Review Article

Measuring improvement in populations: implementing and evaluating successful change in lung cancer care

Xinhua Yu[1], Lisa M. Klesges[1], Mathew P. Smeltzer[1], Raymond U. Osarogiagbon[2]

[1]Division of Epidemiology, Biostatistics, and Environmental Health, School of Public Health, University of Memphis, Memphis, Tennessee, USA; [2]Multidisciplinary Thoracic Oncology Program, Baptist Cancer Center, Memphis, Tennessee, USA

Contributions: (I) Conception and design: X Yu, RU Osarogiagbon; (II) Administrative support: X Yu; (III) Provision of study materials: RU Osarogiagbon; (IV) Collection and assembly of data: RU Osarogiagbon; (V) Data analysis and interpretation: X Yu, RU Osarogiagbon; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Xinhua Yu, MD, PhD. Division of Epidemiology, Biostatistics, and Environmental Health, School of Public Health, University of Memphis, Memphis 38152, Tennessee, USA. Email: xyu2@memphis.edu.

Abstract: Improving quality of care in lung cancer, the leading cause of cancer death worldwide and in the United States, is a major public health challenge. Such improvement requires accurate and meaningful measurement of quality of care. Preliminary indicators have been derived from clinical practice guidelines and expert opinions, but there are few standard sets of quality of care measures for lung cancer in the United States or elsewhere. Research to develop validated evidence-based quality of care measures is critical in promoting population improvement initiatives in lung cancer. Furthermore, novel research designs beyond the traditional randomized controlled trials (RCTs) are needed for wide-scale applications of quality improvement and should extend into alternative designs such as quasi-experimental designs, rigorous observational studies, population modeling, and other pragmatic study designs. We discuss several study design options to aid the development of practical, actionable, and measurable quality standards for lung cancer care. We also provide examples of ongoing pragmatic studies for the dissemination and implementation of lung cancer quality improvement interventions in community settings.

Keywords: Quality improvement; dissemination and implementation; health services research; evidence based practice

Submitted Jul 14, 2015. Accepted for publication Jul 15, 2015.
doi: 10.3978/j.issn.2218-6751.2015.07.13
View this article at: http://dx.doi.org/10.3978/j.issn.2218-6751.2015.07.13

Introduction

Quality of health care has become an increasingly important topic since the release of Institute of Medicine (IOM) report "Crossing the Quality Chasm" in 2001 (1). However, defining the quality of care is not straightforward. The IOM defines quality as "*the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge*", and stipulates six domains of quality of care (*Table 1*): safety, effectiveness, patient-centeredness, timeliness, efficiency, and equity. Health care providers, researchers and policy-makers have devoted significant amount of efforts to measure and improve quality of care according to these six domains, though not all domains have received equal attention. Recent efforts by government agencies such as the Center for Medicare and Medicaid Services in the US to link payment to quality or 'value of care' have stirred a new wave of interest in quality improvement within clinical and community settings.

Efforts to improve quality of cancer care generally lag behind those of other diseases. A recent IOM report "Delivering High-Quality Cancer Care" pointed out several hurdles in improving the quality of cancer care (2). Measuring quality of cancer care is difficult due to complexities in patients' clinical characteristics, diagnostic

Table 1 Domains of quality of care

Domains	Recommendations	Comments
Safety	Avoiding injuries to patients from the care that is intended to help them	Safety spans from preventing errors in care, to health care environment, and to treating complication
Effectiveness	Providing services based on scientific knowledge to all who could benefit, and refraining from providing services to those not likely to benefit	Evidence-based practice should be integrated with best research evidence and patient values. It should avoid underuse, overuse, and misuse of certain services. Misuse of services often results in errors in care
Patient-centeredness	Providing care that is respectful of and responsive to individual patient preferences, needs, and values and ensuring that patient values guide all clinical decisions	Patient-centered care includes dimensions such as: respect for patient's values, preferences, and expressed need; coordination and integration of care; information, communication, and education; physical comfort; emotional support; and involvement of family and friends
Timeliness	Reducing waits and sometimes harmful delays for both those who receive and those who give care	Timeliness is related to the attention to process flow and respect for patients. Delays and barriers during the process could be due to health care systems, providers, and patients
Efficiency	Avoiding waste, in particular waste of equipment, supplies, ideas, and energy	Defining wastefulness is not straightforward. Cost-effectiveness analysis could be used to examine whether certain procedures or techniques are cost effective
Equity	Providing care that does not vary in quality because of personal characteristics such as gender, ethnicity, geographic location, and socioeconomic status	Equity should be at both the population and individual levels, that is, reducing health disparities among population subgroups, and providing care to all individuals based on their needs

Adapted from "Crossing the Quality Chasm" IOM report, 2001 (1).

and staging procedures, treatment options, and follow up care. Variations among practitioners and clinics create additional complexity. There is also the problem of nihilism in cancer care, especially lung cancer care, characterized by generally low expectations for positive outcomes. The association between age and cancer, and the accumulation of age- and tobacco-related comorbidities further complicate care delivery, quality measurement, and reinforces nihilism. Rapid innovations in technology and treatments add another layer of complexity. In addition, fragmented health care systems and lack of coordination among key specialists create extra barriers to quality measurement and improvement. Furthermore, the traditional emphasis on physicians' assessment of patients' clinical status and a focus on survival, and less on patients' psychological well-being and preferences, prevent measuring the full spectrum of quality of care.

According to Donabedian's framework, measuring quality of care consists of three key aspects: structure, process, and outcomes (2-5). Structure measures focus on the infrastructure of health care systems, physician/staffing characteristics, and volumes of care delivery. Process measures evaluate how care is delivered. Examples include utilization rates, such as use of screening tests, non-invasive and invasive staging tests, and receipt of chemotherapy among eligible patients. Because many of the process indicators are based on clinical guidelines, they are objective, comparable across institutions, and easier to interpret by the public. Process measures are thus the most commonly used quality of care measures, as illustrated in the American College of Surgeons Commission on Cancer (CoC) quality surveillance measures. However, process measures are not necessarily directly associated with patient outcomes such as survival. Outcome measures include both clinical outcomes and patient reported outcomes. Examples include survival statistics, complication rates, measures of personal health and functional status, quality of life, symptom burden, and psychological well-being. In comparing these patient outcomes, risk adjustment is needed to take into account variability in patient characteristics, disease severity, and comorbidity (6). On the other hand, patient reported outcomes are often not well

defined and can be difficult to compare across institutions and populations.

Quality measures depend on the perspective of stakeholders such as patients, their caregivers, clinicians, health system administrators, third party payers, large employers (in the US system), and health policymakers. For example, cost measures assessing the resources used in health care will have very different results depending on whether the perspective adopted is that of patients, caregivers, third party payers, or society at large. Efficiency measures may assess the time, effort, or cost to produce a specific output. They may include time from diagnosis to treatment, the relative number of steps required and the relative cost (in dollar terms or patient discomfort) of the cumulative steps involved. Patients' perspective, such as satisfaction with the care provided may differ from care-providers' perspectives (for example, see Kedia *et al.* in this special issue). More research is needed to develop cancer-specific quality measures, especially in the evolving environment of cancer care delivery. For a more detailed discussion of the pros and cons of the various methods of quality improvement, please refer to the paper by Farjah and Detterbeck in this special issue.

Developing validated lung cancer quality of care measures

Although incidence and mortality rates have been declining in the US since 2000, lung cancer remains the top cause of cancer death (7). This burden is even greater in the rest of the world (see the paper by Jemal *et al.* in this special issue). Compared to female breast and colorectal cancer which have some validated quality measures developed by National Comprehensive Cancer Network (NCCN)/ American Society of Clinical Oncology (ASCO) (8), relatively little has been done in measuring and improving the quality of lung cancer care. There is no standard set of quality measures accepted by a majority of stakeholders involved in lung cancer care. Many process and outcome measures focus on patients with early stage lung cancer who have potentially curable disease. Some effort has been devoted to evaluating access to, and outcomes of, surgical resection, which has often shown large variations within and across populations (9,10).

The ideal quality of care measure should be practical, measurable, and actionable (11,12). It should be strongly correlated to patient outcomes, relevant to majority of patients, meaningful across diverse practice settings, and independent of patient characteristics to allow comparisons across heterogeneous populations (13). To develop quality of care measures, health services researchers and organizations often aggregate opinions from expert panels or through the Delphi method, and test them in actual practice to examine whether they are meaningful and responsive (13-17). Examples of recently developed quality of care measures for lung cancer are shown in *Table 2*.

The most commonly used measures are process-based, for example, utilization rates such as CT and PET-CT scans, invasive staging tests such as endobronchial ultrasound (EBUS) and mediastinoscopy, and surgical resection rates (19). Some process measures are more directly linked to patient outcomes than others. For example, mediastinal lymph node examination is associated with better survival in patients who undergo resection, but the optimal extent of mediastinal lymph node examination remains open to debate (20-22).

Growing attention has been devoted to the timeliness of care (23). There is widely acknowledged significant delay from diagnosis to the receipt of treatment in lung cancer (24,25). In one community-based healthcare system, the median duration from abnormal imaging to surgery was 84 days, with interquartile range from 43 to 189 days (26). The two periods of greatest delay were from abnormal imaging to attempt of a diagnostic biopsy, and from the final staging test to surgery (Jinshan Li, in preparation). Similarly, a study based on Medicare claims showed that more than 35 days of delay from diagnosis to treatment initiation was associated with worse survival (27). However, other reports indicate that shorter duration between diagnosis and treatment was related to worse outcomes (28-30), but this is likely due to confounding by indication. Patients with biologically more aggressive disease and a greater symptom burden (and therefore poorer prognosis) are more likely to receive treatment quickly. Overall, the evaluation of the survival implications of delayed care is difficult because it mandates accurate adjustment for key clinical variables that are often unavailable in retrospective studies.

Although 'appropriateness of care' is not one of the six IOM qualities of care domains, it is subsumed in the domains of effectiveness, patient-centeredness, and safety. Receiving appropriate diagnostic and staging tests, and stage-appropriate treatment is important in lung cancer care (31). For example, surgical resection is the most important curative treatment modality in non-small cell lung cancer (NSCLC), but only among physiologically fit patients

Table 2 Examples of quality of care measures for non-small cell lung cancer (NSCLC)

Thoracic Oncology Network (13)
Percent of nonsurgical biopsies in patients with clinical stage IV non-squamous lung cancer who obtained an adequate amount of tissue for molecular testing
Percentage of patients with clinical stage III or IV lung cancer, or neurological symptoms, who have had brain imaging performed within 3 months of the initiation of treatment
Percentage of patients with evidence of 1-3 distant metastases who have had an attempt at biopsy confirmation of a site of metastasis, or documentation of a reason that this was not possible or necessary
Percentage of patients with clinical stage IB or higher, but no evidence of metastatic disease, who have had a mediastinal lymph node sampling procedure performed prior to the initiation of curative-intent therapy
Percentage of patients with lung cancer who have an American Joint Committee on Cancer (AJCC) seventh edition clinical lung cancer stage documented prior to curative-intent therapy
Percentage of active smokers with lung cancer who have had smoking cessation counseling documented
Percentage of patients with lung cancer in whom a performance status measure is documented in the pretreatment phase
Quality indicators used in a Florida study (15)
Preoperative PET scan performed
Surgical resection for stage I, II NSCLC performed
Cisplatin-based adjuvant chemotherapy used for patients with stage II or III NSCLC who had surgical resection
Postoperative radiation not used for patients with stage I, II who had a complete surgical resection
Brain staging in chemoradiation patients performed
Concurrent chemoradiation used for unresected stage III
Standard chemotherapy used for locally advanced stage
Performance status assessment performed in advanced stages
Commission on Cancer (CoC) Quality of Care Measure (18)
At least 10 regional lymph nodes are removed and pathologically examined for AJCC stage IA, IB, IIA, and IIB resected NSCLC
Systemic chemotherapy is administered within 4 months to day preoperatively or day of surgery to 6 months postoperatively, or it is considered for surgically resected cases with pathologic, lymph node-positive (pN1) and (pN2) NSCLC.
Surgery is not the first course of treatment for cN2, M0 lung cases
National Comprehensive Cancer Network (NCCN) guideline (17)
Negative margin for surgical resection
Lobectomy or higher for anatomic extent of resection
Hilar node sampling
Minimum of 3 mediastinal nodal stations sampled
Invasive mediastinal staging before surgical resection for patients with stage I or II cancer
Time from diagnosis to first treatment
Multidisciplinary evaluation involving thoracic surgeons, medical oncologists, radiation oncologists, and pulmonologists
Additional quality of care measures to consider
Resections performed by thoracic surgeons
PET-CT scan before invasive staging test (e.g., mediastinoscopy) (14)
30-day mortality after surgery
Patient reported outcomes such as satisfaction of care, symptoms, pain, psychosocial well-beings
Radiation dose limits to normal tissue for radiation oncology (National Quality Forum)
Pain intensity quantified and plan of care for pain available (National Quality Forum)

with early stage disease. Measuring the appropriateness of surgical resection is significantly impaired by variable thoroughness of clinical staging, evaluation of physiologic function, and pathologic staging.

Outcome measures such as survival and complication rates are critical in lung cancer as well. It seems logical to assume that patients who receive the combination of timely and accurate diagnosis, thorough staging, and appropriate treatment, will have better survival than those whose care is delayed, who are poorly staged, and/or inappropriately treated. However, patient factors such as socioeconomic characteristics, comorbidities, tumor characteristics and treatment regimens need to be considered when comparing survival across populations (32). Finally, certain well-established general quality measures such as surgical complication rates, interval post-operative mortality (30-, 60-, 90-day) or readmission rates are reported by institutions to national organizations and can be used to compare the overall quality of care across institutions (33-35).

Lung cancer quality improvement initiatives

Measurement provides the foundation for quality improvement, which must be continuous and systematic to identify process and outcome variables linked with greater effectiveness. Quality improvement requires strategic planning and implementation of actionable, evidence-based interventions with measurable outcomes. Most quality improvement initiatives target health care system structure or processes of care (36). Key stakeholders, such as administrators, physicians, nurses, ancillary support staff, patients and their caregivers must be identified and engaged in the process from planning to executing to ensure that all meaningful and relevant features are considered. Key components of successful quality improvement include characteristics of innovation (simplicity, practicality, degree of disruption of existing processes, ease of adoption, ready evidence of meaningful outcomes improvement), cultural environment, level of internal and external support.

Measures to quantify the efficacy and effectiveness of quality improvement interventions should also be clinically meaningful, sensitive to change, closely related to the quality of care, and preferably patient-centered, such as measures of patient satisfaction and survival. As summarized in recent reviews, quality improvement research should be relevant, generalizable, and impactful to clinical practice and community (37,38). Rigorous study design should be employed to test the effectiveness of interventions in diverse settings where care is usually provided, including both large academic health centers and community practice settings, in order to enable evaluation of heterogeneity in effectiveness (39).

Dissemination and implementation framework for quality improvement

As measures and effective processes of care are identified, performance can be improved with successful and timely dissemination of quality improvement innovation into real-world care environments. To enhance adoption of these new approaches, a purposeful understanding of conducting dissemination and implementation research studies can improve the pace of change (40,41). Such 'effectiveness' studies should be guided by a dissemination and implementation framework. Frameworks provide a solid scientific basis for planning study designs that are pragmatic to current practice conditions, include relevant outcomes measurement, and allow for population comparisons and estimated impact for local adoption (42).

Various frameworks have been used to help design care delivery effectiveness studies to improve adoption into practice and policy (43), and examples are readily available (44). A commonality of most of these frameworks is the aim to enhance external validity so that relevance to local conditions can be evaluated. As examples, we highlight two that have been applied to work in lung cancer.

The Pragmatic-Explanatory Continuum Indicator Summary (PRECIS) framework (45) includes ten domains that span from participant eligibility criteria, experimental intervention flexibility, to follow up intensity, primary trial outcomes, and participant compliance with prescribed intervention. Each domain has multiple items for evaluating whether and how the trial has considered certain issues. It is helpful during the study planning and design stage.

The reach, effectiveness, adoption, implementation, and maintenance (RE-AIM) framework covers the full spectrum of the dissemination and implementation process, and provides clear and measurable outcomes to help study design (46-48). The 'reach' domain measures the participation rates of individuals in the target population. It estimates the proportion of eligible individuals who actually participate in the study and their characteristics in comparison to the eligible clinical or community population. The 'effectiveness' domain measures the good and bad outcomes of the intervention. The 'adoption' domain measures the representativeness of the care providers and institutional settings that participate in the

intervention. The 'implementation' domain measures the extent to which an intervention is delivered as intended. It assesses any adaptations of an intervention in various environments. The 'maintenance' domain measures the long term sustainability of the innovation at the individual and institutional levels (47,48).

More recently, a practical, robust implementation and sustainability model (PRISM) has been proposed for integrating research findings into practice (49). This model explicitly incorporates organizational and patient perspectives in implementing interventions, and frames implementation and sustainability together with multiple critical evaluation points based on the RE-AIM framework to ensure quality of implementation.

Finally, since most intervention studies target behavioral change in institutions and/or physicians, it is imperative to consider behavioral change theory in quality improvement study design (50). Psychosocial and behavioral theories such as health belief model, stage of change model, and social learning model are instrumental for designing and implementing interventions across multi-layers of health care systems (51-54).

Study designs for implementing quality of care improvement initiatives

Designing research studies for care delivery improvement based on an implementation framework generally requires different approaches than traditional clinical trials designs. Unlike efficacy studies which are to establish the magnitude of an intervention effect (effect size) in an ideal setting, effectiveness studies need pragmatic designs in which interventions are tested in heterogeneous, 'real-world' environments, including community practices (39). Participating institutions and physicians may adopt the intervention with varying levels of faithfulness, and many factors are out of the investigators' control. The rationale and recommendations for these pragmatic designs have been summarized as the 5 R's by Peek *et al.* (38) and include key elements for consideration including: (I) relevance to stakeholders; (II) rapid and recursive in application; (III) redefining rigor; (IV) reporting on resources; and (V) replicable results. Research studies are needed that address the rapid pace of change in healthcare, have clearer application to local health care and community settings, and adapt to new models of performance improvement (38,39,55,56). Common study designs for implementing care improvement are shown in *Table 3*, together with several advantages and disadvantages.

The cluster randomized controlled trial (RCT) is the most common RCT in quality improvement studies. Many care improvement and effectiveness studies target institutions or providers, not patients, as the intervention unit (36,57). Even when the intervention targets individual patients, it may be logistically preferable to assign intervention at the level of institutions, practice settings, or physicians to avoid the risk of intervention contamination among patients within the same practice setting. In these scenarios, the cluster RCT is a more appropriate design. Cluster RCT has certain challenges. Study outcomes may be based on patient outcomes such as rate of certain procedure use, stage distributions, and survival. The correlation among patients within the cluster must be taken into account during data analysis (58). Random effect models or generalized estimate equations with cluster effects are commonly used to adjust for such correlation within clusters. Sample size and statistical power can be challenging because the effective sample size for comparisons is the number of clusters to which the intervention is delivered, not the number of individual patients. Therefore, the cluster RCT is often underpowered for detecting complex relationships such as interactions between main effects (58).

A common alternative to parallel cluster RCT is stepped wedge design (59-64). This design differs from the typical cluster RCT in that interventions are delivered not in parallel to the control (usual care) but staggered. Unlike the cluster RCT in which interventions differ between groups, the stepped wedge design implements the same intervention in all groups, but at different time waves. The order of receiving intervention for each participating site is randomized, and stratification by important factors is also possible. In this regard, the stepped wedge design is an experimental design as well. This design is preferred when the study involves many heterogeneous settings, and simultaneously implementing interventions in many institutions is logistically or financially infeasible (60). Since all groups will receive the intervention, it is important that the intervention has been proven beneficial (or at least harmless) to all participants, thus it may be unethical to withhold the intervention from some participants. An extension of the stepped wedge design is the 'multiple baseline' study design (65,66), in which multiple measures of outcomes before intervention (baseline) are used to compare the outcome changes between pre- and post-intervention.

Table 3 Common study designs for quality of care improvement

Study design	Descriptions	Advantages	Disadvantages
Randomized controlled trials (RCTs)			
Cluster RCTs	Intervention and control (usual care) are randomly assigned to two groups; randomization unit is group (institutions, practice settings, physicians); intervention and control are typically assigned in parallel	Solid statistical foundation to causally identify intervention effect; randomization controls for known and unknown confounders; appropriate for interventions targeting institutions and/or physicians; reduce the risk of contamination among individuals within the group	Control group may switch over to intervention (contamination); need more efforts and cost in study coordination; sometimes unethical to withhold intervention for participating institutions; analysis needs to take account of correlation within the group; sample size is often inadequate for interactions and complex relationship; heterogeneity of participating institutions or physicians may result in imbalance in important factors
Stepped wedge design	All units will receive intervention, but at different time; the sequence of receiving intervention for each unit is randomized	Logistically easy to sequentially implement intervention; avoid ethical concerns when the intervention has been proven effective in clinical practice; larger number of institutions can be recruited over time	Intervention should be proven effective; study duration will be longer than parallel design; multiple measures of outcomes require more resources and add burden to participants; need more sample size; less efficient than parallel cluster RCT; data analysis is more complex
Quasi-experimental design			
Regression discontinuity design	Comparing the outcomes before and after intervention; intervention assignment based on pre-treatment scores; the slope change of outcome against scores at the cutpoint suggests intervention effect	Small number of groups may be needed	Not all outcomes and intervention can use this design; the slope of change may be due to the change of other factors coinciding with the intervention; without external control, causal inference is less solid
Interrupted time series design	Multiple measures of outcomes before and after intervention; the change of time trend in outcomes at the time of intervention suggests intervention effect	Can use only one group; time trend can be estimated	The study period is longer; overall time trend should be detrended; correlations within multiple measurements of outcome need to be accounted for; less clear in causal inference
Controlled pre-post design	Measure outcomes pre and post intervention; control groups are measured at the same time; a significant change between pre- and post intervention, compared with the outcome difference in the controls during the same period suggests intervention effect	Self-matched; control group tells the time trend during the study period; simpler design	Bias and confounding exist; without randomization, multivariate analysis is needed; causal inference is uncertain
Observational study design (natural experiment)	Observe the adoption process in the community; intervention is not assigned by investigators, but adopted by institutions themselves; comparing outcomes before and after intervention adoptions and between adopters and non-adopters	Easy to implement; can utilize routinely collected data such as administrative data or medical claims data	Bias and confounding exist; more complicate models are needed to take account of clusters and adjust for many confounding factors; causal inference is never certain

The challenges of the stepped wedge design include larger sample size requirements due to reduced effect size, more measurements needed (baseline before intervention, multiple measures post intervention), and a longer study period due to staggered implementation (63,67). The data analysis is also more complicated. The comparisons can be conducted vertically between institutions, and horizontally before and after the intervention within each institution. Both can be incorporated in the same mixed effect model with the time trend exploration as well. With the increasing use of stepped wedge designs in quality improvement studies, more research is needed in sample size estimation and analytical method development.

Although RCTs are considered the gold standard, alternative designs may be easier, quicker, less expensive to execute, and preferable under certain conditions. Properly designed, they can provide strong and valid evidence as well. Under certain circumstances, when randomization is infeasible, quasi-experimental or non-randomized designs can be used (68,69). For example, regression discontinuity design can be used to compare changes before and after an intervention. The intervention is assigned based on the cutpoint of pre-treatment scores. An abrupt change of the slope of outcome against the score at the cutpoint suggests the intervention effects. Similarly, interrupted time series design can also be used in evaluating the intervention effects even in a single institution. In this design, outcomes are measured multiple times before and after the interventions. A significant change in the time trend of outcomes at the time of intervention indicates the intervention effects. However, overall time trend should be eliminated, and adjustments are needed for auto-correlations among multiple measurements of outcomes.

A simpler version of quasi-experimental design is the controlled pre-post design (68). One group of institutions receives an intervention with outcomes measured pre- and post-intervention, while the control group receives usual care with outcomes also measured at the same time as the intervention group. A significant change of outcome from pre- to post-intervention in the intervention group, compared with the outcome difference during the same period in the control group, estimates the intervention effect. The challenge in such design is the myriads of confounding factors that need to be considered in the analysis because the intervention is not randomized.

Finally, observational studies, or 'natural experiments', can be used to evaluate intervention effects, in which investigators observe the diffusion process of an intervention in a community without investigators' interference (70). For example, we can compare outcomes before and after the adoption of an intervention among adopters, and also compare the changes between adopters and non-adopters. Routinely collected data such as administrative data and medical claims data can be used to evaluate intervention effects. The main concern in this design is bias and confounding. It is uncertain why some institutions adopt the intervention earlier, and what other factors may affect the quality of care outcomes. Multivariate analysis is needed but causal inference is never certain.

Example: improving mediastinal lymph node collection in a high lung cancer mortality zone of the US

Mediastinal lymph node examination is a key quality of care measure, associated with better survival among curatively resected NSCLC patients (22). Baseline studies in the metropolitan Memphis, TN population revealed a high proportion of resections without lymph node examination (pNX) or without mediastinal lymph node examination, and a low general lymph node count (71). After interaction with the surgeons and pathologists involved in care, it was determined that the problem arose from three processes: events in the operating room during surgery (the surgical lymphadenectomy); the communication between the operating room and pathology laboratory teams (possible loss of specimens in transit, poor identification of the source of lymph node specimens); and events in the pathology laboratory (the retrieval and examination of lymph node specimens) (72,73).

A lymph node collection kit that clearly labels the mediastinal stations, with a checklist to remind surgeons of the recommended sites for lymph node specimen collection, were developed to improve the surgical lymphadenectomy and communication between operating room and pathology laboratory teams (72-74). Pilot studies showed the efficacy of the kit in improving the quality of hilar and mediastinal lymphadenectomy and increasing the rate of detection of hilar and mediastinal lymph node metastasis (72). To disseminate the new method into all institutions within the tri-state (North Mississippi, Eastern Arkansas and West Tennessee) region, we employed a multiple baseline design without randomization (quasi-experimental design) to examine the effectiveness of this intervention in the more heterogeneous communities.

This ongoing National Cancer Institute-funded study

employed the RE-AIM framework to guide the study evaluation and a staggered implementation execution strategy (46,52). Fourteen institutions were stratified into three homogeneous cohorts based on volume of lung cancer resections, teaching hospital status and metropolitan or rural location. The intervention was implemented in three cohort waves. The interval between waves was about 3 months, with all institutions adopting the kit within a year and having at least 1 year of follow up. Details of mediastinal lymph node examination and pathologic stage distribution were measured for 5 years before the intervention and also at least 3 years after the intervention. A generalized linear mixed effect model was proposed to include comparisons for before and after the intervention within the institutions, between institutions and across different waves. The advantages of this study design include relatively easy implementation, active quality improvement in all institutions, allowance for exploration of barriers in the dissemination and implementation process, and exploration of potential heterogeneity of treatment effects between types of patients, surgeons, and institutions.

Discussion and summary

More research is needed to develop quality of care measures for lung cancer beyond process measures. Research should attempt to rigorously link quality of care measures with patient outcomes such as survival (75), and promote determination of appropriate measurement cutpoints. For example, the NCCN recommends examination of lymph nodes from a minimum of three mediastinal stations (17), while the CoC quality surveillance measure stipulates examination of a minimum of 10 lymph nodes in stage I-II lung cancer resections, with no station specification (18) (*Table 2*). Other evidence suggests that approximately 20 lymph nodes are required to achieve the best survival benefit in patients with pN0 (76), and specific examination of mediastinal lymph node has additional survival benefit (22). Each of these possible quality parameters needs prospective validation.

Rigorous research is needed on the outcome implications of recommendations for structural and process of care measures such as multidisciplinary care delivery, timeliness of care, quality of staging, and stage-appropriate treatment parameters. More work is needed on linkages with the quality of follow up care and patient-reported outcomes. Given the heavy worldwide burden of lung cancer, a validated set of actionable and quantifiable quality measures must be developed to compare the quality of care across practice settings. Since public disclosure of quality of care information improves quality of care (77), eventually, a public reporting system on the quality of lung cancer care must be established.

Acknowledgements

This work was (partially) supported through a Patient-Centered Outcomes Research Institute (PCORI) Award (IH-1304-6147) to Dr. Osarogiagbon.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.
Financial Disclosure: X Yu, RU Osarogiagbon.

Disclaimer: All statements in this report, including its findings and conclusions, are solely those of the authors and do not necessarily represent the views of the Patient-Centered Outcomes Research Institute (PCORI), its Board of Governors or Methodology Committee.

References

1. Committee on Quality of Health Care in America, Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academy Press, 2001.
2. Institute of Medicine. Delivering High-Quality Cancer Care: Charting a New Course for a System in Crisis. Washington, DC: The National Academies Press, 2013.
3. Donabedian A. The quality of care. How can it be assessed? JAMA 1988;260:1743-8.
4. Chassin MR, Galvin RW. The urgent need to improve health care quality. Institute of Medicine National Roundtable on Health Care Quality. JAMA 1998;280:1000-5.
5. Campbell SM, Roland MO, Buetow SA. Defining quality of care. Soc Sci Med 2000;51:1611-25.
6. Iezzoni LI. Risk Adjustment for Measuring Health Care Outcomes. 4th ed. Chicago: Health Administration Press, 2013.
7. American Cancer Society. Learn About Cancer? Lung Cancer - Non-Small Cell. [Accessed June 28, 2015]. Available online: http://www.cancer.org/cancer/lungcancer-non-smallcell/index

8. Desch CE, McNiff KK, Schneider EC, et al. American Society of Clinical Oncology/National Comprehensive Cancer Network Quality Measures. J Clin Oncol 2008;26:3631-7.
9. Bach PB, Cramer LD, Warren JL, et al. Racial differences in the treatment of early-stage lung cancer. N Engl J Med 1999;341:1198-205.
10. Lathan CS, Neville BA, Earle CC. Racial composition of hospitals: effects on surgery for early-stage non-small-cell lung cancer. J Clin Oncol 2008;26:4347-52.
11. Mainz J. Defining and classifying clinical indicators for quality improvement. Int J Qual Health Care 2003;15:523-30.
12. Tanvetyanon T. Quality-of-care indicators for non-small cell lung cancer. Cancer Control 2009;16:335-41.
13. Mazzone PJ, Vachani A, Chang A, et al. Quality indicators for the evaluation of patients with lung cancer. Chest 2014;146:659-69.
14. Hermens RP, Ouwens MM, Vonk-Okhuijsen SY, et al. Development of quality indicators for diagnosis and treatment of patients with non-small cell lung cancer: a first step toward implementing a multidisciplinary, evidence-based guideline. Lung Cancer 2006;54:117-24.
15. Tanvetyanon T, Corman M, Lee JH, et al. Quality of care in non-small-cell lung cancer: findings from 11 oncology practices in Florida. J Oncol Pract 2011;7:e25-31.
16. Darling G, Malthaner R, Dickie J, et al. Quality indicators for non-small cell lung cancer operations with use of a modified Delphi consensus process. Ann Thorac Surg 2014;98:183-90.
17. NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®). Non-Small Cell Lung Cancer (version 3.2015). [Accessed June 28, 2015]. Available online: http://www.nccn.org/professionals/physician_gls/f_guidelines.asp#site
18. CoC Quality of Care Measures. CoC Measures for Quality of Cancer Care—Non-Small Cell Lung Cancer. [Accessed June 28, 2015]. Available online: https://www.facs.org/quality%20programs/cancer/ncdb/qualitymeasures
19. Farjah F, Flum DR, Ramsey SD, et al. Multi-modality mediastinal staging for lung cancer among medicare beneficiaries. J Thorac Oncol 2009;4:355-63.
20. Osarogiagbon RU, Darling GE. Towards optimal pathologic staging of resectable non-small cell lung cancer. Transl Lung Cancer Res 2013;2:364-71.
21. Darling GE, Allen MS, Decker PA, et al. Randomized trial of mediastinal lymph node sampling versus complete lymphadenectomy during pulmonary resection in the patient with N0 or N1 (less than hilar) non-small cell carcinoma: results of the American College of Surgery Oncology Group Z0030 Trial. J Thorac Cardiovasc Surg 2011;141:662-70.
22. Osarogiagbon RU, Yu X. Mediastinal lymph node examination and survival in resected early-stage non-small-cell lung cancer in the surveillance, epidemiology, and end results database. J Thorac Oncol 2012;7:1798-806.
23. Olsson JK, Schultz EM, Gould MK. Timeliness of care in patients with lung cancer: a systematic review. Thorax 2009;64:749-56.
24. BTS recommendations to respiratory physicians for organising the care of patients with lung cancer. The Lung Cancer Working Party of the British Thoracic Society Standards of Care Committee. Thorax 1998;53 Suppl 1:S1-8.
25. Billing JS, Wells FC. Delays in the diagnosis and surgical treatment of lung cancer. Thorax 1996;51:903-6.
26. Faris N, Yu X, Sareen S, et al. Preoperative Evaluation of Lung Cancer in a Community Health Care Setting. Ann Thorac Surg 2015;100:394-400.
27. Gomez DR, Liao KP, Swisher SG, et al. Time to treatment as a quality metric in lung cancer: Staging studies, time to treatment, and patient survival. Radiother Oncol 2015;115:257-63.
28. Quarterman RL, McMillan A, Ratcliffe MB, et al. Effect of preoperative delay on prognosis for patients with early stage non-small cell lung cancer. J Thorac Cardiovasc Surg 2003;125:108-13; discussion 113-4.
29. Myrdal G, Lambe M, Hillerdal G, et al. Effect of delays on prognosis in patients with non-small cell lung cancer. Thorax 2004;59:45-9.
30. Salomaa ER, Sällinen S, Hiekkanen H, et al. Delays in the diagnosis and treatment of lung cancer. Chest 2005;128:2282-8.
31. Lennes IT, Lynch TJ. Quality indicators in cancer care: development and implementation for improved health outcomes in non-small-cell lung cancer. Clin Lung Cancer 2009;10:341-6.
32. Kane RL, Radosevich DM. Conducting health outcomes research. Sudbury, Mass.: Jones & Bartlett Learning, 2011.
33. Green A, Hauge J, Iachina M, et al. The mortality after surgery in primary lung cancer: results from the Danish Lung Cancer Registry. Eur J Cardiothorac Surg 2015. [Epub ahead of print].
34. Pezzi CM, Mallin K, Mendez AS, et al. Ninety-day mortality after resection for lung cancer is nearly double 30-day mortality. J Thorac Cardiovasc Surg

2014;148:2269-77.
35. Hu Y, McMurry TL, Isbell JM, et al. Readmission after lung cancer resection is associated with a 6-fold increase in 90-day postoperative mortality. J Thorac Cardiovasc Surg 2014;148:2261-7.e1.
36. Sanson-Fisher RW, D'Este CA, Carey ML, et al. Evaluation of systems-oriented public health interventions: alternative research designs. Annu Rev Public Health 2014;35:9-27.
37. Green LW, Glasgow RE. Evaluating the relevance, generalization, and applicability of research: issues in external validation and translation methodology. Eval Health Prof 2006;29:126-53.
38. Peek CJ, Glasgow RE, Stange KC, et al. The 5 R's: an emerging bold standard for conducting relevant research in a changing world. Ann Fam Med 2014;12:447-55.
39. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. JAMA 2003;290:1624-32.
40. Brownson RC, Colditz GA, Proctor EK, editors. Dissemination and implementation research in health: translating science to practice. New York: Oxford University Press, 2012.
41. Lobb R, Colditz GA. Implementation science and its application to population health. Annu Rev Public Health 2013;34:235-51.
42. Gaglio B, Phillips SM, Heurtin-Roberts S, et al. How pragmatic is it? Lessons learned using PRECIS and RE-AIM for determining pragmatic characteristics of research. Implement Sci 2014;9:96.
43. Tabak RG, Khoong EC, Chambers DA, et al. Bridging research and practice: models for dissemination and implementation research. Am J Prev Med 2012;43:337-50.
44. Dissemination & Implementation Models in Health Research & Practice. [Accessed July 10, 2015]. Available online: http://dissemination-implementation.org/viewAll_di.aspx
45. Thorpe KE, Zwarenstein M, Oxman AD, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. J Clin Epidemiol 2009;62:464-75.
46. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999;89:1322-7.
47. Gaglio B, Shoup JA, Glasgow RE. The RE-AIM framework: a systematic review of use over time. Am J Public Health 2013;103:e38-46.
48. Dzewaltowski DA, Glasgow RE, Klesges LM, et al. RE-AIM: evidence-based standards and a Web resource to improve translation of research into practice. Ann Behav Med 2004;28:75-80.
49. Feldstein AC, Glasgow RE. A practical, robust implementation and sustainability model (PRISM) for integrating research findings into practice. Jt Comm J Qual Patient Saf 2008;34:228-43.
50. Nilsen P. Making sense of implementation theories, models and frameworks. Implement Sci 2015;10:53.
51. Glanz K, Rimer BK, Viswanath K, editors. Health Behavior and Health Education: Theory, Research, and Practice. 4th ed. San Francisco: Jossey-Bass, 2008.
52. Klesges LM, Estabrooks PA, Dzewaltowski DA, et al. Beginning with the application in mind: designing and planning health behavior change interventions to enhance dissemination. Ann Behav Med 2005;29 Suppl:66-75.
53. Eccles M, Grimshaw J, Walker A, et al. Changing the behavior of healthcare professionals: the use of theory in promoting the uptake of research findings. J Clin Epidemiol 2005;58:107-12.
54. French SD, Green SE, O'Connor DA, et al. Developing theory-informed behaviour change interventions to implement evidence into practice: a systematic approach using the Theoretical Domains Framework. Implement Sci 2012;7:38.
55. Glasgow RE, Chambers D. Developing robust, sustainable, implementation systems using rigorous, rapid and relevant science. Clin Transl Sci 2012;5:48-55.
56. Abernethy AP, Etheredge LM, Ganz PA, et al. Rapid-learning system for cancer care. J Clin Oncol 2010;28:4268-74.
57. Murray DM, Pennell M, Rhoda D, et al. Designing studies that would address the multilayered nature of health care. J Natl Cancer Inst Monogr 2010;2010:90-6.
58. Hayes RJ, Moulton LH. Cluster Randomised Trials (Chapman & Hall/CRC Biostatistics Series). Boca Raton: Chapman and Hall/CRC, 2009.
59. Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. Contemp Clin Trials 2007;28:182-91.
60. Hemming K, Haines TP, Chilton PJ, et al. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. BMJ 2015;350:h391.
61. Hemming K, Girling A, Martin J, et al. Stepped wedge cluster randomized trials are efficient and provide a method of evaluation without which some interventions would not be evaluated. J Clin Epidemiol 2013;66:1058-9.
62. Brown CA, Lilford RJ. The stepped wedge trial design: a

systematic review. BMC Med Res Methodol 2006;6:54.
63. Hemming K, Lilford R, Girling AJ. Stepped-wedge cluster randomised controlled trials: a generic framework including parallel and multiple-level designs. Stat Med 2015;34:181-96.
64. Moulton LH, Golub JE, Durovni B, et al. Statistical design of THRio: a phased implementation clinic-randomized study of a tuberculosis preventive therapy intervention. Clin Trials 2007;4:190-9.
65. Hawkins NG, Sanson-Fisher RW, Shakeshaft A, et al. The multiple baseline design for evaluating population-based research. Am J Prev Med 2007;33:162-8.
66. McHugh M, Harvey JB, Kang R, et al. Community-Level Quality Improvement and the Patient Experience for Chronic Illness Care. Health Serv Res 2015. [Epub ahead of print].
67. Kotz D, Spigt M, Arts IC, et al. Use of the stepped wedge design cannot be recommended: a critical appraisal and comparison with the classic cluster randomized controlled trial design. J Clin Epidemiol 2012;65:1249-52.
68. Shadish WR, Cook TD, Campbell DT. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. 2nd ed. Belmont: Wadsworth Publishing, 2001.
69. Handley MA, Schillinger D, Shiboski S. Quasi-experimental designs in practice-based research settings: design and implementation considerations. J Am Board Fam Med 2011;24:589-96.
70. Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins, 2008.
71. Allen JW, Farooq A, O'Brien TF, et al. Quality of surgical resection for nonsmall cell lung cancer in a US metropolitan area. Cancer 2011;117:134-42.
72. Ramirez RA, Wang CG, Miller LE, et al. Incomplete intrapulmonary lymph node retrieval after routine pathologic examination of resected lung cancer. J Clin Oncol 2012;30:2823-8.
73. Osarogiagbon RU, Ramirez RA, Wang CG, et al. Dual intervention to improve pathologic staging of resectable lung cancer. Ann Thorac Surg 2013;96:1975-81.
74. Osarogiagbon RU, Miller LE, Ramirez RA, et al. Use of a surgical specimen-collection kit to improve mediastinal lymph-node examination of resectable lung cancer. J Thorac Oncol 2012;7:1276-82.
75. Detterbeck F. What is quality and does it matter? J Thorac Oncol 2009;4:279-80.
76. Osarogiagbon RU, Ogbata O, Yu X. Number of lymph nodes associated with maximal reduction of long-term mortality risk in pathologic node-negative non-small cell lung cancer. Ann Thorac Surg 2014;97:385-93.
77. Jung K. The impact of information disclosure on quality of care in HMO markets. Int J Qual Health Care 2010;22:461-8.

Cite this article as: Yu X, Klesges LM, Smeltzer MP, Osarogiagbon RU. Measuring improvement in populations: implementing and evaluating successful change in lung cancer care. Transl Lung Cancer Res 2015;4(4):373-384. doi: 10.3978/j.issn.2218-6751.2015.07.13

