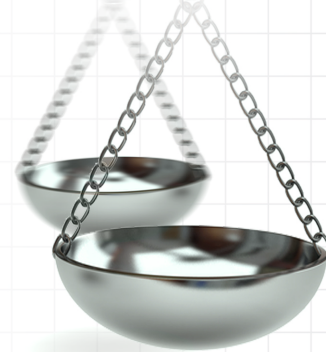


For reprint orders, please contact: reprints@futuremedicine.com



Using big data for quality assessment in oncology

Journal of **Comparative Effectiveness Research**

There is increasing attention in the US healthcare system on the delivery of high-quality care, an issue central to oncology. In the report 'Crossing the Quality Chasm', the Institute of Medicine identified six aims for improving healthcare quality: safe, effective, patient-centered, timely, efficient and equitable. This article describes how current big data resources can be used to assess these six dimensions, and provides examples of published studies in oncology. Strengths and limitations of current big data resources for the evaluation of quality of care are also discussed. Finally, this article outlines a vision where big data can be used not only to retrospectively assess the quality of oncologic care, but help physicians deliver high-quality care in real time.

First draft submitted: 17 December 2015; Accepted for publication: 19 February 2015; Published online: 19 April 2016

Keywords: big data • cancer • quality of care • registries

Healthcare in the USA has continued to change from rewarding 'doing more' to patients to a system that focuses on delivering 'high-quality' care [1]. The Institute of Medicine (IOM) in a report 'Crossing the Quality Chasm', described six aims for high-quality healthcare: safe, effective, patient-centered, timely, efficient and equitable [2]. With increasing emphasis on quality in healthcare overall, there have been parallel efforts in the research community to measure the quality of oncologic care. The goal of this article is to review how big data resources can be used to assess each of the six domains of quality healthcare defined by the IOM, and describe a future vision where big data can move from assessing quality to directly helping physicians deliver high-quality oncologic care.

Safe: avoiding injuries to patients from the care that is intended to help them

Big data are powerful resources for assessing the safety of medical treatments on a large scale, which can reveal novel insights. In

an observational study by Keating *et al.* [3] utilizing data from the Veterans Healthcare Administration, the authors found that prostate cancer patients treated with androgen-deprivation therapy (ADT) had higher incidences of diabetes and cardiovascular disease compared with patients not treated with ADT. Information on diabetes, coronary heart disease and myocardial function were ascertained using International Classification of Diseases, Ninth Revision (ICD-9) codes associated with inpatient and outpatient physician visits. This and other similar studies [4–6] have led to an increasing awareness of these potential effects of a common prostate cancer treatment.

The Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute (NCI) is a collection of population-based cancer registries across the USA, which provides patient information including: demographics, sex, age at diagnoses, tumor site and morphology, treatment and survival [7]. The population-based nature of this database makes it a useful tool for exam-

James R Broughman^{1,2}
& Ronald C Chen^{*1,2,3}

¹Department of Radiation Oncology, the University of North Carolina at Chapel Hill, CB #7512, Chapel Hill, NC 27599, USA

²School of Medicine, the University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

³Lineberger Comprehensive Cancer Center, the University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

*Author for correspondence:

Tel.: +1 984 974 0400

Fax: +1 984 974 8607

ronald_chen@med.unc.edu

Future
Medicine



ining cancer incidence and mortality across the USA. A study by Darby *et al.* [8] used SEER data to assess long-term mortality from heart disease in 308,861 women who received radiotherapy for breast cancer. Comparing the risk of mortality from heart disease in women who received radiotherapy for right-sided breast cancer versus left-sided breast cancer (the latter group of patients more likely to receive incidental radiation dose to the heart), this study found that patients treated from 1973 to 1982 on the left side had a higher risk for cardiac mortality after treatment. This was one of the first large-scale studies to quantify the risk of cardiac sequelae due to incidental radiation doses to the heart from treatment of a common disease (breast cancer).

SEER linked with Medicare data (SEER-Medicare) additionally provide Medicare claims and diagnostic data for patients 65 years and older, and is another commonly used data resource to examine population-based patterns of care and patient outcomes. In addition to the information provided by SEER, linked SEER-Medicare data contain diagnostic and billing codes including tests, procedures, outpatient visits, hospital admissions and home healthcare for Medicare beneficiaries. Several studies summarized in Table 1 demonstrate how SEER-Medicare data can be used to assess safety. The first example, a study by Pinder *et al.* [9] assessed the incidence of congestive heart failure in 43,338 women following anthracycline chemotherapy for breast cancer. Congestive heart failure diagnoses were identified through claims in Medicare inpatient, outpatient and physician files. A second example, by Hershman *et al.* [10] determined the rate of cardiac toxicity in patients treated with doxorubicin-based chemotherapy for diffuse large B-cell lymphoma using methodology similar to the previous study. A study by Sheets *et al.* [11] compared the rates of gastrointestinal and urinary morbidity, erectile dysfunction and hip fractures associated with various types of radiation treatments for localized prostate cancer.

Advantages of using big data to assess treatment safety

The power of big data to examine safety in cancer treatments lies in an ability to access and analyze information from large populations of patients, thus providing insights on sometimes potentially unexpected treatment effects not previously revealed in smaller studies.

Effective: providing services based on scientific knowledge to all who could benefit, & refraining from providing services to those not likely to benefit

In 1999, the IOM published a report describing the discrepancy between the ideal cancer care system and

that which was actually observed across the USA [33]. Among the list of described problems were underuse of cancer screening, lack of adherence to standards for diagnosis (inadequate biopsies, poor reporting of pathology studies), inadequate patient counseling regarding treatment options and underuse of adjuvant radiotherapy and chemotherapy following surgery in various cancers. Acknowledgement of these inconsistencies led to significant investments in research geared toward monitoring quality and clinical guideline adherence. Clinical practice guidelines are tools to help guide physicians and patients in diagnostic and treatment decisions. A commonly referenced set of guidelines in big data research are those created by the National Comprehensive Cancer Network (NCCN). Numerous published studies have utilized data from large cancer registries to examine patterns of care compared with 'standards' published in the NCCN guidelines. A few are summarized in Table 1 and below to illustrate.

In a study published by Gray *et al.* [12], the authors used the National Cancer Data Base (NCDB), a large cancer registry which includes ~70% of incident cancer patients across the USA. This study examined how frequently patients with muscle-invasive bladder cancer received curative-intent treatments based on clinical practice standards set by the NCCN and European Association of Urology. This study showed that curative-intent treatment was significantly underused in elderly patients. This was attributable to older patients not being medically eligible for radical cystectomy, yet not being referred for radiotherapy which is another guideline-recommended treatment option. Another study, by Chen *et al.* [13], used SEER-Medicare data to examine the proportion of Medicare prostate cancer patients across the USA who received treatments consistent with the NCCN guidelines. Similar to the prior study, this one also demonstrated that a sizable proportion of patients received potentially inappropriate care. Specifically, this study found that for patients with high-risk prostate cancer, the most aggressive form of localized disease, guideline-concordance ranged from 66.6% (for patients aged 66–69 years) down to 51.9% (aged 75–79 years). Another example, by Chagpar *et al.* [14], used the NCDB to assess adherence to NCCN guidelines in colon cancer patients. Guideline adherence was based on whether the appropriate treatment had been recommended, independent of whether it was actually received. This study reported that patients with stage I colon cancer received guideline-concordant treatment recommendations more frequently (96%) than those with stage II (low risk 66%; high risk 36%), stage III (71%) and stage IV (73%) disease.

Large cancer registries in some individual US states and in other countries have also been used to examine

patterns of care and guideline adherence. A study by van de Water *et al.* [15] assessed treatments received by breast cancer patients identified using The Netherlands Cancer Registry. Treatment was considered to be guideline-adherent based on Dutch recommendations for breast cancer management. Finally, Bristow *et al.* [16] assessed the frequency with which ovarian cancer patients identified from the California Cancer Registry received stage-appropriate surgical procedures and chemotherapy.

Big data can also be used to compare the effectiveness of different cancer treatments. A recent study used SEER-Medicare data to compare the outcomes of prostate cancer patients who received a robotic prostatectomy versus an older technique open prostatectomy [34]. For this comparison, big data were required because a randomized trial was not possible given the rapid diffusion of robotic surgery across the USA. This study showed that robotic versus open prostatectomy was associated with lower blood transfusion rates and slightly shorter hospital stays, but higher costs.

Advantages of using big data to assess delivery of effective treatment

While patterns of care and guideline concordance from single institutions would be of interest locally to the institution, population-based studies such as those described above can provide results with larger policy implications. For example, the significant underutilization of curative-intent treatments found by studies from Gray *et al.* [12] and Chen *et al.* [13] despite published guidelines calls for additional interventions, perhaps treatment pathways, to mitigate these significant problems on a national level. Big data are also useful for comparative effectiveness research especially when randomized trials are either not possible or impractical.

Patient-centered: providing care that is respectful of & responsive to individual patient preferences, needs & values, & ensuring that patient values guide all clinical decisions

The complexity of the cancer care system can make it difficult for patients and their physicians to coordinate a comprehensive and patient-centered treatment plan. Patient-centered care should seek to involve patients, families and the treatment team including physicians in conversations to develop thorough treatment plans that integrate medical, emotional and social needs. Studies using big data can be used to assess indicators of patient-centeredness in oncologic care.

An example was a study published by Burge *et al.* [17], which utilized the Nova Scotia Cancer Registry to determine if greater family physician continuity of care provided to cancer patients at the end-of-life

decreased emergency department utilization. It has been well documented that cancer patients generally prefer to spend their final days out of hospital [35–37], as trips to the emergency department can be emotionally, physically and financially burdensome. This study found that patients with low and moderate continuity of care were two- to four-times more likely to make visits to the emergency department compared with those with high continuity of care. In another study by Sharma *et al.* [18], the investigators used SEER-Medicare to quantify intensive care unit (ICU) usage during the last 6 months of life among 45,627 patients with advanced lung cancer. This study found that ICU usage during the last 6 months of life increased from 17.5% in 1993 to 24.7% in 2002 ($p < 0.001$); interestingly, there was also an increase in hospice usage during that same period (28.8–49.9%).

A study by Royce *et al.* [19] used data from the population-based National Health Interview Survey (NHIS) to examine the rates of prostate, breast, cervical and colorectal cancer screening in patients with limited life expectancy. The NHIS is an annual, large-scale survey conducted by the National Center for Health Statistics which is part of the Centers for Disease Control and Prevention that collects information on healthcare received by individuals across the USA; NHIS data are recognized as reflective of US population-based healthcare patterns. Although it is widely recognized that cancer screening in patients with limited life expectancy does not benefit (and likely harms) the patient [38], this study showed that a significant proportion of men and women with less than 5-year life expectancy across the USA received prostate (55%), breast (38%), colorectal (41%) and cervical cancer (31%) screening.

Advantages of using big data to assess delivery of patient-centered care

The studies described above have identified a pervasive use of overly aggressive care in patients who are potentially harmed by these interventions. Again, the large-scale examination enabled by big data is a key aspect of the impact of the studies in increasing awareness of these issues and their policy implications.

Timely: reducing waits & sometimes harmful delays for both those who receive and those who give care

There is compelling evidence that delays between cancer diagnosis and treatment initiation are associated with poor patient outcomes [39–46]. It is therefore important to identify if cancer patients are receiving timely treatment, and big data resources are ideal for these studies.

Table 1. Studies utilizing big data to assess healthcare quality.

Study (year)	Design	Patients (n)	Cancer type	Outcome measured	Ref.
Safe					
Keating <i>et al.</i> (2010)	Veterans Healthcare Administration	37,443	Prostate	Diabetes and cardiovascular disease with androgen deprivation therapy	[3]
Darby <i>et al.</i> (2005)	SEER	308,861	Breast	Mortality from heart disease and lung cancer following radiotherapy	[8]
Pinder <i>et al.</i> (2007)	SEER-Medicare	43,338	Breast	Congestive heart failure after chemotherapy	[9]
Hershman <i>et al.</i> (2008)	SEER-Medicare	9438	Non-Hodgkin lymphoma	Cardiac adverse effects after chemotherapy	[10]
Sheets <i>et al.</i> (2012)	SEER-Medicare	12,976	Prostate	Adverse effects of different types of radiotherapy	[11]
Effective					
Gray <i>et al.</i> (2013)	National Cancer Data Base	28,691	Bladder	Treatment guideline compliance	[12]
Chen <i>et al.</i> (2014)	SEER-Medicare	29,001	Prostate	Treatment guideline compliance	[13]
Chagpar <i>et al.</i> (2012)	National Cancer Data Base	173,243	Colon	Treatment guideline compliance	[14]
Van de Water <i>et al.</i> (2012)	Netherlands Cancer Registry	31,520	Breast	Treatment guideline compliance	[15]
Bristow <i>et al.</i> (2013)	California Cancer Registry	13,321	Ovarian	Treatment guideline compliance	[16]
Patient-centered					
Burge <i>et al.</i> (2003)	Nova Scotia Cancer Registry	8702	Multiple	Continuity of care and emergency department usage at end of life	[17]
Sharma <i>et al.</i> (2008)	SEER-Medicare	45,627	Lung	End of life intensive care unit usage	[18]
Royce <i>et al.</i> (2014)	National Health Interview Survey	27,404	Noncancer individuals	Cancer screening in individuals with different life expectancies	[19]
Timely					
Fedewa <i>et al.</i> (2010)	National Cancer Data Base	107,587	Breast	Time from surgery to adjuvant chemotherapy	[20]
Stokes <i>et al.</i> (2013)	SEER-Medicare	23,960	Prostate cancer	Time from diagnosis to start of treatment	[21]
Froud <i>et al.</i> (2000)	Breast Cancer Outcomes Unit Database	1962	Breast	Association between delay in adjuvant radiation and local recurrence	[22]
Hollenbeck <i>et al.</i> (2010)	SEER-Medicare	2084	Bladder	Association between delay in bladder cancer diagnosis and survival	[23]
Hershman <i>et al.</i> (2006)	SEER-Medicare	4382	Colorectal	Association between delay in adjuvant chemotherapy and survival	[24]

NLMs: National Longitudinal Mortality Study; SEER: Surveillance, Epidemiology, and End Result.

Table 1. Studies utilizing big data to assess healthcare quality (cont.).					
Study (year)	Design	Patients (n)	Cancer type	Outcome measured	Ref.
Efficient					
Warren <i>et al.</i> (2002)	SEER-Medicare	28,916	Breast	Treatment cost	[25]
Yabroff <i>et al.</i> (2008)	SEER-Medicare	2,342,558	Multiple	Treatment cost	[26]
Nguyen <i>et al.</i> (2011)	SEER-Medicare	45,636	Prostate	Cost of adopting new technologies in prostate cancer care	[27]
Falchhook <i>et al.</i> (2014)	SEER-Medicare	47,224	Prostate	Cost of unnecessary staging scans in early-stage prostate cancer	[28]
Equitable					
Fedewa <i>et al.</i> (2010)	National Cancer Data Base	312,339	Prostate	Differences in disease severity by race and insurance status	[29]
Fedewa <i>et al.</i> (2012)	National Cancer Data Base	69,739	Cervical	Association between insurance status and stage at diagnosis	[30]
Gross <i>et al.</i> (2008)	SEER-Medicare	143,512	Multiple	Racial differences in treatments received	[31]
Du <i>et al.</i> (2011)	SEER-NLMS	13,234	Multiple	Socioeconomic differences in cancer treatment and survival	[32]
NLMS: National Longitudinal Mortality Study; SEER: Surveillance, Epidemiology, and End Result.					

In a study published by Fedewa *et al.* [20], the delay in receiving adjuvant chemotherapy for breast cancer was assessed using NCDB. Overall, 95.8% of women received adjuvant chemotherapy within 90 days following surgery, which is a commonly recommended time frame. Stokes *et al.* [21] used SEER-Medicare data to examine the time from prostate cancer diagnosis to initiating treatment. Somewhat surprisingly, this study showed that patients with high-risk cancer had a longer time from diagnosis to definitive treatment compared with low-risk cancer.

Studies have also assessed the impact of treatment delays on patient outcomes. For example, a study by Froud *et al.* [22] investigated the effect of time delay between breast-conserving surgery and adjuvant radiation therapy on local recurrence among 1962 women with breast cancer. Patients were identified from the population-based Breast Cancer Outcomes Unit Database which contains demographic, staging, treatment and outcome information for patients in the British Columbia Cancer Agency. This study found no difference in the rate of ipsilateral breast cancer recurrence on univariate or multivariate analysis comparing groups with 0–5, 6–8, 9–12 and ≥ 13 weeks between breast conserving surgery and initiation of radiotherapy.

Hollenbeck *et al.* [23] used SEER-Medicare data to determine the association between delays in bladder cancer diagnosis and mortality. The investigators identified 29,740 patients who had hematuria (the most common presenting symptom for bladder cancer) in the year preceding a bladder cancer diagnosis. These patients were stratified according to the time interval between hematuria and the bladder cancer diagnosis. The study found that patients who experienced a delay of 9 months or more between detection of hematuria and cancer diagnosis had higher mortality compared with patients diagnosed within 3 months. Another study by Hershman *et al.* [24] investigated the association between survival and time delay to adjuvant chemotherapy in colon cancer. This study found that a delay in the initiation of adjuvant chemotherapy by >3 months was associated with higher colon cancer-specific mortality and all-cause mortality.

Advantages of using big data to assess delivery of timely care

Delivery of timely care and potential harms from treatment delay are questions which cannot be answered in randomized clinical trials. Specifically, it would be unethical to randomize patients to timely versus delayed cancer treatment. Big data resources which contain large numbers of patients are ideal for assessing the timeliness of care on a population-level and potential impact of treatment delays on patient outcomes.

Efficient: avoiding waste, including waste of equipment, supplies, ideas & energy

The IOM estimates that US\$765 billion of the health-care budget – roughly 30 cents of every medical dollar – is wasted each year in the USA [47]. Big data can be used to assess cancer-specific medical costs in an attempt to understand the economic impact associated with cancer care. One such study by Warren *et al.* [25] accomplished this by using SEER-Medicare data to assess the cost of treatment for 28,916 elderly women with early-stage breast cancer in a fee-for-service setting. Although the initial cost of treatment was higher for women receiving breast-conserving surgery plus radiotherapy (BCT) versus modified radical mastectomy (MRM), there was no significant difference in long-term total care costs (US\$77,106 for BCT vs US\$75,897 for MRM). Another study by Yabroff *et al.* [26] also used SEER-Medicare data, and analyzed medical care costs of 718,907 cancer patients matched to 1,623,651 control subjects without cancer. This study found that the cost of cancer care varies widely by cancer type, with 5-year costs ranging from less than US\$20,000 for breast cancer to over US\$40,000 for brain, esophageal and ovarian cancers.

One way to assess potential waste in healthcare spending is to examine the adoption of newer, more expensive tests and treatments – specifically those that have not been proven to improve patient outcomes. Many studies have examined cancer care across the USA and consistently reported that newer and more costly technologies and treatments are often rapidly adopted for use before research has demonstrated their effectiveness [11,48–49]. In a study by Nguyen *et al.* [27] using SEER-Medicare data, the cost associated with adoption of newer technologies in prostate cancer led to an increase in Medicare spending of US\$350 million annually. Waste in healthcare can also be assessed by calculating the costs associated with unnecessary tests. Falchok *et al.* [28] examined this question, assessing the use of unnecessary computed tomography (CT), MRI and bone scans during staging work-up in patients with low- and intermediate-risk prostate cancers. For these patients with early prostate cancer diagnoses, staging scans are not recommended by published practice guidelines [50]. This study found that unnecessary staging scans in these patients cost Medicare approximately US\$11.3 million each year.

Advantages of using big data to assess the efficiency of treatment

Large claims-based data sources like SEER-Medicare allow researchers to directly assess health services provided and the costs associated with these services. Potential medical waste can be evaluated by comparing clinical practice with published guidelines.

Equitable: providing care that does not vary in quality because of personal characteristics such as gender, ethnicity, geographic location & socioeconomic status

Access to quality healthcare has a significant impact on cancer outcomes. However, many cancer patients face barriers to receiving the care they need. Big data analyses have been central to demonstrating worrisome health disparities in oncology.

Several studies have demonstrated that certain patient groups are more likely to be diagnosed with more advanced cancers. A study by Fedewa *et al.*, for example, used the National Cancer Data Base to compare the severity of prostate cancer at diagnosis by race/ethnicity and insurance status in 312,339 men [29]. Uninsured and Medicaid patients were more likely to have more aggressive cancers at diagnosis compared with privately insured patients. Similarly, more aggressive cancers were found in African–American, Hispanic and Asian men compared with Caucasians. Similar authors also used the National Cancer Data Base to examine the association between insurance status with cervical cancer stage [30]. Even after adjusting for race/ethnicity and other sociodemographic and clinical factors, this study found that uninsured, Medicaid and Medicare patients were more likely to be diagnosed with advanced-staged cervical cancers compared with patients with private insurance.

Many studies using big data have also demonstrated disparities in cancer treatment. Gross *et al.* [31] used SEER-Medicare data to show that African–American versus Caucasian patients were less likely to receive aggressive treatment for lung, breast, colon and prostate cancers. Du *et al.* [32] investigated the effects of socioeconomic factors on racial disparities in cancer treatment and survival. This study involved 13,234 cancer patients (including breast, colorectal, prostate, lung, cervical, ovarian, melanoma and bladder cancer) identified from the National Longitudinal Mortality Study-SEER linked database, and found that African–American patients with early-stage nonmetastatic cancer had worse 3-year overall survival compared with Caucasians. However, this difference in survival disappeared after adjusting for socioeconomic factors, including health insurance status, education level and income. This suggests that racial disparities in many common cancers may be partly mediated by socioeconomic factors.

Advantages of using big data to assess delivery of equitable treatment

Big data resources are needed to study the issue of potential health disparities in cancer care and outcomes. Use of population-based data sources like SEER facilitates the generalizability of the results to the overall US population.

Additional strengths & limitations

Randomized controlled trials are considered the ‘gold standard’ for comparing efficacy of cancer treatments. Thus, to assess the ‘effective’ aspect of quality cancer care, there are important strengths of using clinical trials over of big data. The most important issue here is the ability of randomization to minimize confounding in measured and unmeasured factors among the comparison groups. While sophisticated analytic methodologies can be used to minimize confounding with big data analyses, residual confounding can remain which may affect the results from these studies [51,52].

On the other hand, there are also important strengths of using big data for comparative effectiveness. Because randomized trials take significant resources and time to complete, it is impractical to use trials to answer all (or even most) of the clinically relevant comparative effectiveness questions. There are other common situations when randomization is not possible, such as when a new treatment or technology is rapidly diffused into common practice to replace older options – as was the case of robotic surgery in prostate cancer [48]. In these situations, big data likely represent the best available method to compare the effectiveness of different options available to a cancer patient, and can also provide results faster than randomized trials; the latter is important so that patients and physicians can have data to guide their decisions before a technology has fully disseminated. Other important strengths of big data include large available sample sizes of patients, which provide sufficient power in these analyses and also subgroup analyses to determine if treatment benefits apply more to certain patient populations.

Another important consideration when evaluating observational studies including analyses of big data is the external validity. External validity is the extent to which the results of a study can be applied to the general population. The results of a study with high external validity will be able to be generalized beyond the study sample to other people and situations. The common threats to external validity include selecting small, homogenous patient samples limited to single geographic locations or institutions. The effect of these threats may be mitigated by using large registries that include diverse patient groups representative of the overall patient population. Population-based big data sources such as SEER and NHIS have external validity (generalizability) as a particular strength. In contrast, patients who enroll on randomized clinical trials are often highly selected and treatments and other clinical care usually provided in a controlled setting; therefore, results from trials may be more

representative of ideal patients and outcomes but less generalizable to the overall population.

For the other IOM-described aspects of quality of care – safety, patient centeredness, timeliness, efficiency and equitability – big data resources represent the best way for assessment. However, a significant limitation of existing big data resources is the lack of detailed clinical information. For example, cancer recurrence data are not available in SEER or NCDB, and few big data sources have patient-reported quality of life. In addition, currently, big data resources which are available for research are at least 2–3 years behind current practice due to the

time needed for data processing and quality assurance. After factoring in the additional time needed for data analysis and publication, the reported quality assessment results are typically at least 5 years old – and thus no longer reflective of ‘current’ practice or quality gaps.

Furthermore, while big data sources can be used to assess the quality of care, it is often unable to answer ‘why’, and lacks information on patient preferences. Therefore, in order to fully understand results from big data analyses, and then to create potential solutions to improve care, primary data collection and prospective studies are needed.

Executive summary

- The Institute of Medicine (IOM) established six aims for high-quality healthcare: safe, effective, patient-centered, timely, efficient and equitable.
- Big data in oncology provide powerful tools for assessing each of these aims.

Safe

- IOM definition: avoiding injuries to patients from the care that is intended to help them.
- Uses and advantages of big data: capturing data from large populations can reveal new findings regarding safety and potential side effects of cancer treatments that might have been harder to detect in smaller studies.

Effective

- IOM definition: providing services based on scientific knowledge to all who could benefit and refraining from providing services to those not likely to benefit.
- Uses and advantages of big data: observing large-scale practice patterns and whether care provided is concordant with published standards can help inform health policy decisions.

Patient-centered

- IOM definition: providing care that is respectful of and responsive to individual patient preferences, needs and values and ensuring that patient values guide all clinical decisions.
- Uses and advantages of big data: the use of large datasets helps detect healthcare practices which may not be patient-centered.

Timely

- IOM definition: reducing waits and sometimes harmful delays for both those who receive and those who give care.
- Uses and advantages of big data: big data can be used to assess whether timely care is delivered and potential harms of care delays.

Efficient

- IOM definition: avoiding waste, including waste of equipment, supplies, ideas and energy.
- Uses and advantages of big data: claims-based data can be used to assess the costs of oncologic care and potential wastes in healthcare spending due to unnecessary tests and treatments.

Equitable

- IOM definition: providing care that does not vary in quality because of personal characteristics such as gender, ethnicity, geographic location and socioeconomic status.
- Uses and advantages of big data: big data resources have played a central role in helping identify disparities in cancer care and outcomes and population-based data resources allow findings to be generalizable to overall population.

Strengths & limitations

- Limitations of current big data resources include lack of certain detailed data elements such as cancer recurrence and quality of life, lack of timeliness in data access and analysis and potential confounding.

Future perspective

- Current status: big data resources are powerful tools to retrospectively examine whether high-quality oncologic care was delivered.
- Future vision: real-time availability of big data for quality assessment and to guide clinical decision-making to facilitate delivery of high-quality care.
- For this vision to become a reality, stakeholders including professional organizations, electronic health record vendors and funding agencies will need to work together.

Conclusion & future perspective

Current big data resources have already been shown to be powerful tools to assess the six dimensions of quality of care as defined by the IOM. However, as mentioned above, there are still significant limitations and challenges to the big data resources available today.

We envision a future where big data will be used to evaluate quality of care in a more timely way, and also to guide clinical decision-making in real-time based on individual patient characteristics and preferences [53]. The continued development and adoption of electronic health record (EHR) systems can help enable this vision to come true. We envision a future where large groups of patient data can be pooled – including detailed information about patient diagnoses and outcomes – from across institutions, so that each patient and their physician can find ‘patients like me’ to help with real-time clinical decision-making. The use of big data to support clinical care can occur at the time of initial treatment decision-making, during treatment for side effect management as well as during follow-up to predict long-term outcomes.

Going from where we are today to making this future vision a reality will not be easy. It will require an ability to combine vast amounts of data across different EHR systems. Currently, data captured in EHRs are incomplete and often not in discrete data elements; the latter important for data analysis. To allow pooling EHR data across a large number of institutions, a set of agreed upon discrete data elements would need to be built into the EHR systems. Furthermore, combining data from various institutions is often hindered by technical and legal hurdles.

References

Papers of special note have been highlighted as:

• of interest; •• of considerable interest

- Chassin MR, Galvin RW. The urgent need to improve health care quality: Institute of Medicine National Roundtable on Health Care Quality. *JAMA* 280(11), 1000–1005 (1998).
- Describes the need for quality improvement in the US healthcare system.
- Institute of Medicine Committee on Quality of Health Care in America. *Crossing the Quality Chasm: A New Health System for the 21st Century*. National Academies Press, Washington, DC, USA (2001).
- Outlines a detailed plan for improving the US healthcare system.
- Keating NL, O'Malley AJ, Freedland SJ, Smith MR. Diabetes and cardiovascular disease during androgen deprivation therapy: observational study of veterans with prostate cancer. *J. Natl Cancer Inst.* 102(1), 39–46 (2010).
- Saigal CS, Gore JL, Krupski TL, Hanley J, Schonlau M, Litwin MS. Androgen deprivation therapy increases cardiovascular morbidity in men with prostate cancer. *Cancer* 110(7), 1493–1500 (2007).
- 5 Tsai HK, D'Amico AV, Sadetsky N, Chen M-H, Carroll PR. Androgen deprivation therapy for localized prostate cancer and the risk of cardiovascular mortality. *J. Natl Cancer Inst.* 99(20), 1516–1524 (2007).
- 6 Nguyen PL, Je Y, Schutz FA *et al.* Association of androgen deprivation therapy with cardiovascular death in patients with prostate cancer: a meta-analysis of randomized trials. *JAMA* 306(21), 2359–2366 (2011).
- 7 Howlader N, Noone A, Krapcho M *et al.* SEER Cancer Statistics Review, 1975–2012, National Cancer Institute, Bethesda, MD, USA (2015). http://seer.cancer.gov/csr/1975_2012/
- 8 Darby SC, McGale P, Taylor CW, Peto R. Long-term mortality from heart disease and lung cancer after radiotherapy for early breast cancer: prospective cohort study of about 300 000 women in US SEER cancer registries. *Lancet Oncol.* 6(8), 557–565 (2005).
- 9 Pinder MC, Duan Z, Goodwin JS, Hortobagyi GN, Giordano SH. Congestive heart failure in older women

The recent Big Data Workshop sponsored by the American Society for Radiation Oncology, American Association of Physicists in Medicine and the NCI specifically discussed this vision, the challenges and potential solutions. For this vision to become a reality, stakeholders including the professional societies, EHR vendors and funding agencies including NCI need to work together. For example, oncologists through professional organizations can contribute by defining core data elements necessary to collect for each patient, which may differ for each type of cancer. Proprietary EHRs can contribute by allowing data to be exportable into a universal format that can be used by all EHR systems, so that the data can be pooled across a large number of institutions, stored in a central repository and made available for clinical use and research. These efforts will require funding to accomplish. Despite these challenges, moving from the current status of using big data to retrospectively assess care quality to a future where big data can be used in real time to help patients receive high-quality care is an important and logical step. This vision is achievable but will require many people throughout oncology to work together.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

- treated with adjuvant anthracycline chemotherapy for breast cancer. *J. Clin. Oncol.* 25(25), 3808–3815 (2007).
- 10 Hershman DL, McBride RB, Eisenberger A, Tsai WY, Grann VR, Jacobson JS. Doxorubicin, cardiac risk factors, and cardiac toxicity in elderly patients with diffuse B-cell non-Hodgkin's lymphoma. *J. Clin. Oncol.* 26(19), 3159–3165 (2008).
 - 11 Sheets NC, Goldin GH, Meyer A-M *et al.* Intensity-modulated radiation therapy, proton therapy, or conformal radiation therapy and morbidity and disease control in localized prostate cancer. *JAMA* 307(15), 1611–1620 (2012).
 - 12 Gray PJ, Fedewa SA, Shipley WU *et al.* Use of potentially curative therapies for muscle-invasive bladder cancer in the United States: results from the National Cancer Data Base. *Eur. Urol.* 63(5), 823–829 (2013).
 - 13 Chen RC, Carpenter WR, Hendrix LH *et al.* Receipt of guideline-concordant treatment in elderly prostate cancer patients. *Int. J. Radiat. Oncol. Biol. Phys.* 88(2), 332–338 (2014).
 - 14 Chagpar R, Xing Y, Chiang Y-J *et al.* Adherence to stage-specific treatment guidelines for patients with colon cancer. *J. Clin. Oncol.* 30(9), 972–979 (2012).
 - 15 Van De Water W, Bastiaannet E, Dekkers O *et al.* Adherence to treatment guidelines and survival in patients with early-stage breast cancer by age at diagnosis. *Br. J. Surg.* 99(6), 813–820 (2012).
 - 16 Bristow RE, Chang J, Ziogas A, Anton-Culver H. Adherence to treatment guidelines for ovarian cancer as a measure of quality care. *Obstet. Gynecol.* 121(6), 1226–1234 (2013).
 - 17 Burge F, Lawson B, Johnston G. Family physician continuity of care and emergency department use in end-of-life cancer care. *Med. Care* 41(8), 992–1001 (2003).
 - 18 Sharma G, Freeman J, Zhang D, Goodwin JS. Trends in end-of-life ICU use among older adults with advanced lung cancer. *Chest* 133(1), 72–78 (2008).
 - 19 Royce TJ, Hendrix LH, Stokes WA, Allen IM, Chen RC. Cancer screening rates in individuals with different life expectancies. *JAMA Intern. Med.* 174(10), 1558–1565 (2014).
 - 20 Fedewa SA, Ward EM, Stewart AK, Edge SB. Delays in adjuvant chemotherapy treatment among patients with breast cancer are more likely in African American and Hispanic populations: a national cohort study 2004–2006. *J. Clin. Oncol.* 28(27), 4135–4141 (2010).
 - 21 Stokes WA, Hendrix LH, Royce TJ *et al.* Racial differences in time from prostate cancer diagnosis to treatment initiation. *Cancer* 119(13), 2486–2493 (2013).
 - 22 Froud PJ, Mates D, Jackson JS *et al.* Effect of time interval between breast-conserving surgery and radiation therapy on ipsilateral breast recurrence. *Int. J. Radiat. Oncol. Biol. Phys.* 46(2), 363–372 (2000).
 - 23 Hollenbeck BK, Dunn RL, Ye Z *et al.* Delays in diagnosis and bladder cancer mortality. *Cancer* 116(22), 5235–5242 (2010).
 - 24 Hershman D, Hall MJ, Wang X *et al.* Timing of adjuvant chemotherapy initiation after surgery for stage III colon cancer. *Cancer* 107(11), 2581–2588 (2006).
 - 25 Warren JL, Brown ML, Fay MP, Schussler N, Potosky AL, Riley GF. Costs of treatment for elderly women with early-stage breast cancer in fee-for-service settings. *J. Clin. Oncol.* 20(1), 307–316 (2002).
 - 26 Yabroff KR, Lamont EB, Mariotto A *et al.* Cost of care for elderly cancer patients in the United States. *J. Natl Cancer Inst.* 100(9), 630–641 (2008).
 - 27 Nguyen PL, Gu X, Lipsitz SR *et al.* Cost implications of the rapid adoption of newer technologies for treating prostate cancer. *J. Clin. Oncol.* 29(12), 1517–1524 (2011).
 - 28 Falchook AD, Salloum RG, Hendrix LH, Chen RC. Use of bone scan during initial prostate cancer workup, downstream procedures, and associated Medicare costs. *Int. J. Radiat. Oncol. Biol. Phys.* 89(2), 243–248 (2014).
 - 29 Fedewa SA, Etzioni R, Flanders WD, Jemal A, Ward EM. Association of insurance and race/ethnicity with disease severity among men diagnosed with prostate cancer, National Cancer Database 2004–2006. *Cancer Epidemiol. Biomarkers Prev.* 19(10), 2437–2444 (2010).
 - 30 Fedewa SA, Cokkinides V, Virgo KS, Bandi P, Saslow D, Ward EM. Association of insurance status and age with cervical cancer stage at diagnosis: National Cancer Database, 2000–2007. *Am. J. Public Health* 102(9), 1782–1790 (2012).
 - 31 Gross CP, Smith BD, Wolf E, Andersen M. Racial disparities in cancer therapy. *Cancer* 112(4), 900–908 (2008).
 - 32 Du XL, Lin CC, Johnson NJ, Altekruse S. Effects of individual-level socioeconomic factors on racial disparities in cancer treatment and survival. *Cancer* 117(14), 3242–3251 (2011).
 - 33 National Cancer Policy Board and the Institute of Medicine. *Ensuring Quality Cancer Care*. National Academies Press, Washington, DC, USA (1999).
 - Describes the discrepancy between the ideal cancer care system and that which is actually observed.
 - 34 Gandaglia G, Sammon JD, Chang SL *et al.* Comparative effectiveness of robot-assisted and open radical prostatectomy in the postdissemination era. *J. Clin. Oncol.* 32(14), 1419–1426 (2014).
 - 35 Brazil K, Howell D, Bedard M, Krueger P, Heidebrecht C. Preferences for place of care and place of death among informal caregivers of the terminally ill. *Palliat. Med.* 19(6), 492–499 (2005).
 - 36 Beccaro M, Costantini M, Rossi PG, Miccinesi G, Grimaldi M, Bruzzi P. Actual and preferred place of death of cancer patients. Results from the Italian survey of the dying of cancer (ISDOC). *J. Epidemiol. Community Health* 60(5), 412–416 (2006).
 - 37 Tang ST. When death is imminent: where terminally ill patients with cancer prefer to die and why. *Cancer Nurs.* 26(3), 245–251 (2003).
 - 38 Published Recommendations. U.S. Preventive Services Task Force. www.uspreventiveservicestaskforce.org
 - 39 Gorin SS, Heck JE, Cheng B, Smith SJ. Delays in breast cancer diagnosis and treatment by racial/ethnic group. *Arch. Intern. Med.* 166(20), 2244–2252 (2006).

- 40 Smith EC, Ziogas A, Anton-Culver H. Delay in surgical treatment and survival after breast cancer diagnosis in young women by race/ethnicity. *JAMA Surg.* 148(6), 516–523 (2013).
- 41 McLaughlin JM, Anderson RT, Ferketich AK, Seiber EE, Balkrishnan R, Paskett ED. Effect on survival of longer intervals between confirmed diagnosis and treatment initiation among low-income women with breast cancer. *J. Clin. Oncol.* 30(36), 4493–4500 (2012).
- 42 Jung SY, Sereika SM, Linkov F, Brufsky A, Weissfeld JL, Rosenzweig M. The effect of delays in treatment for breast cancer metastasis on survival. *Breast Cancer Res. Treat.* 130(3), 953–964 (2011).
- 43 Murphy C, Galloway T, Handorf E *et al.* Survival impact of increasing time to treatment initiation for head and neck cancer in the United States. *Int. J. Radiat. Oncol. Biol. Phys.* 93(3), S129–S130 (2015).
- 44 Gomez DR, Liao K-P, Swisher SG *et al.* Time to treatment as a quality metric in lung cancer: staging studies, time to treatment, and patient survival. *Radiother. Oncol.* 115(2), 257–263 (2015).
- 45 Nguyen PL, Whittington R, Koo S *et al.* The impact of a delay in initiating radiation therapy on prostate-specific antigen outcome for patients with clinically localized prostate carcinoma. *Cancer* 103(10), 2053–2059 (2005).
- 46 Freedland SJ, Kane CJ, Amling CL *et al.* Delay of radical prostatectomy and risk of biochemical progression in men with low risk prostate cancer. *J. Urol.* 175(4), 1298–1303 (2006).
- 47 McGinnis JM, Goolsby WA, Olsen L. *Leadership Commitments to Improve Value in Health Care: Finding Common Ground: Workshop Summary.* National Academies Press, Washington, DC, USA (2009).
- 48 Hu JC, Gu X, Lipsitz SR *et al.* Comparative effectiveness of minimally invasive vs open radical prostatectomy. *JAMA* 302(14), 1557–1564 (2009).
- 49 Goozner M. The proton beam debate: are facilities outstripping the evidence? *J. Natl Cancer Inst.* 102(7), 450–453 (2010).
- 50 National Comprehensive Cancer Network Guidelines. National comprehensive cancer network. www.nccn.org/professionals/physician_gls/pdf/prostate.pdf
- 51 Hamstra DA, Chen RC. RE: Mortality after radical prostatectomy or external beam radiotherapy for localized prostate cancer. *J. Natl Cancer Inst.* 106(3), djt463 (2014).
- 52 Chen RC, Zietman AL. Prostatectomy vs radiotherapy: a study for cautious interpretation. www.ascp.org/issues/april-15-2014
- 53 Chen RC, Gabriel PE, Kavanagh BD, McNutt TR. How will big data impact clinical decision making and precision medicine in radiation therapy? *Int. J. Radiat. Oncol. Biol. Phys.* doi:10.1016/j.ijrobp.2015.10.052 (2015) (Epub ahead of print).