



Comparative Effectiveness Research and Personalized Medicine: From Contradiction to Synergy

Prepared for: Personalized Medicine Coalition

Clifford Goodman, PhD
The Lewin Group Center for Comparative Effectiveness Research

October 28, 2009

ABOUT THE LEWIN GROUP CENTER FOR COMPARATIVE EFFECTIVENESS RESEARCH

The Lewin Group Center for Comparative Effectiveness Research manages, conducts, interprets and supports the use of comparative effectiveness research to make informed decisions regarding patient outcomes, health care policy and the optimal use of health care resources.

The Lewin Group is a premier national health care and human services consulting firm that has delivered objective analyses and strategic counsel to prominent public agencies, nonprofit organizations, industry associations and private companies across the United States for nearly 40 years. In keeping with our tradition of objectivity, The Lewin Group is not an advocate for or against any legislation. The Lewin Group is part of Ingenix, Inc., which is a wholly owned subsidiary of UnitedHealth Group. To assure the independence of its work, The Lewin Group has editorial control over all of its work products. Go to www.lewin.com/integrity for more information.

The Lewin Group is widely recognized for its expertise in the U.S. health care system, its record in health technology assessment and evidence-based medicine; health care policy and other issues, and its independence and objectivity. In the Lewin Center, we are combining these strengths with the unique capabilities of our sister company, i3, in clinical trials and study design, health economics and outcomes research; and through Ingenix, access to one of the world's largest and most robust patient data sets including de-identified, integrated diagnosis, medical, and pharmacy claims data for millions of individuals. Learn more at www.lewin.com/CER.

Comparative Effectiveness Research and Personalized Medicine: From Contradiction to Synergy

October 28, 2009

Prepared for:
Personalized Medicine Coalition

Submitted by:
The Lewin Group
Center for Comparative Effectiveness Research

EXECUTIVE SUMMARY

Two high-profile health care issues are the emerging delivery paradigm of personalized medicine (PM) and the form of evaluation known as comparative effectiveness research (CER). While the purpose of CER is to determine which health care intervention works best for a given health care problem, the purpose of PM is to ensure that health care delivers ‘the right treatment to the right patient at the right time.’ Both are intended to support high-quality, evidence-based decisions for optimal patient care. However, like most other forms of evaluation of health care interventions, CER is usually oriented toward evaluating treatment effects across study populations, while PM focuses on using individuals’ genomic information and other personal traits to inform decisions about their health care.

Interventions that have a statistically significant treatment effect across a population on average do not necessarily work for all treated patients; they may be ineffective for some patients and harmful for others. Other interventions that do not have a statistically significant treatment effect across a study population—and that may be dismissed as ineffective—may work for certain subgroups of the population.

If CER does not investigate important differences in patient response to interventions—for example, whether patient response to a cancer drug varies by certain genetic characteristics—its findings may be inadequate or misleading for patient care. This could have extended consequences if these findings are incorporated into product labeling, practice guidelines, reimbursement policies, or utilization management that could curtail PM. If CER does examine how well interventions work in patient subgroups, the resulting evidence can be used in more flexible, adaptive guidelines and policies that would better enable PM.

For CER to contribute to PM, it must account for patient differences that influence the impact of interventions on health outcomes. These characteristics can include severity of disease, comorbidities and risk factors, genetic characteristics, sociodemographic characteristics, health-related behaviors, environmental factors, and more. The variable impacts on patient outcomes, including health benefits and harmful side effects that can arise from these different patient characteristics, are sometimes known as “heterogeneity of treatment effects” (HTEs).

Aligning CER and PM means that PM is subject to prevailing evidence requirements for screening, diagnostic, therapeutic, and other interventions. For genetic and genomic testing, health professional groups, guideline panels, and payers are calling not only for rigorous evidence of test accuracy but for evidence of clinical utility, that is, impact of test results on clinical decisions and, ultimately, patient outcomes. This applies, for example, to gene expression profile testing to predict breast cancer outcomes, pharmacogenomic testing for guiding treatment for depression, and selecting treatments for colorectal cancer.

Full alignment of PM and CER depends on adoption of health information technology (HIT). In conducting CER, electronic health records (EHRs) can capture patients’ genetic and other individual health information in the course of routine health care, clinical trials, and other studies. In translating CER to practice, clinical decision support systems, EHRs, and other HIT can ensure that evidence pertaining to PM is present and actionable at the point of health care decisions. HIT’s considerable potential for serving PM is far from being realized.

The American Recovery and Reinvestment Act of 2009 provided an unprecedented \$1.1 billion boost to the federal investment in CER. This major national commitment to CER has certain encouraging implications for PM. CER priority setting reports of the Federal Coordinating Council for Comparative Effectiveness Research and Institute of Medicine and pending legislation emphasize the need for subgroup analyses. The Agency for Healthcare Research and Quality is sponsoring an analysis of how well comparative effectiveness studies conducted by CER agencies in the U.S. and abroad have accounted for HTEs.

CER is influencing innovation in PM, including enabling new opportunities and diminishing prospects for some less likely to fare well in a market informed by head-to-head comparisons. Federal support for CER and related methods, infrastructure, and training could reduce development costs of some PM interventions. Led by global pharmaceutical and biotechnology companies that have been responding to evolving evidence requirements in international markets, many in industry are incorporating CER and PM considerations into their R&D.

Communications and applications of CER findings and other evidence must be adaptive and targeted to clinicians, patients, payers, and the public. These messages should address strengths and limitations of this evidence, how specific it is for patient subgroups, and evidence gaps that are priorities for further CER.

The signals approaching the intersection of CER and PM are clear:

- The design and conduct of CER must consider and account for potential differences in response by subgroups of patients.
- The strengths and limitations of CER findings and other evidence, including whether it accounts for HTEs as opposed to an average effect across a population, must be accurately reflected in product labeling, guidelines, payment policies, utilization management, and other gatekeeping policies.
- To enable evidence-based PM, these gatekeeping policies must be flexible, adaptive, and updated as needed.
- Generally higher and more specific evidence requirements for health technologies apply to PM interventions as well, with implications for their adoption, use, and payment.
- The ability of CER to contribute to PM on any systematic and ongoing basis depends on HIT, particularly in the form of EHRs and clinical decision support systems.
- CER offers opportunities for innovation in PM, along with inevitable shakeouts. Funding for CER and related methods development, data sources, and infrastructure should boost innovation. Technologies that achieve prevailing evidence requirements and demonstrate comparative or superior effectiveness will gain market advantages.
- Current promising signs for CER and PM alignment include explicit attention to PM in recommended national priorities for CER, pending legislation to sustain the national investment in CER, and development of CER methods and research infrastructure.

Whether CER and PM will be aligned or opposed is now unfolding. CER and PM offer complementary advantages of great potential. In a stressed health care system poised for reform, a continued, concerted effort is necessary to ensure that this potential is realized.

INTRODUCTION

Two high-profile health care issues are the emerging delivery paradigm of personalized medicine (PM) and the form of evaluation known as comparative effectiveness research (CER). While the purpose of CER is to determine which health care intervention works best for a given health care problem, the purpose of PM is to ensure that health care delivers ‘the right treatment to the right patient at the right time.’ Both are intended to support high-quality, evidence-based decisions for optimal patient care. However, like most other forms of evaluation of health care interventions, CER is usually oriented toward evaluating treatment effects across study populations, while PM focuses on using individuals’ genomic information and other personal traits to inform decisions about their health care.

Federal Initiative in CER

The Medicare Prescription Drug, Improvement, and Modernization Act of 2003 (MMA) authorized \$50 million for “comparative clinical effectiveness” at the federal Agency for Healthcare Research and Quality (AHRQ) and “such sums as necessary” for later years. However, actual appropriations were lower. This initiative was funded at \$15 million annually for the fiscal years 2005, 2006 and 2007. Funding rose to \$30 million for fiscal year 2008 and to \$50 million as a result of the Omnibus Appropriations Act of 2009. In February 2009, the American Recovery and Reinvestment Act of 2009 (ARRA) greatly increased federal funding for CER, appropriating \$1.1 billion to “accelerate the development and dissemination of CER of health care treatments and strategies.” This included \$300 million for the Agency for Healthcare Research and Quality (AHRQ), \$400 million for the National Institutes of Health (NIH), and \$400 million for the Secretary of HHS.

ARRA mandated the Secretary of HHS to contract with the Institute of Medicine (IOM) to produce and submit a report to the Congress and the Secretary by June 30, 2009, that included recommendations on national priorities for CER to be conducted or supported with the ARRA funds.^{1,2} ARRA also established a Federal Coordinating Council for Comparative Effectiveness Research (FCC) comprising senior federal officers from HHS and other agencies and called for it to submit a report by June 30, 2009, to the President and the Congress containing information describing current federal activities on CER and recommendations for CER to be conducted or supported from ARRA funds.³

In support of their deliberations, the IOM and FCC held public events and provided other means for gaining stakeholder comments on CER priorities and processes. Among the many issues raised, some stakeholders called for CER to account for and serve the needs of PM. Appropriately, the June 30, 2009, reports on CER from the IOM and FCC reflected this input:

¹ H.R. 1, American Recovery and Reinvestment Act of 2009. Washington, DC: 111th Congress of the United States of America, House of Representatives. Accessed March 15, 2009 http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=111_cong_bills&docid=f:h1enr.pdf.

² Institute of Medicine. Initial National Priorities for Comparative Effectiveness Research. Washington, DC: The National Academies Press, 2009.

³ Federal Coordinating Council for Comparative Effectiveness Research. Report to the President and the Congress. Washington, DC: U.S. Department of Health and Human Services, 2009.

With the growing knowledge of disease mechanisms, systems biology, genomics, and other sciences that create the potential for more targeted therapies, patients and providers are increasingly seeking evidence not only from representative populations, but also from relevant subgroups. Increasing emphasis on patient-level attributes that may modify the balance of benefits or harms can lead to more personalized medicine, reducing the pressure to try alternatives found to be ineffective in similar subgroups. — Institute of Medicine⁴

In addition, comparative effectiveness should complement the trend in medicine to develop personalized medicine – the ability to customize a drug and dose based on individual patient and disease characteristics. One of the advantages of large comparative effectiveness studies is the power to investigate effects at the sub-group level that often cannot be determined in a randomized trial. This power needs to be harnessed so personalized medicine and comparative effectiveness complement each other. — Federal Coordinating Council for Comparative Effectiveness Research⁵

The relationship between CER and PM is the subject of pending legislation as well. The Patient-Centered Outcomes Research Act of 2009, in support of the conduct of comparative effectiveness research, introduced by Senators Max Baucus and Kent Conrad, specifies the importance of incorporating PM into CER:

Taking into account potential differences. – Research shall – (i) be designed, as appropriate, to take into account the potential for differences in the effectiveness of health care treatments, services, and items as used with various subpopulations, such as racial and ethnic minorities, women, age, and groups of individuals with different comorbidities, genetic and molecular subtypes, or quality of life preferences; and (ii) include members of such subpopulations as subjects in the research as feasible and appropriate.⁶

Although the federal initiative in CER clearly recognizes the need for CER to address PM, considerable work remains to adapt the methods used in CER for this purpose.

THE CONCEPTS OF CER and PM

CER is intended to address some of the substantial evidence gaps that exist in U.S. health care. It can involve all types of interventions, including drugs, biologics, tests, imaging, and medical and surgical procedures, as well as health care organization, delivery, and financing. In some respects, CER represents the latest form of generating statistically-driven health care evidence for health care decisions and policies. By establishing what is the most effective intervention for a given health problem, CER has the potential to improve patient outcomes and improve cost-effectiveness of health care.

Among the various definitions of CER is the one used by the IOM in its June 2009 report on recommended national priorities for CER:

⁴ Institute of Medicine. Initial National Priorities for Comparative Effectiveness Research, 2009.

⁵ Federal Coordinating Council for Comparative Effectiveness Research. Report to the President and Congress. U.S. Department of Health and Human Services, June 30, 2009.

⁶ S. 1213, Patient-Centered Outcomes Research Act of 2009. Washington, DC: 111th Congress of the United States of America, Senate.

[T]he generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care. The purpose of CER is to assist consumers, clinicians, purchasers, and policy makers to make informed decisions that will improve health care at both the individual and population levels. CER's distinguishing characteristics include informing a specific clinical or policy decision, comparing at least two approaches or interventions, describing results at the subgroup level, measuring benefits in real-world populations, and applying appropriate methods and data sources. — Institute of Medicine⁷

The IOM definition addresses PM, i.e., “improve health care at both the individual and population level” and “describing results at the subgroup level.” The definition of CER used by the FCC addresses evidence about “which interventions are most effective for which patients under specific circumstances” and “diverse patient populations and subgroups.”⁸

PM refers to the use of information about individuals’ personal traits—including their genomes, health states, and behavioral, environmental, socioeconomic, cultural, and other personal determinants of response to health care interventions—to better manage their disease or disease risk. Among the various definitions of CER is the one used by the President’s Council of Advisors on Science and Technology:

“Personalized medicine” refers to the tailoring of medical treatment to the individual characteristics of each patient. It does not literally mean the creation of drugs or medical devices that are unique to a patient but rather the ability to classify individuals into subpopulations that differ in their susceptibility to a particular disease or their response to a specific treatment. Preventive or therapeutic interventions can then be concentrated on those who will benefit, sparing expense and side effects for those who will not. — President’s Council of Advisors on Science and Technology⁹

In some respects, physicians always have sought to practice PM by treating one patient at a time, with more or less consideration for each patient’s personal and family history and other relevant individual circumstances. Drug regimens have been tailored to individual patients’ biomarkers, such as weight, blood pressure, and cholesterol levels. But for the most part, physicians’ tools have been constrained, as has their ability to account for all of patients’ personal factors that might affect management of their health care problems.

An important emerging tool for PM is pharmacogenomics (PGx), which is the study of how individual genetic differences affect drug response.¹⁰ Although the sequencing of the human genome has had only a modest impact on clinical practice to date, PGx has begun to offer powerful tools for applying information about individual genetic variations and drug response for health care decisions.^{11,12} By offering clinicians a new set of diagnostic tools and tests to

⁷ IOM 2009.

⁸ Federal Coordinating Council 2009.

⁹ President’s Council of Advisors on Science and Technology. *Priorities for Personalized Medicine*. September 2008. Available at: http://www.ostp.gov/galleries/PCAST/pcast_report_v2.pdf

¹⁰ PGx encompasses inter-individual genetic differences such as variation in DNA sequence, gene expression, and copy number related to an individual’s metabolism of drugs (pharmacokinetics) or physiological response to drugs (pharmacodynamics). DHHS, Secretary’s Advisory Committee on Genetics and Health in Society. *Realizing the Potential of Pharmacogenomics: Opportunities and Challenges*. May 2008. http://www4.od.nih.gov/oba/SACGHS/reports/SACGHS_PGx_Report.pdf.

¹¹ Hopkins MM, Ibarreta D, Gaisser S, et al. Putting pharmacogenetics into practice. *Nat Biotechnol* 2006;24(4):403-10.

assess risks and benefits associated with existing medicines for particular patients, PGx can help physicians determine what may be the most suitable interventions for a specific patient.^{13,14} Further, the sequencing of the human genome and advances in PGx are being used in design and development of new therapies and regimens.

PM is essential for pursuing the goals of CER. For example, some commonly used drugs prescribed in the U.S. today are effective in fewer than 60% of treated patients.¹⁵ The fact that a large number of widely used drugs work well in some people and not in others suggests that personal or subgroup differences are a large contributor to effectiveness, and that the concept of comparative effectiveness is of limited relevance without considering these differences.

THE TROUBLE WITH AVERAGES

[W]hen you're talking about personalized medicine, what's important is not an average effect that you estimate for a whole population, but personalized evidence for an individual patient or type of patients. Personalized medicine requires personalized evidence. — Mark McClellan¹⁶

Like other research on health care technologies, CER typically focuses on average treatment effects. The estimated average treatment effect reported in a clinical trial or other study is the average of the individual treatment effects across the population of patients in the study. However, for a given health problem, individuals may vary in the magnitude and type of response to a particular treatment. Interventions that yield a statistically significant treatment effect across a study population may not necessarily work for all treated patients; they may be ineffective for some patients and harmful for others¹⁷ In some instances, virtually all patients may benefit from a treatment, though in varying degrees. In other instances, only some subgroups of patients benefit and some experience only side-effects.¹⁸ Interventions that do not yield a statistically significant treatment effect across a study population—and that may be dismissed as ineffective—actually may work for certain subgroups of the population.

[T]he benefit or harm of most treatments in clinical trials can be misleading and fail to reveal the potentially complex mixture of substantial benefits for some, little benefit for many, and harm for few.”¹⁹

¹² Schmedders M, van Aken J, Feuerstein G, et al. Individualized pharmacogenetic therapy: a critical analysis. *Community Genet* 2003;6(2):114-9.

¹³ Melzer D, Raven A, Detmer DE, et al. *My very own medicine: what must I know? Information policy for pharmacogenetics.* Cambridge, UK: University of Cambridge, Department of Public Health and Primary Care, 2003.

¹⁴ Ginsburg GS, Angrist M. The future may be closer than you think: a response from the Personalized Medicine Coalition to the Royal Society's report on personalized medicine. *Fut Med* 2006;3(2):119-23.

¹⁵ Spear BB, Heath-Chiozzi M, Huff J. *Trends Mol Med.* 2001 May; 7(5):201-4.

¹⁶ McClellan M. Meeting the Healthcare Challenge: The Potential of Personalized, Evidence-Based Medicine. Conference: 21st Century Medicine: Personalized and Evidence-Based. Washington, DC, September 18, 2007.

¹⁷ To assume that a relationship between variables found at an aggregate level (e.g., for a given population) also applies at an individual level is known as the “ecological fallacy.”

¹⁸ Gabler NB, Duan N, Liao D, et al. Dealing with heterogeneity of treatment effects: is the literature up to the challenge? *Trials* 2009;10:43.

¹⁹ Kravitz RL, Duan N, Braslow J. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *Milbank Q* 2004;(82):661-87.

While a randomized controlled trial (RCT) can establish which of two alternative ACE inhibitors is most likely to benefit the average patient with hypertension, a substantial percentage of patients who are prescribed that drug might not respond, or could experience adverse reactions. Reflecting differences in patient response, most drugs work well for some people but not for others, and drugs that have similar average effects often have very different benefit and risk profiles across patients. Many widely prescribed medications that are effective on average in large patient populations are not effective for all who use them.^{20,21,22,23}

Clinical practice guidelines that incorporate such trial results without accounting for these differences run the risk of standardizing interventions that are suboptimal or potentially harmful to some patient subgroups. On the other hand, decisions based on average effects derived from rigorous RCTs are likely to be better founded, on average, than the many decisions based on inadequate or absent evidence today.²⁴

CER studies to date have rarely accommodated the collection and reporting of genomic, behavioral, environmental, and other individual patient differences, such as adverse effects, and real-world issues of patient noncompliance. This omission stems, in part, from the inherent nature of various methodological approaches used to produce evidence of comparative effectiveness and the time, costs, and other hurdles associated with collecting and analyzing sufficient data at the subpopulation level to yield clinically and statistically significant findings.

Heterogeneity of Treatment Effects: Accounting for Subgroup and Individual Differences

The variable response to a given treatment by patients with different characteristics is known as heterogeneity of treatment effects (HTEs). These characteristics can include severity of the disease under study, sociodemographic characteristics, genetic characteristics, and health-related behaviors. HTEs arise when one or more of these characteristics interact with a treatment to influence a treatment effect. If this influence is great, then the treatment effect for the subgroup of patients with those characteristics may vary substantially from the average treatment effect across the broader patient population. Failure to recognize HTEs can undermine the interpretation of clinical trials and the generalizability of those findings to corresponding patient populations in real-world practice.²⁵

Different patient characteristics can affect HTEs in several main ways, including patient diversity in disease risk or prognosis without treatment, responsiveness to treatment, vulnerability to adverse effects, and patient utility (or preference) for different outcomes.²⁶ Patient utility is an important component to PM. Whether for breast cancer, prostate cancer, or

²⁰ Spear 2001.

²¹ Mangravite LM, Thorn CF, Krauss RM. Clinical implications of pharmacogenomics of statin treatment. *Pharmacogenomics J.* 2006;6(6):360-74.

²² Rieder MJ, Reiner AP, Gage BF, et al. Effect of VKORC1 haplotypes on transcriptional regulation and warfarin dose. *N Engl J Med* 2005;352:2285-93.

²³ Terra SG, Hamilton KK, Pauly DF et al. Beta1-adrenergic receptor polymorphisms and left ventricular remodeling changes in response to beta-blocker therapy. *Pharmacogenet Genomics* 2005;15(4):227-34.

²⁴ Kravitz 2004.

²⁵ Greenfield S, Kravitz R, Duan N, Kaplan S. Heterogeneity of treatment effects: implications for guidelines, payment, and quality assessment. *Am J Med* 2007;120(4A):S3-9.

²⁶ Kravitz 2004.

joint disease, different patients can have substantially different utilities for the same treatment, even when it produces the same health outcomes.

Reports of clinical trials often fail to account for HTEs. When RCTs or other studies investigate the effectiveness of interventions in narrowly defined patient groups (e.g., that over-represent one sex, particular age groups, particular racial or ethnic groups, limited comorbidities, or specific laboratory values), they can underestimate the true HTEs that exist in the target population of interest. This can lead investigators, as well as clinicians, payers, and quality assurance groups, to conclude that the results of the study are more broadly applicable than they truly are.²⁷ Not considering how trial results can fail to represent patient response in real practice, physicians may over-treat or under-treat patients.

Most CER to date has not focused on patient subgroups, even large, aggregated ones. An analysis conducted by the Congressional Research Service found that only 13% of comparative clinical effectiveness studies published in the peer-reviewed literature during the period January 2004-August 2007 focused on effectiveness of treatments in subpopulations other than white middle-age adults (or females for diseases that only occur in females), e.g., children, the elderly, and non-white populations. Only about 5% of these CER studies included patients with comorbidities, even though nearly 60% of hospitalized patients have one comorbidity and more than a third have at least two comorbidities.^{28,29} To examine a large number of potential HTEs in a patient population in a prospective study could require large, expensive study designs. Implications for study designs, including some alternative types, are described below.

Knowing that patients are likely to respond differently to the same treatment has important implications for planning clinical trials and other studies in CER. The number of patients required to demonstrate statistically significant treatment effects will be smaller for those subgroups that are more responsive to a given treatment. Therefore, exploratory analyses of existing data (such as from subgroup analyses of completed clinical trials or registries or other observational studies) to identify what appear to be more responsive patient subgroups can be used to plan prospective clinical trials using smaller sample sizes and facilitate more efficient (shorter, less costly) confirmatory trials.³⁰

Beyond HTEs among patient subgroups and individuals, even the responses of individual patients may vary over time for treatments for certain diseases. In such instances, it may be incorrect to assume that patients respond consistently to treatment, e.g., that 60% of patients are responders and 40% are non-responders. It could be, for example, that, on a random basis, all patients respond 60% of the time and fail to respond 40% of the time. If so, seeking a genetic

²⁷ Ibid.

²⁸ Jacobson, GA. Comparative clinical effectiveness and cost-effectiveness research: background, history, and overview. Washington, DC: Congressional Research Service. October 15, 2007.

²⁹ Many analyses of subgroups report patient response by individual subgroups, which can suggest differences from group to group. However, these analyses for individual subgroups do not provide quantitative information about how patient responses actually differ *across* subgroups; doing so involves using certain statistical tests or interval estimation. A recent analysis of RCTs reported in leading medical journals found that only 28% reported HTE analysis, that another 28% reported subgroup analyses only without formal examination of HTE, and the rest reported neither. See: Gabler 2009.

³⁰ Gabler 2009.

basis for the observed 40% non-response to treatment may be futile.³¹ For some disorders, there may be changes in individuals' gene expression that are time-dependent.³² Such variable response in individuals cannot be assessed in parallel group RCTs in which patients are randomly assigned to one treatment or another. Certain alternative trial designs, such as sequential cross-over and matched pairs, can be used to detect and estimate these individual treatment effects.^{33, 34}

Mainstream application of PM will depend on evidence of the relative effectiveness of interventions in subgroups. The ability of CER to align with PM will derive from how well it generates evidence of treatment effectiveness at the subgroup level. Recognizing that systematic reviews and comparative effectiveness often "do not or inadequately address clinically meaningful differences between individuals," AHRQ is sponsoring an analysis of how well these studies have accounted for HTEs in study populations. This analysis is examining CER and related reports from AHRQ, the Cochrane Collaboration, the Oregon Health & Science University's Drug Effectiveness Review Project (DERP), Australia's National Health and Medical Research Council, and the UK's National Institute for Clinical Excellence (NICE). The findings will be a starting point for a working group and guidance document for improving the ability of CER to account for and report on HTEs.³⁵

NEED FOR MORE EVIDENCE TO SUPPORT PM

Interventions used in PM are subject to prevailing requirements for evidence demonstrating how well they work compared to standard care. Increasingly, this means showing that an intervention has some direct, or least demonstrably indirect, favorable impact on health outcomes in real-world practice settings. In PM, the interventions that are subject to greater evidence demands include not only genetic and genomic testing and targeted therapeutics, but the services and systems that may influence the use and impact of PM, including various approaches to health care management, delivery, benefit design, and payment.

For genetic and genomic testing and other aspects of molecular-based PM, this means demonstrating not only that a test can accurately detect a particular gene or biomarker, but that the test result identifies or predicts a corresponding disease or disorder, and that this information has an impact on clinical decisions and patient outcomes in practice.³⁶ These concepts are known as analytic validity, clinical validity, and clinical utility (Box 1).

³¹ Senn S. Individual response to treatment: is it a valid assumption? *BMJ* 2004;329(7472):966-8.

³² Kalow W. Pharmacogenetics and pharmacogenomics: origin, status, and the hope for personalized medicine. *Pharmacogenomics J* 2006;6(3):162-5.

³³ *Ibid.*

³⁴ Kravitz 2004.

³⁵ Agency for Health Research and Quality. Comparative Effectiveness Methods – Clinical Heterogeneity. <http://effectivehealthcare.ahrq.gov/healthInfo.cfm?infotype=rr&ProcessID=93>.

³⁶ Garber AM, Tunis SR. Does comparative-effectiveness research threaten personalized medicine? *N Engl J Med* 2009;360(19):1925-7.

While genetic tests for detection of variant genes typically are highly accurate, interactions among genes and among genes and environmental factors often limit the clinical validity and clinical utility of these tests. Even a test with proven analytic validity and clinical validity may not provide clinical utility unless it yields a favorable net balance of risks and benefits in routine practice. For clinical utility, a test has to have the potential to augment what is already known about an individual's condition (or provide similar information in a more efficient or cost-effective manner) and inform a decision regarding an available intervention, behavior change, life planning or other option that may affect health outcomes, life events, or quality of life.³⁷

Box 1. Validity and Utility of Genetic Tests

- **Analytic validity:** A test's ability to accurately and reliably measure the genotype of interest. Analytic validity focuses on the laboratory components of testing, including analytic sensitivity, analytic specificity, laboratory quality control and assay robustness.
- **Clinical validity:** A test's ability to detect or predict the associated disorder (phenotype), including clinical sensitivity (or the clinical detection rate), clinical specificity and positive and negative predictive values. Clinical validity is affected by the prevalence of the disorder, penetrance, analytic sensitivity and genetic and environmental modifiers.
- **Clinical utility:** A test's ability to affect clinical decisions and patient outcomes in practice. Other elements or contextual factors to be considered include the natural history of the disorder, availability and effectiveness of interventions, quality assurance, health risks of testing or resulting interventions, financial impacts of testing, adequacy of facilities to provide services, availability of patient and provider education and monitoring and evaluation of test performance in practice.

Adapted from: Evaluation of Genetic Testing. ACCE: A CDC-Sponsored Project Carried Out by the Foundation of Blood Research. Centers for Disease Control and Prevention, National Office of Public Health Genomics. <http://www.cdc.gov/genomics/gtesting/ACCE.htm>.

For most tests, the availability of adequate supporting evidence decreases markedly from demonstrating analytical validity to demonstrating clinical validity to demonstrating clinical utility. Evidence gaps on the clinical utility of genetic tests have important consequences for patient health and resource use. Incomplete evidence of clinical utility for tests can lead to false expectations on the part of patients and families.

Credible evidence on the clinical utility of these tests, particularly in comparison to standard approaches for identifying and managing genetic-based health risks, would help to distinguish between truly beneficial tests and those that currently have little or no impact on patient health and well-being. Although they may be able to detect tens or hundreds or even thousands of genetic variations, many of the available and emerging genomic test panels have no known impact to date on clinical decisions or health outcomes.³⁸ For the many instances in which a disease or condition is identifiable via an accurate test, but for which validated interventions (treatments, behavior change, or life planning decisions) are unavailable or impractical, a test

³⁷ The Lewin Group 2007.

³⁸ Secretary's Advisory Committee on Genetics and Health in Society. U.S. System of Oversight of Genetic Testing: A Response to the Charge of the Secretary of Health and Human Services. April 2008. Available at: http://www4.od.nih.gov/oba/SACGHS/reports/SACGHS_oversight_report.pdf.

will have little no clinical utility. The example of testing for genes that predict patient response to anticoagulation therapy with warfarin highlights the demand for evidence of clinical utility by payers, even when regulators acknowledge analytic and clinical validity (Box 2). Similarly, the instance of genotype-guided tamoxifen therapy for early stage breast cancer suggests how better evidence on the impact of testing on health outcomes is needed to support testing and treatment decisions (Box 3).

Box 2. CMS Findings on Genetic Testing for Warfarin Anticoagulation Response

A current example of how a payer coverage policy can reflect a determination of inadequate evidence on clinical utility is the proposed decision by the Center for Medicare and Medicaid Services (CMS) regarding coverage of PGx testing for warfarin response for Medicare beneficiaries who are candidates for anticoagulation therapy. Anticoagulation therapy with warfarin involves administering the drug in a narrow therapeutic range that varies from patient to patient, providing enough of the drug to minimize the chances that patients will form dangerous blood clots while not providing so much that could lead to potentially life-threatening bleeding. Dosing is especially important when initiating therapy, when problems in adjusting the dose can lead to bleeding and other complications.

In 2007, the Food and Drug Administration determined that available PGx information warranted a change in the labeling of warfarin to call attention to the potential relevance of genetic information to prescribing of warfarin.³⁹ However, regarding Medicare coverage, in 2009, CMS stated:

“CMS found no evidence that genetic testing can replace PT/INR [prothrombin time/International Normalized Ratio Testing] for titrating and monitoring warfarin therapy... [W]e propose that the evidence is insufficient to determine that pharmacogenomic testing to predict warfarin responsiveness improves patient oriented health outcomes related to the underlying indication for warfarin anticoagulation or adverse events related to warfarin therapy itself. In addition, we propose that the evidence is insufficient to determine that pharmacogenomic testing to predict warfarin responsiveness leads to changes in physician management of beneficiaries’ anticoagulation therapy that would result in positive outcomes.”⁴⁰

Noting that the testing is promising, CMS proposed a “coverage with evidence development” arrangement in which Medicare would cover the test only for beneficiaries enrolled in an RCT meeting certain specifications, one of which is “The research study protocol must explicitly discuss subpopulations affected by the treatment under investigation.” Such RCTs would provide further evidence that could inform a revised coverage determination. Other research continues on how genotypes affect sensitivity to warfarin and how well genetic tests predict safer and more effective doses of warfarin, including a large, multicenter RCT designed to determine whether genetic information provides additional benefit to what can be accomplished with traditional clinically-based warfarin information alone.⁴¹

Population-based research with sufficient power for subgroup analyses, including certain RCT designs and other methods, is needed to identify and quantify the relationships among genomic traits, biomarkers, therapies, and health outcomes to establish the evidence base for informing PM.⁴² Consistent with the purpose of CER, clinical utility of genetic testing and other PM

³⁹ FDA approves updated warfarin (Coumadin) prescribing information. Press release of the Food and Drug Administration, Rockville, MD, August 16, 2007. <http://www.fda.gov/bbs/topics/NEWS/2007/NEW01684.html>.

⁴⁰ Centers for Medicare and Medicaid Services. Proposed Decision Memo for Pharmacogenomic Testing for Warfarin Response (CAG-00400N) May 4, 2009.

⁴¹ Shurin SB, Nabel EG. Pharmacogenomics--ready for prime time? *N Engl J Med* 2008;358(10):1061-3.

⁴² The Lewin Group 2007.

interventions should be supported with evidence from applicable patient subgroups and routine health care settings.

Box 3. Genotype-Guided Tamoxifen Therapy for Early Stage Breast Cancer

Tamoxifen is the most widely used treatment for hormone-dependent (hormone receptor positive) breast cancer. The pharmacological activity of tamoxifen depends on its conversion by a drug-metabolizing enzyme, cytochrome P450 2D6 (CYP2D6), to the potent anti-estrogen endoxifen. About 7-10% of Caucasian patients have reduced CYP2D6 activity resulting from a non- or under-functioning polymorphism of the *CYP2D6* gene. Also, certain drugs, e.g., selective serotonin-reuptake inhibitors (SSRIs), can inhibit the function of CYP2D6. *CYP2D6* gene tests are being used to predict tamoxifen response and personalization of therapies, including use of alternative therapies such as aromatase inhibitors, in women who are taking or considering taking tamoxifen for early stage breast cancer.

The available evidence on the relationship between reduced CYP2D6 activity and health outcomes with tamoxifen is not strong and the findings are highly variable. Prospective studies on this relationship are lacking; the available studies of long-term outcomes have been mostly retrospective and based on stored tissues or those obtained previously in prospective studies of tamoxifen. In some studies, women with *CYP2D6* polymorphisms have a higher risk of recurrence than women with normal (wild type) *CYP2D6* genes, though apparently not better overall survival.⁴³ A recent systematic review of available epidemiological studies of the association between *CYP2D6* genotype and breast cancer recurrence reported widely heterogeneous results. Further, studies of the relationship between tamoxifen dosing and biomarker and health outcomes, of the association between blood concentrations of tamoxifen and its metabolites and clinical outcomes, and other analyses suggest that tamoxifen and metabolites should reach concentrations sufficient to achieve the therapeutic effect, regardless of CYP2D6 inhibition.⁴⁴

Stronger evidence from adequately powered prospective studies or retrospective analyses of large prospective trials would help to support practice guidelines and payment policies regarding, e.g., whether CYP2D6 genotype testing should be routine for patients with hormone receptor positive breast cancer. Ongoing research seeks to identify other genetic variations in tumors and in germ-lines (inherited genes) that may affect the efficacy and safety of tamoxifen, other hormone-based therapies, and chemotherapies for treating breast cancer.

The AHRQ Effective Health Care Program, which administers most of the CER activity of that agency, has commissioned several evidence reports pertaining to PM topics involving genetic or genomic testing for guiding treatment for depression, breast cancer, and diabetes and for prognosis of outcomes of patients with breast cancer. As exemplified in the instance of PGx testing to inform treatment of adults with depression (Box 4), these reports include careful examination of the quality of available evidence, its relevance to clinical practice, and identification of research needs to fill important evidence gaps. Unlike CMS, AHRQ is solely a research agency that does not make coverage determinations or other policy.

Much work is needed to develop evidence of how well these technologies work in comparison to standard alternatives for informing decisions and affecting outcomes. Notably, one of the top

⁴³ Schroth W, Goetz MP, Hamann U, et al. Association between CYP2D6 polymorphisms and outcomes among women with early stage breast cancer treated with tamoxifen. *JAMA*. 2009 Oct 7;302(13):1429-36.

⁴⁴ Lash TL, Lien EA, Sørensen HT, Hamilton-Dutoit S. Genotype-guided tamoxifen therapy: time to pause for reflection? *Lancet Oncol* 2009;10(8):825-33.

CER priorities recommended by the IOM encompasses comparative effectiveness of genetic and biomarker testing:

*Compare the effectiveness of genetic and biomarker testing and usual care in preventing and treating breast, colorectal, prostate, lung, and ovarian cancer, and possibly other clinical conditions for which promising biomarkers exist.*⁴⁵

Box 4. Evidence Report on CYP450 Testing for Adults with Depression Treated with SSRIs

AHRQ commissioned one of its Evidence-based Practice Centers to conduct a systematic review of *cytochrome P450* gene testing for adults with depression. The review found only mixed evidence regarding the association between *CYP450* genotypes and SSRI metabolism, efficacy, and tolerability in the treatment of depression, mainly from a series of heterogeneous studies in small samples. The review found no data regarding whether: (1) testing for *CYP450* polymorphisms in adults entering SSRI treatment for non-psychotic depression leads to improvement in outcomes versus not testing, or if testing results are useful in medical, personal, or public health decisionmaking; (2) *CYP450* testing influences depression management decisions by patients and providers in ways that could improve or worsen outcomes; (3) there are direct or indirect harms associated with testing for *CYP450* polymorphisms or with subsequent management options. The report concluded that:

*“There is a paucity of good-quality data addressing the questions of whether testing for CYP450 polymorphisms in adults entering SSRI treatment for non-psychotic depression leads to improvement in outcomes, or whether testing results are useful in medical, personal, or public health decisionmaking.”*⁴⁶

The potential for proven testing and targeted therapies to have their desired impacts depends on how well various approaches to health care organization, delivery, management, and payment can support and enable PM. Whether for their comparative effectiveness or other types of evidence, these approaches, which affect behavioral, environmental, and social aspects of PM as well as genetic or genomic ones, should be subject to assessment for their impact on adoption, use, and health outcomes of PM. Examples are: alternative delivery models (e.g., combined clinical and social interventions and comprehensive care coordination), health literacy programs, medication adherence programs, shared decision-support systems for patients and clinicians, targeted methods for dissemination and translation of PM findings, electronic health records (EHRs) and personal health records, health benefits design, and patient cost-sharing strategies. Regarding health benefits design, even PM interventions that are validated for screening or preventive applications will not be accessible to patients whose health benefits provide limited or no coverage for screening or preventive services.

CER METHODS AND ADAPTATION FOR PM

No single study design can answer the range of evidence questions that arise in CER. To address the needs of PM in particular, CER must draw from a broad methods portfolio. Different methods may be required, for example, to assess short- and long-term comparative effectiveness of alternative therapies, identify subgroups with variable treatment responses,

⁴⁵ IOM 2009.

⁴⁶ Duke Evidence-based Practice Center. Testing for *Cytochrome P450* Polymorphisms in Adults With Non-Psychotic Depression Treated With Selective Serotonin Reuptake Inhibitors (SSRIs). Contract No. 290-02-0025. Evidence Report/Technology Assessment Number 146, January 2007.

identify short-term as well as rare or delayed adverse effects, and compare alternative tests for accuracy and ultimate impact on health outcomes. Of key relevance to PM is the extent to which one or a combination of these methods can generate clinically and statistically significant findings at sufficiently discrete levels to inform decisions about using interventions whose outcomes are sensitive to individual differences. Certainly, the need for subgroup-specific CER findings varies for different health care problems and interventions. This section briefly describes the portfolio of CER methods and their strengths and limitations for informing PM.

Clinical Trials

Randomized Controlled Trials

RCTs are considered the gold standard of scientific evidence for establishing causality of a given intervention on particular health outcomes. Their strength derives from their ability to minimize sources of bias and other factors that might confound determination of that causal relationship. For new drugs and some medical devices, RCT evidence is used to establish efficacy and safety, often compared to placebo or no treatment rather than a standard of care or other active intervention, for gaining market approval in the U.S. by the FDA. While this is essential evidence for determining whether an intervention works under some set of carefully defined conditions, it may not generate the evidence needed to establish real-world effectiveness of an intervention in patient populations with diverse characteristics under different condition. In contrast to real-world conditions, RCTs typically involve narrowly defined patient groups (e.g., in a specific age range with a single disease with narrowly defined or no comorbidities), practice settings, and outcomes, and can be too short in duration to capture certain important outcomes and adverse events.^{47,48} As such, RCTs are not always appropriate for CER, and can have limitations for generating the evidence needed for PM.

RCTs that enroll particular subgroups in sufficient numbers to achieve the statistical power to detect treatment effects will have the capacity to yield results that can inform more personalized decisions for patients with these characteristics. However, without some prior basis for identifying such subgroups, these RCTs may have to be very large and costly.

Subgroup analyses of RCT data for genetic traits that are associated with favorable or unfavorable response can be used to design smaller trials that enroll only those patients who are expected to do better. If such patient subgroups truly are better responders or are less subject to adverse effects, these trials can better detect those outcomes if they truly exist, rather than having them diluted in the larger population of mixed responders. Subgroup analyses of RCT data can be strongly suggestive (though not necessarily definitive) of the relationship between a subgroup trait and an outcome from a given intervention,⁴⁹ as in the instance of *KRAS* gene testing for targeted therapy for colorectal cancer (Box 5).

⁴⁷ Horn SD, Gassaway J. Practice-based evidence study design for comparative effectiveness research. *Medical Care* 2007;45 (10 Suppl 2):S50-7.

⁴⁸ Moon M, Smith B, Gustafson S. Creating a center for evidence-based medicine. Washington DC: MedPAC. February 2008.

⁴⁹ Retrospective subgroup analyses of RCT data are not, themselves, RCTs or other prospective trials. They can generate hypotheses that can be tested using RCTs or PCTs.

Practical Clinical Trials

Practical (or pragmatic) clinical trials (PCTs), sometimes known as effectiveness trials, address some of the disadvantages inherent in RCTs and other types of studies in CER. PCTs compare alternative interventions that are relevant to clinicians and their patients, focus on more heterogeneous patient populations and practice settings, and collect data on a broad range of health outcomes. PCTs may be randomized; types include large simple trials, cluster-randomized studies, and time-series analyses of planned changes in care. In principle, PCTs should generate more results for use in PM, although the extent to which PCTs have captured genomic data from their study populations remains to be determined. PCTs can be expensive and require large sample sizes and long follow-up periods. While the difference in effectiveness between two active treatments may be clinically meaningful, that difference may be considerably smaller than the difference between each active treatment and placebo. As such, detecting clinically meaningful differences with statistical significance between the alternatives in a “head-to-head” PCT can require much larger sample sizes, which can increase the costs and time to complete these trials.⁵⁰ PCTs may be designed for prospective comparison of particular subgroups, or, as noted above for RCTs, they may be subject to retrospective analyses for subgroup differences that can be studied in subsequent prospective trials.

Adaptive Clinical Trials and Other Trial Designs

Other clinical trial designs offer improved ways of “stratifying” (identifying and testing subgroups of) population response to interventions. Of particular note are adaptive clinical trials, which are “learn-as-you-go” approaches to conducting clinical trials. In adaptive clinical trials, one or more decision points are built into the trial design for analysis of outcomes and associated patient or disease characteristics to identify subgroups who are responding favorably to an investigational treatment. Planning for such mid-course corrections can help to focus trial resources on enrolling more patients with attributes (e.g., particular genetic traits) that are more likely to have favorable results and enrolling fewer patients who are less likely to respond or more likely to experience adverse effects. This can increase the chances of detecting statistically significant treatment effects in a population subgroup that otherwise would have been statistically lost in a broader pool of patients with a higher proportion of non-responders.

Some adaptive trial designs allow for stopping a trial early or later than expected based on results that indicate how effective a treatment is after a limited number of patients have been tested. Other designs enable dropping a treatment arm that appears to be ineffective, modifying sample sizes to discern statistically significant treatment effects, and rebalancing treatment assignments using adaptive randomization.⁵¹ In postmarket studies of drugs, adaptive trial designs could inform PM with findings that are useful in tailoring treatment regimens for patients with the predictive attributes identified in these trials. Additional types of

⁵⁰ Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA* 2003;290(12):1624-32.

⁵¹ Institute of Medicine. *Improving the Quality of Cancer Clinical Trials: Workshop Summary*. Washington, DC: The National Academies Press, 2008.

trials that can generate evidence pertaining to PM are randomized consent trials,⁵² regression discontinuity trials,⁵³ and combined single-subject (“n of 1”) trials.⁵⁴

Box 5. KRAS Gene Mutation Testing for Patients with Metastatic Colorectal Cancer

Colorectal cancer is the third most commonly diagnosed cancer and third-highest cause of cancer death for men and women in the U.S.⁵⁵ Up to 20% of patients with colorectal cancer will present with metastases, with a 5-year survival of less than 10%. Two powerful drugs against colorectal cancer are cetuximab (Erbix®) and panitumumab (Vectibix®). These are monoclonal antibodies that bind to the epidermal growth factor receptor (EGFR) on cells, inhibiting growth of metastatic colorectal cancer. These EGFR inhibitors also have considerable side effects. The effectiveness of these therapies in colorectal cancer varies depending on the genetic makeup of the tumor. In some patients, the tumor cells have a mutation in a gene known as *KRAS* (*v-Ki-ras2 Kirsten rat sarcoma viral oncogene*) that may cause a tumor to be non-responsive to cetuximab and panitumumab. Retrospective analyses of data from several RCTs involving patients receiving combination cetuximab and chemotherapy demonstrated that individuals with normal (or “wild-type”) *KRAS* had significant improvements in tumor response and that few or none of those with mutated-type *KRAS* responded to cetuximab.⁵⁶ Therefore, determination of *KRAS* status can help to avoid ineffective treatment and unnecessary exposure to the side effects of these drugs, and redirect treatment decisions to alternative therapies. In July 2009, the FDA announced revisions to the prescribing information for EGFR inhibitors and colorectal cancer, requiring inclusion of information on variations in the *KRAS* gene that may affect patient response to the drugs. As some *wild-type KRAS* patients do not respond to EGFR inhibitors, further research is needed to improve identification of those patients (perhaps using a combination of genetic tests), in order to better direct their treatment selection. Although *KRAS* testing to inform treatment decisions for these patients is strongly indicated based on available evidence, the actual impact of the testing on clinical decisions or on patient outcomes has not yet been assessed in prospective studies.

Observational Studies

Observational studies include analyses of various sources, including insurance claims and other administrative data sets, medical/health records, integrated health system databases, patient registries, and other clinical databases. Observational studies often enable including populations and particular subgroups that tend to be underrepresented in clinical trials.⁵⁷ In patients with given diseases or conditions, particularly those with stable or steadily progressing courses, these studies can be used to examine relationships among certain interventions, patient characteristics, provider characteristics and differences in biomarkers, outcomes, and adverse events. They can be used to generate hypotheses about these relationships that can be tested in prospective clinical trials. Because they use such existing sources as claims data and de-identified EHRs, observational databases typically are less costly than clinical trials.

⁵² See, e.g.: Institute of Medicine. *Small Clinical Trials: Issues and Challenges*. Washington, DC: National Academy Press, 2001.

⁵³ See, e.g.: Cappelleri JC, Trochim WM. Ethical and scientific features of cutoff-based designs of clinical trials: a simulation study. *Med Decis Making* 1995;15(4):387-94.

⁵⁴ See, e.g.: Gabler 2009.

⁵⁵ Ng K, Zhu AX. Targeting the epidermal growth factor receptor in metastatic colorectal cancer. *Crit Rev Oncol Hematol* 2008;65(1):8-20.

⁵⁶ Brown HM, Aronson N, Ziegler KM, et al. *KRAS* mutations and epidermal growth factor receptor inhibitor therapy in metastatic colorectal cancer. Report No. 23(6). Technology Evaluation Center, Blue Cross Blue Shield, 2009.

⁵⁷ AHRQ. *Methods reference guide for effectiveness and comparative effectiveness reviews*. Rockville, MD: Agency for Healthcare Research and Quality. October 2007.

Observational studies can be prospective or retrospective, but they are not experimental and cannot demonstrate causality. They are more likely to be subject to certain biases and confounding factors that RCTs and PCTs are designed to diminish. Data that would be available in an experimental research design to adjust for factors such as disease severity that can bias findings may be missing from claims and other administrative data sets. While observational studies can find associations between interventions and outcomes, including for patient subgroups, they cannot establish causality. They are limited by data elements collected for their original purposes, e.g., payment of claims.

Patient registries are structured inventories of data on patients who have received particular interventions (e.g., a particular drug, implanted device, or surgical procedure). Among the types of evidence gaps that registries can fill are the types and frequency of adverse events that are not detected in clinical trials, which may have had too few participants to detect rare adverse events or were not long enough to detect delayed adverse events. Certain public and private databases contain substantial genomic data that can be used to analyze group and individual patient data on associations between genotypes and drug-response that could yield significant insights about the most effective interventions for specific patient groups.⁵⁸ Some of the nation's integrated health care systems are using their large clinical and administrative data systems and registries to conduct CER to determine the most effective interventions for subgroups of patients.⁵⁹

As described above for retrospective analyses of RCT data, certain observational studies can be used to identify subgroups with potentially important HTEs. Such subgroups can then be enrolled in RCTs or other prospective studies to confirm whether these HTEs truly exist.

Syntheses of Existing Evidence

Systematic Literature Reviews

Systematic literature reviews are prospectively designed, comprehensive literature reviews that are focused on well-defined evidence questions. (In the context of CER, systematic reviews are sometimes known as “comparative effectiveness reviews.”) They are intended to identify, appraise, and synthesize all relevant high-quality research evidence. They may incorporate meta-analyses in instances where the available evidence from primary studies is sufficiently homogeneous with respect to the populations, interventions, and outcomes studies.⁶⁰ While they can yield more robust findings from existing evidence, systematic reviews do not generate new data, and they are limited by availability, quality, and heterogeneity of available evidence

⁵⁸ Gurwitz D, et al., A call for the creation of personalized medicine data bases. *Nat Rev Drug Discov* 2006;5:23-6.

⁵⁹ For example, Kaiser Permanente, Group Health, and Geisinger Health have initiated programs that use their large databases to complement clinical trial findings for purposes consistent with CER. The Health Maintenance Research Network (HMORN) includes 15 managed care organizations covering more than 15 million individuals and working cooperatively on effectiveness research. See: IOM. *Learning What Works Best*. September 2007; and HMO Research Network. Research Projects. Available at: <http://hmoresearchnetwork.org/projects.htm>.

⁶⁰ Meta-analysis refers to statistical techniques for combining results from multiple existing studies. This combination may produce a stronger conclusion than can be provided by any singular study. It is generally most appropriate when there are not definitive studies on a topic and non-definitive studies are in some disagreement (e.g., regarding direction or magnitude of a treatment effect).

from clinical trials and other primary studies.⁶¹ Given the lack of clinical trials of head-to-head comparisons of alternative interventions, there are few systematic reviews or meta-analyses of such direct comparisons. However, if each of two interventions of interest have been compared to placebo (or a third intervention in common) for the same health problem in similar populations in separate RCTs, then a systematic review may be able to generate findings about the indirect comparative effectiveness of the two interventions. Systematic reviews can identify and compile available evidence on subgroups from multiple studies, which can be used to generate hypotheses about treatment effects in these groups that can be tested in new RCTs or PCTs. Systematic literature reviews generally are far less costly to conduct than clinical trials.⁶²

One of the applications of systematic reviews that is particularly useful for PM is characterizing the effects on health outcomes of a range or diversity of factors. These might include varying levels or intensities (e.g., dosages or duration) of interventions, different patient characteristics (age, sex, other demographic factors, comorbidities, etc.), or different care settings. Such findings may provide more specific information for patients with these characteristics. Because they involve identifying and compiling available relevant evidence on a given research question, systematic reviews can be very useful for identifying needs for further CER using primary data sources.

Modeling

Modeling refers to quantitative representations or simulations of health care. It may be used when available evidence insufficient to answer a research question or to project risks, benefits, or costs of alternative care scenarios. The validity of models depends on the source data and evidence, and the adequacy of understanding of the physiological pathways or decision processes pertaining to patients and interventions being modeled. Modeling does not generate new primary data. However, supported by new analytical techniques and advances in computing power, modeling can tap findings from clinical trials and other primary research as well as existing data sources to simulate head-to-head comparisons of alternative treatments.

Using large sets of de-identified data from multiple health plans, and drawing on known causal relationships established in clinical trials, modeling can detect varying levels of association between interventions and outcomes in stratified patient populations. Further, models can help to predict clinical outcomes that might have important implications for decisions about whether to use particular interventions.⁶³ It can help inform decisions about using screening and diagnostic tests by estimating their likely benefits and harms when applied to populations and subgroups with particular risk factors or other characteristics, such as in screening for breast, cervical, or colorectal cancer. As more patient-specific (including genomic) information is captured in clinical registries, EHRs, and other databases, modeling is being used to simulate clinical decisions and outcomes for individual patients.

⁶¹ AHRQ. Methods reference guide for effectiveness and comparative effectiveness reviews. Rockville, MD: Agency for Healthcare Research and Quality. October 2007.

⁶² CBO. Research on the comparative effectiveness of medical treatments. Washington, DC: Congressional Budget Office. December 2007.

⁶³ Eddy, DM, Linking electronic medical records to large-scale simulation models: can we put rapid learning on turbo? *Health Affairs* 2007;26(2):w125-36.

CER and Innovation in PM

Especially insofar as it accounts for HTEs and other subgroup and individual traits, CER is likely to alter value propositions in health care innovation. It will provide new opportunities and hasten some shakeouts in new product pipelines. CER findings that favor an innovation versus a standard of care, either because the innovation is superior or because it delivers the same outcomes at a lower cost, can provide an immediate market advantage. CER is expected to strengthen incentives for developing medical interventions with demonstrable advantages over existing options, and may diminish incentives for investing in “me too” products that offer no comparative advantages. Led by global pharmaceutical and biotechnology companies that have been responding to evolving evidence requirements in international markets, many in industry are incorporating CER considerations into their pipelines. New product sponsors will weigh strategies for investing in CER of their own as they consider the prospects of their products being subject to CER sponsored by government or by competitors. Many of these companies also are advancing PM by incorporating PGx and related aspects of PM into their R&D. Greater emphasis on PM in CER may alter certain attributes of innovation, including selection of study outcomes reflecting patient interests (e.g., more patient-centered outcomes) and extending to options or differentiation in ease of use, dosing, and packaging.

Federal support of CER could reduce development costs of some new interventions.⁶⁴ Analyses of linked databases may help to identify new genetic determinants of drug response and related biomarkers. Evolution and expansion of the CER methods portfolio will offer more options for validating innovations in PM, including the alternatives to traditional RCTs noted above. These efforts will better define the types of evidence sought by health professionals, payers, and quality standards organizations for establishing clinical utility of genetic and genomic tests and evidence expectations for other types of PM interventions.

Role of Health Information Technology

Full alignment of PM and CER depends on adoption of health information technology (HIT). There are two main ways in which HIT can enable CER to contribute to PM. First, in support of conducting CER, EHRs can capture patients’ genetic and other individual health information in the course of routine health care, clinical trials, and other studies. Secure, ongoing collection of gene-based and other molecular test data from EHRs and clinical laboratories and linking these with population-focused patient registries can support research on relationships among personal traits, interventions, and outcomes. Second, in translating CER to practice, HIT can ensure that evidence pertaining to PM is present and actionable at the point of decision-making, enabling patients and their physicians to consider patients’ personal traits when weighing the risks and benefits of alternative treatment options. Through computerized clinical alerts and reminders and ready access to relevant clinical practice guidelines, quality standards, and research findings, clinical decision support systems can help clinicians to identify interventions or regimens that are more likely to benefit patients with particular characteristics.⁶⁵

⁶⁴ Garber 2009.

⁶⁵ Clinical decision support encompasses, e.g., computerized alerts and reminders, means to bring care into compliance with clinical guidelines, generation of order sets and patient data reports, advice to promote accurate and

HIT's considerable potential for serving PM is far from being realized. It is subject to the rate of adoption of EHRs, interoperability of HIT systems, and development and adoption of clinical decision-support systems, all of which are in early stages of development or adoption. Expert groups such as the Personalized Health Care Workgroup of the DHHS American Health Information Community (AHIC) have recommended ways to incorporate genetic and genomic test results in EHRs and harmonize related standards. This includes developing a core data set pertaining to outcomes of clinical interventions influenced by PGx tests that need to be captured in EHRs.⁶⁶ Recognizing the demand for evidence-based, point-of-care decision-making and the considerable federal investment in HIT, vendors are adapting EHR systems and clinical decision support systems in ways that will serve CER, other research, and health care delivery.

COMMUNICATING CER FINDINGS TO ENABLE PATIENT-SPECIFIC DECISIONS

CER is augmenting bodies of evidence that must be considered in the context of individual patient needs at the point of care by clinicians and patients. Also weighing this evidence are payers, developers of guidelines, and others who might influence adoption and use of health care interventions.

As CER further reflects HTEs and other individual factors that can affect the use and outcomes of health care interventions, communications and applications of these findings must be more adaptive. Diagnostic and treatment decisions will present more options, and there may be greater need for communication between clinicians and patients, and more considerations for payment, utilization management, and other administrative functions.

Beyond clinical and health care management and administration settings, CER findings must be communicated to the public in an accurate, comprehensible manner, including their limitations. Much public information regarding clinical interventions, including emerging findings from CER and PM, can be confusing, if not contradictory. The IOM recommends more effective use of communication and marketing principles "to make validated determinations more accessible to the consumer public and to help inform and educate the public about how evidence continually evolves and how to judge its state of play for a given issue at a given point."^{67,68}

timely diagnoses, and tools that enhance clinical workflow. See: Osheroff JA, et al. A roadmap for national action on clinical decision support. *J Am Med Inform Assoc* 2007;14:141-5.

⁶⁶ Glaser J, Henley DE. Personalized Healthcare Workgroup Recommendations Letter. American Health Information Community. June 3, 2008.

http://healthit.hhs.gov/portal/server.pt/gateway/PTARGS_0_10731_848353_0_0_18/Personalized%20Healthcare%20Workgroup%20Recommendations%20Letter.pdf.

⁶⁷ IOM 2007.

⁶⁸ The American Medical Association recently expressed that "CER evidence cannot adequately address the wide array of patients with their unique clinical characteristics, co-morbidities, and certain genetic characteristics. In addition, patient autonomy and choice may play a significant role in both CER findings and diagnostic/treatment planning in the clinical setting. As a result, sufficient information must be included concerning the limitations and exceptions of CER studies so that physicians who are making individualized treatment plans will be able to differentiate patients to whom the study findings apply from those for whom the study is not representative." Maves MD, Executive Vice President, CEO, American Medical Association. Letter to Sens. Max Baucus and Kent Conrad, September 23, 2009.

Communicating CER findings to enable patient-specific decisions has multiple elements. Certainly, the data pertaining to PM must be valid, integrated, and accessible in the physician's office, specialty clinic, hospital, and other settings. Clinical encounters should allow for use of personalized information and patient preferences, with ready access to practice guidelines and quality standards that account for evidence-based personalized care options. Shared clinician-patient decision-making should be routine. Further, this communication must provide for changes in patient status, preferences, and new evidence.⁶⁹

Progressing from current health care systems to those that can optimize the use of PM will require overcoming certain hurdles with direct or indirect implications for the role of CER in PM. As described above, CER and other research that is designed to yield only average treatment effects may be misleading. More evidence linking genomic and other testing information to treatment decisions and outcomes is needed. Implementation of PM is constrained by the rate of adoption of EHRs, clinical decision support systems, and other information systems. Health technology assessment, clinical practice guidelines, and payment policies must adapt to more differentiated bodies of evidence, particularly for making distinctions between average effects and HTEs. Better means are needed for targeted translation of PM-relevant CER findings for clinicians, patients, payers, and others. Approaches to PM, including how to interpret and apply evidence from CER and other sources to individual patients, must be incorporated into health professional education.

The federal commitment to CER recognizes the importance of effective communication of CER findings, as reflected in the June 2009 reports of recommended national CER priorities from the FCC and the IOM and in pending legislation. For example, the America's Healthy Future Act of 2009, introduced by Senator Baucus, describes how a Patient-Centered Outcomes Research Institute would:

*"...disseminate the findings of research to clinicians, patients, and the public in a comprehensible manner and form so that they are useful to patients and providers in making health care decisions. The dissemination of the research would (1) discuss conclusions and considerations specific to certain subpopulations, comorbidities, or risk factors, as appropriate, and (2) include considerations such as limitations of the research and discussions about what further research might be needed, as appropriate."*⁷⁰

AHRQ, via its John M. Eisenberg Clinical Decisions and Communications Science Center and other means, has a key role in communicating CER findings. AHRQ uses multiple media, including websites, publications, conferences, public meetings, and newsletters, to disseminate or communicate CER findings. Further work on disseminating CER findings for use in patient care decisions is needed, as well as advancing the state-of-the-art of disseminating tailored evidence-based information to its intended target users.

⁶⁹ Stewart WF, Jones JB, Paulus R, Selna M. Personalized health management: a Geisinger view. In *Personalized Health Care Pioneers, Partnerships, Progress*. U.S. Department of Health and Human Services. November 2008. http://www.hhs.gov/myhealthcare/news/phc_2008_report.pdf.

⁷⁰ America's Healthy Future Act of 2009. Washington, DC: 111th Congress of the United States of America, Senate.

SUMMARY AND CONCLUSIONS

Population-based evidence must be complemented by personalized evidence that accounts for how patients' genomic and other personal traits affect their responses to health care.

Considered alone, neither population-based evidence derived from CER nor personalized evidence derived from PGx and other research suffice. Research priorities, design and conduct of data collection, reporting of results, and translation of CER and PM into practice and policy should be fully integrated. This can achieve alignment, and even synergy, of CER and PM.

1. CER has been largely oriented toward population-based evaluations and applications. In contrast, PM focuses on using individuals' genomic information and other personal traits to inform their health care decisions.
2. Like other forms of evaluation of health care interventions, CER generally has focused on identifying interventions that are effective, on average, across a broad patient population.
 - Interventions that yield a statistically significant treatment effect across a study population may not necessarily work for all treated patients; they may be ineffective for some patients and even harmful for others.
 - Other interventions that do not yield a statistically significant treatment effect across a study population—and that may be dismissed as ineffective—may work for certain subsets of the population.
3. The absence of PM considerations in CER could be suboptimal for patient interests, particularly to the extent that CER findings are used to support gatekeeping or other authoritative functions, such as product labeling, clinical practice guidelines, coverage policies, and quality measures and criteria. To the extent that PM is incorporated into CER, the resulting evidence will be more relevant and useful for these same functions.
4. For CER to contribute to PM, it will have to emphasize priorities and study designs that account for individuals' genetic, behavioral, environmental, and other personal traits that mediate the impact of screening, diagnostic, therapeutic, and other interventions on patient outcomes.
 - To date, only a small percentage of published comparative effectiveness studies have focused on treatment effectiveness in patient subgroups.
5. Aligning CER and PM depends on several key factors, including: the research questions being addressed; the type of interventions being studied; study design and implementation; the ways in which findings are communicated to and applied by patients, clinicians, payers, and others; and the ability of health care organization, delivery, management, and payment to support and enable PM.
6. The extent to which population-based evidence can be used to inform health care decisions for specific individuals depends not only on how well the study population represents those individuals; it also depends on whether the study designs and analytical methods used are capable of detecting important treatment effects and adverse outcomes for the patient subgroups representing those individuals.

7. The interventions used in PM are subject to prevailing requirements for rigorous evidence demonstrating how well they work compared to standard care. Increasingly, this means showing that an intervention has some direct, or least demonstrably indirect, favorable impact on health outcomes in real-world practice settings.
 - For genetic and genomic testing and other aspects of molecular-based PM, this means demonstrating not only technical accuracy of a test, but further downstream impact on health care decisions and outcomes.
8. HIT can help align CER and PM in two main ways. First, through EHR capture of genetic and other personal health information in clinical trials and clinical practice, it can support CER to augment the evidence base for PM. Second, clinical decision support systems and other forms of HIT can ensure that evidence pertaining to PM is present and actionable at the point of decision-making by patients and clinicians.
9. CER offers an evolving portfolio of methods with great potential for meeting the needs of PM, including those arising from the CER methods development being supported by AHRQ and ongoing work in the public and private sectors on data mining and analysis of claims and other administrative and observational data. Adaptive clinical trials designs and other variations on clinical trials that focus on deriving evidence efficiently for responsive vs. nonresponsive patient subgroups are of particular promise for PM.
10. There are encouraging developments in the adaptation of CER for PM and policy makers' commitment to ensure that PM is integrated into CER. CER priority setting reports of the FCC and IOM (including their recommended CER topics) and pending legislation emphasize the need for subgroup analyses and consideration of patient-level attributes.
11. CER is likely to alter value propositions for innovation in PM. It will provide new opportunities and hasten some shakeouts.
 - The need to generate comparative evidence at more discrete levels raises the risk of innovation and forces choices about its direction and sequence. Targeted therapies that can demonstrate comparative effectiveness may gain market advantages.
 - Federal support of comparative effectiveness trials and other studies could reduce development costs of some new interventions. Analyses of linked databases may help to identify new genetic determinants of drug response and related biomarkers.
12. As CER further reflects patient risk factors, comorbidities, HTEs, and other individual factors that can affect the use and outcomes of health care interventions, communications and applications of these findings must be more adaptive and targeted to clinicians, patients, payers, and the public, accordingly. These messages should address limitations of this evidence for decision-making and evidence gaps that are priorities for further CER.

The signals approaching the intersection of CER and PM are clear:

- The design and conduct of CER must consider and account for potential HTEs.
- The strengths and limitations of CER and other evidence, including whether it accounts for HTEs as opposed to an average effect across a population, must be accurately reflected in

product labeling, guidelines, payment policies, utilization management, and other gatekeeping policies.

- To enable evidence-based PM, these gatekeeping policies must be flexible, adaptive, and updated accordingly.
- Generally higher and more specific evidence requirements that prevail for most health care interventions apply to PM diagnostics and therapeutics as well, with implications for adoption, use, and payment for PM.
- The ability of CER to contribute to PM on any systematic and ongoing basis depends on broad adoption of HIT, which has been slow to date.
- CER offers opportunities for innovation in PM, along with inevitable shakeouts. Funding for CER and related methods development, data sources, and infrastructure should boost innovation. Technologies that achieve prevailing evidence requirements and demonstrate comparative or superior effectiveness will gain market advantages.
- Current promising signs for CER and PM alignment include explicit attention to PM in recommended national priorities for CER, pending legislation to sustain the national investment in CER, and development of CER methods and research infrastructure.

Whether CER and PM will be aligned or in opposition is starting to unfold. Although they originated with different orientations, CER and PM offer complementary advantages of great potential for patient health. In a stressed health care system poised for reform, a continued, concerted effort is necessary to ensure that this potential is realized.