device manufacturers; physician groups; hospitals and hospital systems; academicians; and others. These diverse organizations share DMAA: The Care Continuum Alliance's vision of aligning all stakeholders toward improving the health of populations. Our members seek to improve health care quality and contain health care costs by providing targeted interventions and services to individuals who are well, at-risk or managing one or more chronic conditions.

General Comment:

The Patient Protection and Affordable Care Act (PPACA), PL 111-148, directs the Department of Health and Human Services (HHS) to require that health insurance issuers annually report on the percentages of premiums spent on clinical services and activities that improve health care quality. Further, PPACA directs the National Association of Insurance Commissioners (NAIC) to establish uniform definitions for expense categories that comprise the medical loss ratio (MLR) for certification by HHS.

As HHS moves toward certifying NAIC uniform/standard definitions of MLR, DMAA: The Care Continuum Alliance urges the Department to recognize the continuous innovation and evolution of health care delivery systems toward the goals of improved quality and health outcomes and greater efficiency. To that end, HHS should establish a transparent and ongoing process with state regulators and other stakeholders that recognizes the

need for flexibility in classifying services for the calculation of the MLR and that avoids stifling the development, continued evolution and broader adoption of population health improvement initiatives. Specifically, as the Department develops a regulatory definition of "activities that improve quality," DMAA: The Care Continuum Alliance cautions against defining too narrowly the services in this category. The definition should recognize the evolving nature of the delivery and financing of health care and not inadvertently suppress innovation by restricting the category to a limited set of activities.

DMAA: The Care Continuum Alliance has previously defined "quality improvement" as an essential element and fundamental goal of population health management. The definition we adopted states: quality improvements flow from population health management strategies that "support the physician or practitioner/patient relationship and plan of care; emphasize prevention of exacerbations and complications utilizing evidence-based practice guidelines and patient empowerment strategies; and evaluate clinical, humanistic, and economic outcomes on an ongoing basis with the goal of improving overall health."ⁱⁱ We recommend HHS consider this definition and the associated population health management strategies in the development of criteria for inclusion in the category of "activities that improve health care quality."

Wellness, disease and case management services improve and support the health of populations and are important activities that improve quality and result in better health outcomes. These services support an efficient and effective health care delivery system that facilitates the engagement and support of providers and patients to mitigate illness and improve long-term health. Wellness, disease and case management services are built on a foundation of evidence-based clinical care and measured by their clinical impact on health status. Such programs and services educate patients and promote self-management skills and behavior change; provide coaching and nurse support; ensure safe transitions in care; improve medication adherence and management; coordinate care between providers and care settings; and enhance quality through evidence-based decision support, data analytics, disease registries and other technologies. These services are primarily provided by licensed, clinical health care practitioners in and across numerous health care delivery settings and offer benefits far beyond cost containment and claims adjustment activities.

DMAA: The Care Continuum Alliance believes costs associated with these services should be classified as either "medical expenses" or "quality improvement expenses" for the purpose of calculating a health plan's MLR under the requirements of PPACA. A recent paper developed on minimum loss ratios by the American Academy of Actuaries describes "case management, disease management, 24-hour nurse hotlines, wellness programs" as more "akin to benefits than administrative expenses" and appropriately factored into the value of benefits for the calculation of medical loss ratio (American Academy of Actuaries, February 2010).

These services have proved efficient and effective in improving the quality and cost of care, as demonstrated by the continued investment in population health management by commercial insurers. Further, DMAA: The Care Continuum Alliance has worked with numerous external stakeholders to develop measurement methodologies to guide evaluation of population health management programs and services. (*Outcomes Guidelines Report Volume 4*. DMAA: The Care Continuum Alliance, 2009; attached) We recommend that the Department refrain from setting evaluation guidelines and, instead, support the private sector's continued work on measurement design and evaluation of

these services. DMAA: The Care Continuum Alliance is glad to serve as a resource for the Department to share its experience developing scientifically sound, consensus measures for population health management programs.

Specific comments:

(B)(1) What Definitions and Methodologies Do States and Other Entities Currently Require When Calculating MLR-Related Statistics?

Current NAIC guidance regarding the classification of expenses, including quality improvement activities such as wellness, disease and case management, is not binding on state insurance regulators. This has led to significant variation in state legislative and regulatory requirements for MLR calculations. Existing NAIC guidance issued in 2002 – Statement of Statutory Accounting Principles (SSAP) 85 – identifies case management and disease management programs as “cost containment expenses.” NAIC defines “cost containment expenses” as “expenses that actually serve to reduce the number of health services provided or the cost of such services.” Additional NAIC guidance directs “cost containment expenses” to be allocated as “administrative expenses” in the MLR calculation.

DMAA: The Care Continuum Alliance has previously communicated its belief that NAIC SSAP 85 is outdated and fails to appropriately account for and reflect the significant positive impact on quality and health outcomes that wellness, disease and case management programs provide. (See attached letter to Randall Stevenson, NAIC, 13 Sept. 2007.) DMAA: The Care Continuum Alliance has posited that the above-mentioned activities should more appropriately be classified as costs related to clinical care.

(B)(3)(b) What, if any, lists of activities that improve health care quality currently exist? What are the pros and cons associated with including various kinds of activities on these lists (for example disease management and case management)?

DMAA: The Care Continuum Alliance is unaware of existing lists of “activities that improve health care quality.” At a minimum, HHS should be consistent with other statutory references in PPACA with respect to “activities that improve health care quality.” In Section 1311, which establishes Health Benefit Exchanges, PPACA requires exchanges to implement quality improvement strategies, including:

(g) REWARDING QUALITY THROUGH MARKET-BASED INCENTIVES.—

(1) STRATEGY DESCRIBED.—A strategy described in this paragraph is a payment structure that provides increased reimbursement or other incentives for—

(A) improving health outcomes through the implementation of activities that shall include quality reporting, effective case management, care coordination, chronic disease management, medication and care compliance initiatives, including through the use of the medical home model, for treatment or services under the plan or coverage;

(B) the implementation of activities to prevent hospital readmissions through a comprehensive program for hospital discharge that includes patient-centered education and counseling, comprehensive discharge planning, and post discharge reinforcement by an appropriate health care professional;

(C) the implementation of activities to improve patient safety and reduce medical errors through the appropriate use of best clinical practices, evidence based medicine, and health information technology under the plan or coverage;

(D) the implementation of wellness and health promotion activities; and

(E) the implementation of activities to reduce health and health care disparities, including through the use of language services, community outreach, and cultural competency trainings.

(2) GUIDELINES.—The Secretary, in consultation with experts in health care quality and stakeholders, shall develop guidelines concerning the matters described in paragraph (1).

(3) REQUIREMENTS.—The guidelines developed under paragraph (2) shall require the periodic reporting to the applicable Exchange of the activities that a qualified health plan has conducted to implement a strategy described in paragraph (1).

Section 2717 of PPACA, provides additional guidance by establishing quality reporting requirements for all plans with respect to benefits and provider reimbursement structures that:

(A) improve health outcomes through the implementation of activities such as quality reporting, effective case management, care coordination, chronic disease management, and medication and care compliance initiatives, including through the use of the medical homes model as defined for purposes of section 3602 of the Patient Protection and Affordable Care Act, for treatment or services under the plan or coverage;

(B) implement activities to prevent hospital readmissions through a comprehensive program for hospital discharge that includes patient-centered education and counseling, comprehensive discharge planning, and post discharge reinforcement by an appropriate health care professional;

(C) implement activities to improve patient safety and reduce medical errors through the appropriate use of best clinical practices, evidence based medicine, and health information technology under the plan or coverage; and

(D) implement wellness and health promotion activities.

DMAA: The Care Continuum Alliance does not suggest that the MLR definition of “activities that improve health care quality” be limited to the activities referenced in this language, but rather be inclusive of them.

It also is important to note that PPACA establishes a National Strategy for Quality Improvement in Health Care (Section 3011) “to improve the delivery of health care services, patient health outcomes, and population health.” Priorities identified as key to the strategy’s development include many common to population health management programs: an emphasis on improved outcomes and efficiency; enhanced use of health care data; reduction of disparities across populations and geographic areas; awareness of high-cost chronic conditions; research on and dissemination of best practices to improve patient safety and reduce medical errors and preventable admissions.

(B)(3)(c) To what extent do current calculations of medical loss ratios include the amount spent on improving health care quality? Is there any data available relating to how much this amount is?

Section 2718, for the first time, introduces new terminology focused on “activities that improve health care quality” and attempts to ensure a broader commitment to delivering high-quality health care. However, data is not readily available at the state or federal levels with regard to current spending on health care quality activities. The lack of uniform definitions further inhibits efforts to collect such data.

DMAA: The Care Continuum Alliance stands ready to serve as a resource for HHS on population health management, medical claims costs and costs associated with activities that improve health care quality, including those regarding the efficacy or measurement of such services for the purposes of determining appropriate classification as either medical or quality expenses.

Sincerely,

A handwritten signature in black ink, appearing to read "Tracey Moorhead". The signature is written in a cursive, flowing style.

Tracey Moorhead
President and CEO

ⁱ Dictionary of Disease Management Terminology, Second Edition. p. 152. DMAA: The Care Continuum Alliance. 2006.

Line 5 – Improving Health Care Quality Expenses Incurred

Expenses, other than those billed or allocated by a provider for care delivery (i.e., claims costs), that are designed to improve health care quality, reduce medical errors, reduce health disparities, and advance the delivery of patient-centered medical care. The following shown in lines 5.1 and 5.2 are the items that will be considered quality of care expenses if they are designed to improve health care quality, **reduce or mitigate risk factors or onset of disease**, reduce medical errors, reduce health disparities, and advance the delivery of patient-centered medical care.

Exclude: Cost containment expenses that do not directly relate to the quality of health care. These are reported in line 7.1.

Line 5.1 – Type A: Expenses for Health Improvements other than Health Information Technology

Expenses, other than those billed or allocated by a provider for care delivery (i.e. claims costs), that are designed to improve **or maintain the quality of a patient's health**, reduce medical errors, **increase patient compliance with treatment plans and medication adherence, mitigate or reduce the risk factors or onset of disease, to medication or medical treatment plans**, reduce health disparities, and advance the delivery of patient-centered medical care in ways that can be objectively measured and verified. In addition,

The following are items that will **be** included as quality of care expenses meeting these criteria:

1. **Care coordination** (not just general care management) - the active **hands on** participation **(e.g. face-to-face, telephonic or web-based interactions, or other modalities with patients and their providers)** to coordinate a patient's care between multiple providers (such as making sure medical records are shared between all the patient's physicians, making/verifying appointments, and medication compliance) and arranging and managing transitions from one setting to another (such as hospital discharge to home or to a rehabilitation center and prevention of hospital readmissions).
2. **Chronic Disease Management** **Hands on** individually tailored programs for specific chronic conditions that interact with the insured **(e.g. face-to-face, telephonic or web-based interactions, or other modalities with patients and their providers)** to (a) remind insured of doctor appointment, (b) check that insured is following a medically effective prescribed regimen for dealing with the specific disease/condition, (c) incorporating feedback from insured in the management program, (d) provide coaching on dealing with the disease/condition.
3. **Preventive Care and Wellness Programs:** **Hands on p**Programs that interact with the insured **(e.g. face-to-face, telephonic or web-based interactions, or other modalities with patients and their providers)** related to: Wellness assessment, wellness / lifestyle coaching programs, coaching programs designed to educate **individual** members on **managing clinically effective for dealing with a specific** chronic diseases, **and coaching or and** education programs designed to change **individual** members' behavior (e.g. smoking, obesity).
4. **24 Hour Nurse Hotlines:** Expenses for 24 hour nurse hotlines should be included in care coordination, chronic disease management, and preventive care and wellness programs to the extent they meet those expense requirements. Any other expenses for 24 hour nurse hotlines should be

excluded from Improving Health Care Quality Expenses and instead included in Claims Adjustment Expenses.

5. **Other costs** approved by the Secretary, in consultation with the NAIC, which in her discretion, upon an adequate showing that the costs improve the quality of healthcare; the burden shall be on the proponent to show that the costs improve the quality of healthcare.

The following items are broadly excluded as not meeting this criteria:

- Utilization Review
- Fraud Prevention activities
- Any function not expressly included in Type A items 1 through 5, above.

Line 5.2

– Type B: Health Information Technology Expenses Related to Health Improvement

Expenses for Health Information Technology (HIT), consistent with the purposes described in A, above, defined as depreciation on hardware and expenses for software, integrated technologies or related licenses, intellectual property, upgrades, or packaged solutions sold as services that are designed for use by health plans, health care providers, or patients for the electronic creation, maintenance, access, or exchange of health information and the personnel costs associated with implementing those technologies or licenses, but limited to the following expenses;

1. Monitoring or reporting clinical effectiveness;
2. Advancing the ability of providers, insurers or other systems to communicate patient centered clinical or medical information rapidly, accurately and efficiently;
3. Tracking whether a specific class of medical interventions or a bundle of related services leads to better patient outcomes;
4. Costs associated with the transition to ICD-10 that are required by federal law and designed to promote quality and safety.
4. ~~Costs directly related to upgrades in HIT that are required to be made in order to comply with new administrative simplification standards and code sets adopted pursuant to the Health Insurance Portability and Accountability Act (HIPAA), 42 U.S.C. 1320d-2, as amended. (Discuss — Exclude as administrative or include as Fed requirement)~~
5. Other costs approved by the Secretary, in consultation with the NAIC, which in her discretion, upon an adequate showing that the costs improve the quality of healthcare; the burden shall be on the proponent to show that the costs improve the quality of healthcare.

SEC. 1311 – Affordable Choices of Health Benefit Plans

(g) REWARDING QUALITY THROUGH MARKET-BASED INCENTIVES.—

(1) STRATEGY DESCRIBED.—A strategy described in this paragraph is a payment structure that provides increased reimbursement or other incentives for—

(A) improving health outcomes through the implementation of activities that shall include quality reporting, effective case management, care coordination, chronic disease management, medication and care compliance initiatives, including through the use of the medical home model, for treatment or services under the plan or coverage;

(B) the implementation of activities to prevent hospital readmissions through a comprehensive program for hospital discharge that includes patient-centered education and counseling, comprehensive discharge planning, and post discharge reinforcement by an appropriate health care professional;

(C) the implementation of activities to improve patient safety and reduce medical errors through the appropriate use of best clinical practices, evidence based medicine, and health information technology under the plan or coverage;

(D) the implementation of wellness and health promotion activities; and

(E) *oAs added by section 10104(g)*. the implementation of activities to reduce health and health care disparities, including through the use of language services, community outreach, and cultural competency trainings.

(2) GUIDELINES.—The Secretary, in consultation with experts in health care quality and stakeholders, shall develop guidelines concerning the matters described in paragraph (1).

(3) REQUIREMENTS.—The guidelines developed under paragraph (2) shall require the periodic reporting to the applicable Exchange of the activities that a qualified health plan has conducted to implement a strategy described in paragraph (1).

Sec. 2717 - ENSURING THE QUALITY OF CARE.

“(a) QUALITY REPORTING.—

“(1) IN GENERAL.—Not later than 2 years after the date of enactment of the Patient Protection and Affordable Care Act, the Secretary, in consultation with experts in health care quality and stakeholders, shall develop reporting requirements for use by a group health plan, and a health insurance issuer offering group or individual health insurance coverage, with

respect to plan or coverage benefits and health care provider reimbursement structures that—

“(A) improve health outcomes through the implementation of activities such as quality reporting, effective case management, care coordination, chronic disease management, and medication and care compliance initiatives, including through the use of the medical homes model as defined for purposes of section 3602 of the Patient Protection and Affordable Care Act, for treatment or services under the plan or coverage;

“(B) implement activities to prevent hospital readmissions through a comprehensive program for hospital discharge that includes patient-centered education and counseling, comprehensive discharge planning, and post discharge reinforcement by an appropriate health care professional;

“(C) implement activities to improve patient safety and reduce medical errors through the appropriate use of best clinical practices, evidence based medicine, and health information technology under the plan or coverage; and

“(D) implement wellness and health promotion activities.

“(2) REPORTING REQUIREMENTS.—

“(A) IN GENERAL.—A group health plan and a health insurance issuer offering group or individual health insurance coverage shall annually submit to the Secretary, and to enrollees under the plan or coverage, a report on whether the benefits under the plan or coverage satisfy the elements described in subparagraphs (A) through (D) of paragraph (1).

“(B) TIMING OF REPORTS.—A report under subparagraph (A) shall be made available to an enrollee under the plan or coverage during each open enrollment period.

“(C) AVAILABILITY OF REPORTS.—The Secretary shall make reports submitted under subparagraph (A) available to the public through an Internet website.

“(D) PENALTIES.—In developing the reporting requirements under paragraph (1), the Secretary may develop and impose appropriate penalties for non-compliance with such requirements.

“(E) EXCEPTIONS.—In developing the reporting requirements under paragraph (1), the Secretary may provide for exceptions to such requirements for group health plans and health insurance issuers that substantially meet the goals of this section.

“(b) WELLNESS AND PREVENTION PROGRAMS.—For purposes of subsection (a)(1)(D), wellness and health promotion activities may include personalized wellness and prevention services, which are coordinated, maintained or delivered by a health care provider, a wellness and prevention plan manager, or a health, wellness or prevention services organization that conducts health risk assessments or offers ongoing face-to-face, telephonic or web-based intervention efforts for each of the program’s participants, and which may include the following wellness and prevention efforts:

- “(1) Smoking cessation.
- “(2) Weight management.
- “(3) Stress management.
- “(4) Physical fitness.
- “(5) Nutrition.
- “(6) Heart disease prevention.
- “(7) Healthy lifestyle support.
- “(8) Diabetes prevention.

PART 2—NATIONAL STRATEGY TO IMPROVE HEALTH CARE QUALITY

SEC. 3011. NATIONAL STRATEGY.

Title III of the Public Health Service Act (42 U.S.C. 241 et seq.) is amended by adding at the end the following:

“PART 5—HEALTH CARE QUALITY PROGRAMS

“Subpart I—National Strategy for Quality Improvement in Health Care

“SEC. 399HH. NATIONAL STRATEGY FOR QUALITY IMPROVEMENT IN HEALTH CARE.

“(a) ESTABLISHMENT OF NATIONAL STRATEGY AND PRIORITIES.—

“(1) NATIONAL STRATEGY.—The Secretary, through a transparent collaborative process, shall establish a national strategy to improve the delivery of health care services, patient health outcomes, and population health.

“(2) IDENTIFICATION OF PRIORITIES.—

“(A) IN GENERAL.—The Secretary shall identify national priorities for improvement in developing the strategy under paragraph (1).

“(B) REQUIREMENTS.—The Secretary shall ensure that priorities identified under subparagraph (A) will—

“(i) have the greatest potential for improving the health outcomes, efficiency, and patient-centeredness of health care for all populations, including children and vulnerable populations;

“(ii) identify areas in the delivery of health care services that have the potential for rapid improvement

in the quality and efficiency of patient care;
“(iii) address gaps in quality, efficiency, comparative effectiveness information (taking into consideration the limitations set forth in subsections (c) and (d) of section 1182 of the Social Security Act), and health outcomes measures and data aggregation techniques;
oAs revised by section 10302.
“(iv) improve Federal payment policy to emphasize quality and efficiency;

“(v) enhance the use of health care data to improve quality, efficiency, transparency, and outcomes;
“(vi) address the health care provided to patients with high-cost chronic diseases;
“(vii) improve research and dissemination of strategies and best practices to improve patient safety and reduce medical errors, preventable admissions and readmissions, and health care-associated infections;
“(viii) reduce health disparities across health disparity populations (as defined in section 485E) and geographic areas; and
“(ix) address other areas as determined appropriate by the Secretary.

“(C) CONSIDERATIONS.—In identifying priorities under subparagraph (A), the Secretary shall take into consideration the recommendations submitted by the entity with a contract under section 1890(a) of the Social Security Act and other stakeholders.

“(D) COORDINATION WITH STATE AGENCIES.—The Secretary shall collaborate, coordinate, and consult with State agencies responsible for administering the Medicaid program under title XIX of the Social Security Act and the Children’s Health Insurance Program under title XXI of such Act with respect to developing and disseminating strategies, goals, models, and timetables that are consistent with the national priorities identified under subparagraph (A).

“(b) STRATEGIC PLAN.—

“(1) IN GENERAL.—The national strategy shall include a comprehensive strategic plan to achieve the priorities described in subsection (a).

“(2) REQUIREMENTS.—The strategic plan shall include provisions for addressing, at a minimum, the following:

“(A) Coordination among agencies within the Department,

which shall include steps to minimize duplication of efforts and utilization of common quality measures, where available. Such common quality measures shall be measures identified by the Secretary under section 1139A or 1139B of the Social Security Act or endorsed under section 1890 of such Act.

“(B) Agency-specific strategic plans to achieve national priorities.

“(C) Establishment of annual benchmarks for each relevant agency to achieve national priorities.

“(D) A process for regular reporting by the agencies to the Secretary on the implementation of the strategic plan.

“(E) Strategies to align public and private payers with regard to quality and patient safety efforts.

“(F) Incorporating quality improvement and measurement in the strategic plan for health information technology required by the American Recovery and Reinvestment Act of 2009 (Public Law 111–5).

“(c) PERIODIC UPDATE OF NATIONAL STRATEGY.—The Secretary shall update the national strategy not less than annually. Any such update shall include a review of short- and long-term goals.

“(d) SUBMISSION AND AVAILABILITY OF NATIONAL STRATEGY AND UPDATES.—

“(1) DEADLINE FOR INITIAL SUBMISSION OF NATIONAL STRATEGY.—Not later than January 1, 2011, the Secretary shall submit to the relevant committees of Congress the national strategy described in subsection (a).

“(2) UPDATES.—

“(A) IN GENERAL.—The Secretary shall submit to the relevant committees of Congress an annual update to the strategy described in paragraph (1).

“(B) INFORMATION SUBMITTED.—Each update submitted under subparagraph (A) shall include—

“(i) a review of the short- and long-term goals of the national strategy and any gaps in such strategy;

“(ii) an analysis of the progress, or lack of progress, in meeting such goals and any barriers to such progress;

“(iii) the information reported under section 1139A of the Social Security Act, consistent with the reporting requirements of such section; and

“(iv) in the case of an update required to be submitted

on or after January 1, 2014, the information reported under section 1139B(b)(4) of the Social Security Act, consistent with the reporting requirements of such section.

“(C) SATISFACTION OF OTHER REPORTING REQUIREMENTS.—

Compliance with the requirements of clauses (iii) and (iv) of subparagraph (B) shall satisfy the reporting requirements under sections 1139A(a)(6) and 1139B(b)(4), respectively, of the Social Security Act.

“(e) HEALTH CARE QUALITY INTERNET WEBSITE.—Not later than January 1, 2011, the Secretary shall create an Internet website to make public information regarding—

“(1) the national priorities for health care quality improvement established under subsection (a)(2);

“(2) the agency-specific strategic plans for health care quality described in subsection (b)(2)(B); and

“(3) other information,

Outcomes Guidelines Report



Outcomes Guidelines Report Volume 4

DMAA: The Care Continuum Alliance
701 Pennsylvania Ave. N.W., Suite 700
Washington, D.C. 20004-2694
(202) 737-5980
(202) 478-5113 (fax)
dmaa@dmaa.org
www.dmaa.org

Contents

Click on a heading to jump directly to the item

| | |
|------------------------------------------------------------------|-----------|
| FOREWORD | 2 |
| ACKNOWLEDGMENTS | 3 |
| OVERVIEW | 7 |
| INDUSTRY UPDATE | 7 |
| OUTCOMES INITIATIVE UPDATE | 10 |
| GUIDELINE INDEX | 11 |
| POPULATION HEALTH MANAGEMENT - EXPANDED METHODOLOGY | 13 |
| WELLNESS | 19 |
| METHODOLOGICAL CONSIDERATIONS | 21 |
| MEASURE SETS | 32 |
| FIVE CORE CHRONICS | 36 |
| METHODOLOGICAL CONSIDERATIONS | 36 |
| MEASURE SETS | 79 |
| REFERENCES | 98 |

Tables, Figures and Charts

Click on a heading to jump directly to the item

| | |
|--------------------------------------------------------------------------------------|----|
| Table I, DMAA Guidelines Locator | 11 |
| Table II, PHM Capabilities | 14 |
| Table III, Comparison of Current State of Disease Management/Wellness Programs | 20 |
| Table IV, Control/Comparison Groups Methods for Wellness Program Evaluation | 27 |
| Table V, Study Design | 38 |
| Table VI, Methods to Define Outliers | 43 |
| Table VII, Denominator Specification for Diabetes | 54 |
| Table VIII, Denominator Specification for Heart Failure | 55 |
| Table IX, Denominator Specification for CAD | 56 |
| Table X, Denominator Specification for Persistent Asthma | 57 |
| Table XI, Denominator Specification for COPD | 58 |
| Table XII, DMAA Trend Recommendation Results | 60 |
| Table XIII, Risk Adjustment Examples | 62 |
| Table XIV, Small Populations, Part I | 71 |
| Table XV, Comparison of Various Utilization Measures | 83 |

Figures

| | |
|-----------------------------------------------------------|----|
| Figure 1, PHM Program Framework | 15 |
| Figure 2, Impacts Model | 29 |
| Figure 3, Predictive Risk Score Calculation | 67 |
| Figure 4, Program Impact Over Time | 74 |
| Figure 5, Achieving Optimal Balance in Study Design | 79 |
| Figure 6, Diabetes Persistence Curve | 91 |
| Figure 7, Operational Measures Flow Diagram | 95 |

Charts

| | |
|-------------------------------------------------------------|----|
| Chart I, Prevalence Test Results | 48 |
| Chart II, True Positive Test Results | 49 |
| Chart III, True Negative Test Results | 50 |
| Chart IV, DMAA vs. Proprietary Criteria, Venn Diagram | 51 |
| Chart V, DMAA vs. Proprietary Criteria, Bar Chart | 52 |
| Chart VI, Small Populations, Part II | 71 |
| Chart VII, Diabetes Example | 91 |

Foreword

Now in its fourth year, the DMAA Outcomes Guidelines process continues to lead population health management in the standardization of evaluation methodology for programs that serve the needs of all populations along the health care continuum. Regardless of the ultimate outcome of health care system reform, the need for standardized program evaluation methodology will continue and this work has come a long way toward meeting this need.

The Outcomes Guidelines process reflects consensus among more than 130 experts from a majority of the stakeholders in chronic care management, wellness and population health management program delivery in the United States and abroad. DMAA's robust and diverse membership, combined with a transparent, consensus-driven process, has fostered this initiative's continued success. Collaborations with organizations seen as experts in this area, including the National Committee for Quality Assurance and URAC, have only added to the success and credibility of the Outcomes Guidelines process. This year is no exception.

This fourth volume of the DMAA Outcomes Guidelines Report combines the work completed and approved this year with recommendations published in the first three volumes. In addition, this year the Volume 4 Outcomes Guidelines Report will be available for download through the DMAA Web site (www.dmaa.org) at no charge. This decision demonstrates our commitment to and necessity of an evaluation methodology that is easily accessible to all health care stakeholders.

The consensus-driven process used to develop these recommendations can succeed only through successful leadership and volunteer support and dedication. Steering Committee co-chairs Sue Jennings, PhD, and Donald Fetterolf, MD, MBA, FACP, have provided the leadership and direction necessary to continue the work and to move forward and expand to areas that coincide with market trends. The DMAA Board of Directors continues to play an active role in recommendation review and approval. The acknowledgments listed on the next several pages include members of the DMAA Outcomes Steering Committee, as well as workgroup members, workgroup leaders and contributing authors. The continued success of this project is due to the dedication and support of these volunteers. To all DMAA volunteers: Thank you for your time and effort to ensure the success of this important work.

Tracey Moorhead
President and CEO

DMAA: The Care Continuum Alliance

Jeanette C. May, PhD, MPH
Vice President, Research & Quality

DMAA: The Care Continuum Alliance

Acknowledgments

Outcomes Steering Committee Members

Committee Co-Chairs:

Donald Fetterolf, MD, MBA, FACP, Fetterolf Healthcare Consulting
Sue Jennings, PhD, Independent Health Care Consultant

Committee Members:

Tariq Abu-Jaber, WellPoint Inc.
Osman Ahmed, OptumHealth
Tamim Ahmed, PhD, OptumHealth
Andy Amster, Kaiser Permanente
Mark Anderson, BS, SHPS
Janice Anderson, URAC
Rajni Aneja, MD, OptumHealth
David Anstatt, MPH, MBA, Capital Blue Cross
William Appelgate, PhD, Iowa Chronic Care Consortium
Avery Ashby, Blue Cross Blue Shield of Kansas City
Eric Bargman, PhD, Blue Cross Blue Shield of Minnesota
Chris Behling, Hooper Holmes Health & Wellness
Steve Bennett, MS, HealthMedia Inc.
Gregory Berg, PhD, McKesson Health Solutions
Christian Birkmeyer, MS, Solucia Consulting
Aree Bly, FSA, MAAA, Actuary, Milliman, Inc.
Rebecca Booth, MA, AHMP, HMC
Brenda Briden, RN, Blue Cross Blue Shield Rhode Island
Pawel Brocki, ArztPartner Alameda AG
David Brumley, MD, MBA, Blue Cross Blue Shield Massachusetts
Griselda Chapa, MPH, MS, HMC
Daniel Checkman, Blue Cross Blue Shield of Louisiana
Chesley Chen, Health Dialog Inc.
Jean Chen, PhD, University of Michigan
Jen Clair, MPH, HMC
Michael Connor, DrPH, Alere
Greg Conrad, Hooper Holmes
Jason Cooper, MS, CIGNA HealthCare and CareAllies
Carrie Cowdin, MBA, Blue Cross Blue Shield Kansas City
Cindy Deming, LifeMasters Support SelfCare
Christopher Dezii, RN, MBA, CPHQ, Bristol-Myers Squibb
Matthew Diment, HealthMedia Inc.
Ian Duncan, FSA, FIA, FCIA, Solucia Consulting
D.W. Edington, PhD, University of Michigan
Richard Feifer, MD, MPH, FACP, Medco Health Solutions Inc.
Leonard Fensterheim, MPH, APS Healthcare Inc.
Helene Forte, RN, MS, PAHM, Aetna
Brian Foy, Blue Cross Blue Shield of Illinois
Sharon Glave Frazee, PhD, Take Care Health Systems
Feride Frech-Tamas, Novartis
R. Allen Frommelt, PhD, Nurtur Health Inc.
Joanne Garcia, RN, MHSM, FACHE, Accordant,
A CVS Caremark Company

Nancy Garrett, Blue Cross Blue Shield of Minnesota
Patricia Gladowski, RN, MSN, Highmark Blue Cross
Daniel Gold, PhD, Mercer Health and Benefits Consulting LLC
Frederic Goldstein, U.S. Preventative Medicine
Neil Gordon, MD, PhD, MPH, Nationwide Better Health
Raena Grant, National Committee for Quality Assurance
Sandi Greenawalt, RN, URAC
Janet Greenhut, MD, MPH, HealthMedia Inc.
Stephen Griffiths, PhD, MS, OptumHealth
Virginia Gurley, MD, MPH, Healthways Inc.
Benjamin Hamlin, MPH, National Committee for
Quality Assurance
Pamela Hara, RN, MBA, Healthways
Ron Harris, MD, FACE, Geisinger Health Plan
Cathy Hartman, Health Dialog
Michelle Hawkins, RN, MSN, MBA, CCM, Johns Hopkins HealthCare
Debra Hicks, Aetna
Laura Higginbotham, RN, MS, Accordant,
A CVS Caremark Company
Cindy Hochart, RN, MBA, Blue Cross Blue Shield Kansas City
Sarah Hoerauf, HealthMedia, Inc.
Lisa Hu, Caremark Inc.
Jeryl Janson, RN, MPH, Principal Financial Group
Herman Jenich, MPP, HMC
Anju Joglekar, PhD, Verisk Health, Inc.
Patricia Jones, RN, MBA, Regence BlueShield
Iver Juster, MD, ActiveHealth Management Inc.
Anthony Kalams, Merck
Vicki Karlan, MPH, CHES, Pfizer Health Solutions Inc.
Yuki Kawabata, MPhil, Sompo Japan Research Institute Inc.
Patricia Kirkpatrick, RN, BS, CPHQ, LifeMasters Supported
SelfCare, Inc.
Thomas Knabel, MD, Ingenix Inc.
Mike Kornhauser, MD, Alere
Mark Kosinski, QualityMetric Inc.
Thomas Kotsos, HMC
Stefan Kottmair, MD, Alameda GmbH
Andrew Krueger, MD, Accordant, A CVS Caremark Company
Megan LaRoche, Kaiser Permanente
Katrina Lane, RN, BSN, Accordant, A CVS Caremark Company
Jeanne Lehn, RN, MSN, Anthem
Erik Lesneski, AllOne Health Management Solutions
Melodi Licht, RN, MSN, Avivia Health from Kaiser Permanente
Eric Lloyd, Health Integrated Inc.
Katrina Lubbers, MA, CHES, CHD Meridian Healthcare,
An I-trax Company
Rose Maljanian, RN, MBA, Strategic Health Equations, LLC

Acknowledgments cont.

Jeremy Margolis, PhD, Accordant, A CVS Caremark Company
Pam Mays, CIGNA HealthCare and CareAllies
Patricia McDermott, RN, Aetna Health Analytics
Mary Jane McKendry, KidneyTel
Kristen McKenna, ActiveHealth Management
Stephen McMurray, MD, Fresenius Medical Care – North America
Syed Mehmud, Milliman Inc.
R.Douglas Metz, MD, American Specialty Health
Karen Michaeli, MSW, Pharos Innovations
Adrienne Mims, MD, MPH, American Specialty Health
Elaine Mischler, MD, Mischler Consultants LLC
Ameeta Mistry, MBA, CPA, LifeMasters Supported SelfCare Inc.
Brenda Motheral, BPharm, MBA, PhD, CareScientific LLC
Carolyn Mustille, RN, MSN, The Permanente Federation
Michael Mustille, MD, Kaiser Permanente
Charles Nagurka, OptumHealth
Craig Nelson, MS, DC, American Specialty Health
Kejian Niu, MS, HMC
Gordon Norman, MD, MBA, Alere
Mark Norton, MPA, PAHM, HMC
Samuel Nussbaum, MD, WellPoint Inc.
Eric Olmsted, PhD, Health Dialog
Steven Ortner, Hooper Holmes
Brian O’Shields, FSA, MAAA, Regence Blue Shield
William Pace, PhD, Aetna
Indira Paharia, MBA, Regence BlueCross BlueShield of Oregon
Urvashi Patel, PhD, Care Management Technologies
Carol Peckham, RD, CDE, Nurtur Health Inc.
Andy Perez, MD, Healthways Inc.
Janet Pernarelli, Reden & Anders Ltd.
Gary Persinger, National Pharmaceutical Council
Michele Peterson, Blue Cross Blue Shield of Illinois
Ruth Phillip, ArztPartner Alameda AG
Jennifer Pitts, PhD, Independent Healthcare Consultant
David Plocher, MD, Blue Cross Blue Shield Minnesota
James Pope, MD, FACC, Healthways Inc.
Julie Pung, Blue Cross Blue Shield Association
Robin Rager, PhD, Optimum Health Management
Jean Range, MS, RN, CPHQ, Joint Commission
Phil Renner, MBA, National Committee for Quality Assurance
Joann Richards, PhD, RN, URAC
Lori Richards, RN, CCM, CPCE, BS, Capital Blue Cross
Shirley Roberts, Iowa Chronic Care Consortium
Tina Ross-Knapp, CCP, Capital Blue Cross
Annette Ruby, RN, SummaCare
Elizabeth Rula, PhD, Healthways
Robert Sack, MD, APS Healthcare Inc.
Tehseen Salimi, MD, MHA, sanofi-aventis
Jay Salvio, CareGuide
Joe San Filippo, Nationwide Better Health
Karen Sanderson, Capital Blue Cross
Tony Sarthou, Alere
Stephen Saunders, MD, MPH, APS Healthcare Inc.
Emily Schifrin, MS, URAC
Steven Schwartz, HealthMedia Inc.
Christobel Selecky, LifeMasters Supported SelfCare Inc.
Cary Sennett, MD, PhD, MedAssurant, Inc.
Seth Serxner, PhD, MPH, Mercer Health and Benefits Consulting LLC
Dwight Seward, OptumHealth
Tatiana Shnaiden, MD, ActiveHealth Management Inc.
Kelly Shreve, Capital Blue Cross
Dexter Shurney, MD, MBA, MPH, Vanderbilt Medical Center
Jaan Sidorov, MD, FACP, CMCE, Sidorov Health Solutions
Ivan Sison, Caremark Inc.
Elizabeth Skinner, Glaxo SmithKline
Julie Slezak, MS, Caremark Inc.
Alison Soucy, BS, Accordant, A CVS Caremark Company
Vic Spain, Merck & Co., Inc.
Claire Spettell, PhD, Aetna
Suzanne Stearns, United Healthcare
Steele Stewart, FSA, MAAA, Blue Cross Blue Shield Kansas City
Matt Stiefel, MPA, Kaiser Permanente Health Plan, Inc.
Scott Stratton, Medco Health Solutions
Teresa Szczepanski, Nationwide Better Health
Nina Taggart, MD, MBA, AllOne Health Management Solutions
Michael Taitel, PhD, Alere
Pei Tang, MS, MA, Alere
Earl Thompson, Health Fitness
Janet L. Thomson, Aetna
Chris Tourville, CIGNA HealthCare and CareAllies
Julie Trocchio, RN, MS, Catholic Health Association
Travis Tucker, Alere
Wendy Turenne, MS, Alere
Howard Underwood, MD, FSA, MAAA, Deloitte Consulting LLP
Mark Ungvarsky, MBA, BA, AllOne Health Management Solutions
Troy Vander Pas, APS Healthcare Inc.
David Veroff, MPP, Health Dialog Inc.
Paul Wallace, MD, Kaiser Permanente
Sheng Wang, Caremark, Inc.
Chunmei Wang, MS, Humana Inc.
Daryl Wansink, Blue Cross and Blue Shield of North Carolina
Ann Watt, MBA, RHIA, The Joint Commission
Susan Weber, RN, CCM, MHP, OptumHealth
Allan Weiland, Healthcare Management Solutions - Kaiser
David Wennberg, MD, MPH, Health Dialog Inc.
Steve Wigginton, Health Integrated Inc.
Kevin Wildenhaus, PhD, HealthMedia, Inc.
Amy Wilson, Blue Cross Blue Shield of Minnesota
Steven Wright, ASA, MAAA, LifeMasters Supported SelfCare, Inc.
Richard Zabinski, OptumHealth
Devin Zatorski, Blue Cross Blue Shield Association
Maria Ziolkowski, MBA, Aetna

Workgroups

EXPANDED METHODS

Leader, Jennifer Pitts, PhD, Independent Health Care Consultant*

Leader, David Veroff, MPP, Health Dialog Inc.*

Osman Ahmed, OptumHealth

Gregory Berg, PhD, McKesson Health Solutions

David Brumley, MD, MBA, Blue Cross Blue Shield Massachusetts

Daniel Checkman, Blue Cross Blue Shield of Louisiana

Greg Conrad, Hooper Holmes

Helene Forte, RN, MS, PAHM, Aetna

Sharon Glave Frazee, PhD, Take Care Health Systems

Daniel Gold, PhD, Mercer Health and Benefits Consulting LLC

Raena Grant, National Committee for Quality Assurance

Sue Jennings, PhD, Independent Health Care Consultant

Patricia Kirkpatrick, RN, BS, CPHQ, LifeMasters Supported SelfCare, Inc.

Andrew Krueger, MD, Accordant, A CVS Caremark Company

Jeremy Margolis, PhD, Accordant, A CVS Caremark Company

Elaine Mischler, MD, Mischler Consultants LLC

Shirley Roberts, Iowa Chronic Care Consortium

Annette Ruby, RN, SummaCare

Elizabeth Rula, PhD, Healthways Inc.

Steven Schwartz, HealthMedia Inc.

Cary Sennett, MD, PhD, MedAssurant, Inc.

Seth Serxner, PhD, MPH, Mercer Health and Benefits Consulting LLC

Dwight Seward, OptumHealth

Tatiana Shnaiden, MD, ActiveHealth Management Inc.

Alison Soucy, BS, Accordant, A CVS Caremark Company

Suzanne Stearns, AmeriChoice

Pei Tang, MS, MA, Alere

Janet L. Thomson, Aetna

Steven Wright, ASA, MAAA, LifeMasters Supported SelfCare, Inc.

WELLNESS

Leader, Craig Nelson, MS, DC, American Specialty Health*

Leader, Kejian Niu, MS, HMC*

David Anstatt, MPH, MBA, Capital Blue Cross

Donald Fetterolf, MD, MBA, FACP, Fetterolf Healthcare Consulting

Helene Forte, RN, MS, PAHM, Aetna

R. Allen Frommelt, PhD, Nurtur Health Inc.

Joanne Garcia, RN, MHSM, FACHE, Accordant, A CVS Caremark Company

Frederic Goldstein, U.S. Preventative Medicine

Benjamin Hamlin, MPH, National Committee for Quality Assurance

Cindy Hochart, RN, MBA, Blue Cross Blue Shield Kansas City

Sarah Hoerauf, HealthMedia, Inc.

Sue Jennings, PhD, Independent Health Care Consultant

Thomas Kotsos, HMC

Charles Nagurka, OptumHealth

Eric Olmsted, PhD, Health Dialog

Steven Ortner, Hooper Holmes

Indira Paharia, MBA, Regence BlueCross BlueShield of Oregon

Jennifer Pitts, PhD, Independent Health Care Consultant

Robin Rager, PhD, Optimum Health Management

Emily Schiffrin, MS, URAC

Kelly Shreve, Capital Blue Cross

Michael Taitel, PhD, Alere

Earl Thompson, Health Fitness

Amy Wilson, Blue Cross Blue Shield of Minnesota

*Contributing author

MEDICATION ADHERENCE

Leader, Jason Cooper, MS, CIGNA HealthCare and CareAllies*

Janice Anderson, URAC

Richard Feifer, MD, MPH, FACP, Medco Health Solutions *

Sandi Greenawalt, RN, URAC

Benjamin Hamlin, MPH, National Committee for Quality Assurance

Ron Harris, MD, FACE, Geisinger Health Plan

Cindy Hochart, RN, MBA, Blue Cross Blue Shield Kansas City

Sue Jennings, PhD, Independent Health Care Consultant

Anthony Kalams, Merck

Jeanne Lehn, RN, MSN, HMC

Jeremy Margolis, PhD, Accordant, A CVS Caremark Company

Urvashi Patel, PhD, Care Management Technologies

Gary Persinger, National Pharmaceutical Council

Lori Richards, RN, CCM, CPCE, BS, Capital Blue Cross

Elizabeth Skinner, Glaxo SmithKline

Kevin Wildenhaus, PhD, HealthMedia, Inc.

Richard Zabinski, OptumHealth

OPERATIONAL MEASURES

Leader, Susan Weber, RN, CCM, MHP, OptumHealth*

Leader, Cindy Hochart, RN, MBA, Blue Cross Blue Shield

Kansas City *

Matthew Diment, Health Media

Joanne Garcia, RN, MHSM, FACHE, Accordant, A CVS Caremark Company

Nancy Garrett, Blue Cross Blue Shield of Minnesota

Daniel Gold, PhD, Mercer Health and Benefits Consulting

Benjamin Hamlin, MPH, National Committee for Quality Assurance

Debra Hicks, Aetna

Katrina Lane, RN, BSN, Accordant, A CVS Caremark Company

Ameeta Mistry, MBA, CPA, LifeMasters Supported SelfCare Inc

Michele Peterson, Blue Cross Blue Shield of Illinois

Karen Sanderson, Capital Blue Cross

Seth Serxner, PhD, MPH, Mercer Health and Benefits Consulting LLC

Travis Tucker, Alere

Cindy Deming, LifeMasters Support SelfCare, Inc.

SELECTION CRITERIA

Leader, Iver Juster, MD, ActiveHealth Management Inc.*

Leader, Steele Stewart, FSA, MAAA, Blue Cross Blue Shield Kansas City*

Christian Birkmeyer, MS, Solucia Consulting

Jason Cooper, MS, CIGNA HealthCare and CareAllies

Benjamin Hamlin, MPH, National Committee for Quality Assurance

Sue Jennings, PhD, Independent Healthcare Consultant

Kristen McKenna, ActiveHealth Management

Kejian Niu, MS, HMC

Ivan Sison, Caremark Inc.

Pei Tang, MS, MA, Alere

Sheng Wang, Caremark, Inc.

LONG TERM EVALUATION

Leader, Mark Norton, MPA, PAHM, Health Management Corporation

Donald Fetterolf, MD, MBA, FACP, Fetterolf Healthcare Consulting*

Thomas Knabel, MD, Ingenix Inc. *

Steven Wright, ASA, MAAA, LifeMasters Supported SelfCare, Inc.

*Contributing author

Population Health Industry Overview

Health care reform seems to be on everyone's mind these days. The complexities of our health care system, coupled with the current economic climate, present challenges to all stakeholders on both the supply and demand sides of health care. Regardless of these specific complexities and challenges, the pressing need for high-quality, cost-effective health care continues – in fact, it becomes even more significant with efforts to extend coverage to those who now lack it. The ultimate goal of improving health status for all Americans must not be lost in the sometimes heated debate about how to reform the funding and delivery of health care services. This requires changing our “sick care” orientation to one that values and supports achieving better health in a cost-effective manner.

Our population health industry continues to meet the needs of diverse populations through partnerships and collaborations focused on helping individuals achieve 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 range and scope of health improvement programs, from wellness and prevention, to chronic care support, to end-of-life care management. Those who arrange and pay for health care benefits and services have never had more options for sourcing and designing flexible programs to support prevention, healthful lifestyles, chronic care support, etc. Increasingly, many are seeking more-integrated programs that consolidate outreach, communication, intervention, transitions, reporting and assessment across a wide portion of the 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 no single plug-and-play model for all settings exists. Rather, flexible models developed and aligned with the needs of the target population and existing services have enjoyed success. What professionals in the “medical home” provide and what is available via other modalities through partnerships to create a “virtual medical home” has yet to be tested, evaluated and improved, but this work is underway now. 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 technology 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 their 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 due to its expertise and experience in key areas of service delivery and outcomes improvement.

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 or diagnostic devices or smartphone apps – will connect to the care system to populate providers' and care managers' electronic health records (EHRs) and individuals' own personal health records 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 allow 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 standalone 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.

Participant Engagement, Incentives, Personalization

As DMAA reported in 2008, in its “DMAA Disease Management and Wellness: Results of a Market Research Survey,” the top concern of many purchasers of population health services remains improving rates of enrollment, engagement and sustained 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: The “VOI” perspective

A perennial issue for population health management is that of quantifying the “value of investment” of its activities. DMAA: The Care Continuum Alliance has worked for several years with thought leaders to define consensus guidelines and best practices for evaluating program outcomes. Much has been accomplished, yet much remains to be done to satisfy the diverse needs of sophisticated researchers, policymakers and business decision makers. Just as no one plug-and-play model of health improvement works best in all settings, no single measure or method of value assessment is appropriate to all circumstances. Examining and detailing the advantages and disadvantages of different approaches continues to be a significant part of this work.

DMAA is pleased to offer this year’s contribution to the ongoing outcomes challenge in the form of a fourth volume of the Outcomes Guidelines Report. DMAA welcomes feedback and further work into next year and beyond.

Outcomes Initiative Update

DMAA 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 fourth 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, disease management and population health management programs. Currently, there are more than 130 volunteers dedicating time to this effort. These volunteers represent all areas of population health management, from both the provider and purchaser communities.

The recommendations developed since the 2006 release of Volume 1 follow these general guiding principles:

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; and
- An intermediate step in evolving practical and reliable methods to facilitate 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.

Initially, efforts focused on evaluation recommendations for programs targeting the five most common chronic conditions: diabetes, asthma, heart failure, coronary artery disease and chronic obstructive pulmonary disease. The process in 2007 began to examine wellness program evaluation, and those efforts continued this year and have expanded to programs that move beyond a single condition toward a population health management-type approach.

This year's Volume 4 report includes the recommendations from Volumes 1, 2 and 3, in addition to the recommendations developed and approved this year. The topics addressed in the guidelines continue to be refined as new research becomes available. The guidelines index on pages 11 and 12 identifies the development stage of each of the recommendations included in this report.

TABLE I – DMAA GUIDELINES LOCATOR

| POPULATION HEALTH MANAGEMENT | | | | |
|--------------------------------------------|--------|-----------|--------------|----------|
| GUIDELINE | PAGE # | BEGINNING | INTERMEDIATE | ADVANCED |
| Methodological Considerations | | | | |
| Definition of Population Health Management | 13 | X | | |
| General Guiding Principles | 13 | X | | |
| PHM Capabilities | 14 | X | | |
| PHM Program Framework | 15 | | X | |
| Program Evaluation Core Measures | 18 | X | | |
| WELLNESS | | | | |
| GUIDELINE | PAGE # | BEGINNING | INTERMEDIATE | ADVANCED |
| Methodological Considerations | | | | |
| Wellness Program Definition | 19 | | | X |
| Program Comparison | 20 | | X | |
| Methodological Considerations | 21 | X | | |
| Use of Control/Comparison Group | 25 | | | X |
| Use of Claims Data | 28 | | | X |
| Model of Program Impacts | 30 | | | X |
| Measure Sets | | | | |
| Primary Outcomes Measures | 32 | | | |
| Timeline | 32 | | | |
| Other Outcomes Measures | 33 | X | X | |
| • Quality of Life | | | | |
| • Presenteeism/Absenteeism | | | | |
| Productivity Measure Recommendation | 33 | | | |

TABLE I – DMAA GUIDELINES LOCATOR

| DISEASE MANAGEMENT/FIVE CORE CHRONICS | | | | |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|-----------|--------------|----------|
| GUIDELINE | PAGE # | BEGINNING | INTERMEDIATE | ADVANCED |
| Methodological Considerations | | | | |
| Evaluation Design | 36 | | | X |
| Identification Methodology | 39 | | | X |
| Defining the Population | 41 | | | X |
| <ul style="list-style-type: none"> • Inclusion Criteria • Look-Back Period • Claims Run Out • Measurement Period • Member Month | | | | |
| Outliers/Exclusions | 42 | | | X |
| Selection Criteria | 44 | | X | |
| Trend | 59 | | | X |
| Risk Adjustment | 61 | | X | |
| Small Populations | 72 | | X | |
| Long-Term Evaluation Key Concepts | 74 | X | | |
| Measure Sets | | | | |
| Financial | 80 | | | X |
| Utilization | 81 | | | X |
| Clinical | | | | |
| <ul style="list-style-type: none"> • Group 1 | 84 | | | X |
| <ul style="list-style-type: none"> • Group 2 | 85 | | | |
| <ul style="list-style-type: none"> • Self Management | 86 | X | | |
| <ul style="list-style-type: none"> • Medication Adherence | 88 | | X | |
| <ul style="list-style-type: none"> • MPR (Refined as of 2009) | | | | X |
| <ul style="list-style-type: none"> • Measure of Persistence | | | | X |
| Operational Measures: Operational Flow Diagram (Refined 2009) | 95 | | X | |
| Health Status | 96 | | | |
| Quality Measure - Satisfaction | 97 | | | X |

Population Health Management - Expanded Methodology

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 disease 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 DMAA 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.

In addition to the guiding principles in Volume 3, this new report defines population health management and the essential ingredients of a PHM program. A program framework for understanding how population health management drives outcomes and other impacts is also included. Last, these guidelines provide a set of core measures upon which every population health management program should focus. These measures are intended to be used across populations along the health care continuum. DMAA recognizes that additional measures will be necessary to address targeted program elements delivered under the PHM umbrella. An example of this could be clinical measures specific to a subset of program participants who received education for management of their diabetes.

DMAA will continue work in population health management in 2010 with the goal of developing a flexible evaluation methodology that could be used to evaluate these programs regardless of the mix of services delivered to the population. In addition, DMAA will begin to align its work in wellness and disease management underneath the PHM strategy.

DMAA 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 and/or improve the physical and psychosocial well-being of individuals through cost-effective and tailored health solutions.

DMAA 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 this.

GUIDING PRINCIPLE II: METHODOLOGY TRANSPARENCY

DMAA 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

DMAA 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.^{1,2}

GUIDING PRINCIPLE IV: USE OF UTILIZATION MEASURES

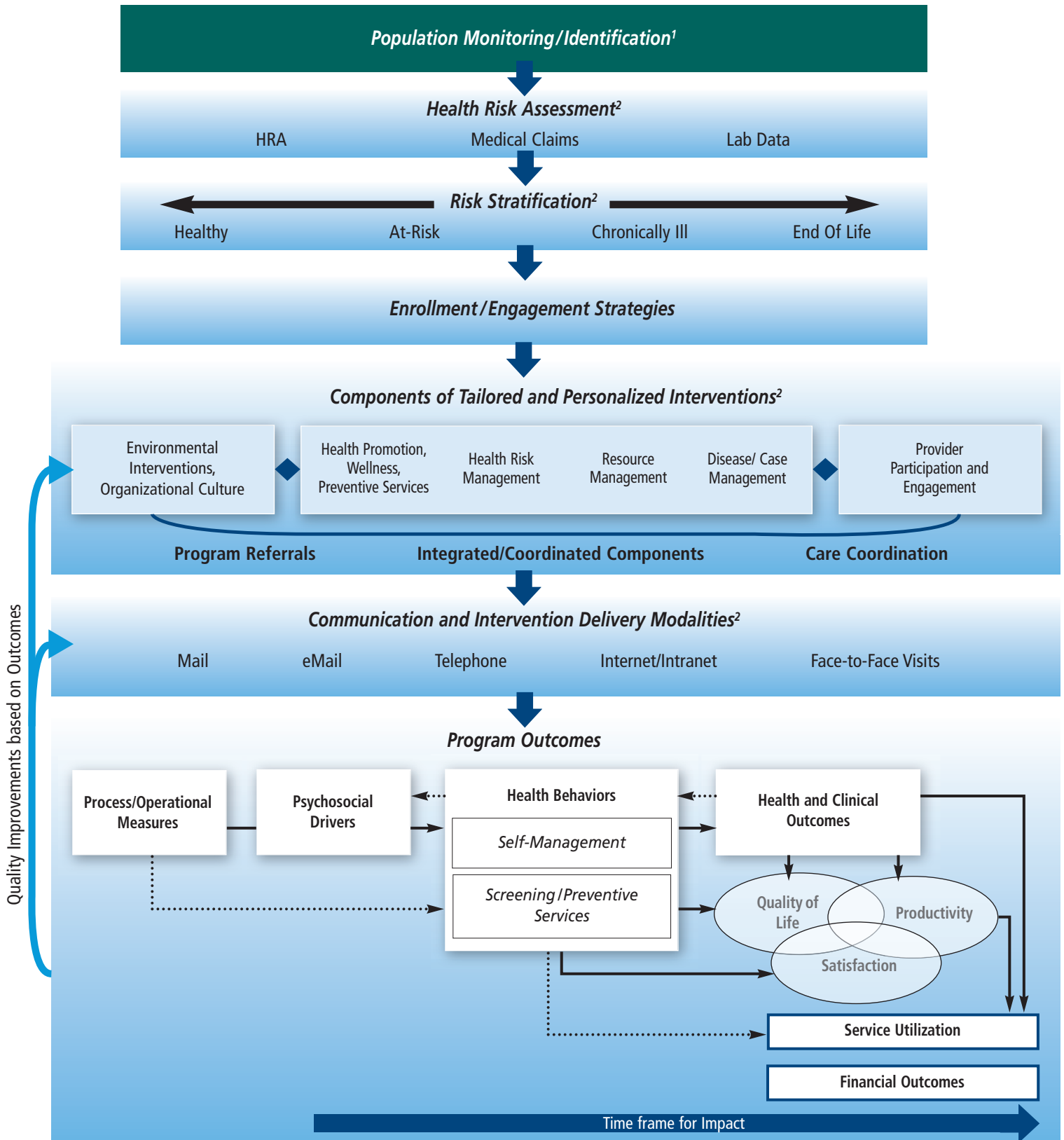
The inclusion of utilization measures in outcomes analyses is increasingly important to those individuals who believe that utilization measures represent a real proxy for program effectiveness. Given that multiple variables exist in clinical program delivery and that many are changing over time (such 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.

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

| TABLE II – PHM CAPABILITIES | | |
|-------------------------------------|-----------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------|
| CAPABILITIES | BASIC CAPABILITIES | ENHANCED CAPABILITIES |
| Engagement Strategies | Active or Passive | Combination of Active and Passive |
| Health Risk Assessment | Single data source | Multiple data sources |
| Risk Stratification | Single dimension | Multiple dimensions |
| Tailored Interventions | Population oriented | Member oriented |
| Program Components | Multiple components along the continuum that address needs specific to healthy, at risk, and/or chronically ill | Multiple Components Integrated and Coordinated |
| Program Delivery Modalities | Single Modality | Multiple Modalities |
| Information Exchange with Providers | One way – Outbound | Two way – Inbound and Outbound |
| Outcomes Measurement | Evaluate Effect on Outcome Domains (separately) | Assess Root Cause Impact on Outcomes (such as individual behavior changes) |

Figure 1 – Population Health Management Program Framework

DMAA recommends the following population health management Program Framework Model:



¹ For a more detailed discussion of monitoring and identification flow please refer to the work of the Operational Measures Workgroup

To help improve our understanding of how the essential elements of a PHM program (included in Table II, page 14) relate to one another, Figure 1 presents a very high-level program framework. This framework outlines the process flow associated with operating a PHM program, beginning with monitoring the population and identifying individuals who are eligible for the program. Once identified, some type of individual health risk assessment is applied and potential participants are stratified based on the results of their risk assessment. Next, proactive methods are used to enroll and engage population members into tailored and personalized interventions based on their level of risk and/or preference for delivery modality. Also depicted is a conceptual outcomes framework that represents hypothesized relationships among outcome domains relevant for a PHM program. While a more complete discussion of the monitoring and identification process can be found in the Operational Measures section of this report (page 94), each of the other essential elements for PHM are described more fully in the following paragraphs.

Another critical source of support for participants in population health management programs is their regular health care providers (primary care physicians, specialists and other health care professionals). Population health management programs complement the provider/patient relationship by offering additional resources to address gaps in patient health care literacy, knowledge of the health care system and timeliness of treatment. At a minimum, programs should encourage participants to share their program experience with their health care providers. Ideally, programs also would include open communication channels with providers and even coordination of care between various health care professionals.

An essential consideration when designing PHM programs is that interventions should be included for participants across the continuum of care from wellness to disease and case management, and even end-of-life support if the population contains individuals in need of such services. As outlined in the framework, program components might include health promotion activities, wellness and preventive services, health risk management, resource management activities and disease and/or case management. While not all population health management programs will offer all these services, some type of intervention should be offered to all eligible members of the population. It is also important that the components of population health management programs are integrated and coordinated and that appropriate referrals are made, both within programs and to available resources in the communities in which they operate.

Health Risk Assessment

Once eligible members of a population have been identified, there are a number of ways to assess the health of individuals in a population. Health risk 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/or 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 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 risk assessments. This stratification should include categories that represent the continuum of care in the population. While some organizations use complicated mathematical proprietary 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

Once individuals in a population are identified and risk 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 processes are 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.

Components of Tailored and Personalized Interventions

A program intended to manage the health of a population can comprise many types of services. 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 importance of environment is depicted in the framework by double-ended arrows between the *Environmental Interventions/Organizational Culture* box and both the *Enrollment/Engagement Strategies* and *Program Components* boxes, indicating that engagement strategies are strongly related to a supportive environment and, vice versa, that PHM program components form an integral part of the culture and environment.

Communication/Intervention Modalities

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

Program Outcomes

The final essential element of the PHM framework is a conceptual outcomes framework. The outcomes framework in Figure 1 begins by considering 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 more full list of specific psychosocial, health behavior and health-related outcomes, please refer to the Model of Wellness Program Impacts on page 29. For health-related outcomes specific to chronic conditions, refer to the measures sets in the Five Core Chronic section of this report, pages 79 to 97.

As outlined in the outcomes section of the PHM framework, quality of life, productivityⁱ and satisfactionⁱⁱ are overlapping constructs that will all be positively impacted by, and may have a reciprocal positive impact on, participants’ behavioral and

ⁱ Productivity often includes both absenteeism and presenteeism.

ⁱⁱ Satisfaction can include life satisfaction, satisfaction with care, satisfaction with the program, etc.

health-related outcomes. 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. DMAA plans to continue to further develop the PHM framework to include a detailed overlay of health information technology, as well as a program evaluation strategy.

To complement the outcomes section of the PHM framework, the workgroup developed a core list of measures to support the overall evaluation of a PHM program. These measures will be developed more completely in 2010 with the possible inclusion of standard measure recommendations currently used in population health management.

DMAA 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

Introduction

Over the past decade, a fundamentally new benefit has arisen alongside, or from within, traditional disease management and population health management programs: the wellness program. Disease 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.

While wellness programs are a relatively new benefit, their use is expanding rapidly. The 2008 DMAA report, “Disease Management and Wellness: Results of a Market Research Survey,” further substantiates the claim that wellness programs are on the rise. The report states that 84 percent of health plans and employers currently offer wellness programs and another 7 percent plan to offer these programs within the next 12 months. Additionally, 21 percent of program providers responding to the survey plan to begin offering wellness programs within the next 12 months.

DMAA considers wellness programs and disease management programs to be subsets of population health management programs. As such, DMAA 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 disease management programs, wellness program outcome measures and standards are less well-developed.

Wellness areas in the DMAA guidelines include both methodological and measure set recommendations. New guidelines to the report this year include recommendations concerning the appropriate use of control/comparison groups, key measures set and the use of claims data for program evaluation.

Wellness Program Definition

DMAA 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

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

| TABLE III – COMPARISON OF CURRENT STATE OF DISEASE MANAGEMENT/WELLNESS PROGRAMS | | |
|----------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------|
| | <i>Disease Management Program</i> | <i>Wellness Program</i> |
| <i>Health status of target population</i> | Always have chronic illness | Total population which varies from optimal health to chronically ill |
| <i>Primary focus of intervention</i> | Optimization of chronic condition management | Modification of health risk behaviors |
| <i>Primary outcome metrics of interest</i> | Health care costs and utilization rates (hospitalizations, ER visits and use of specific diagnostic and therapeutic procedures) | Specific health risk behaviors, indirect health-related costs (productivity, etc.) |
| <i>Unit cost of program</i> | Moderate | Low to moderate |
| <i>Claims data availability (for purposes of evaluating program impacts)</i> | Usually available | Variable availability, less likely than with Disease Management program |
| <i>Time frame for impact on participant behavior</i> | Near term | Near term |
| <i>Time frame for impact on health status</i> | Near to medium term | Near to long term depending upon program elements and aims |
| <i>Time frame for impact on health care costs</i> | Near to medium term | Medium to long term |
| <i>Use of health risk assessment for baseline assessment</i> | Variable, less likely than with wellness program | Nearly always |
| <i>Use of biometric screening data for baseline assessment</i> | Variable | Variable |
| <i>Status/availability of narrative coaching/encounter notes for analysis</i> | Variable | Usual (depending upon program type) |

TABLE III (cont.) – COMPARISON OF CURRENT STATE OF DISEASE MANAGEMENT/ WELLNESS PROGRAMS

| | <i>Disease Management Program</i> | <i>Wellness Program</i> |
|-----------------------------------------------------------|----------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------|
| <i>Availability of process measure metrics</i> | Always | Always |
| <i>Approach to evaluating financial impact of program</i> | Direct analysis of claims data | Modeling of risk factor changes and their effects on health care costs, productivity and direct analysis of claims data |
| <i>Availability/use of case reports</i> | Usually available and appropriate to <i>illustrate</i> program effects, not to <i>measure</i> them | Usually available and appropriate to <i>illustrate</i> program effects, not to <i>measure</i> them |

Methodological Considerations

The logic of wellness programs, as expressed graphically in the Wellness Program Impacts Model (page 29), 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
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:

DMAA 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).

DMAA 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 70.)

The above discussion would apply equally to a traditional disease 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 20, 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 38 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 and 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 DMAA 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 DMAA recommendations in evaluating wellness programs:

1. The program is designed to address modifiable health risk factors;
2. A tool is used to assess the 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 32. 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 assessments (HRA) 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 achieves the desired outcome and proffers the ability to provide additional important lessons learned for future 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

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

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 is a modification of Table II: Study Design, on page 16 of the DMAA Outcomes Guideline Report, Volume 3. This modification reflects the differences between the disease management programs and populations 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 a randomized control group – 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 relative 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 IV – CONTROL / COMPARISON GROUPS METHODS FOR WELLNESS PROGRAM EVALUATION

| METHOD TO DEVELOP A COMPARISON GROUP | Randomized Control | Non-Randomized Control | | Comparison Group | | Own Control |
|---------------------------------------------------|---------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------|
| | | Matched Control | Un-Matched Control | Comparable Employee Population | Book-of-Business | |
| General description of method | Intervened population compared with individuals randomly selected to have services withheld | Intervened population compared with non-participants matched to have similar demographic, clinical and behavioral characteristics | Intervened population compared with non-participants | Intervened population compared with similar employer group(s) in same program | Intervened population compared with book-of-business population in same program | Pre/post comparison of intervened population. |
| Comparison time frame | Concurrent to intervention | Concurrent to intervention | Concurrent to intervention | Prior period | Prior period | Concurrent to intervention |
| Primary outcome measures | Modifiable risk factors, biometric variables | Modifiable risk factors, biometric variables | Modifiable risk factors, biometric variables | Modifiable risk factors, biometric variables | Modifiable risk factors, biometric variables | Modifiable risk factors, biometric variables |
| Population selection bias | None | Somewhat significant | Significant | None | None | |
| Source of comparison group | Population for whom program was implemented, randomly selected group withheld from program | Population for whom program was implemented, purchaser decision to not participate | Population for whom program was implemented, purchaser decision to not participate | Vendor data | Vendor data | Intervened population |
| Credibility of causal statements | Extremely strong | Moderate | Poor | Very poor | Very poor | Very poor |
| Control of confounding variables | Controls known and unknown, measured and unmeasured confounding variables. | Controls known and measured confounding variables. | Does not control confounding variables | Does not control confounding variables | Does not control confounding variables | Does not control confounding variables |
| Program sponsor resistance to approach | High | Moderate/High | Moderate | None | None | None |
| Ease of implementation | Very difficult | Difficult | Somewhat Difficult | Easy | Easy | Easy |
| Clarity of method to lay audience | Very clear | Very unclear | Somewhat unclear | Clear | Clear | Clear |
| Multiyear application vs. single year application | Much harder | Much harder | Somewhat harder | Same | Same | Same |
| Method availability | Rarely possible | Occasionally possible | Occasionally possible | Usually possible | Always possible | Always possible |
| Bleed of interventions to comparison population | Control group may get provider-based interventions, other vendor interventions, secular interventions and self-care | Control group may get provider-based interventions, other vendor interventions, secular interventions and self-care | Control group may get provider-based interventions, other vendor interventions, secular interventions and self-care | Comparison population already intervened | Comparison population already intervened | NA |
| Key strengths | Gold standard evaluation method | Possible to infer reasonable level of causality without experimental design | Least costly and easiest method of compared intervened to non-intervened population | Low cost, availability of data, ability to answer question asked: (How does my group compare to other groups that have been in the program?) | Low cost, availability of data, ability to answer question asked: (How does my group compare to other groups that have been in the program?) | Low cost, availability of data, ability to answer the question asked: (Did the intervened group change over time.) |
| Key problems/biases | Cost, sponsor resistance, IRB imperatives, low generalizability | Cost, availability of control group HRA data | Limited ability to make causal inferences, | No inferences on causality are possible | No inferences on causality are possible | No inferences on causality are possible; No information on relative performance of intervened group. |

Use of Claims Data in Program Evaluation

DMAA 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 tend to 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.

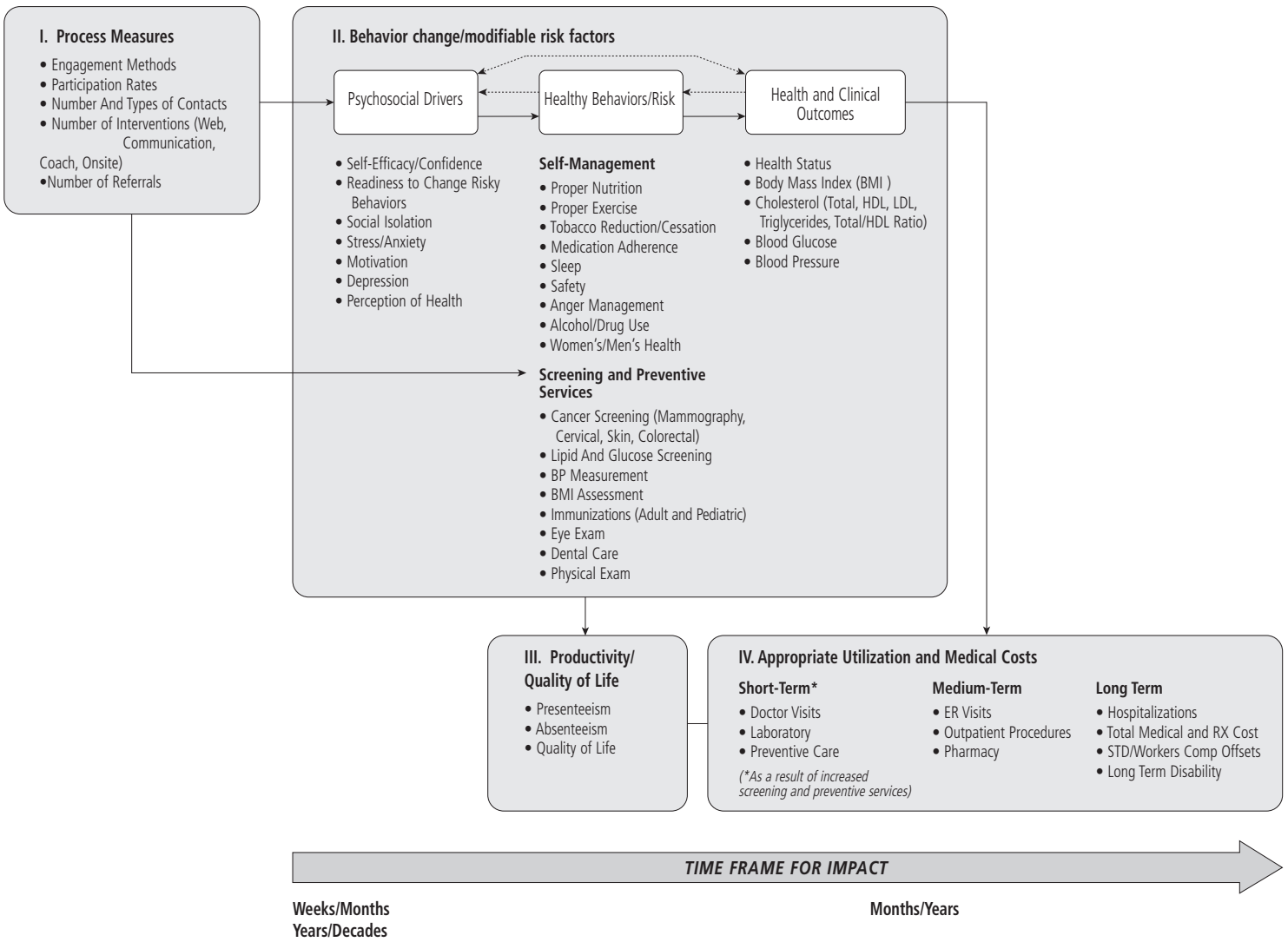
3. There is a minimum of three years of program exposure. As discussed previously in Outcomes Guidelines Report, Volume 2, and shown in the Impacts Model, (figure 2, next page), the time frame of the anticipated cascade of events (behavior change > biometric change > health status change > health cost 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 specific of program scope and intensity, an even longer time frame may 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 disease 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.

FIGURE 2 – IMPACTS MODEL



Wellness Program Model of Impacts

The Model of Wellness Program Impacts is designed to outline the range of relevant outcomes for the evaluation of wellness programs (see figure 2, page 29). The framework of this model also is designed to convey two important points. First, outcomes associated with wellness programs (and disease 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.

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 may 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 may be years (or even decades) before significant impact is shown in medical costs, workers' compensation offsets or long-term disability.

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 Model of Wellness Program Impacts in figure 2 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-efficacy³ is a major tenet of social cognitive theory⁴ 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^{5,6}, 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 model, improvements in psychosocial drivers can positively influence health behaviors^{7,8}. Proposed, are two categories of health behaviors: 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 Behavior Change/Modifiable Risk Factors is health and clinical outcomes. According to the model, improvements in health behaviors can positively impact clinical indicators of health^{9,10}. 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 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 depression¹¹, 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.

QUALITY OF LIFE/PRODUCTIVITY

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 is 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 are also acceptable.

UTILIZATION AND MEDICAL COSTS

As outlined in the model, improvements in behavior change and modifiable risk factors will positively impact utilization and medical costs^{12,13}, the fourth domain in the model of wellness program impacts. This domain encompasses appropriate utilization to maintain wellness, including regular physician visits for recommended physical examinations, laboratory tests and preventive services. Also included is utilization and costs for avoidable emergency department visits and hospitalizations, as well as costs associated with workers' compensation and short- and long-term disability.

Measure Sets

DMAA 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

DMAA 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.

DMAA 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.

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

While the ability of wellness and disease 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.

DMAA 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.

DMAA 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 Expanded

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 disease management programs on workplace productivity, as it may 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)¹⁴ and in a negative sense (being unproductive while being present at work) in more recent studies¹⁵. 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.”¹⁶

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. To assist in this assessment, DMAA will publish a white paper that provides:

- expanded discussions of the key issues identified in this section;
- a bibliography that contains general reviews of and comparisons between instruments and more focused reviews on the monetization of presenteeism; and
- information on the most commonly used instruments.

Five Core Chronics

Much of the work completed for the guidelines, excluding wellness, applies to a traditional disease management program. Although application to other 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 following guidelines in the Five Core Chronics section have been developed following this operating principle.

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 disease 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 disease 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 disease 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 that one can attribute changes to the disease 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. DMAA acknowledges the desirability and value of the randomized controlled trial model to reach conclusions about disease management value, but also recognizes the impracticality of expecting all business purchasers of disease management services to set up this model for evaluation purposes. Therefore, DMAA 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.

DMAA 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, DMAA 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 disease management program outcomes. With this in mind, a matrix (Table V) that compares several of the most commonly used study designs is included on page 38. Both disease management service providers and purchasers are encouraged to review the information included in the matrix when selecting a study design that differs from the DMAA-recommended design. DMAA 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 V – STUDY DESIGN

| METHOD TO DEVELOP A COMPARISON GROUP | Randomized concurrent control | Matched control | Non-participating individuals comparison | Historical control (as defined by DMAA guidelines) |
|---------------------------------------------------------------|---------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| General description of method | Intervened population compared with individuals randomly selected to have services withheld | Intervened population compared with individuals from non-participating groups matched to have similar characteristics as intervened individuals | Intervened participating individuals compared with intervened non-participating individuals | Intervened population compared with similarly identified population in a baseline period (with costs trended forward) |
| Comparison time frame | Concurrent to intervention | Concurrent to intervention | Concurrent to intervention | Prior period |
| Population selection bias | None | Somewhat significant | Significant | None |
| Source of comparison group | Population for whom program was implemented, randomly selected group withheld from program | Population for whom program was implemented, purchaser decision to not participate | Population for whom program was implemented, individual decision to not participate | Population for whom program was implemented, in prior period |
| Trend factor | Not required | Not required | Not required | Individuals without measured conditions in population for whom program was implemented |
| "Regression to mean" issues | None | None | None | Low |
| Credibility of causal statements | Extremely strong (gold standard) | Strong (with proper design) | Poor | Moderate |
| Applicability to all types of programs/program designs | Strong | Strong | Strong | Uncertain |
| Program sponsor resistance to approach | Strong | Moderate (confidentiality of non-participating groups can be an issue) | None | Low |
| Ease of implementation | Very difficult | Difficult | Easy | Somewhat difficult |
| Clarity of method to lay audiences | Very clear | Very unclear | Clear | Somewhat unclear |
| Multiyear application vs. single year application | Much harder | About the same | About the same | Somewhat harder |
| Method availability | Rarely possible | Occasionally possible | Frequently possible | Usually possible |
| Bleed of interventions to comparison population | Control group likely to get provider-based interventions | Control group likely to get provider-based interventions | Control group likely to get provider-based interventions and softer member-based interventions | Control group unlikely to get any interventions |
| Key strengths | Generally seen as gold standard evaluation method | Potential for a concurrent comparison population with many of the same characteristics as the intervened population | Easy to implement | Relatively universally available method |
| | Ease/strength of interpretation | Credibility of causal statements can be very strong | Method is easy to understand | Application of well-accepted actuarial processes |
| Key problems/biases | Sponsor resistance to implementation | Difficult to get access to non-participating groups of sufficient size in same geographic area, with similar benefit structure | Very significant bias associated with differences in motivation between participants and non-participants | Difficult to ensure equivalence between baseline and intervention year (particularly in populations with many shifts in size, composition, and/or benefit structure) |
| | Difficulty of multiyear assessment | "Black box" approach difficult to understand | Likely not possible in "opt-out" program models | Difficulty in deriving credible trend factor |

Population Identification Methodology

Developing a claims-based methodology that identifies appropriate patients for inclusion in a disease 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 44). 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. Disease 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 are also 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 disease 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,

DMAA 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 disease 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 disease management program evaluation,

DMAA 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,

DMAA 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.

DMAA 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 disease management programs have information on enrollment dates,

DMAA 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. DMAA 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 disease 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 disease 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 disease 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 VI, Methods to Define Outliers, page 43. DMAA recommends a review of the information in the table prior to selection of a method.

TABLE VI – METHODS TO DEFINE OUTLIERS

| METHOD | Stop-Loss Method | Percentile Distribution Method | Standard Deviation (SD) Method |
|---------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| General description of method | 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. | A threshold value is determined based on the X percentile of claims costs (e.g., 99.5%); costs above this threshold are excluded. | A threshold value is determined based on X (e.g., 3SD) standard deviations from population mean. Costs above this threshold are excluded. |
| Trend factor | Should be used to account for medical cost trends year to year. | Not required. | Not required. |
| Applicability to all types of programs/program designs | May require lower threshold to offset variability in small populations. | N/A | May not be appropriate for populations with non-normal distributions. |
| Ease of implementation | Very Easy | Easy | More difficult. |
| Clarity of method to lay audiences | Very Clear | Clear | Requires knowledge of simple statistics. |
| Multiyear application vs. single year application | Trend adjustment should be considered for multiyear assessment. | Easy to apply. | Should recalculate SD for each year. |
| Key strengths | Better for evaluations where two groups are expected to have different frequency cost distributions rather than unit cost variance. | Better for evaluations where two groups are expected to have unit cost variance rather than different frequency cost distributions. | Better for evaluations where two groups are expected to have unit cost variance rather than different frequency cost distributions. |
| | Ease of application and most sensitive to varying rates of catastrophic claims. | Can be used without adjustment on small populations. | Can be used without adjustment on small populations. |
| Key limitations | Stop-loss threshold is arbitrary and not data-sensitive. | Threshold is also arbitrary but data-sensitive to varying percentages of shock-loss claims. | Threshold is also arbitrary but data-sensitive, and is only somewhat sensitive to a varied percentage of claims categorized as shock-loss. |
| | 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. | Handles variant cost distributions with variant thresholds, which may not be desirable with variant rates of catastrophic claims. | 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. |
| | Lack of threshold trend adjustment will deflate overall trended results proportional to the presence of catastrophic cases. | | Handles variant cost distributions with variant thresholds, which may not be desirable with variant rates of catastrophic claims. |

Selection Criteria

Background

The DMAA Outcomes Guidelines Report Volume 3, published in 2008, stated that, “Achieving consensus on how to select populations for the evaluation of disease management programs remains an important goal of DMAA. Standard selection criteria will contribute to improving standardized program evaluation, as well as help facilitate rigorous performance comparisons and increase transparency to purchasers. 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.”

“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 disease management programs for the five chronic conditions. DMAA approved denominator specifications for extensive testing in 2008-2009.

This year’s scope of work for Selection Criteria focused on the use of denominators for program comparison, recognizing that internal (program-specific) evaluation of how well the program achieved its goals (e.g., for quality or outcomes improvement) might differ from one program to another, based on the kinds of populations served, and certain specifics of program content and delivery.

Following are:

- Review of prior years’ work (from volumes 2 and 3), including development of a philosophical framework and an overview of denominator specifications.
- This year’s scope of work, which involved testing our Selection Criteria.
- The results, interpretation and implications of the tests.
- Resulting updates (which are minor) to the detailed Selection Criteria specifications.
- Next year’s scope of work – field testing the denominators for suitability for their intended use: measurement of clinical and financial outcomes.

Prior Work

In volumes 2 and 3, DMAA specified that it would define selection criteria for the five core chronic conditions, using administrative data (medical and pharmacy claims, coded as ICD-CM-9 for diagnoses, CPT-4 for procedures and encounters, and NDC for pharmaceuticals; as well as revenue or place of service codes generally provided on claims forms).

DMAA reviewed various potential validity risks of using administrative data, such as variability in coding and a propensity of claims data to generate false positive identifications.

Recognizing that no algorithm for identification of individuals with a chronic condition could identify everyone who in fact had that condition (100 percent sensitivity) with no false positive identifications (100 percent specificity), DMAA established a philosophical framework by which any proposed identification algorithm could be constructed and evaluated. Constructing selection criteria for the five chronic conditions was the work of Guidelines Volume 3; evaluation of these criteria with potential modification was this year's scope of work.

Guiding principles adopted for the work include: fairness and relevance; accuracy and balance; and efficiency.

Fairness and relevance implies that most programs can embrace the criteria with the understanding that these criteria might not perfectly match the case identification criteria for any specific program; and that the criteria should be generally recognized as aligned with clinical definitions for the specified conditions.

Accuracy and balance implies a practical and useful balance of sensitivity and specificity, and honors the principle of equivalence. Criteria should have high specificity without overly sacrificing sensitivity (i.e., should focus somewhat more on avoiding false-positive identifications than on including everyone with the condition, but should be sensitive enough to identify the great majority of individuals who would be targeted for management by a disease management program).

The *principle of equivalence* states that a person who actually has condition X in two measurement periods should have the same probability of being identified by the selection criteria in both periods. Likewise, a person who does not have X in two measurement periods should have the same probability of not being identified by the selection criteria in both periods.

Efficiency implies that the data used for selection should be widely available, requiring elements that can generally be found in Medicaid, commercial and Medicare administrative data sets; and that the “recipe” or algorithm used for denominator inclusion should be transparent and straightforward to implement. Because most health plans and programs are familiar with the HEDIS, our overall approach is similar enough to that of HEDIS to be readily understood and to facilitate implementation.

Overview of Denominator Specifications

Our orientation has been to construct denominator cores for the five conditions that can be implemented (in some cases, with appropriate modifications) across multiple metric types: clinical, functional status, utilization and financial. DMAA therefore specifies common building blocks for each of the five condition denominators.

The DMAA Outcomes Guidelines Report Volume 3 included draft specifications for the five chronic conditions. These are reprinted at the end of this section with updates (approved by the Outcomes Methodology and Measurement Committee). The draft specifications are:

- Required benefits (e.g. medical, pharmacy).
- Types of administrative claims data to be used.
- Minimum eligibility (covered or insured status) in the identification period.
- Demographic characteristics: specifies the minimum and maximum allowable age as of the last day of the measurement period.
- Identification time frame: The period during which all identification criteria must be met to qualify for selection.
- Type and number of codes that must occur in the identification time frame.
- Condition-independent denominator exclusions.

This Year's Scope of Work

The Selection Criteria Workgroup focused on testing the selection criteria developed in Guidelines Volume 3, and included on pages 54 to 58 of this report, for suitability for program comparison. "Suitability" was operationally defined so as to permit testing of the guiding principles expressed in Volume 3.

Specifically, the Selection Criteria Workgroup used (a) the specific identification criteria for five chronic conditions, using (b) the principle of equivalence (annual qualification using identical criteria in each identification period) with (c) varying identification time frames and (d) minimum eligibility requirements for meeting:

- Accuracy and balance:
 - Principle: Reasonable results compared with general industry standards.
 - Prevalence Test: Are the prevalence rates reasonable compared with work group member's experiences?
 - Principle: Reasonable specificity without unduly sacrificing specificity.
 - True Positive Test: Are individuals identified in time 1 (T1) also identified in time 2 (T2)?
 - True Negative Test: Are individuals not identified in T2 also not identified in T1?
- Fairness and relevance.
- Principles: A) DMAA criteria should identify most people identified by programs (programs want to be compared on people they identify); B) DMAA criteria should not identify many people not identified by programs (programs do not want to be compared on people they do not identify).
 - Overlap Test: Overlap in pool of identified individuals between DMAA and vendor criteria:
 - Individuals identified by both DMAA and vendors' criteria.
 - Individuals identified by vendors' criteria but not DMAA's.
 - Individuals identified by DMAA's criteria but not vendors'.

Testing the selection criteria against the guiding principles

Three disease management program providers tested the Volume 3 Selection Criteria using commercial population data sets. The table below presents the number of members included in aggregate. The impact of age selection criteria requirements and the need for having enrollment for the 24- or 48-month test periods reduced the number of members available to apply each disease's claim code criteria.

| | Total Members in Data Set | Members Meeting Age and Eligibly Requirements |
|-------------------|---------------------------|-----------------------------------------------|
| 24 Months of Data | 1.5 Million | 0.4 to 0.7 Million |
| 48 Months of Data | 1.9 Million | 0.4 to 0.7 Million |

Member-count ranges in the "Members Meeting Age and Eligibility Requirements" represent age variations in disease identification and whether continuous vs. six months' eligibility was required in a specific test.

Diabetes, CHF and CAD age criteria are 18 and older; asthma, between ages 5 and 56; and COPD, 40 and older. The vendors tested both a) requiring members to be continuously enrolled for both identification periods; and b) with at least six months enrollment in each identification period.

To test the accuracy and balance principle, the work group started with a 12-month identification period. Two consecutive non-overlapping 12-month identification periods (24 months total with three months of run out after the second year) were established. To ensure the Principle of Equivalence was met, True Positive and True Negative tests required that identified individuals re-qualify for each identification period. The following tests were performed:

- **Prevalence Test:** Are the *prevalence rates* for the five diseases reasonable and consistent across testers and across the consecutive time periods? This was accomplished by summing and dividing a) the number of individuals identified with the disease in 12 months; by b) the total number of individuals who met each disease's age and enrollment criteria in the same 12 months.
- **True Positive Test:** Are individuals *identified* in the first 12 months (T1) also identified in the second 12 months (T2)? This was accomplished by summing and dividing a) the number of individuals identified in both T1 and T2; by b) the number of individuals identified in T1.
- **True Negative Test:** Are individuals not identified in T2 also *not identified* in T1? This was accomplished by summing and dividing a) the number of individuals not identified in T2 or T1; by b) the number of individuals not identified in T2.

The testers applied the selection criteria independently and provided the aggregate number of members by the summations listed in the above paragraph to the workgroup. The data were then aggregated across the testers, and the workgroup reviewed both the individual test results and the aggregate numbers. For the most part, the testers' results were similar. Since the demographics of the three test populations were different, DMAA expected and found some slight differences across all test results.

Next, to reveal the impact on the three test scenarios of doubling the identification time frame, DMAA expanded the testing to include two continuous, non-overlapping, 24-month periods (48 months total with three months of run-out after the fourth year).

To test fairness and relevance, the three program providers applied their proprietary identification criteria and DMAA's to a 24-month identification period with three months of run-out. The testers then performed the overlap test by comparing the number of individuals identified using the DMAA and their own criteria to determine those identified by:

- both DMAA and their criteria;
- their criteria only; or
- DMAA criteria only.

To protect the proprietary nature of each vendor's criteria, the testers provided their results to an independent third party on the work group. The independent third party then took the straight average of the results, asked each tester to explain any significant variance that might exist from DMAA criteria and shared the appropriate and de-identified results with the workgroup.

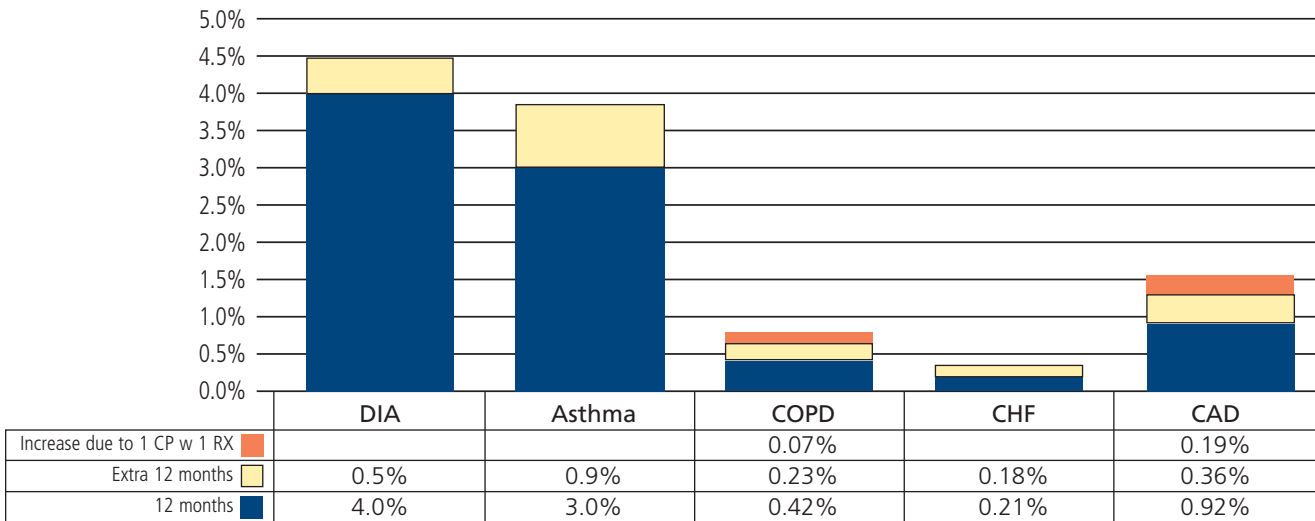
Test results: interpretations and implications

Prevalence Test Results:

The graph below presents the Prevalence Test results for the five diseases and using three sets of identification criteria:

- A 12-month identification period.
- A 24-month identification period.
- A 24-month identification period with an addition to the selection algorithm of an “OR” criterion for one physician encounter AND at least one disease-specific prescription for CAD and COPD.

Chart I – Prevalence using 12 and 24 months of data For members with 12 or 24 months of coverage, respectively

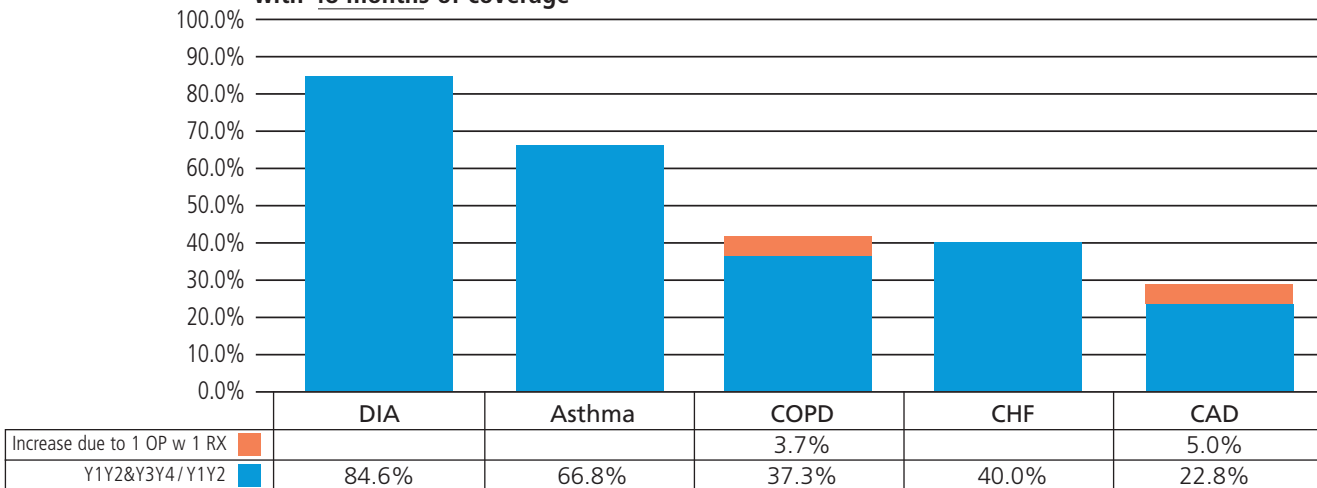


The impact of adding an “OR” criterion for one emergency department, outpatient facility or office visit with a CAD or COPD diagnosis AND a disease-specific prescription is a relative 12 percent and 15 percent increase in CAD and COPD’s prevalence rates, respectively. More will be said below about the rationale for adding this criterion to CAD and COPD.

True Positive Test Results

The graph below presents the True Positive Test results for a 24-month identification period. As mentioned earlier, the testers first tested a 12-month identification period by using two continuous, non-overlapping 12-month periods (24 months). To perform the 24-month identification period True Positive Test, two consecutive 24-month periods (48 months) were used. The graphs of the 12- and 24-month True Positive Test results were very similar, so only the 24-month True Positive graph is shown below. Prior to performing the tests, the workgroup had set a plausibility-based target True Positive value of 90 percent. However, the results for the five conditions were all below this target, with COPD, CHF and CAD significantly below.

Chart II – “% True Positive” = (Selected in Both T1 and T2) / (Selected in T1) For members with 48 months of coverage



Each tester produced disease-specific results consistent with the other testers. One implication of having a selection criteria True Positive value below 40 percent is that at least 60 percent of the individuals will be different from one non-overlapping identification period to the next. To help mitigate this issue, feedback was obtained from others outside the workgroup about possible reasons for the low True Positive test and ways to increase the True Positive results. Two potential methods were discussed and then tested:

- **Increase the number of months needed in the identification period.** This was done by testing a 24-month identification period. While prevalence did rise with 24 months, the True Positive test showed nearly the same results as the 12-month period.
- **Find additional ways to identify those with COPD, CHF and CAD.** A common characteristic of diabetes and asthma that contributes to their much higher True Positive Test is that their selection criteria include maintenance drugs highly specific to the diseases. A re-run of the 24-month True Positive test, excluding the maintenance drugs for diabetes and asthma, resulted in a decline in diabetes and asthma True Positive tests to 38 percent – consistent with COPD and CHF.

Accordingly, the workgroup tested adding drugs with high specificity for COPD and CAD (there are no commonly prescribed drugs with high specificity for CHF): tiotropium for COPD and nitrates or lipid-lowering agents for CAD. To increase these drugs' disease-identification specificity, the test requirement specified, in addition to a prescription fill, at least one emergency department, outpatient or physician visit in the identification period with the diagnosis of COPD or CAD, respectively. The impact of adding these "OR" criteria was to increase the True Positive values by 10 percent for COPD and 22 percent for CAD.

The workgroup determined that there were no additional claims codes (beyond those already specified in Guidelines 3) that could be used to increase values of the True Positive Tests.

While additional methods for identifying COPD, CHF and CAD more consistently may include using more than 24 months of data, other diagnostic or procedure codes or additional associated prescription drugs, the workgroup came to a general conclusion that the selection criteria outlined in Volume 3 are fairly conservative (meaning the criteria are, based on clinical experience, to be unlikely to identify someone without the condition). Our interpretation of the results from the True Positive testing is (at least in part) that many individuals who meet the Volume 3 criteria for COPD, CHF and CAD may not seek or receive follow-up care in the year or two after a qualifying event. A chart or claim review would be required to validate this interpretation.

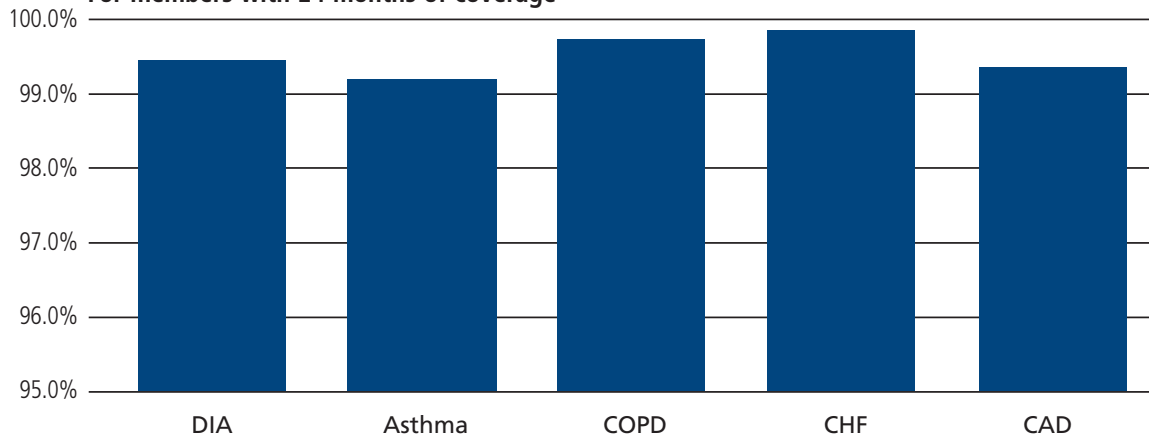
True Negative Test Results

The graph below presents the True Negative Test results for a 12-month identification period. The results for all five diseases are well above the 90 percent target the workgroup agreed to prior to testing. The very high True Negative test results can be attributed to the very low incidence rates of each of the five diseases.

The workgroup agreed the True Negative Test should use individuals "Not Selected in Year 2" as the basis to compare those "Not selected in both identification periods." Of those included in "Year 1 Not Selected" are those who develop a condition in Year 2. Thus, using "Year 1 Not Selected" as the denominator would contribute a bias in the True Negative Test results.

Due to the 12-month identification period's high True Negative test results, the workgroup decided that testing the 24-month identification period's True Negative value was not needed.

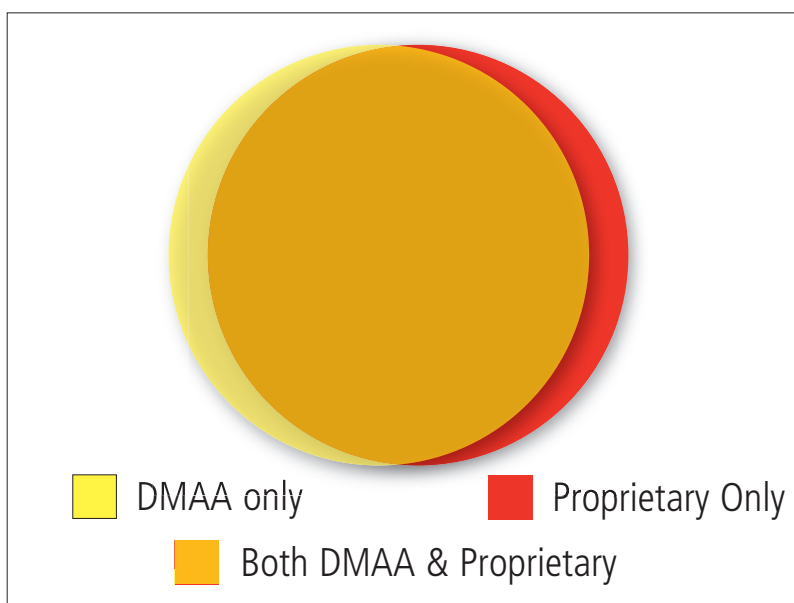
**Chart III – "% True Negative" = (Not Selected in Both Y1 and Y2) / (Not Selected in T2)
For members with 24 months of coverage**



Overlap Test Results

The diagram below presents a hypothetical distribution of members identified using the DMAA criteria and each vendor's proprietary criteria for identifying diseased members, with both using 24 months of claims with three months of run-out. To keep vendor criteria confidential, the workgroup evaluated the average across the three testers of the ratio of non-overlapping areas (DMAA-only and proprietary-only) with the overlapping.

Chart IV – DMAA vs. Proprietary Criteria

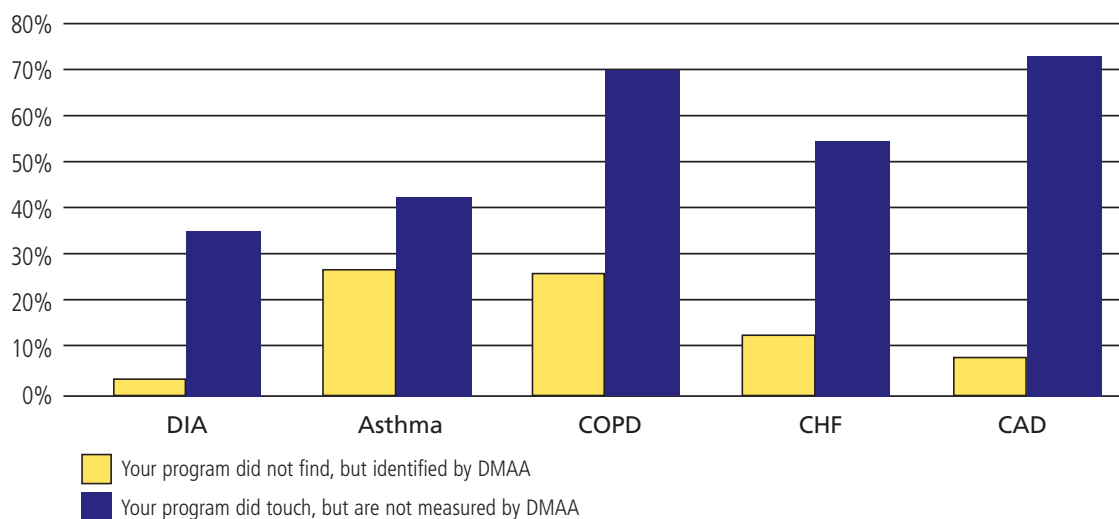


The graph below presents the average ratio of the three vendors' non-overlapping individuals to those individuals who were identified by both the DMAA and the proprietary criteria. The yellow bars represent the ratio of individuals found by DMAA only divided by those found by both criteria. The purple bars represent the ratio of those individuals found by the vendors' criteria divided by those individuals found by both criteria. In general, the DMAA criteria found fewer people than the proprietary criteria. The overlaps vary by condition, with the DMAA finding between 3 percent and 12 percent more people with diabetes, CAD and CHF than the overlap group. For asthma and COPD, the DMAA found 25 percent to 27 percent more people than the overlap group of individuals.

While the graph does not show it, the overlap vs. non-overlap areas varies significantly when using the vendors' proprietary methods. The vendors' criteria also found 30 percent to 72 percent more diseased individuals than the overlap area due to the vendors having broader definitions of disease classes. For example, cardiovascular disease was included by one vendor in the same criteria set as CAD, thus, significantly overstating the proprietary-only ratio to the overlap area.

As there was considerable variability in the vendor-specific criteria, it will be necessary to assess the question of overlap in more extensive field testing in 2010.

Chart V – DMAA vs. Proprietary Criteria



Updates to the detailed Selection Criteria specifications

The tables in this section show the selection criteria—codes and timing—used for identifying individuals as “having” the five chronic conditions. Updates to the criteria tables in Guidelines Volume 3 are highlighted and summarized below:

- CAD: Add additional qualifying criteria: one outpatient or emergency department encounter plus at least one prescription fill for a nitrate or lipid-lowering drug.
- COPD: Add additional qualifying criteria: one outpatient or emergency department encounter plus at least one claim for Tiotropium.
- For CAD and COPD, members must have both medical and pharmacy benefits
- Minor code updates from Guidelines Volume 3:
 - diabetes: Change ICD-9 “362.0” to “362.0x.”
 - asthma: Change ICD-9 “493” to “493.xx.”
 - CAD: Add ICD-9 “412.xx” for inpatient.
- Note also that DMAA recommends using current-year HEDIS Inpatient definition, e.g., from HEDIS clinical measure specification CMC-B to identify inpatient claims.
- DMAA recommends using current-year HEDIS NDC code lists for the Rx portion of identification.

In addition, DMAA recommends:

- ***Minimum eligibility (insured or covered): any six months in the 24 months of the identification frame (this requirement may be altered for certain purposes, such as a specific clinical metric concerning an annual service).***
- ***Identification time frame: 24 months incurred claims (ending at the last day of the identification frame) with at least three months' paid run-out (standard run-out may be greater for certain types of analyses, e.g., financial).***

Next Year's Scope of Work: Field Testing

While the selection criteria have been tested for conformance to the guiding principles of accuracy and balance, they now need to be tested for suitability to their purpose: measuring outcomes for the purpose of program comparison. In addition, DMAA will continue to assess variability across vendors in the overlap between DMAA criteria and vendor criteria. Finally, DMAA will address the question of condition-independent denominator exclusions, including codes indicating residential treatments, such as hospice.

Accordingly, in 2010, DMAA will work with population health management programs to explore how well-suited are the denominators for measuring clinical and financial outcomes. A panel of six to eight testers will perform the field tests, which will be evaluated by the Selection Criteria Workgroup, including a representative from the National Committee for Quality Assurance.

Selection criteria for specific conditions

The accompanying tables on the following pages show the core denominator specifications for each of the five conditions.

TABLE VII – 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: NDC (Using HEDIS CDC-A drug list)

Eligibility

- Same definition across all conditions – although may differ for financial and clinical metrics
- To be completed next year after comparative testing and consultation

Condition-Independent Exclusions

- Common across all conditions
- During empirical testing next year, will consider additional common exclusions, including codes representing residential treatments, such as hospice, and different exclusions depending on whether denominator is to be used with a clinical or a financial metric.

ID Timeframes

- Same definition across all conditions—although may differ for financial and clinical metrics
- To be completed next year after comparative testing and consultation with methods/finance workgroup

Demographics

- Age is as of the last day of the measurement period (note that the measurement period may differ for financial and clinical metrics)
- 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
 - Eligible visit codes – see appendix
- 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
- In empirical testing, test the frequency of individuals who would qualify for selection under the diabetes selection criteria who also have any claims with any of the following diagnoses: 962.0 (poisoning by adrenal cortical steroids), 962.4 (poisoning by anterior pituitary hormones) or 256.4 (polycystic ovaries)

TABLE VIII – 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

Eligibility

- Same definition across all conditions – although may differ for financial and clinical metrics
- To be considered next year after comparative testing and consultation

Condition-Independent Exclusions

- Common across all conditions
- During empirical testing next year, will consider additional common exclusions, including codes representing residential treatments, such as hospice, and different exclusions depending on whether denominator is to be used with a clinical or a financial metric.

ID Timeframes

- Same definition across all conditions – although may differ for financial and clinical metrics
- To be completed next year after comparative testing and consultation with methods/finance workgroup

Demographics

- Age is as of the last day of the measurement period (note that the measurement period may differ for financial and clinical metrics)
- 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
 - Eligible visit codes – see appendix

Condition-Specific Exclusions

- None

TABLE IX – DENOMINATOR SPECIFICATION FOR CAD

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, HCPCS
- Pharmacy: NDC. Note: May need pharmacy claims for some of the metric numerators but not for denominator criteria described here.

Eligibility

- Same definition across all conditions – although may differ for financial and clinical metrics
- To be completed next year after comparative testing and consultation

Condition-Independent Exclusions

- Common across all conditions
- During empirical testing next year, will consider additional common exclusions, including codes representing residential treatments, such as hospice, and different exclusions depending on whether denominator is to be used with a clinical or a financial metric.

ID Timeframes

- Same definition across all conditions – although may differ for financial and clinical metrics
- To be completed next year after comparative testing and consultation with methods/finance workgroup

Demographics

- Age is as of the last day of the measurement period (note that the measurement period may differ for financial and clinical metrics)
- 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, 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)
 - Eligible visit codes – see appendix

Condition-Specific Exclusions

- None

TABLE X – 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: NDC (using HEDIS ASM-C drug list)

Eligibility

- Same definition across all conditions – although may differ for financial and clinical metrics
- To be completed next year after comparative testing and consultation

Condition-Independent Exclusions

- Common across all conditions
- During empirical testing next year, will consider additional common exclusions, including codes representing residential treatments, such as hospice, and different exclusions depending on whether denominator is to be used with a clinical or a financial metric.

ID Timeframes

- Same definition across all conditions – although may differ for financial and clinical metrics
- To be completed next year after comparative testing and consultation with methods/finance workgroup

Demographics

- Age is as of the last day of the measurement period (note that the measurement period may differ for financial and clinical metrics)
- 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
 - Eligible emergency department visit codes – see appendix
- OR 4 OR MORE OFFICE/OUTPATIENT VISITS OR ENCOUNTERS OR EMERGENCY DEPARTMENT VISITS WITH
 - ICD9 code in any position: 493
 - Eligible visit codes – see appendix
- 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

TABLE XI – DENOMINATOR SPECIFICATION FOR COPD

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, HCPCS
- Pharmacy: NDC. Note: May need pharmacy claims for some of the metric numerators but not for denominator criteria described here.

Eligibility

- Same definition across all conditions – although may differ for financial and clinical metrics
- To be completed next year after comparative testing and consultation

Condition-Independent Exclusions

- Common across all conditions
- During empirical testing next year, will consider additional common exclusions, including codes representing residential treatments, such as hospice, and different exclusions depending on whether denominator is to be used with a clinical or a financial metric.

ID Timeframes

- Same definition across all conditions – although may differ for financial and clinical metrics
- To be completed next year after comparative testing and consultation with methods/finance workgroup

Demographics

- Age is as of the last day of the measurement period (note that the measurement period may differ for financial and clinical metrics)
- 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
 - Eligible visit codes – see appendix

Condition-Specific Exclusions

- None

APPENDIX: ELIGIBLE VISIT CODES: For Heart Failure, CAD, COPD and Asthma

| Description | CPT | UB-92 Revenue |
|-------------|--------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------|
| Outpatient | 99201-99205, 99211-99215, 99217-99220, 99241-99245, 99341-99345, 99347-99350, 99382-99386, 99392-99396, 99401-99404, 99411, 99412, 99420, 99429, 99499 | 051x, 052x, 057x- 059x, 077x, 0982, 0983 |
| ED | 99281-99285 | 045x, 0981 |

APPENDIX: ELIGIBLE VISIT CODES: For Diabetes

| Description | CPT | UB-92 Revenue |
|-------------|-----------------------------------------|-----------------|
| Outpatient | 92002-92014, 99387, 99397, 99455, 99456 | 082x-085x, 088x |

Trend

The most difficult part about evaluating the financial performance of disease 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 consequently, etc., and no one’s response could be tossed aside as wrong. Similarly, to calculate a trend that would represent what “might have been” without a disease 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.

DMAA 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 disease management program.

DMAA 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 disease 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 DMAA future research project.
- This would allow disease 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 disease 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 XII – DMAA TREND RECOMMENDATION RESULTS

| <i>Client #1, Annual Re-qualification</i> | | | | 1 Year Trend | | 2 Year Trend | | 3 Year Trend | |
|-------------------------------------------|-----------|--------------|-----------|--------------|------|--------------|------|--------------|------|
| Year | MM | PMPM Medical | PMPM RX | Med | RX | Med | RX | Med | RX |
| Non-Chronic | | | | | | | | | |
| 2004 | 2,345,429 | \$ 114.65 | \$ 38.14 | | | | | | |
| 2005 | 2,933,300 | \$ 123.29 | \$ 40.61 | 1.08 | 1.06 | | | | |
| 2006 | 3,950,584 | \$ 135.39 | \$ 43.10 | 1.10 | 1.06 | 1.09 | 1.06 | | |
| 2007 | 5,536,021 | \$ 153.05 | \$ 44.35 | 1.13 | 1.03 | 1.11 | 1.04 | 1.10 | 1.05 |
| Chronic | | | | | | | | | |
| 2004 | 181,316 | \$ 454.61 | \$ 178.61 | | | | | | |
| 2005 | 246,631 | \$ 498.19 | \$ 188.72 | 1.10 | 1.06 | | | | |
| 2006 | 329,180 | \$ 566.01 | \$ 207.41 | 1.14 | 1.10 | 1.12 | 1.08 | | |
| 2007 | 469,303 | \$ 664.90 | \$ 217.06 | 1.17 | 1.05 | 1.16 | 1.07 | 1.14 | 1.07 |
| All | | | | | | | | | |
| 2004 | 2,526,745 | \$ 139.04 | \$ 48.22 | | | | | | |
| 2005 | 3,179,931 | \$ 152.36 | \$ 52.10 | 1.10 | 1.08 | | | | |
| 2006 | 4,279,764 | \$ 168.51 | \$ 55.74 | 1.11 | 1.07 | 1.10 | 1.08 | | |
| 2007 | 6,005,324 | \$ 193.05 | \$ 57.84 | 1.15 | 1.04 | 1.13 | 1.05 | 1.12 | 1.06 |

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 disease 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 DMAA-recommended annual qualification method was used to identify the population for testing.

Table XII 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.

Risk Adjustment

Risk adjustment, as applied to evaluation of disease 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 disease management intervention on the financial outcomes for that population.

Performing disease 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 disease management program, as all other relevant factors that contribute to the financial outcome are equivalent.

General considerations for using risk adjustment for disease 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 disease management programs is to adjust for these exogenous confounders to the greatest extent possible, while not altering or distorting disease 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 disease management program evaluations:

- Many approaches to adjusting for exogenous variables have some risk of inadvertently adjusting for variables that are positively impacted by disease management programs; this can result in combined adjustment for confounding factors, as well as target factors at the same time, potentially discounting the desired disease 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, the data available.
- Performing risk adjustment is neither simple nor formulaic; no single approach to risk adjusting disease 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 disease 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 disease 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 disease management programs positively impact.

The application of risk adjustment to disease 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 XIII offers examples of how risk adjustment could be used for various situations.

TABLE XIII – RISK ADJUSTMENT EXAMPLES

| <i>Examples</i> | <i>Use</i> | <i>Category 1</i> | <i>Category 2</i> | <i>Recommended Approach</i> |
|---------------------------------------------------------|----------------------------|-------------------|-------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Non-chronic trend for adjusting financial outcome | Program evaluation | X | | Independent of disease management program; recommended for adjustment using historical relationship of chronic to non-chronic trend. |
| Comparative clinical measures across different programs | Comparison across programs | X | | Severity adjustment for baseline differences (pre-program) in differing populations is recommended. |
| Condition-specific utilization for insured population | Program evaluation | | X | As demographic shifts over time may significantly alter the prevalence of chronic conditions and associated utilization, prevalence adjustment is recommended. |
| Case mix of diseased or eligible population | Program evaluation | | X | Risk adjustment optional; if desired for significant pre/post differences, do it cautiously and conservatively. |

To provide more detail and to highlight some advantages and disadvantages of specific techniques, DMAA offers two case study examples of risk adjustment below. In both examples, the methods described were applied using actual health claims experiences. DMAA hopes these case studies can help guidelines users understand the importance and complexity of risk adjustment. DMAA is not, at this point, advocating either technique as a single appropriate methodology.

Case Study Example I

Use of Risk Adjustment Approaches in Disease Management Evaluation

Introduction: A widely used method of performing disease management program savings evaluations is the DMAA 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 at in the absence of the disease 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 disease 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 disease 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.

| GROUP | AVERAGE COST BASELINE PMPM | NUMBER OF MEMBER MONTHS BASELINE | PERCENT OF MEMBER MONTHS | AVERAGE COST YEAR 1 PMPM | NUMBER OF MEMBER MONTHS YEAR 1 | PERCENT OF MEMBER MONTHS |
|------------------|----------------------------|----------------------------------|--------------------------|--------------------------|--------------------------------|--------------------------|
| Terminating | \$929.75 | 31,407 | 11% | \$721.73 | 53,938 | 18% |
| Continuing | \$706.53 | 186,918 | 65% | \$623.80 | 180,522 | 60% |
| Newly Identified | \$601.71 | 70,571 | 24% | \$528.11 | 64,582 | 22% |
| Total | \$705.19 | 288,896 | 100% | \$620.80 | 299,042 | 100% |

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):

| | |
|----------------------------|----------|
| Baseline Chronic PMPM | \$705.19 |
| x Benchmark Trend | 1.05 |
| = Expected PMPM | \$740.45 |
| - Actual Intervention PMPM | \$620.80 |
| = Estimated Savings PMPM | \$119.65 |

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.

| GROUP | AVERAGE COST BASELINE PMPM | YEAR 1 MEMBER MONTH PERCENT | AVERAGE COST YEAR 1 PMPM |
|------------------|-------------------------------|--------------------------------|-----------------------------|
| Terminating | \$929.75 | 18% | \$721.73 |
| Continuing | \$706.53 | 60% | \$623.80 |
| Newly Identified | \$601.71 | 22% | \$528.11 |
| Total | \$705.19 | 100% | \$620.80 |
| Re-weighted | \$724.15 | | |

Accordingly, the savings concluded changes, as shown below.

Savings Calculation (adjusted):

| | |
|----------------------------|----------|
| Baseline Chronic PMPM | \$724.15 |
| x Benchmark Trend | 1.05 |
| = Expected PMPM | \$760.36 |
| - Actual Intervention PMPM | \$620.80 |
| = Estimated Savings PMPM | \$139.56 |

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:

| CHRONIC CONDITION | AVERAGE COST BASELINE PMPM | NUMBER OF MEMBERS BASELINE | AVERAGE COST YEAR 1 PMPM | NUMBER OF MEMBERS YEAR 1 |
|--------------------------------------|-------------------------------|-------------------------------|-----------------------------|-----------------------------|
| Asthma | \$587.43 | 5,426 | \$606.75 | 6,073 |
| CAD | \$521.51 | 2,151 | \$525.47 | 1,857 |
| CHF | \$574.26 | 533 | \$553.47 | 480 |
| COPD | \$595.08 | 1,696 | \$525.82 | 1,618 |
| Diabetes | \$495.50 | 7,554 | \$521.55 | 6,995 |
| Asthma & CAD | \$698.99 | 237 | \$688.11 | 196 |
| CAD & CHF & COPD & Diabetes | \$1,640.84 | 584 | \$1,440.18 | 472 |
| Asthma & CAD & CHF & COPD & Diabetes | \$1,799.27 | 382 | \$1,716.11 | 269 |
| Total | \$725.99 | 30,513 | \$697.04 | 28,590 |
| Re-weighted | \$709.94 | | | |

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 disease 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 disease 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 disease management program.

Case Study Example II

Lagged Prospective Approach

Introduction: To evaluate the financial outcome of a disease 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.

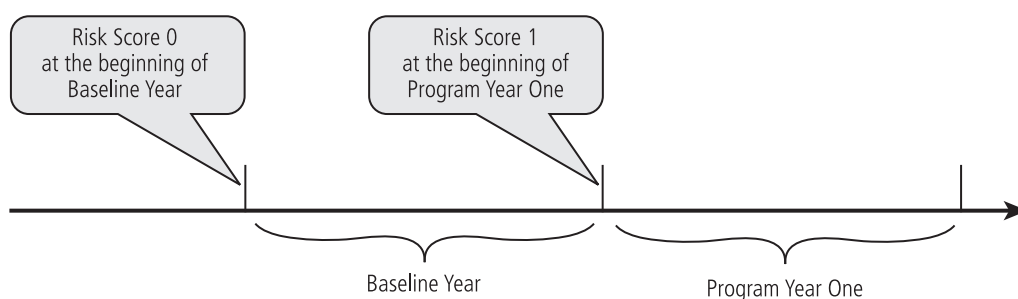
Methodology:

- 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 disease 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.

Figure 3 – Predictive Risk Score Calculation



This risk adjustment method can be used to evaluate the effect of disease 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.

| | <i>PDMPM COSTS</i> | <i>AVERAGE RISK SCORE</i> |
|-----------------|--------------------|---------------------------|
| Baseline Period | \$686.84 | 2.24 |
| Program Period | \$773.54 | 2.35 |

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 that there is 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 PDMPD in the baseline period is calculated this way:

$$\text{Risk-adjusted PDMPM in baseline} = \$686.84 * 2.35 / 2.24 = \$720.57$$

Then, this risk-adjusted PDMPM 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:

| | |
|-----------------------------|----------|
| Baseline Chronic PDMPM | \$720.57 |
| x Benchmark Trend | 1.10 |
| = Expected PDMPM | \$792.63 |
| - Actual Intervention PDMPM | \$773.54 |
| = Estimated Savings PDMPM | \$ 19.09 |

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.

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 disease 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 disease 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 disease management interventions. In larger populations, the impact of a few cost outliers on the measure variance does not have as significant an impact.

Table XIV, Small Populations, 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 disease management program, then repeatedly taking various sample sizes and computing the economic impact of the intervention for each sample.

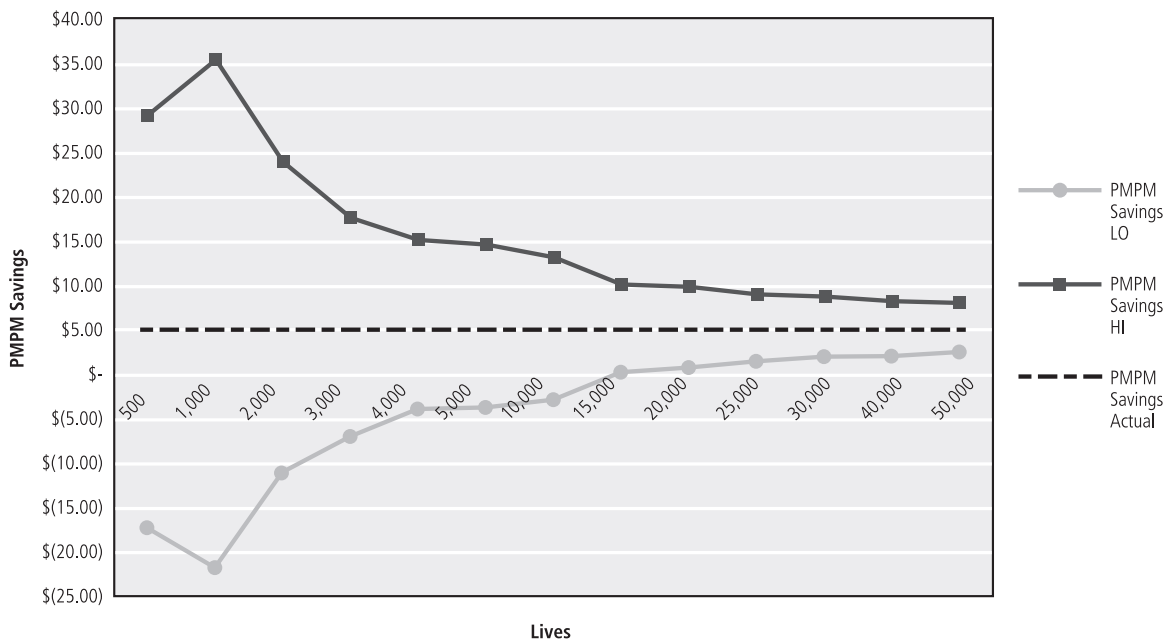
- The table shows values for the upper and lower confidence intervals on first year PMPM medical cost savings estimates in a population with a robust chronic condition management program in place.
- The savings estimates were derived from methods compliant with the DMAA Outcomes Guidelines Report.
- 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 XIV – SMALL POPULATIONS, PART I

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

*Savings for the population with chronic conditions is divided by the total population member months

CHART VI – SMALL POPULATIONS, PART II



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 versus 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 disease 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 samples 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 sized 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 disease 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 in later estimating outcomes.

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

Alternative 1 - 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 disease management program appears to save approximately \$5 per member per month. If another program using the identical system only enrolled 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 disease 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.

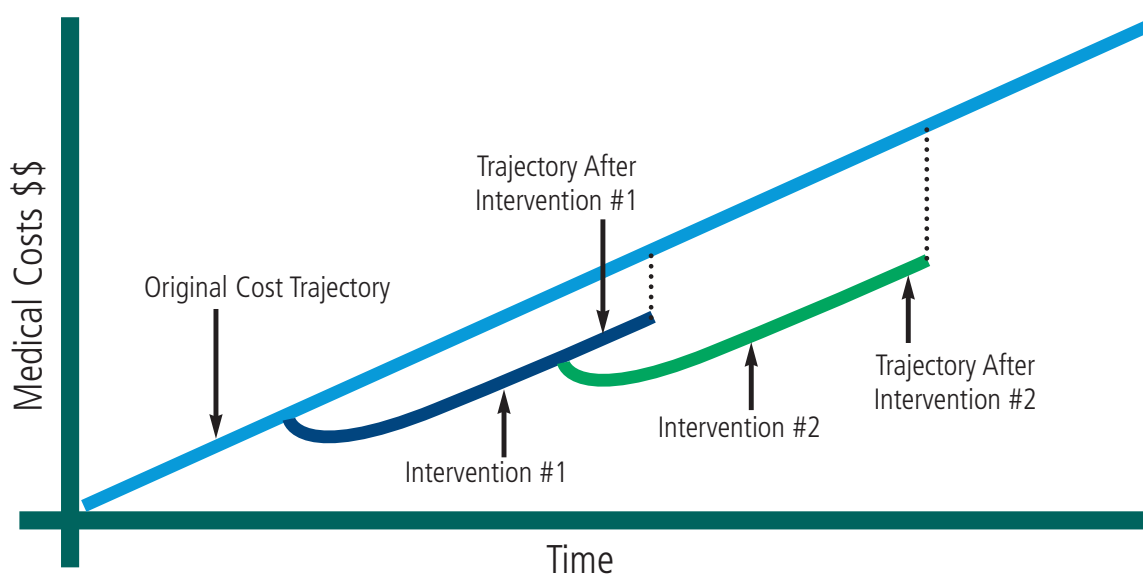
Long-Term Results Evaluation In Medical Management Programs

The DMAA 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 DMAA Long-Term Evaluation Workgroup developed a frequently asked questions (FAQ) section that reviews the topic in depth; it is included below. In addition, the workgroup included a graphic to illustrate the considerations reviewed in the FAQ section. The graphic is a simple example, not based on data analysis, but general experience and is intended to illustrate that a specific intervention does not continually keep bending 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 4 – 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 there and treat the new vendor's first year as "year one" for the program and have substantial ROI expectations?

A: Replacing a vendor with the intent of having "substantial ROI expectations" can be a treacherous decision. Replacing a vendor makes sense if:

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

If the vendor is ineffective in the basic task it sought to contract to perform, it indeed makes sense to change vendors. 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 may 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 vendor 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 vendor 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 vendor represents existing capabilities plus substantial paradigmatic change. If the first vendor was effective over several years and the second vendor 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 may 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 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 DMAA 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 quicker 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 disease 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.

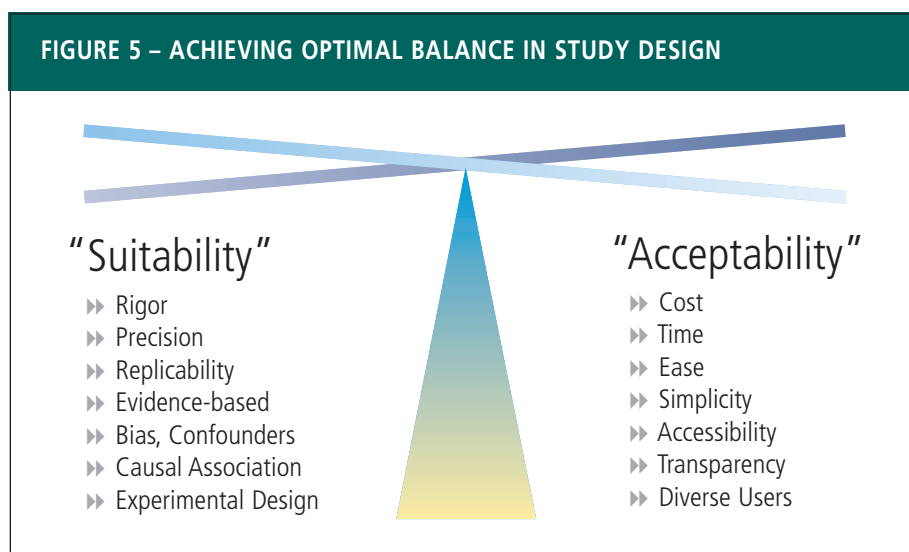
Measure Sets

Financial Measure

DMAA recognizes that disease management 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 disease management services, as each party may 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 DMAA 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?



The DMAA 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, DMAA 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 disease management settings, DMAA'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,

DMAA 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 disease management program per predefined eligibility criteria) should be used to represent gross and net health care cost savings.
- DMAA recognizes that return on investment will inevitably be computed by decisionmakers, but that this should not be the primary financial metric for disease management 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 disease 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

DMAA 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 the 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 disease 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 disease 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 disease 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 disease management programs, but insensitive to other possible program impacts.
 - Does not measure disease management impact on comorbidities.
 - Sensitive to condition prevalence changes, so prevalence adjustment required.
 - May serve as “end to end” test of disease management identification, outreach, enrollment, engagement, impact, retention, etc.

Other utilization measures which may be of interest to purchasers and suppliers of disease management programs may be collected and reported at the option and agreement of the parties. Table XV, page 83, “Comparison of Various Utilization Measures,” elaborates on several of these measures.

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.
- Disease management 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 the population health management. Group II includes measures that are in wide use in disease management programs, but for which national consensus on measure specification is not yet available (e.g., sodium intake self-monitoring by heart failure patients).

TABLE XV – COMPARISON OF VARIOUS UTILIZATION MEASURES

| <i>Measure</i> | <i>Numerator</i> | <i>Denominator</i> | <i>Utility/Comments</i> |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| All-cause utilization of diseased population | All admissions for any cause for the period | Member mos. for the defined disease management-eligible or diseased population | <ul style="list-style-type: none"> • Relatively sensitive for major cost driver impacted by disease 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 disease management-eligible or diseased population | <ul style="list-style-type: none"> • More sensitive for expected key impact of disease management programs but does not measure disease 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 for the period for which principal diagnosis is condition in question | Member mos. for the entire insured or covered population | <ul style="list-style-type: none"> • Specific for one expected impact of disease management programs but insensitive to other possible program impacts • Does not measure disease management impact on comorbidities • Sensitive to condition prevalence changes, so prevalence adjustment required • May serve as “end to end” test of disease management identification, outreach, enrollment, engagement, impact, retention, etc. |
| Condition-specific utilization of entire population – known as “plausibility indicators” by DMPC (Variant 2 – two separate measures) | Admissions and ER visits (separately) for the period for which principal diagnosis is condition in question | Member mos. for the entire insured or covered population | <ul style="list-style-type: none"> • Same as for Variant 1 • Adds ER visits, which reflect additional disease management impact beyond admissions reductions |
| Condition-specific utilization of entire population – (Variant 3 – one combined measure) | Combined Admissions and ER 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 | <ul style="list-style-type: none"> • Same as for Variant 1 • Adds ER visits to admissions in a single weighted aggregate measure that may reflect additional disease management impacts beyond just 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 | <ul style="list-style-type: none"> • The most inclusive measure of utilization but with attribution problems • May be insensitive to disease 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) | <ul style="list-style-type: none"> • Relatively sensitive for major cost driver impacted by disease management programs • Measures impact on comorbidities as well as primary conditions of interest • Does not test accuracy of identification of diseased population |

GROUP I MEASURES

Asthma

- ▶▶ Flu vaccination
- ▶▶ Pneumococcal vaccination
- ▶▶ Smoking cessation identification & advice
- ▶▶ Current medication use
 - ▶▶ Controller medication

COPD

- ▶▶ Flu vaccination
- ▶▶ Pneumococcal vaccination
- ▶▶ Smoking cessation identification & advice
- ▶▶ Spirometry evaluation
- ▶▶ Medication use
 - ▶▶ Bronchodilator

Heart Failure

- ▶▶ Flu vaccination
- ▶▶ Pneumococcal vaccination
- ▶▶ Smoking cessation identification & advice
- ▶▶ Medication persistence
 - ▶▶ beta blockers
 - ▶▶ ACE/ARB
 - ▶▶ anticoagulants (with chronic or proxysmal AF)

Coronary Artery Disease

- ▶▶ Flu vaccination
- ▶▶ Pneumococcal vaccination
- ▶▶ Smoking cessation identification & advice
- ▶▶ LDL testing & 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 & advice
- ▶▶ Daily aspirin use
- ▶▶ LDL testing & control
 - ▶▶ annual test
 - ▶▶ LDL < 100; < 130
- ▶▶ HbA1c testing & control
 - ▶▶ annual test
 - ▶▶ HbA1c < 7.0; > 9.0
- ▶▶ Blood pressure
 - ▶▶ blood pressure < 130/80; < 140/90
- ▶▶ Eye Exam
- ▶▶ Nephropathy testing

GROUP II MEASURES

▶▶ Continue to work with NCQA on development of measures identified for the upcoming year. 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

DMAA 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 disease 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.

DMAA 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 disease 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. While acknowledging their importance, the decision was made at this time to focus only on the three metrics identified by the group and supported by the literature. These are highlighted in the list below and will be developed through a white paper on the topic that will be released in early 2009 on the DMAA Web site.

- 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

For many years, the health care system has recognized the importance of adherence to pharmacotherapy regimens. Many studies have been published that emphasize the consequences associated with non-adherent behaviors, such as higher overall health care costs, reduced quality of life and increased mortality rates. Despite the topic's magnitude, health management was once absent a standardized methodology to measure and compare medication adherence across managed populations.

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

MEDICATION POSSESSION RATIO

Medication possession ratio (MPR) is an operational definition of adherence: a retrospective assessment of days supplied over an evaluation period¹⁶. 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 \leq 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, following the methodology descriptions.

-
- **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 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.

MEDICATION PERSISTENCE

Medication persistence can be defined as the “amount of time that an individual remains on chronic drug therapy.”¹⁷ 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, following the methodology descriptions.
 - **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.

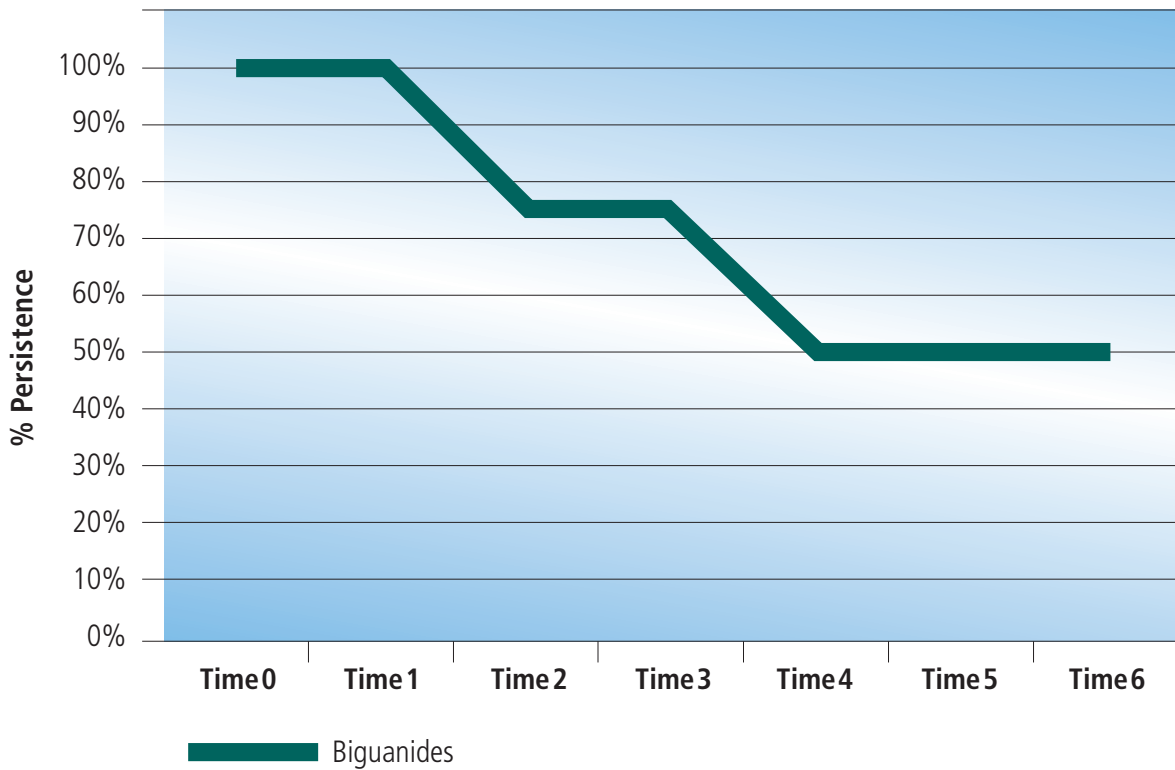
EXAMPLES

Chart VII – Persistence Example

| | Time0 | Time1 | Time2 | Time3 | Time4 | Time5 | Time6 | |
|------|-------|-------|-------|-------|-------|-------|-------|-----------------------------------------|
| Mbr1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | Eligible at year start; 100% persistent |
| Mbr2 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | Eligible at year start; 50% persistent |
| Mbr3 | 1 | 1 | 1 | 1 | – | – | – | Eligible mid-year; 100% persistent |
| Mbr4 | 1 | 1 | 0 | 0 | – | – | – | Eligible mid-year; 33% persistent |
| P% | 100% | 100% | 75% | 75% | 50% | 50% | 50% | |

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

Figure 6 – Diabetes Persistence Curve



Frequently Asked Questions

Question: What are medication possession ratio (MPR) and persistence meant to measure?

Answer: 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.

.....

Question: What are commonly recurring barriers to optimal MPR and persistence?

Answer: 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.

.....

Question: Why is measurement more variable with inhaled medications (e.g. asthma)?

Answer: 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 days supply field's accuracy is quite 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 may explain periods of adherence versus non-adherence (when a person is “off-season”).

.....

Question: When might be considered a “triggering point” for interventions regarding MPR and persistence?

Answer: Many of the effectiveness studies performed by the pharmaceutical industries have focused on MPR \geq 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 may represent a point to determine adherent/nonadherent cohorts while retaining parity between sensitivity and specificity.¹⁸ 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.

.....

Question: Although the recommendations call for annual measurement with appropriate claims run-out, could my organization calculate on an ongoing basis or quarterly?

Answer: 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.

.....

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

Answer: 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, prior blood glucose readings, etc. The adherence workgroup, from a complexity reduction standpoint, decided to approach the more straightforward forms to vet these standardized methods.

.....

Question: How do you define drug class?

Answer: 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.

.....

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

Answer: 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.

.....

Question: How do discount drug programs impact these measurements?

Answer: The growth of discounted prescription drug programs may 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 where a benefit card is not presented and an administrative claim is not generated.

.....

Question: How does the drug class-level MPR methodology account for concomitant therapy and medication switches?

Answer: 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.

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

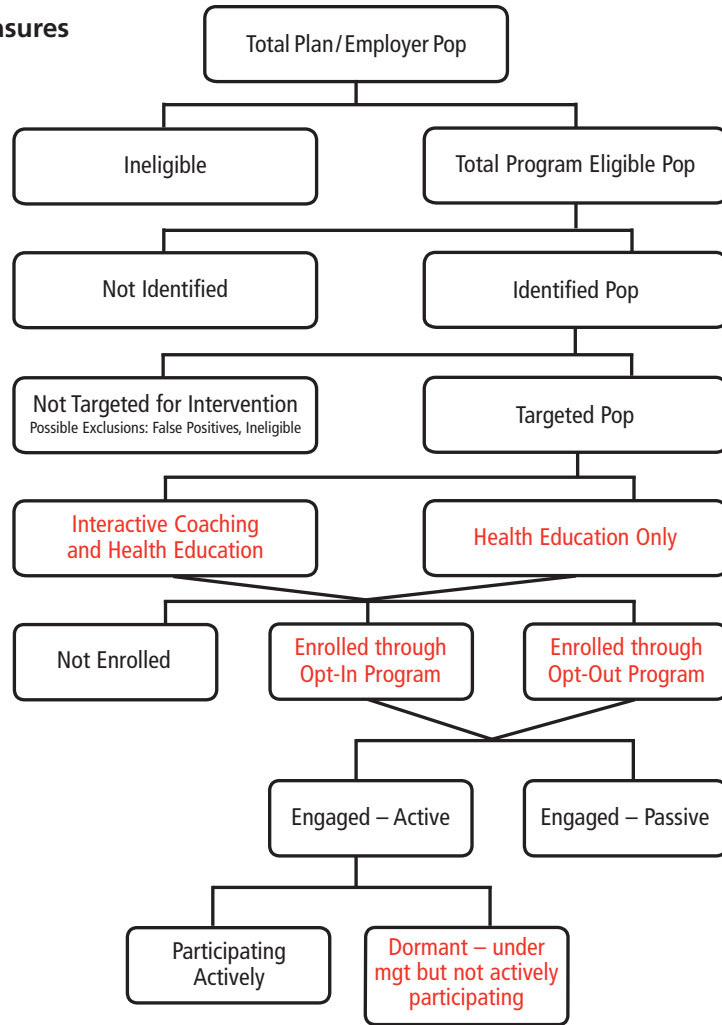
Answer: If an individual has claims for two drug classes for a condition, the calculation of condition-level MPR must make an assumption regarding whether the situation involves concomitant therapy or a switch in therapy. If the methodology assumes concomitant therapy, it will accumulate each drug's 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.

.....

Operational Measures

Understanding how a program identifies, enrolls, outreaches, intervenes and engages participants is important to customers of these programs and an important part of any program evaluation. Although important, definitions of these concepts vary significantly throughout population health management. DMAA has refined the operational flow diagram released in Outcomes Guidelines Report Volume 3 (page 58). The goal of the flow diagram is to accurately represent a typical program process along with definitions and possible measures for each of the identified stages of the process. The majority of disease management programs will follow this flow, but DMAA recognizes that, based on the population, program goal and other variables, programs may follow a slightly different operational process. In addition, possible measures identified are the most common and may or may not be used in an evaluation process. DMAA intends to continue this work in 2010 with the goal of additional refinement to the diagram and a recommended set of clear and concise measures that represent each stage in the operational program process.

Figure 7 – Operational Measures Flow Diagram



| | | |
|-----------------------|---|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ELIGIBLE POPULATION | → | Total program eligible population – Individuals meeting benefit eligibility above and eligibility requirements established by the plan to be considered for the DSM care management program (e.g. covered members > 18 yrs old) Total DSM program eligible population = health plan eligible and/or program eligible |
| IDENTIFIED POPULATION | → | Identified Population – DSM program-eligible members who are identified as having a qualified condition. Possible measures: number of identified members by condition, stratification, prevalence rate, incidence rate |
| TARGETED POPULATION | → | Targeted Population – Members who are identified as having a qualified condition, and who are targeted for program intervention. Possible measures: number of targeted members, target rate |
| ENROLLMENT | → | <i>Opt-in program – Enrolled</i> = members consenting to participate in the program <i>Opt-out program – Enrolled</i> = targeted population considered enrolled, unless action taken to disenroll Possible measures: <i>Opt-in program</i> – reach/contacted rate, consented rate, declined rate, approached/uncertain number, unable to reach rate; <i>Opt-out program</i> – opt-out rate. |
| ENGAGEMENT | → | Engagement: <i>Passive engagement</i> strategies do not require active member consent, or acknowledgement of program participation; <i>Active engagement</i> strategies do require active member consent, or acknowledgement of program participation. Possible measures: <i>Active engagement</i> – active inbound interventions, active outbound interventions; <i>Passive engagement</i> – passive outbound interventions |
| PARTICIPATION | → | <i>Opt-in program – Participating</i> = Enrolled members who are actively participating in DSM program, and typically broken down by stratification level and/or level of engagement and intervention. <i>Opt-out program – Participating</i> = Enrolled members who have not taken action to discontinue intervention, and typically broken down by stratification level and/or level of engagement and intervention. Possible measures: total participation rate, active participation rate, average amount of time as active participant, percentage moving from active to passive engagement, percentage of participants moving to another level of stratification |

Additional Measures

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

Additional metrics addressed 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 disease management programs.

Health Status

DMAA 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 onscreen. 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

DMAA recommends that participant satisfaction measures be incorporated into the evaluation of disease 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 disease management programs. As a result, DMAA 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 disease management program.
 - Nature of condition.
 - Respondent demographics.
- Program-Specific Module (optional)
 - Biometric Monitoring.
 - Other program-specific assessments.

There are several versions of the Participant Satisfaction Survey available. More information on the DMAA Participant Satisfaction Survey can be found on DMAA’s Web site, www.dmaa.org.

References

- ¹ Campbell D, Stanley J. *Experimental and quasi-experimental design for research*. Boston. Houghton Mifflin Company. 1963.
- ² Cook T, Campbell D. *Quasi-experimentation: design & analysis issues for field settings*. Boston, New York. Houghton Mifflin Co. 1979.
- ³ Bandura A, Strunk B. Cultivating competence, self-efficacy and intrinsic interest through proximal self-motivations. *J Pers Soc Psychol*. 1981;41:586-598.
- ⁴ Bandura A. Social cognitive theory: an agenic perspective. *Annual Review of Psychology*. 2001;52:1-26.
- ⁵ Prochaska JO, DiClemente CC, Norcross JC. In search of how people change. *Am Psychol*. 1992;47:1102-1114.
- ⁶ De Vet E, De Nooijer J, De Vries NK, Brug J. The transtheoretical model for fruit, vegetable and fish consumption: associations between intakes, stages of change and stage transition determinants. *Int J Behav Nutr Phys Act*. 2006 Jun 19;3(1):13.
- ⁷ Allison MJ, Keller C. Self-efficacy intervention effect on physical activity in older adults. *Western Journal of Nursing Research*. 2004;26(1):31-46.
- ⁸ Perkins S, Jenkins LS. Self-efficacy expectation, behavior performance, and mood status in early recovery from percutaneous transluminal coronary angioplasty. *Heart and Lung*. 1998;27:37-46.
- ⁹ Thomas D, Elliott E, Naughton G. Exercise for type 2 diabetes mellitus. *Cochrane Database Syst Rev*. 2006 Jul 19;3:CD002968.
- ¹⁰ Pelkman CL, Fishell VK, Maddox DH, Pearson TA, Mauger DT, Kris-Etherton PM. Effects of moderate-fat (from monounsaturated fat) and low-fat weight-loss diets on the serum lipid profile in overweight and obese men and women. *Am J Clin Nutr*. 2004 Feb;79(2):204-12.
- ¹¹ Hoffman MD, Hoffman DR. Does aerobic exercise improve pain perception and mood? A review of the evidence related to healthy and chronic pain subjects. *Curr Pain Headache Rep*. 2007 Apr;11(2):93-7.
- ¹² Holme I, Haaheim LL, Tonstad S, Hjermann I. Effect of dietary and antismoking advice on the incidence of myocardial infarction: A 16-year follow-up of the Oslo Diet and Antismoking Study after its close. *Nutr Metab Cardiovasc Dis*.
- ¹³ Barefoot C, Helms MJ, Mark DB, Blumenthal JA, Califf RM, Haney TL, et al. Depression and long-term mortality risk in patients with coronary artery disease. *Am J Cardiol*. 1996;78:613-617.
- ¹⁴ Koopman C, et al. Stanford Presenteeism scale: health status and employee productivity. *Journal of Occupational and Environmental Medicine*. 2002;44(1):14-20.
- ¹⁵ For example, see Schultz AB and Edington DW. Employee health and presenteeism: a systematic review. *Journal of Occupational Rehabilitation*. 2007;17(3):547-579.
- ¹⁶ Cramer JA, Roy A, Burell A, et al. Medication compliance and persistence: terminology and definitions. *Value Health*. 2008;2(1):44-47.
- ¹⁷ Sikka R, Xia F, Aubert RE. Estimating medication persistency using administrative claims data. *Am J Manag Care*. 2005 Jul;11(7):449-456.
- ¹⁸ Hansen R, et al. Comparison of methods to assess medication adherence and classify nonadherence. *The Annals of Pharmacotherapy*. 2009;43(3):413-422.