ClinicoEconomics and Outcomes Research

Dovepress

open access to scientific and medical research

Open Access Full Text Article

PERSPECTIVES

# Use of electronic medical records in oncology outcomes research

This article was published in the following Dove Press journal:
ClinicoEconomics and Outcomes Research
23 February 2010
Number of times this article has been viewed

Gena Kanas[1]
Libby Morimoto[1]
Fionna Mowat[1]
Cynthia O'Malley[2]
Jon Fryzek[3]
Robert Nordyke[2]

[1]Exponent, Inc., Menlo Park, CA, USA; [2]Amgen, Inc., Thousand Oaks, CA, USA; [3]MedImmune, Gaithersburg, MD, USA

**Abstract:** Oncology outcomes research could benefit from the use of an oncology-specific electronic medical record (EMR) network. The benefits and challenges of using EMR in general health research have been investigated; however, the utility of EMR for oncology outcomes research has not been explored. Compared to current available oncology databases and registries, an oncology-specific EMR could provide comprehensive and accurate information on clinical diagnoses, personal and medical histories, planned and actual treatment regimens, and post-treatment outcomes, to address research questions from patients, policy makers, the pharmaceutical industry, and clinicians/researchers. Specific challenges related to structural (eg, interoperability, data format/entry), clinical (eg, maintenance and continuity of records, variety of coding schemes), and research-related (eg, missing data, generalizability, privacy) issues must be addressed when building an oncology-specific EMR system. Researchers should engage with medical professional groups to guide development of EMR systems that would ultimately help improve the quality of cancer care through oncology outcomes research.

**Keywords:** medical informatics, health care, policy, outcomes

## Introduction

The term "EMR" generally refers to an electronic medical record and can indicate the establishment and use of standardized electronic databases for health care, with a digital record as its basis. Current EMR systems contain a record of all clinical, administrative, and laboratory encounters between a patient and a provider, including medical notes, results, billing claims, and insurance information, typically linked with a unique patient identifier (Figure 1). The most comprehensive EMR databases would be those that have a stable population (eg, low rate of patients moving out of the medical system) and include records from all in- and out-patient care, laboratory and radiological results, and all prescribed and over-the-counter (OTC) medications, with a unique identifier for each patient to link them to different medical care providers and services, and to maintain confidentiality.

Current EMR systems in the United States (US) that fit this description exist at a few large, integrated health providers, such as Kaiser Permanente, Harvard Pilgrim Health System, and the Department of Veterans Affairs (VA), but are generally lacking. With only 25% of Americans belonging to any type of health maintenance organization (HMO),[1] this type of health care structure is not typical. While European and socialized countries may have existing platforms (eg, registries, health care systems) that are more readily suitable to EMR, with some exploring EMRs in oncology,[2] the primary focus of the current article is on the US.

Correspondence: Gena Kanas
Exponent, Inc., 149 Commonwealth Drive, Menlo Park, CA 94025, USA
Tel +1 (650) 688-7363
Fax +1 (650) 688-1799
Email gkanas@exponent.com

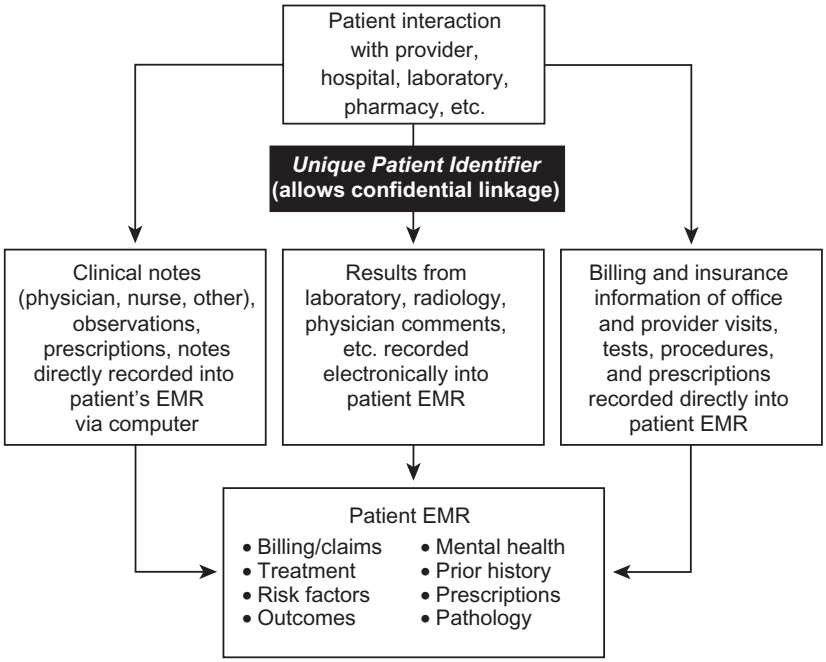


**Figure 1** Schematic of current electronic medical record (EMR) components.

The potential utility of EMR to enhance clinical medicine was recognized in the 1990s, primarily due to the increasing use of the Internet. The widespread use of the Internet fostered more cooperative relationships among doctors, patients, and other care providers, allowing for merging of data across different treatment settings. Several government initiatives and programs have been developed over the past several years to increase the use and development of EMR systems in the US (Table 1).

The use of an EMR in clinical medicine offers numerous potential benefits, including an improvement in the quality of patient care, a reduction in the cost of health care, and improvement in the ease of transferring patient information among providers.[3] In the US, maintenance of electronic records of continuous care is often hindered by frequent changes in providers and plan members, due to job changes or changes in coverage, but EMR adoption across a wide range of health systems could potentially provide a seamless record even across such changes and patient migration. The adoption of EMR, however, has generally been slow due to the cost of system implementation and user training. While the US government has been promoting the adoption of a national EMR system, little guidance has been provided on how it should look or function.[4] To date, less than 20% of group physician practices in the US have adopted EMRs.[5] In particular, small practices, which make up approximately 80% of all physician practices, have been slower to adopt EMR than physician groups associated with larger hospitals and health systems.[6] The general challenges of adopting and using EMR have been discussed elsewhere.[3,7]

While the benefits and challenges for adoption and use of EMR in clinical medicine and general health research have been investigated, the utility of EMR for oncology outcomes research has not been explored. In this paper, we describe how oncology outcomes research can benefit from the use of a hypothetical oncology-specific EMR network that incorporates widespread coverage, both across provider types and networks, as well as geographically. Although this level of coverage may not currently exist, even in best practice, among existing EMR systems, the purpose of this paper is to demonstrate the breadth of research questions and beneficial parties who will be affected when and if such a system is implemented. In addition, we describe the specific challenges that must be addressed when building such a system.

## EMR for oncology research

With comprehensive and accurate information on clinical diagnoses, personal and medical histories, planned and actual treatment regimens, and post-treatment outcomes, EMR systems could provide a wealth of data for outcomes research. In a comprehensive literature review of the number and types of EMR systems used for outcomes research from 2000 through 2006, the number of EMR-based research

**Table 1** Brief history of government initiatives to develop and use EMR

| Date | Government entity | Initiative | Description |
|---|---|---|---|
| 2004 | President Bush | Establishment of the position of National Health Coordinator for Information Technology (ONC)[41] | Position was charged with coordinating federal and private-sector health information initiatives to achieve the widespread adoption of intra- and interoperable electronic health records within 10 years |
| 2004 | FDA | Critical Path Initiative[42] | Designed to stimulate and facilitate a national effort to modernize the scientific process of moving a drug or device through discovery into a medical product |
| May 2008 | FDA | Sentinel Initiative[43] | Creation and implementation of the Sentinel system, a national, integrated, electronic system of existing data sources that will be maintained separately by their owners, with strong privacy and security safeguards, can be queried to monitor the performance of a product throughout its life cycle |
| 2006 | CMS | Strategic plan for 2006 to 2009[44] | Written that secure electronic records and electronic prescriptions (collectively, EHRs) would promote reliable and affordable health care, streamline billing and delivery of health care to patients, increase the ability of diverse EMR systems to work together (eg, interoperability), decrease transcription and other errors, and promote education of patients and care providers |
| 2009 | President Obama | American Recovery and Reinvestment Act of 2009 ("the Stimulus Package")[4] | Includes the Health Information Technology Extension Program with $19 billion in grants and loans set aside for infrastructure, and incentive payments for providers who adopt certified EHR technology |

**Abbreviations:** CMS, Center for Medicare and Medicaid Services; EHR, electronic health record; EMR, electronic medical record; FDA, Food and Drug Administration; ONC, Office of the National Coordinator.

studies conducted annually has increased 6-fold.[8] Although the use of computerized clinical information has been demonstrated to improve the efficiency of oncology clinical practice (eg, with computerized chemotherapy producing significantly fewer errors than manual prescriptions,[9] only 10% of 126 published studies of outcomes research using EMR from the US focus on oncology outcomes.[8] In addition, several limitations exist in using EMR data, which must be considered (discussed later).

Implementation of comprehensive EMR systems that focus specifically on the treatment of oncology patients could answer many questions that are unique to oncology clinical practice (eg, see Table 2). The relevance of research from such a resource would be broad, with the potential to benefit parties at each level within the US health care structure from oncology patients, researchers, and clinicians; to policy makers, health economists, and administrators; to pharmaceutical and other health companies. These various perspectives are described in more detail below (see also Figure 2).

## Utility of oncology EMR compared to other available data sources

Several clinical and research databases are currently being used to conduct clinical outcomes research in oncology. Table 3 provides a summary of some of these data sources and their advantages and disadvantages when conducting outcome research, and how an "ideal" EMR could improve upon such databases. These data sets vary in breadth of coverage (national to regional), funding source (private to national government funded), mechanism of data collection (continuous, post hoc), and scope of data collected. While each contains unique elements that allow it to answer specific oncology research questions among certain populations, each is also limited by specific constraints that may be improved upon by a widespread, comprehensive, and provider/insurance-neutral EMR network of oncology practices.

An ideal oncology EMR would provide several advantages over current research, observational studies, and claims databases. Clinical data gathered in the course of routine medical care, if systematically collected and routinely stored, would provide a data source less costly in labor and operational funding than current research databases; would be more accessible than integrative HMO data; would prospectively collect patient data and characteristics, permitting analyses of factors from many points during the course of disease both pre- and post-diagnosis (unlike most population-based case-control studies); and would follow the entire patient cancer experience from diagnosis onward, collecting information that may affect outcomes. Further, this information would

**Table 2** Examples of outcomes research questions that may be addressed with an oncology-specific EMR

| Relevance to | Question/outcome | Background | How oncology EMR could address |
|---|---|---|---|
| Patient | Does using postmenopausal hormones after being diagnosed with colorectal cancer improve chances of survival? | The use of postmenopausal hormones has consistently been associated with reduced risk of CRC among women,[45,46] including results from a randomized clinical trial evaluating combined hormone therapy with conjugated equine estrogen plus medroxyprogesterone acetate compared with placebo.[47] Although some evidence indicated that hormone use at diagnosis may be associated with improved survival from CRC,[48,49] other studies have observed no effect.[50] To date, only one study has evaluated the use of postmenopausal hormones after diagnosis;[48] improved survival was observed among hormone users after diagnosis, although very limited treatment information was available | This question is difficult to assess in observational epidemiologic studies, because post- and prediagnosis use are highly correlated,[48] and because hormone use after diagnosis may be additionally influenced by factors such as burden of disease, response to cancer treatments, and physician recommendations. It is also unclear whether any observed result is due to a modifying effect that estrogens have on cancer progression, or whether the presence of exogenous estrogens during carcinogenesis contributes etiologically to a less aggressive form of disease.[49] A comprehensive EMR system that includes information on medications and treatments, survival, as well as some common health-related factors (eg, smoking, BMI), within one linked record could effectively be used to explore these inter-relationships |
| Researcher | Does response to treatment for breast cancer differ with specific molecular characteristics of tumors? Can these molecular signatures help predict who will have a favorable response to specific regimens? | While cancers of individual organs were once thought to be homogeneous and distinct entities, it is now apparent that these cancers are, in reality, heterogeneous entities that can be further classified by histologic subtype, and beyond that, by molecular subtype. Further, it appears that multiple subtypes of cancer can have distinct etiologic pathways and clinical relevance, with differing presentation at diagnosis and different prognoses.[51] For example, while breast cancers have long been divided into histopathologic subtypes (eg, ductal, lobular, inflammatory), and more recently, characterized by tumor marker expression status (ER/PR positive or negative), breast cancers are now classified into five clinically relevant subtypes based on gene expression profiles.[52] Although it is fairly well accepted that differences exist in the clinical relevance of these subtypes in terms of tumor stage, grade, and survival probabilities,[52–55] these observations have frequently been based on data from clinical case series. Further, while there have been suggestions that response to treatment may differ based on these molecular profiles,[56,57] more clinical research is needed to fully elucidate the complex pathways that dictate drug sensitivity and response | By combining the planned and actual treatment, side effects, and response to treatment with histopathological and molecular subtyping obtained from tumor tissue samples obtained during biopsy and/or surgical resection, differences in response to treatment may be detected. While this type of research is highly dependent on whether and what kind of molecular analyses are performed, EMR systems could be capable of collecting and linking results of any such testing. Further, EMR could assist in tracking the collection and storage of biological specimens, such that they could be located easily and retrieved by researchers to conduct future analyses |

**Abbreviations:** BMI, body mass index; CRC, colorectal cancer; EMR, electronic medical record; ER, estrogen receptor; PR, progesterone receptor.

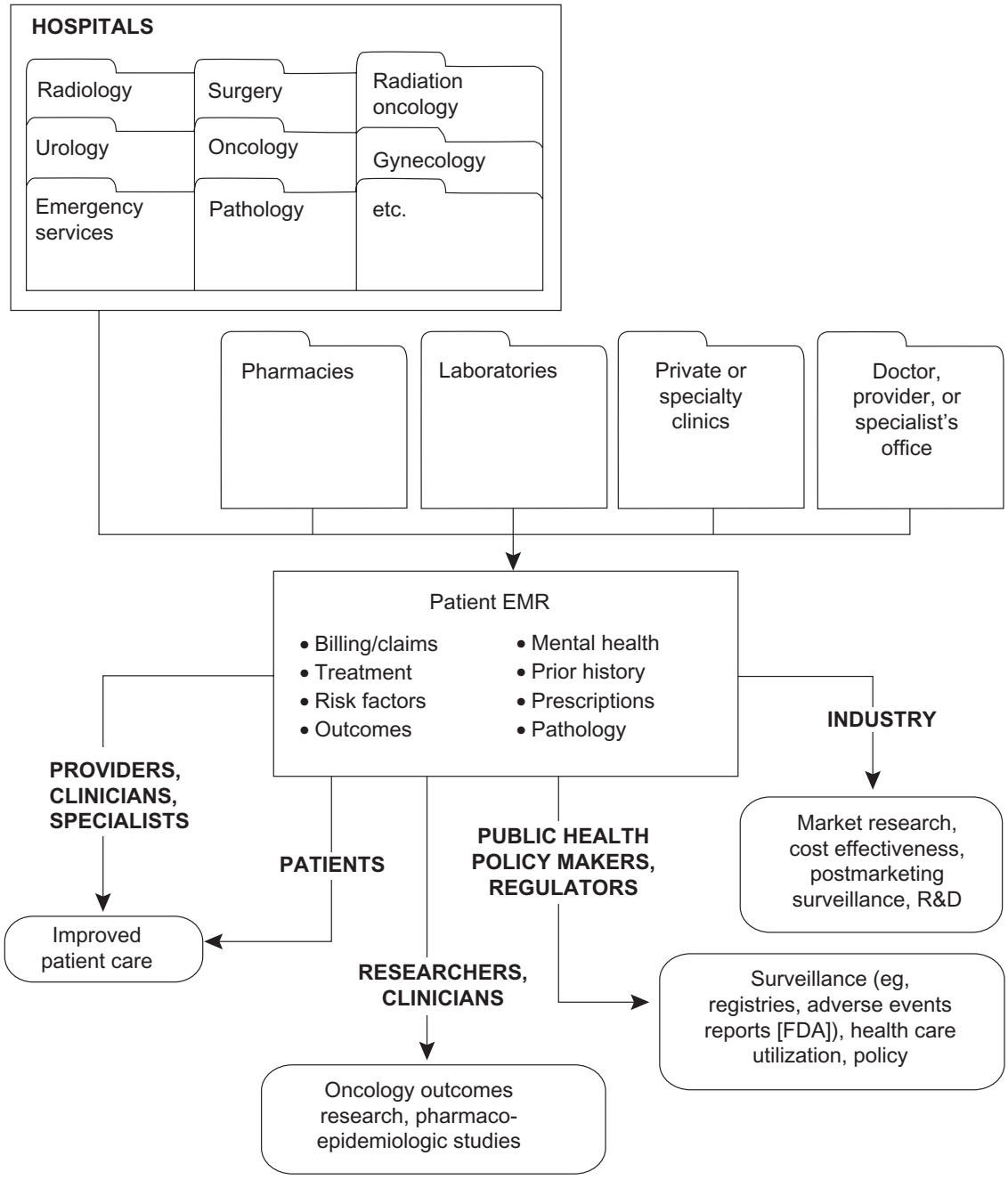


**Figure 2** Utility of electronic medical record (EMR) to various groups for outcomes research.

be more suitable for real-world research than data obtained in clinical trials.

## Relevance to the patient, clinician, and medical researcher

For patients, providers, and researchers, the most relevant questions focus on improving health and quality of life (QOL), providing better quality of care, and advancing disease knowledge. For example, an oncology patient would be interested in their chances of surviving existing disease and preventing future tumor recurrence. While many population-based observational studies have examined the relation of pre-diagnosis lifestyle factors – such as weight, diet, exercise, nonsteroidal anti-inflammatory drug (NSAID) use, alcohol consumption, and smoking – with cancer risk, the effect of risk factors on survival after diagnosis has been more difficult to study. In particular, studies that attempt to elicit self-reported risk factors occurring after diagnosis are susceptible

**Table 3** Summary of selected existing databases and improvements provided by oncology EMR

| Data type | Description | Examples | Advantages | Disadvantages | How could oncology EMR improve? |
|---|---|---|---|---|---|
| Registries | Compiled to collect information on incidence, mortality, prevalence. Can be public or private, national or regional | Surveillance Epidemiology End Results (SEER) (other examples include individual state cancer registries; C-CFR, B-CFR) | Population-based census of all cancer incidence in a given area. For example, the 17 SEER registries provide near-national coverage of cancers in the US (26% of US population, 74 million persons) from 1973 to present | Generally does not include detailed risk factor information, such as detailed SES, screening history, lifestyle factors (eg, obesity, smoking) | Provide detailed risk-factor information |
| | | | Comprehensive<br>– Includes all incident cancers (excluding basal and squamous cell skin cancer)<br>– Includes information on patient demographics, tumor site, histopathology, stage at diagnosis, tumor behavior/laterality, some information about first course of treatment, follow-up for vital status<br>– Includes mortality data | For publicly accessible data, treatment information limited to surgery and radiation (general information on receipt of chemotherapy, immunotherapy, and hormonal therapy may be requested from individual contributing registries from POC studies) | Provide detailed information on all actual and planned treatments |
| | | | Some tumor marker information for certain cancers (ER/PR for breast cancer; AFP, hCG, and LDH for testicular cancer) | No information on side effects, co-morbidities, recurrences | Provide information on outcomes other than death, including side effects, co-morbidities, recurrences<br>Could provide detailed follow-up information for longitudinal/prospective analyses |
| | | | Geographically and ethnically diverse coverage; includes all payers and all ages | Limited to first course of treatment (typically first 6 months) | Could provide information on all courses of treatment |
| | | | Standardized electronic data content and comparability between contributing registries allow data pooling; standardized quality control and quality check | Not prospective | |
| Health care system databases | Can contain national and international data; public, private | Kaiser Permanente (KP) Medical Care Program (other examples include Group Health Cooperative, Harvard Pilgrim Health Care, Henry Ford Health System, and other member organizations of the HMO Research Network; Medicare; and the Veterans Administration Databases) | Includes linked data from both in- and out-patient care, mental health, laboratory and radiographic results, all prescribed and OTC medications | Data not publicly accessible (generally accessible only to organizational researchers); some vendors (eg, Premier, Ingenxi) capture data from large health plans, but data must be purchased on a case-by-case basis and can be costly | Data could be made publicly available if certain criteria met (eg, patient confidentiality) |
| | | | Stable patient population | Patient coverage may not be representative of general population | Depending on how participating providers are selected, may provide more representative sample of general cancer population with health insurance coverage |

| | | | | | |
|---|---|---|---|---|---|
| | | | Large coverage (approximately 8.2 million subscribers) | Quality of data can vary considerably from one plan to another depending on, eg, how much effort is spent abstracting the information from physicians' notes, laboratory results | |
| | | | Depending on system, data are collected automatically during the course of conducting clinical practice – no data entry, validation, and post ad data collection required | | |
| | | | Wide geographic coverage | | |
| Research studies | Public, private, national, regional, specific population groups (eg, Iowa Women's Health Study; Black Women's Cohort; California Teachers Study) | Nurses' Health Study (other examples might include Women's Health Initiative, Framingham Heart Study) | Large enrollment population | Labor intensive (information collected via questionnaire and transcribed manually); regular follow-ups | Data collected during the course of clinical treatment and care on active basis |
| | | | Provides updates in changes of risk factors, modifiable behaviors | Extremely expensive | More streamlined costs once recording process established |
| | | | Active follow-up for incident medical outcomes | Large study staff required | Could reduce potential for manual transcription and other errors |
| | | | Detailed information on actual treatments | Data not automatically collected; some data are selective | Data could be made publicly available if certain criteria met (eg, patient confidentiality) |
| | | | | Generally not publicly accessible; access generally requires the cooperation of a co-investigator of the study and/or a lengthy application process | |
| Administrative claims data | Large federally funded health insurance program for specific demographic groups that includes coverage for, among other services, prescription drug coverage | Medicare (other examples include Medicaid, also administered by CMS) | Large enrollment population (all persons >65 years old) | Limited to the elderly population | Could cover all ages, demographics, and payers |
| | | | Allows for longitudinal/prospective analyses using unique BEF ID | Limited to one specific payer | Data could be made publicly available if certain criteria met (eg, patient confidentiality) |

(*Continued*)

**Table 3** (*Continued*)

| Data type | Description | Examples | Advantages | Disadvantages | How could oncology EMR improve? |
|---|---|---|---|---|---|
| | | | Nationwide coverage | No prescription information (limited to those not self-administered drugs); also do not have clinical information (eg, test results) or patient attributes (eg, age, sex, race) | Include prescription and OTC information |
| | | | Includes linked data from both in- and out-patient care, physicians, skilled nursing home, hospice, home health agency, and DME | Data not publicly accessible (but can be purchased; approvals required for research purposes) | |
| | | | Outcomes can be tracked at various levels (patient, physician, institution) | | |
| Surveys | Public and private surveys, national coverage | National Health and Nutrition Examination Survey (NHANES) (other examples include NIS, NHAMCS, NAMCS, KID, NHDS, CDC BRFSS) | Nationwide coverage | Snapshot picture of pharmaceutical use and disease risk factors taken once every ~10 years; useful for prevalence estimates and changes in prevalence of behaviors/usage over time (eg, for NAMCS, drug information limited to office visits where medical therapy is administered; certain physicians (anesthesiology, radiation) excluded) | Risk factors, behaviors, and pharmaceutical use linked by individual patient identifiers to health outcomes |
| | | | For some surveys (eg, NHANES), medical questionnaires are included as well as epidemiologic follow-up for older surveys (NHANES I, II, and III) | Follow up epidemiologic studies are somewhat limited (eg, NHANES II and III only link mortality and cause of death; NHANES I does include some additional information) | Continuous coverage of medical history (rather than a snapshot) |
| | | | Data set is publicly available after NCHS publication | Not prospective in nature and cannot be linked across years or across surveys; behaviors not linked to any health outcomes, including cancer | |

**Abbreviations:** AFP, alpha fetal protein; BEF, beneficiary encryption file; B-CFR, Breast Cancer Family Registry; BRFSS, Behavioral Risk Factor Surveillance System; C-CFR, colon cancer family registry; CDC, Centers for Disease Control and Prevention; CMS, Centers for Medicare and Medicaid Services; DME, durable medical equipment; ER/PR, estrogen receptor/progesterone receptor; hcG, human chorionic gonadotropin; KID, Kids Inpatient Database; KP, Kaiser Permanente; LDH, lactate dehydrogenase; NAMCS, National Ambulatory Medical Care Sample; NCHS, National Center for Health Statistics; NHAMCS, National Hospital Ambulatory Medical Care Samples; NHANES, National Health and Nutrition Examination Survey; NHDS, National Hospital Discharge Survey; NIS, National Inpatient Sample; OTC, over the counter; POC, patterns of care; SEER, Surveillance Epidemiology End Results; SES, socioeconomic status.

to lead-time bias and recall bias,[10] particularly if the time between diagnosis and interview is protracted, and only certain individuals survive long enough to be interviewed. Further, because the experience of cancer treatment, as well as the presence of disease itself, can affect behaviors, pre-diagnosis risk-factor information may not accurately reflect behaviors after diagnosis. An EMR system that collects information on some common health-related behaviors both pre- and post-diagnosis, medications and treatments, and response to treatment in a standardized fashion could aid in elucidating the complex inter-relationships among these factors. In addition, patient choice – such as the type of therapy or the decision to treat their disease at all – is a primary predictor of treatment patterns in several cancers,[11–15] and information collected by EMR on treatment efficacy and side effects may directly influence the decision-making process.

A practicing oncologist might be interested in studies of various chemotherapeutic agents, dosages, and regimens to prolong patient survival with the least risk of toxicity, side effects, and other detriments to QOL. Although clinical trial data provide some evidence of efficacy, they are conducted within a controlled treatment structure and among patients that are highly selected by health status, tumor type, and previous response to treatment. Thus, the data from these trials may not provide a realistic view of treatment response within a "real-world" administration. In addition to wide variation among cancer patients with respect to personal and tumor characteristics, differences in psychological factors and lifestyle choices may influence their course of disease and response to treatment. Even among patients with similar prescribed regimens and dosages, individual adherence to recommended treatment, and differences in patient choices of treatment type, may further influence response.

In an EMR-based oncology practice, the adherence to prescribed regimen, reported treatment side effects, psychological concerns, and possibly other modifiers of treatment efficacy would be recorded with each clinic visit. In combination with quantifiable outcomes data (eg, disease progression or survival), an oncology EMR can provide clinicians with a more representative data source by which to develop their medical recommendations. These data from an EMR could also help researchers investigate ways to improve guideline adherence and implementation of evidence-based medicine for follow up and treatment that may improve the care received by the oncology patient.

Advances in biotechnology and bioinformatics have permitted the examination of etiologic and outcomes research in oncology with increasing detail, with molecular, cellular, and other biological measurement often incorporated to characterize cancer and disease progression (see Table 2). EMR systems that record such biomarker data, when available, over the course of clinical practice could provide a potentially cost-effective[16] and efficient means for conducting this research, and are a valuable tool in pharmacoepidemiologic studies.

## Relevance to the policy maker

From the policy makers' perspective, outcomes research in oncology can provide clues to identify deficits in outreach and programmatic spending and current health care resource utilization and accessibility. The recent Institute of Medicine (IOM) report on optimizing the effectiveness of health care services[17] recommends conducting systematic reviews of current published research on the effectiveness of health services to create a link between scientific evidence and clinical decision-making. New research studies using comparative effectiveness study methods (eg, comparing new drugs, devices, and procedures to current treatment options)[18,19] can also provide answers on what treatments are most effective from both the cost and clinical standpoints for specific populations, such as addressing whether health care resources are being spent on the most effective treatments for oncology patients. EMR systems could be a valuable resource for conducting these research studies in oncology, because they contain more comprehensive health histories and outcome information than claims databases or other disease registries for researchers to address a variety of treatments and outcomes.

Research and public health measures aimed at identifying deficits and disparities in medically underserved groups, and mitigating these differences, have become high priorities of government health care program policies.[20–23] Rates of cancer incidence, mortality, and survival may differ by age, race/ethnicity, socioeconomic status (SES), educational attainment level, and geographic location,[24] and it is thought that access to health care, screening, and treatment resources, and the quality of treatment given, may underlie a large proportion of these differences.[25] For example, analyses of a population-based cancer registry of patients aged 66 to 85 years with a primary diagnosis of colorectal, breast, lung, or prostate cancer during 1992 to 2002, researchers found that black patients were significantly less likely than white patients to receive therapy for their cancer, and there was no evidence that the magnitude of racial disparities had lessened during the study period.[26] Disparities in cancer treatment can vary across cancer types, with differential impacts on access to

care, ability to absorb costs, and cultural and personal biases, all of which are factors not currently captured in most available data. EMR systems that contain detailed insurance, personal, and clinical information may form a more complete picture of the many factors that influence treatment patterns, providing policy makers with better direction when directing health care policies.

## Relevance to the pharmaceutical industry/manufacturer

EMR systems could provide answers to research questions relevant to several groups within the pharmaceutical and biotechnology industries. With comprehensive and accurate information on various aspects of patient-care information, EMR could potentially provide a set of clinical data to answer questions related to marketing, outcomes research, research and development, and phase IV post-marketing surveillance (eg, adverse events identification, targeted evaluation of treatment groups, observational studies), among others. For example, information in EMR (received by the pharmaceutical industry data de-identified of personal information) could be mined to determine the size of therapeutic markets for rare cancers or conditions, or in populations that are resistant to or cannot take conventional medication. EMR could also aid industry in evaluating the prevalence or incidence of certain conditions and associated co-morbidities, in various subpopulations, in certain types of hospitals, or in different areas, to aid in marketing drugs and supporting new drug applications and other regulatory documentation.

Economic analyses and health care resource use evaluations would also be possible with EMR systems. Cost-effectiveness strategies of pharmacologic therapy could be evaluated, including for example, population of predictive health economics models for forecasting costs and market share. Forecasting models could also be developed from these data, with particular focus on specific drugs, disease stage, line of therapy, or clinic type, provided that issues of generalizability were accounted for (see later section). From a marketing point of view, EMRs could also be used to develop differentiation strategies among competing firms by evaluating patient profiles and drug use.

Because EMRs represent a "real-world" patient cohort, these data could also be used to aid in preparation of regulatory applications using actual population-based data, as well as in identification of possible safety issues, and could therefore also be of use in pharmacovigilance studies. A clear EMR record of all medications prescribed and taken, whether for medications, biological products, or herbal supplements could aid in understanding and preventing adverse events with certain therapeutics. EMRs may be beneficial for process-of-care studies, where insight into patient care, disease endpoints, and therapies could be evaluated in tandem in a single EMR and could aid in beneficial results for patients (eg, shorter lengths of stay in hospital, preventive or palliative care). Detailed clinical and laboratory history of patients undergoing treatment contained in EMR may reveal new indications or contraindications of existing therapeutics. Effectiveness of therapeutics could also be tracked in certain situations through clinical practice and reporting of patient outcomes.

# Challenges of using EMR for oncology research

Despite the promise that EMR can provide for oncology outcomes research, several challenges must be considered to develop a comprehensive, valid, and useful data resource. While many of these challenges of EMR are general, some are oncology specific. Challenges can include structural,[27,28] clinical,[29,30] and public health research-related issues,[31,32] all of which should be considered in building a useful EMR system.

## Structure

At present, a variety of options are available in EMR systems, and physician groups and health systems are free to choose the system best suited to their needs.[3] For example, the American Society of Clinical Oncology (ASCO) has developed a set of recommendations for oncologists to use in choosing an EMR.[33] ASCO identified several important elements that should be included in any EMR used by an oncologist: ability to enter staging information, work flow, chemotherapy doses and administration, toxicity assessment, clinical trial management, drug inventory management, and survivor care.

### Interoperability

Regardless of the type of EMR system chosen, it will need to be interoperable, that is, having the ability to interact and exchange data across departments and clinics (eg, laboratory, pathology, clinic).[34] Without interoperable interfaces, information cannot be exchanged easily, which can affect the work flow for the clinic and could lead to missing information within the medical record and major problems with data quality and completeness.

### Incorporation of different data formats

Information available in an EMR may include a variety of data formats, ranging from text for office visits to images

from radiology visits. In addition, results can include text reports, numerical data such as laboratory values, graphs, and images such as CT scans or MRIs. As such, the results can be entered using a wide variety of methods, including direct data uploads from the processing laboratory, scanning of text using optical character recognition, or simply capturing the electronic readout as an image. Physician orders can be recorded in several ways, with varying degrees of automation, including selecting individual chemotherapy drugs with pre-populated fields or customizing automated orders with free-text entry. However, treatment offered outside the institutions that contribute to the EMR (eg, radiation therapy) may be available only from information scanned into the EMR system, which may not be readily searchable by computer algorithms if they contain non-text fields such as images.

Due to the variety of sources, data collection may be time intensive, requiring manpower and database management.[35] Because some of the information in the EMR may be text based,[27] rather than consisting of a data set or collection of raw values, natural language algorithms may be necessary to extract the information of interest. Studies using natural language algorithms have been able to identify a majority of potential study participants by disease type. For example, a study designed to identify cases of heart failure from an EMR found 100% sensitivity and 97.8% specificity when using this method.[36] Other limitations may include not being able to extract information from reports that have been scanned into the medical record using these algorithms, and not having access to previous paper records that have not been incorporated into the EMR.[3]

## Clinical

### Variety of diagnostic coding vocabulary

The coding vocabulary used to identify and define cancer cases within the medical record varies by physician type. Standard codes include the American Joint Committee on Cancer (AJCC) codes for tumor staging, Systematized Nomenclature of Medicine (SNOMED) used primarily by pathologists, and various International Classification of Diseases (ICD) codes (ICD-9-CM, ICD-10, and ICD-O), and Current Procedural Terminology (CPT) for coding of diagnoses and procedures performed during diagnosis and treatment. In some cases, the clinicians are not the ones directly recording the codes, but they provide a narrative of the disease history or clinic visit that is subsequently coded by medical record coders at a later time, strictly for billing purposes. Disease progression of cancer, in particular metastasis or recurrence, may follow a variety of pathways, and there is no standard for recording progression in the medical record, nor are cancer-specific codes available for identifying cancers that have progressed. Because several standardized codes are used over the course of an oncology patient's care by different departments and treating physicians, with some used primarily for billing purposes, the researcher will need to consider all sources of codes, and possibly lack of codes, to identify cancer cases or they may miss identifying all cases of a particular cancer.

### Linking of departments and continuity of care

Cancer, as a diagnosis, encompasses a variety of tumor types that are treated by physicians in a variety of specialties, including general surgery, hematology, oncology, urology, and endocrinology. Within a comprehensive health care system, these specialties may be handled with an interconnected EMR; however, outside these health systems, the specialties may not be linked in a standardized fashion, or some specialty department may not use EMR. In addition, practices are often challenged in the amount of retrospective information that they receive from prior health care providers and institutions, and are also constrained by costs and logistics in the amount of data they can enter into the EMR due to inadequate time and staffing. Longitudinal observations of study subjects may be difficult if data prior to the adoption of the EMR have not been added to the database. The concerns of not incorporating records from all specialties providing care are 2-fold.

First, patients with some cancers, such as prostate cancer or early-stage breast cancer, may be diagnosed and treated primarily within one or two departments (eg, urology, surgery) or, as is the case for the majority of patients, different outpatient clinics entirely. If the EMR does not link these departments, these patients and essential elements of their clinical history may be missed. This may also make the construction of the oncology patient's cancer experience incomplete. Similarly, for childhood cancers, specialized hospitals and clinics that deal primarily with pediatric cancers may or may not be connected with the EMR where the children receive their primary care. For researchers interested in other subpopulations, such as those investigating rare cancers or conditions that are difficult to diagnose in some cases (eg, Ewing's sarcoma), the low adoption rates of EMR may present a barrier to identifying these small groups, because the number of cases may be too small from available EMR data, thereby preventing a study from being adequately powered for research.

Second, outcomes of procedures that are not considered clinically relevant to a particular department may not be captured in their EMR system. For example, persisting post-surgery incision pain that affects QOL may not be recorded within a surgical department EMR as an outcome of surgery, particularly if there is a long lag time after the procedure or if the patient seeks care from their primary care physician.

## Research-specific

The most obvious problem with using EMR systems for research is that they collect information and are designed for the care of each individual patient; the data are not collected specifically for research. For example, demographic data collected for patient care, rather than research purposes, may not be collected systematically with standardized definitions. Thus, the rigors of data collection for research, including validation, quality checking, and reproducibility, are not inherent in the EMR-collected data. Issues related to the reliability and validity of a study, such as dealing with missing data and generalizability to the source population, as well as ethics and privacy concerns and general accessibility when using an EMR, must be considered.

### Missing data

Missing information in the EMR may lead to misclassification of exposure and/or outcome in epidemiologic and other research.[31] Another consideration is the extent of the data collected within the EMR (eg, personal/medical histories, planned versus actual treatment regimens, psychosocial parameters, follow-up on treatment outcomes). Because various physicians or practices may record information differently, the researcher will have to make decisions on whether the collected data set is complete. A study performed in a network of primary-care clinics determined that clinicians reported missing clinical information in almost 14% of visits.[37] In addition, there may be a lag time between the clinic visit and data entry into the EMR. The researcher will also need to distinguish whether information is missing or was just never followed through or recorded. For example, the researcher may need to determine whether a record is requesting a treatment or recording that treatment was given. Insurance claims data sets capture treatments that are given; thus, an EMR linked to an insurance claims database could help ascertain the difference. If paper records are available, they could also be used to identify missing data. Researchers will need to account for missing data in their analyses to ensure robustness and accuracy.

### Generalizability

The population covered within an EMR may not be representative of the underlying population in the region, thus limiting the generalizability of potential research studies to larger populations or to the US in general.[31] Due to the low adoption rates of EMR mentioned previously, there may be few clinics or hospitals in a geographic area that collect medical records electronically. Clinics and hospitals that have adopted EMR may be different in terms of the patient population that uses the facilities compared to those clinics and hospitals that have not adopted EMR. Even within clinics, there may be differences in patient participation and enrollment, depending on physician and patient preference, thereby possibly skewing the population characteristics. Researchers should keep generalizability in mind when interpreting the results of studies conducted with EMR data.

### Ethical and privacy concerns

Any large database of identifiable personal health information (PHI), such as that from an EMR, is subject to the Common Rule[38] and the Health Insurance Portability and Accountability Act (HIPAA), which also has a privacy rule embedded within it.[39] The researcher must work with the Institutional Review Board (IRB) associated with the EMR to ensure that the research activity protects the privacy and confidentiality of the patients included. Additional safeguards to protect the privacy and confidentiality of subject data can include de-identification of PHI, password protection, data encryption, and use of firewalls if there is potential for Internet access to the information.[40]

# Conclusions and future directions

Cancer continues to be a leading cause of morbidity and mortality, and there is an increasing need for new data sources to address questions regarding oncology outcomes, particularly when current data systems used in research are limited. The drive to achieve a nationwide system of electronic medical records, and the recommendation by the IOM to optimize health services through systematic reviews is a call for researchers to engage with medical professional groups to help guide development of EMRs and EMR-based systems that have utility for several groups, including patients and policy makers. This interaction will ensure that secondary uses of the data are secure and protect patient privacy, are useful to researchers and others, and will help improve the quality of cancer care through oncology outcomes research.

## Acknowledgments/disclosures

Funding for this manuscript provided by Amgen, Inc. At the time of writing Drs O'Malley, Fryzek and Nordyke were employees and stockholders of Amgen, Inc.; they have no relevant conflicts of interest. Drs Kanas, Morimoto, and Mowat are employees of Exponent, Inc. and have no relevant conflicts of interest.

## References

1. National Center for Health Statistics. *Health, United States, 2007 with Chartbook on Trends in the Health of Americans*. Hyattsville, MD 2007.
2. Galligioni E, Berloffa F, Caffo O, et al. Development and daily use of an electronic oncological patient record for the total management of cancer patients: 7 years' experience. *Ann Oncol*. 2009;20(2):349–352.
3. Mehta NB, Partin MH. Electronic health records: a primer for practicing physicians. *Cleve Clin J Med*. 2007;74(11):826–830.
4. American Recovery and Reinvestment Act of 2009. http://www.gpo.gov/fdsys/pkg/PLAW-111publ5/content-detail.html. accessed Feb 23, 2009.
5. Burt CW, Sisk JE. Which physicians and practices are using electronic medical records? *Health Aff (Millwood)*. 2005;24(5):1334–1343.
6. Thakkar M, Davis DC. Risks, barriers, and benefits of EHR systems: a comparative study based on size of hospital. *Perspect Health Inf Manag*. 2006;3:5.
7. Baron RJ, Fabens EL, Schiffman M, Wolf E. Electronic health records: just around the corner? Or over the cliff? *Ann Intern Med*. 2005;143(3):222–226.
8. Dean BB, Lam J, Natoli JL, Butler Q, Aguilar D, Nordyke RJ. Use of electronic medical records for health outcomes research: A literature review. *Med Care Res Rev*. 2009;66(6):611–638.
9. Huertas Fernandez MJ, Baena-Canada JM, Martinez Bautista MJ, Arriola Arellano E, Garcia Palacios MV. Impact of computerised chemotherapy prescriptions on the prevention of medication errors. *Clin Transl Oncol*. 2006;8(11):821–825.
10. Schmier JK, Halpern MT. Patient recall and recall bias of health state and health status. *Expert Rev Pharmacoecon Outcomes Res*. 2004;4(2):159–163.
11. Cardoso F, Di LA, Lohrisch C, Bernard C, Ferreira F, Piccart MJ. Second and subsequent lines of chemotherapy for metastatic breast cancer: what did we learn in the last two decades? *Ann Oncol*. 2002;13(2):197–207.
12. Donovan KA, Greene PG, Shuster JL, Partridge EE, Tucker DC. Treatment preferences in recurrent ovarian cancer. *Gynecol Oncol*. 2002;86(2):200–211.
13. Penson RT, Dignan F, Seiden MV, et al. Attitudes to chemotherapy in patients with ovarian cancer. *Gynecol Oncol*. 2004;94(2):427–435.
14. Osoba D. Lessons learned from measuring health-related quality of life in oncology. *J Clin Oncol*. 1994;12(3):608–616.
15. Weeks JC, Cook EF, O'Day SJ, et al. Relationship between cancer patients' predictions of prognosis and their treatment preferences. *JAMA*. 1998;279(21):1709–1714.
16. Schmier JK, Kane DW, Halpern MT. Practical applications of usability theory to electronic data collection for clinical trials. *Contemp Clin Trials*. 2005;26(3):376–385.
17. Institute of Medicine (IOM). *Knowing What Works in Health Care: A Roadmap for the Nation*. Washington, DC: The National Academies Press; 2008.
18. Agency for Healthcare Research and Quality. Effective Health Care. URL: effectivehealthcare.ahrq.gov, accessed February 17, 2009.
19. Congressional Budget Office. Research on the Comparative Effectiveness of Medical Treatments: Issues and Options for an Expanded Federal Role; 2007 December. http://www.cbo.gov/doc.cfm?index=8891. accessed Feb 23, 2009.
20. Health Disparities Collaboratives (HDC) URL: http://www.healthdisparities.net/hdc/html/collaborativesOverview.aspx. Accessed June 25, 2008.
21. Healthy People 2010: Understanding and Improving Health. 2nd ed. In: *US Department of Health and Human Services*, 23 Aug, 2000. US Government Printing Office; 2000.
22. Strategic research plan to reduce and ultimately eliminate health disparities, fiscal years 2002–2006. 2002. http://ncmhd.nih.gov/our_programs/strategic/pubs/VolumeI_031003EDrev.pdf. Accessed June 25, 2008.
23. Lisovicz N, Wynn T, Fouad M, Partridge EE. Cancer health disparities: what we have done. *Am J Med Sci*. 2008;335(4):254–259.
24. The NCI Strategic Plan for Leading the Nation to Eliminate the Suffereing and Death Due to Cancer. In: Institute NC, ed NIH Publication No. 06-5773. US Department of Health and Human Services, National Institutes of Health; 2006.
25. Li CI, Malone KE, Daling JR. Differences in breast cancer stage, treatment, and survival by race and ethnicity. *Arch Intern Med*. 2003;163(1):49–56.
26. Gross CP, Smith BD, Wolf E, Andersen M. Racial disparities in cancer therapy: did the gap narrow between 1992 and 2002? *Cancer*. 2008;112(4):900–908.
27. Hayrinen K, Saranto K, Nykanen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *Int J Med Inform*. 2008;77(5):291–304.
28. Ambinder EP. A history of the shift toward full computerization of medicine. *J Oncol Pract.* 2005;1(2):54–56.
29. Ambinder E. electronic health records. *J Oncol Pract*. 2005;1(2):57–63.
30. ASCO Electronic Health Record Roundtable. Ensuring continuity of care through electronic health records. *J Oncol Pract*. 2007;3(3):137–142.
31. Aschengrau A, Seage GR. *Essentials of Epidemiology in Public Health*. Boston: Jones and Bartlett Pub; 2003.
32. Goldwein J. Using an electronic health record for research. *J Oncol Pract*. 2007;3(5):278–279.
33. ASCO Develops Practical Guide for Selecting, Implementing Electronic Health Records. *ASCO Daily News*. 2008:24B.
34. Detmer D, Bloomrosen M, Raymond B, Tang P. Integrated personal health records: transformative tools for consumer-centric care. *BMC Med Inform Decis Mak*. 2008;8:45.
35. Weiner MG, Lyman JA, Murphy S, Weiner M. Electronic health records: high-quality electronic data for higher-quality clinical research. *Inform Prim Care*. 2007;15(2):121–127.
36. Pakhomov S, Weston SA, Jacobsen SJ, Chute CG, Meverden R, Roger VL. Electronic medical records for clinical research: application to the identification of heart failure. *Am J Manag Care*. 2007;13(6 Part 1):281–288.
37. Smith PC, Araya-Guerra R, Bublitz C, et al. Missing clinical information during primary care visits. *JAMA*. 2005;293(5):565–571.
38. Department of Health and Human Services. Protection of Human Subjects. 45 CFR 46. 2008;128–140. http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.htm. accessed Feb 23, 2009.
39. Health Insurance Portability and Accountability Act of 1996 (HIPAA), PL 104-191-2003.
40. Wolf LE, Walden JF, Lo B. Human subjects issues and IRB review in practice-based research. *Ann Fam Med*. 2005;3 Suppl 1:S30–S37.
41. Bush GW. Exec. Order No. 13335: Incentives for the Use of Health Information Technology and Establishing the Position of the National Health Information Technology Coordinator40. Weekly Compilation of Presidential Documents; 2004:702–704.
42. Food and Drug Administration. FDA's Critical Path Initiative. URL: http://www.fda.gov/oc/initiatives/criticalpath/. Accessed February 23, 2009.
43. Food and Drug Administration. *The Sentinel Initiative, National Strategy for Monitoring Medical Product Safety*: Department of Health and Human Services; May 2008.
44. US Department of Health and Human Services. *Achieving a Transformed and Modernized Health Care System for the 21st Century: CMS Strategic Action Plan 2006–2009.* October 16, 2006.

45. Grodstein F, Newcomb PA, Stampfer MJ. Postmenopausal hormone therapy and the risk of colorectal cancer: a review and meta-analysis. *Am J Med*. 1999;106(5):574–582.
46. Hebert-Croteau N. A meta-analysis of hormone replacement therapy and colon cancer in women. *Cancer Epidemiol Biomarkers Prev*. 1998;7(8):653–659.
47. Chlebowski RT, Wactawski-Wende J, Ritenbaugh C, et al. Estrogen plus progestin and colorectal cancer in postmenopausal women. *N Engl J Med*. 2004;350(10):991–1004.
48. Chan JA, Meyerhardt JA, Chan AT, Giovannucci EL, Colditz GA, Fuchs CS. Hormone replacement therapy and survival after colorectal cancer diagnosis. *J Clin Oncol*. 2006;24(36):5680–5686.
49. Slattery ML, Anderson K, Samowitz W, et al. Hormone replacement therapy and improved survival among postmenopausal women diagnosed with colon cancer (USA). *Cancer Causes Control*. 1999;10(5):467–473.
50. Ritenbaugh C, Stanford JL, Wu L, et al. Conjugated equine estrogens and colorectal cancer incidence and survival: The Women's Health Initiative Randomized Clinical Trial. *Cancer Epidemiol Biomarkers Prev*. 2008;17(10):2609–2618.
51. Chen YC, Hunter DJ. Molecular epidemiology of cancer. *CA Cancer J Clin*. 2005;55(1):45–54; quiz 57.
52. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*. 2001;98(19):10869–10874
53. Bidard FC, Conforti R, Boulet T, Michiels S, Delaloge S, Andre F. Does triple-negative phenotype accurately identify basal-like tumour? An immunohistochemical analysis based on 143 'triple-negative' breast cancers. *Ann Oncol*. 2007;18(7):1285–1286.
54. Carey LA, Perou CM, Livasy CA, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA*. 2006;295(21):2492–2502.
55. Kim MJ, Ro JY, Ahn SH, Kim HH, Kim SB, Gong G. Clinicopathologic significance of the basal-like subtype of breast cancer: a comparison with hormone receptor and Her2/neu-overexpressing phenotypes. *Hum Pathol*. 2006;37(9):1217–1226.
56. Cleator S, Heller W, Coombes RC. Triple-negative breast cancer: therapeutic options. *Lancet Oncol*. 2007;8(3):235–244.
57. Sorlie T, Perou CM, Fan C, et al. Gene expression profiles do not consistently predict the clinical treatment response in locally advanced breast cancer. *Mol Cancer Ther*. 2006;5(11):2914–2918.