



HHS Public Access

Author manuscript

Ann Intern Med. Author manuscript; available in PMC 2015 June 24.

Published in final edited form as:

Ann Intern Med. 2014 July 1; 161(1): 20–30. doi:10.7326/M13-2231.

Inter-Institutional Variation in Management Decisions for Treatment of Four Common Cancers: A Multi-Institutional Cohort Study

Jane C. Weeks, MD, MSc*, Hajime Uno, PhD, Nathan Taback, PhD, Gladys Ting, MS, Angel Cronin, MS, Thomas A. D'Amico, MD, Jonathan W. Friedberg, MD, MSc, and Deborah Schrag, MD, MPH

Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA (JCW, HU, GS, AC, DS), University of Toronto, Toronto, ON (NT), Department of Surgery, Duke University Medical Center, Durham, NC (TAD), Department of Medicine, University of Rochester School of Medicine and Dentistry, Rochester, NY (JWF)

Abstract

Background—When clinical practice is governed by evidenced-based guidelines and there is consensus regarding their validity, practice variation should be minimal. Where evidence gaps exist, greater variation is expected.

Objective—To systematically assess inter-institutional variation in management decisions for 4 common cancers.

Design—Multi-institutional observational cohort study of cancer patients diagnosed between July 2006 through May 2011 and observed through December 31, 2011.

Setting—18 cancer centers participating in the formulation of treatment guidelines and systematic outcomes assessment through the National Comprehensive Cancer Network.

Corresponding author: Deborah Schrag MD MPH, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215, Ph: 617-632-2509, F: 617-582-8560, deb_schrag@dfci.harvard.edu.

*Dr Weeks is deceased as of 9/10/2013 however, she wrote and approved the the originally submitted manuscript prior to her death.

Address for reprint requests: Deborah Schrag MD MPH, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215

Current postal address for authors:

Jane C. Weeks, MD, MSc, Department of Medical Oncology, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215

Hajime Uno, PhD, Department of Medical Oncology, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215

Nathan Taback, PhD, Statistical Sciences, University of Toronto, 6th Floor, Sidney Smith Hall, 100 St. George St, Toronto, Ontario, Canada M5S 3G3

Gladys Ting, MS, Department of Medical Oncology, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215

Angel Cronin, MS, Department of Medical Oncology, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215

Thomas A. D'Amico, MD, Duke University Medical Center, Box 3496, Duke South, White Zone, Room 3589, Durham, North Carolina 27710

Jonathan W. Friedberg, MD, MSc, 601 Elmwood Avenue Box 704, Rochester, NY 14642

Deborah Schrag, MD, MPH, Department of Medical Oncology, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215

Disclaimer: This is the prepublication, author-produced version of a manuscript accepted for publication in *Annals of Internal Medicine*. This version does not include post-acceptance editing and formatting. The American College of Physicians, the publisher of *Annals of Internal Medicine*, is not responsible for the content or presentation of the author-produced accepted version of the manuscript or any version that a third party derives from it. Readers who wish to access the definitive published version of this manuscript and any ancillary material related to this manuscript (e.g., correspondence, corrections, editorials, linked articles) should go to Annals.org or to the print issue in which the article appears. Those who cite this manuscript should cite the published version, as it is the official version of record.

Patients—25,589 patients with incident cancer of the breast, colorectum, lung, or non-Hodgkin's lymphoma (NHL).

Measurements—Inter-institutional variation for 171 binary management decisions with varying levels of supporting evidence. For each decision, variation was characterized by the median absolute deviation (MAD) of the center-specific proportions.

Results—Inter-institutional variation was high (MAD >10%) for 35/171 (20%) oncology management decisions. This included: 9/22 (41%) for NHL, 16/76 (21%) for breast, 7/47 (15%) for lung, and 3/26 (12%) for colorectal. Decisions involving imaging and/or diagnostic procedures accounted for 46% and chemotherapy regimen choice for 37% of high variance decisions. The evidence grade underpinning the 35 high variance decisions was level I for 0%, 2A for 49% and 2B/other for 51%.

Limitations—Physician identifiers were unavailable, and results may not generalize outside of major cancer centers.

Conclusions—The substantial variation in institutional practice manifest among cancer centers reveals a lack of consensus about optimal management for common clinical scenarios. For clinicians, awareness of management decisions with high variation should prompt attention to patient preferences. For health systems, high variation can be used to prioritize comparative effectiveness research, patient/provider education or pathway development.

Introduction

There is a long-standing commitment to randomized trials as the definitive source of evidence to support clinical decision making in cancer. Publicly- and privately-supported trials have helped to establish evidence-based standards for many of the most important treatment choices. Nonetheless, substantial gaps in the evidence base remain. Cancer is not one disease, but dozens, with complex natural history. Even when the efficacy of a particular treatment has been established in high quality randomized trials, questions of generalizability may remain (1). The result is that the quality of the evidence is uneven across the myriad decisions that arise in the care of cancer patients (2).

The goal of observational studies is to address these gaps in the evidence. To date, topic selection has been left largely to the discretion of investigators, who may or may not consider the non-empirically-based priority lists generated through stakeholder consensus (3). An alternative strategy is to develop an empirically driven research agenda through systematic identification of variation in care. Greater variation in the care of similar patients signals a lack of consensus about what constitutes optimal care, suggesting important gaps in the evidence base where research might have an impact.

We therefore sought to systematically assess inter-institutional variation in management decisions for 4 common cancers, using a data source in which variation is an especially reliable signal of evidence gaps and lack of consensus. The National Comprehensive Cancer Network (NCCN) Outcomes Database is a prospective registry with comprehensive information about patients receiving their cancer care in participating member institutions (4). Criteria for NCCN membership include a demonstrated multidisciplinary focus on

cancer treatment and a reputation as a regional leader in transdisciplinary oncology treatment research and education. We hypothesize that inter-institutional variation among NCCN centers may be an especially useful marker for evidence gaps for several reasons. First, physicians in these centers are sub-specialists who know the available evidence well and indeed collaborate to develop practice guidelines. Second, patient characteristics are reasonably similar and well measured across the centers and it is unlikely that there would be substantial differences in the preferences of patients seen at these different centers. Finally, by focusing on inter-institutional rather than patient level variation and assessing multiple decisions across the disease trajectory, our analysis targets lack of consensus among expert clinicians, areas where the return on CER investment should be high.

Methods

Study Subjects and Data Sources

The analysis was conducted using data from the 4 most mature NCCN Outcomes Databases: breast, colorectal, non-Hodgkin's lymphoma (NHL), and non-small cell lung cancer (NSCLC). All patients receiving some or all of their primary oncologic care at a participating NCCN center were included in the database for that disease. The number of participating centers varied by disease. For this analysis we included patients presenting relatively recently but with sufficient follow-up to fully characterize the care received.

The primary data source for the NCCN Outcomes Database is medical record abstraction, performed by dedicated clinical research associates, with initial and on-going training from centralized project staff. We have shown that the reliability of data abstracted by study staff is excellent (5). Patients in most sites complete a baseline survey that is the primary source for sociodemographic and risk factor data; chart review is used to supplement survey data as needed. Vital status is confirmed for all patients using Social Security Death Index and National Death Index searches. The institutional review boards (IRB) at each center approved the data collection, transmission, and storage protocols; where required, we received informed consent for use of all patient data. The IRB at Dana-Farber Cancer Institute gave approval for the current study.

Elements in the database include: sociodemographics; baseline clinical characteristics (stage, histology, margins, tumor markers and other pathology, menopausal status); comorbidity as assessed by the Charlson/Katz Index (6, 7); diagnostic procedures; treatment; follow-up testing; and outcomes (response to treatment, recurrence, vital status, and cause of death). Stage was assigned according to the version of the American Joint Committee on Cancer (AJCC) Staging Manual applicable at the time of diagnosis (8, 9). Detailed data were collected longitudinally until death for all patients who continued being seen at the participating center; patients who transferred their care were followed for vital status only.

Development of the List of Binary Management Decisions

Our goal was to identify a comprehensive list of dichotomous management choices for which estimates of inter-institutional variation could be generated from the available data. We first catalogued all management decisions mentioned in the 4 disease-specific NCCN

guidelines, using the version of the guideline in effect during the midpoint of the diagnosis date interval for each analytic cohort (10–13). For example, the lung cancer cohort included patients diagnosed from 1/2007–12/2010. The midpoint of this interval is 1/2009, and therefore the guidelines in effect as of January 2009 were used for lung cancer. We reviewed those lists for completeness and accuracy with 2 clinical experts for each cancer type, and asked them to suggest any relevant cross-cutting, i.e., generic rather than disease-specific, management decisions. They proposed 8 additions to the guideline-based list of decisions.

We then conducted analyses of the NCCN outcomes data, identifying for each decision the patients for whom the decision was relevant based on clinical characteristics (i.e., the “denominator” population) and length of follow up required, and among those, the proportion who received each management option, at each institution, and, overall. In select cases, the data for two decisions were combined. Criteria for this aggregation included similar population definitions and management choices, similar overall and institution-specific practice patterns, and small sample sizes.

In order to calculate standardized metrics for inter-institutional variation our goal was to dichotomize management options for each decision. To convert multi-option decisions into dichotomous ones, we used one of two approaches. When appropriate, we disaggregated a decision into two sequential decisions (e.g., by replacing a decision regarding use of chemotherapy where the initial options were drug X, drug Y, or no chemotherapy, with two decisions, one reflecting whether chemotherapy was given at all, and a second on the use of drug X vs drug Y among patients receiving chemotherapy). In other cases, when one of the options was used very rarely, we subsumed it in another option by slightly broadening the definition (e.g., by replacing options of CT, PET, or PET-CT, with options of CT vs PET or PET-CT).

All management decisions were classified by whether they dealt primarily with imaging and/or diagnostic procedures, systemic therapy, surgery, radiation therapy, supportive care, or observation/palliation.

Evidence Grading

For each management decision in the final list, we reviewed the relevant NCCN guideline to determine whether an evidence grade was provided for that choice. If so, that evidence/consensus grade was assigned using the standard NCCN grading system (14): Category 1 -- the recommendation is based on high-level of evidence (e.g. randomized controlled trials) and there is uniform NCCN consensus; Category 2A-- the recommendation is based on lower-level evidence and there is uniform NCCN consensus; Category 2B -- the recommendation is based on lower-level evidence and there is nonuniform NCCN consensus (but no major disagreement); and Category 3 – the recommendation is based on any level of evidence but reflects major disagreement.

When the NCCN guidelines provided a specific recommendation but the criteria required to determine concordance were not available (e.g., the resectability of a lesion), we classified that decision as “Depends on clinical characteristics” (DCC). When the guidelines did not provide a definitive recommendation (e.g., “consider radiotherapy”), we classified that

decision as “No definitive recommendation” (NDR). For management decisions that were included based on expert suggestion and that did not appear in the guidelines, we classified the decision as “Not guideline derived” (NGD).

Statistical Analysis

Calculation of inter-institutional variation was performed for each binary management decision with at least 100 patients after excluding institutions with fewer than 10 patients for that decision. The median absolute deviation (MAD) was used to quantify inter-institutional variation, given by:

$$\text{MAD} = \text{median}_i \{ |p_i - \text{median}_j(p_j)| \},$$

where p_i is the proportion of patients receiving a particular management in institution i . The MAD is a widely used measure of variation and is particularly appropriate for this analysis because absolute deviation is robust to outliers (unlike variance or standard deviation). As a benchmark reference value, we calculated an expected MAD for each binary management decision in the case that inter-institution variation does not exist, by simulating the data from binomial distributions, where the success probability of the binomial distribution was set to the same value for each institution (specifically, the observed proportion calculated across all institutions). For each simulated dataset, we calculated a MAD. We repeated this 1000 times and took the average of the 1000 MADs.

We used linear regression with the MAD as the response variable to determine whether disease type, treatment/management modality, and evidence grade were associated with inter-institutional variation. Several categories were collapsed in this analysis to avoid small cell sizes (for example, 'supportive care' and 'observation/palliation' were collapsed into a single category because only 3 decisions consisted of supportive care). For each dependent variable, we performed a test of the null hypothesis of no difference between any of the categories using a Wald test. In these analyses, we excluded all overlapping decisions (i.e., those describing a subset of patients included in another, less narrowly defined decision) such that conditional on the independent variables, the response values would be statistically independent of each other. The normality assumption of the distribution of the residuals appeared to be satisfied for each regression model. Statistical analyses were conducted using SAS (version 9.2; SAS Institute, Cary, NC) and Stata (version 11.1, StataCorp LP, College Station, TX). All reported p-values are two-sided, with $p < 0.05$ considered to be statistically significant.

Role of the Funding Source

This study was funded by a grant from the National Cancer Institute (NCI) to the Dana-Farber Cancer Institute (RC1CA146196). The NCCN provided support for the outcomes database infrastructure and management. The funding sponsors had no such involvement in the design of the study; the collection, analysis, and interpretation of the data; and in the decision to approve publication of the finished manuscript.

Results

Our breast cancer cohort included 11,293 women with incident cancer presenting to 17 NCCN institutions from 7/2008–6/2010. The colorectal cancer cohort included 4,564 patients with incident tumors or new metastatic disease presenting to 8 NCCN institutions from 9/2007–5/2011. The NHL cohort consisted of 3,014 patients with a new diagnosis of any of the 6 histologic subtypes included in the database (small lymphocytic, follicular, non-gastric MALT, nodal marginal zone, mantle cell, and diffuse large B cell lymphoma) presenting to 7 institutions from 7/2006–6/2010. The NSCLC cohort included 6,718 patients with incident tumors who presented to one of 8 NCCN centers from 1/2007–12/2010. Details of the centers included in the analysis of each cancer type are in the appendix (Appendix Table 1). For each cancer type, inter-institution variations in age and comorbidity were modest. Mean ages by institutions were within 6.3 years of range, and the range of the proportions of high comorbidity was less than 15% for all cancer types.

Ranking of Management Decisions by Inter-Institutional Variation

Of 229 management decisions included in our initial list, 58 had a sample size of less than 100 patients. The remaining 171 decisions (76 breast, 26 colorectal, 22 NHL, and 47 NSCLC) were then ranked by the magnitude of inter-institutional variation. The 10 decisions across all four diseases with the greatest variation are shown in Table 1; a complete ranked, annotated list of all 171 decisions, separately by disease, is in the appendix (Appendix Tables 2–5). Each decision involves two alternative management strategies for a clinically-defined “denominator” population. The table provides the aggregate proportion of patients treated according to each of the two alternative strategies (i.e., all patients across all centers), and the institution range (i.e., the highest and lowest center-specific proportions receiving each strategy). Each decision is also annotated with: the calculated inter-institutional variation for that measure, and an indicator of whether one strategy was recommended by the NCCN guidelines, and if so, the evidence grade supporting that recommendation. Institution-specific proportions of the binary management decisions, ordered within NCCN evidence grade category by MAD, are shown in Figure 1 for breast cancer, Figure 2 for colorectal cancer, Figure 3 for lung cancer, and Figure 4 for non-Hodgkin's lymphoma. The numbers on the x-axis for Figures 1–4 correspond to the decision number (i.e. the first column) in Appendix Tables 2–5.

Three management decisions (2%) had MAD >0.20, which corresponded to institution ranges of 5%–62%, 21%–99%, and 0%–88%. Among the 20% (35/171) of decisions with MAD >0.10, the median width of the institution range was 0.55. Highly variable decisions (those with MAD exceeding 10%) represented 9/22 (41%) for NHL, 16/76 (21%) for breast, 7/47 (15%) for lung, and 3/26 (12%) for colorectal. Decisions involving imaging and/or diagnostic procedures accounted for 46% of high variance decisions; whether to administer chemotherapy accounted for 6%, whether to administer radiation or perform surgery accounted for 11%; and regimen choice represented 37% of high variance decisions. The grade of evidence underpinning high variance decisions was level I for 0%, 2A for 49% and 2B/other for 51% of the 35 high variance decisions. In 79% of the management decisions,

the observed MAD exceeded what would be expected from our simulations, and in 41%, the observed MAD was more than twice as large as expected.

Factors Associated with Variation

The analysis of factors associated with inter-institutional variation was conducted among the subset of management decisions that were completely non-overlapping with other decisions (142/171). The attributes of those decisions are shown in Table 2. Figure 5 shows the MAD by disease type, modality, and evidence grade. None of these associations met conventional levels of statistical significance ($p=0.125$ for disease, $p=0.071$ for modality, and $p=0.058$ for evidence grade). There was slightly greater inter-institutional variation for NHL than other diseases (mean MAD of 0.09 for NHL, versus 0.07 for breast cancer, 0.07 for colorectal cancer, and 0.06 for lung cancer); for imaging/diagnostic procedures than for other modalities (mean MAD of 0.09, versus 0.07 for surgery/radiation therapy, 0.06 for chemotherapy, and 0.05 for supportive care/observation/palliation); and for NCCN evidence grade 2A (mean MAD of 0.08, versus 0.07 for category 2A/other and 0.05 for grade 1).

Discussion

In a systematic assessment of a comprehensive set of management decisions for 4 common cancers, we found substantial inter-institutional variation in practice patterns among expert centers. Our goal was to identify practice patterns suggestive of a lack of consensus among expert clinicians rather than to characterize regional differences in the tendency to practice in a particular style. For many of these highly variable decisions, the alternatives are strategies that differ substantially in toxicity, cost, or both. Category 1 evidence was available for only 11% of the decisions assessed. Of note, decisions supported by high level evidence did not demonstrate statistically significantly lower levels of inter-institutional variation, which suggests that variation cannot simply be inferred from an evidence grade. We found differences based on modality and disease, with greater inter-institutional variation for decisions involving imaging and/or diagnostic procedures, and for management of non-Hodgkin's lymphoma, although this was not statistically significant.

Although we used NCCN guidelines as a starting point in generating a list of decisions to examine, this analysis is not an assessment of the quality of care in the major cancer centers that comprise NCCN institutions. For over half the decisions we assessed, the guidelines make no definitive recommendations about the preferred strategy, so there is no standard against which to compare the care we observed. Instead, our goal was to use this variation to identify clinical decisions for which there is a lack of consensus in the field. We found substantial variation related to the institution delivering the care across many cancer types, throughout the disease trajectory, and across therapeutic modalities. This observation is consistent with the literature demonstrating that institutional practice patterns are highly influenced by local practice culture that likely stems from the beliefs/preferences of local opinion leaders, especially with respect to adoption of new technologies (15–17).

Despite a shared focus on variation, this analysis is quite different in nature and intent from well-known efforts using administrative data to compare regions, e.g., the Dartmouth Atlas (18–21), or to evaluate specific clinical decisions (22–25). Rather than looking at broad,

overall patterns of care, our richly detailed clinical data allowed us to examine very specifically-defined binary management decisions. Outcomes databases with rich clinical detail have frequently been used to determine that variation is attributable to non-clinical factors (26, 27) but often cannot specify whether the variation is inherent to ecological factors specific to a region, clinician views, or structural aspects of care such as the availability of radiation therapy. Our unit of analysis was the institution, specifically, a set of cancer centers with similar structures and ample access to the resources necessary to deliver each of the binary treatment options. Each of the cancer centers in our sample has ample facilities and resources to administer both alternatives in the binary management decisions we evaluated. Moreover, we considered management decisions that involve therapies that are routinely covered by health care payors. As a result, the likelihood that ecological factors stemming from payor mix, or resource availability account for a large component of variation is attenuated, albeit not entirely eliminated.

Should we be concerned that decisions for management of common clinical situations in oncology vary so much across structurally similar institutions? For example, is it a problem that depending on where a patient receives care, the probability of initial cytotoxic treatment for indolent lymphoma rather than observation varies from 51 to 100%, or that the probability of receiving sphincter-sparing surgery for distal rectal cancer varies from 30–90%? Without a solid evidence base demonstrating the superiority of one strategy over the alternative, it's not possible to make a judgment about which option represents higher quality care. However, it's quite likely that for most of these decisions, one alternative is, in fact, more effective, less toxic, and/or less expensive; we're just lacking the data needed to identify which one that is. By giving priority to research on decisions that exhibit greater degrees of variation and affect larger numbers of patients, we can optimize the public health return on the investment in that research

Substantial variation should prompt further investigation to pinpoint whether it derives from patient preferences, availability of resources, training and expertise, various financial incentives, a limited evidence base, or a combination of these factors. The presence of patient or clinical attributes that predict treatment decisions should be identified and if indeed, important, should be made explicit. In some instances, surveys to ascertain the role of patient preferences may be necessary. In others, physician surveys may reveal whether structural aspects of care delivery such as surgeon availability or insurance coverage influence practice patterns. Where neither patient attributes, preferences or access to care are heterogeneous, it is likely that clinician uncertainty about the optimal management strategy drives variation. In this context, the revealed variation in management decisions highlights areas where clinical trials are most likely to have an impact. In an era of constrained resources, prioritizing clinical research in areas of greatest uncertainty is likely to yield the greatest return on investment.

Our study has a number of strengths, including the richness of the available clinical data, allowing us to ensure that we were comparing patterns of care among truly similar patients, and the setting in major cancer centers, reducing the influence of factors other than evidence gaps on variation. Our design also has several potential limitations. Notwithstanding the granularity of our data source, we cannot entirely rule out subtle variation in patient

attributes or ecological factors as contributors to variation. Second, because the data source we used does not contain physician identifiers, we could not conduct analyses that took into account the clustering of patients by physician in addition to center. However, these clinicians are salaried employees, and are therefore less likely than physicians in fee-for-service settings to be influenced by financial incentives. Finally, because the practice patterns we observed were not population-based, our estimates of the proportion of patients treated in a particular fashion for each decision are not generalizable. However, our primary goal was not to characterize population-based practice patterns but rather to assess inter-institutional variation as a signal suggesting lack of consensus among knowledgeable clinicians, and using a national sample of academic cancer centers as our “laboratory” provided important advantages in that respect.

There are a growing number of clinically rich multi-institutional disease-specific registries that could be similarly interrogated to identify clinically important gaps in the evidence for other conditions (28–30). A systematic approach of this kind could contribute a much-needed empiric perspective to priority setting in the emerging discipline of CER and the research agendas of organizations such as PCORI. The striking degree of inter-institutional variation we observed for decisions involving two established management choices is compelling evidence of the need for a research agenda in cancer that is not exclusively focused on the testing of new agents. Our results also have more immediate implications for patients and payers. As unsettling as it may be, patients should understand that even for very common conditions, the recommended approach may differ dramatically depending on where they choose to receive their care. Payers need to recognize that the overall costs of care may be much more dependent on how clinicians choose to manage very similar patients than on the unit costs of specific services. Ranking management decisions based on the magnitude of observed inter-institutional variation is a strategy to prioritize research based on revealed lack of consensus. In an era of increasingly constrained research funding, such systematic approaches may help maximize the return on investment in effectiveness research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding: Analytic work was supported by a grant from the National Cancer Institute (NCI) to the Dana-Farber Cancer Institute (RC1CA146196). The National Comprehensive Cancer Network and 18 of its member cancer centers provided funding for abstraction of medical records and curation of the cancer outcomes database. Funding agencies had no role in interpretation of data or final manuscript editing.

References

1. Hutchins LF, Unger JM, Crowley JJ, Coltman CA Jr, Albain KS. Underrepresentation of patients 65 years of age or older in cancer-treatment trials. *N Engl J Med.* 1999; 341(27):2061–2067. [PubMed: 10615079]
2. Poonacha TK, Go RS. Level of scientific evidence underlying recommendations arising from the National Comprehensive Cancer Network clinical practice guidelines. *J Clin Oncol.* 2011; 29(2): 186–191. [PubMed: 21149653]

3. Medicine Io.. Initial national priorities for comparative effectiveness research. Washington, DC: Institute of Medicine; 2009.
4. Weeks JC. Outcomes assessment in the NCCN. *Oncology (Williston Park)*. 1997; 11(11A):137–140. [PubMed: 9430183]
5. Kho ME, Lepisto EM, Niland JC, Friedberg JW, Lacasce AS, Weeks JC. Reliability of staging, prognosis, and comorbidity data collection in the National Comprehensive Cancer Network (NCCN) non-Hodgkin lymphoma (NHL) multicenter outcomes database. *Cancer*. 2008; 113(11): 3209–3212. [PubMed: 18855918]
6. Charlson M, Pompei P, Ali=es K, MacKenzie C. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J Chron Dis*. 1987; 40(5):373–383. [PubMed: 3558716]
7. Katz JN, Chang LC, Sangha O, Fossel AH, Bates DW. Can comorbidity be measured by questionnaire rather than medical record review? *Med Care*. 1996; 34(1):73–84. [PubMed: 8551813]
8. Greene, F. American Joint Committee on Cancer., American Cancer Society. *AJCC cancer staging handbook: from the AJCC cancer staging manual*. ed 6th. New York: Springer; 2002.
9. Edge, S. *AJCC cancer staging handbook: from the AJCC cancer staging manual*. ed 7th. New York: Springer; 2010. American Joint Committee on Cancer, American Cancer Society.
10. NCCN. *NCCN Clinical Practice Guidelines in Oncology, (Breast Cancer ed v.1.2009)*. Fort Washington, PA: The National Comprehensive Cancer Network; 2009. www.nccn.org.
11. NCCN. *NCCN Clinical Practice Guidelines in Oncology, (Colon Cancer ed v.3.2009)*. Fort Washington, PA: The National Comprehensive Cancer Network; 2009. www.nccn.org.
12. NCCN. *NCCN Clinical Practice Guidelines in Oncology, (Non-Hodgkin's Lymphomas ed v. 1.2007)*. Fort Washington, PA: The National Comprehensive Cancer Network; 2007. www.nccn.org.
13. NCCN. *NCCN Clinical Practice Guidelines in Oncology, (Non-Small Cell Lung Cancer ed v. 2.2009)*. Fort Washington, PA: The National Comprehensive Cancer Network; 2009. www.nccn.org.
14. NCCN Categories of Evidence and Consensus. http://www.nccn.org/professionals/physician_gls/categories_of_consensus.asp.
15. Soumerai SB, McLaughlin TJ, Gurwitz JH, Guadagnoli E, Hauptman PJ, Borbas C, et al. Effect of local medical opinion leaders on quality of care for acute myocardial infarction: a randomized controlled trial. *JAMA*. 1998; 279(17):1358–1363. [PubMed: 9582043]
16. Greco PJ, Eisenberg JM. Changing physicians' practices. *N Engl J Med*. 1993; 329(17):1271–1273. [PubMed: 8413397]
17. Flodgren G, Parmelli E, Doumit G, Gattellari M, O'Brien MA, Grimshaw J, et al. Local opinion leaders: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev*. 2011(8):CD000125. [PubMed: 21833939]
18. Venkatraman G, Likosky DS, Morrison D, Zhou W, Finlayson SR, Goodman DC. Small area variation in endoscopic sinus surgery rates among the Medicare population. *Arch Otolaryngol Head Neck Surg*. 2011; 137(3):253–257. [PubMed: 21422309]
19. Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder EL. The implications of regional variations in Medicare spending. Part 1: the content, quality, and accessibility of care. *Ann Intern Med*. 2003; 138(4):273–287. [PubMed: 12585825]
20. Donohue JM, Morden NE, Gellad WF, Bynum JP, Zhou W, Hanlon JT, et al. Sources of regional variation in Medicare Part D drug spending. *N Engl J Med*. 2012; 366(6):530–538. [PubMed: 22316446]
21. Stukel TA, Fisher ES, Alter DA, Guttman A, Ko DT, Fung K, et al. Association of hospital spending intensity with mortality and readmission rates in Ontario hospitals. *JAMA*. 2012; 307(10):1037–1045. [PubMed: 22416099]
22. Shapiro MF, Morton SC, McCaffrey DF, Senterfitt JW, Fleishman JA, Perlman JF, et al. Variations in the care of HIV-infected adults in the United States: results from the HIV Cost and Services Utilization Study. *JAMA*. 1999; 281(24):2305–2315. [PubMed: 10386555]

23. Nattinger AB, Gottlieb MS, Veum J, Yahnke D, Goodwin JS. Geographic variation in the use of breast-conserving treatment for breast cancer. *N Engl J Med.* 1992; 326(17):1102–1107. [PubMed: 1552911]
24. Bach PB, Cramer LD, Warren JL, Begg CB. Racial differences in the treatment of early-stage lung cancer. *N Engl J Med.* 1999; 341(16):1198–1205. [PubMed: 10519898]
25. Schrag D, Cramer LD, Bach PB, Begg CB. Age and adjuvant chemotherapy use after surgery for stage III colon cancer. *J Natl Cancer Inst.* 2001; 93(11):850–857. [PubMed: 11390534]
26. Matlock DD, Peterson PN, Wang Y, Curtis JP, Reynolds MR, Varosy PD, et al. Variation in use of dual-chamber implantable cardioverter-defibrillators: results from the national cardiovascular data registry. *Arch Intern Med.* 2012; 172(8):634–641. discussion 41. [PubMed: 22529229]
27. O'Hare AM, Rodriguez RA, Hailpern SM, Larson EB, Kurella Tamura M. Regional variation in health care intensity and treatment practices for end-stage renal disease in older adults. *JAMA.* 2010; 304(2):180–186. [PubMed: 20628131]
28. Cohen ME, Bilimoria KY, Ko CY, Richards K, Hall BL. Variability in length of stay after colorectal surgery: assessment of 182 hospitals in the national surgical quality improvement program. *Ann Surg.* 2009; 250(6):901–907. [PubMed: 19953710]
29. Arnold SV, Spertus JA, Tang F, Krumholz HM, Borden WB, Farmer SA, et al. Statin use in outpatients with obstructive coronary artery disease. *Circulation.* 2011; 124(22):2405–2410. [PubMed: 22064595]
30. Subherwal S, Peterson ED, Dai D, Thomas L, Messenger JC, Xian Y, et al. Temporal trends in and factors associated with bleeding complications among patients undergoing percutaneous coronary intervention: A report from the National Cardiovascular Data CathPCI Registry. *J Am Coll Cardiol.* 2012; 59(21):1861–1869. [PubMed: 22595404]

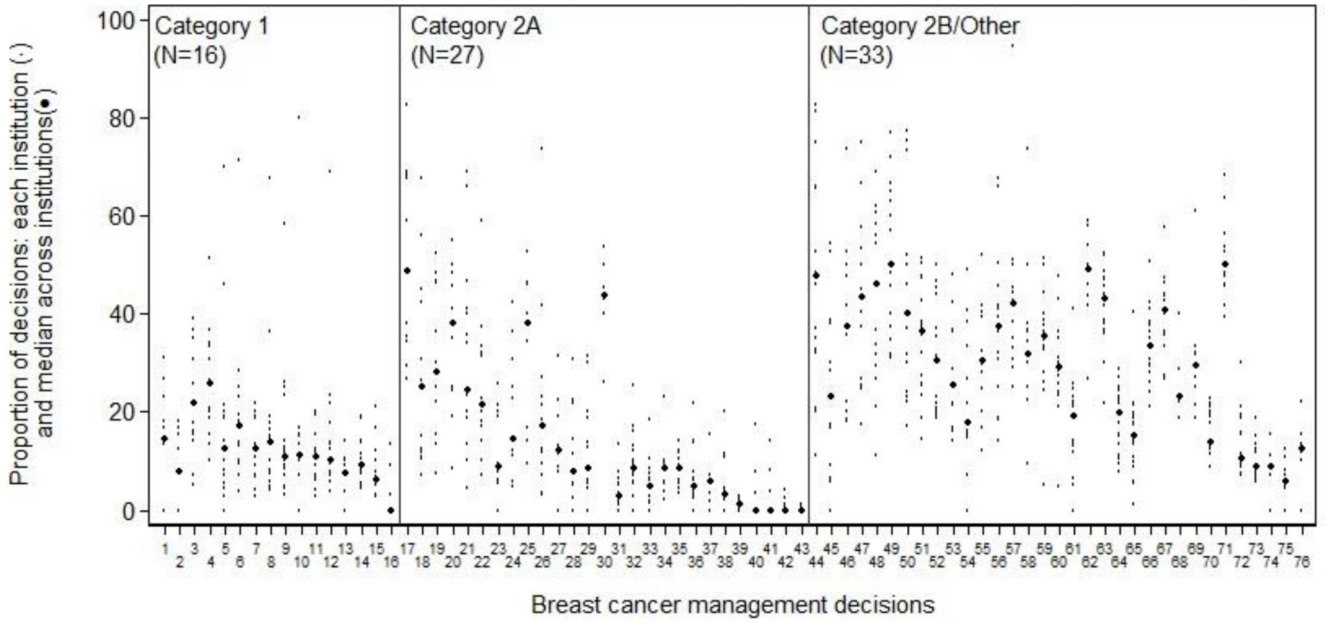


Figure 1. Breast cancer: inter-institution variation of management decisions. The small dots represent the institution-specific proportions for institutions contributing at least 10 patients to the decision, and the more prominent dots represent the between-institution medians for each management decision. Management decisions are ordered within NCCN evidence grade category by the median absolute deviation from the median (MAD) with the largest on the left. The numbers on the x-axis correspond to the decision number in Appendix Table 2.

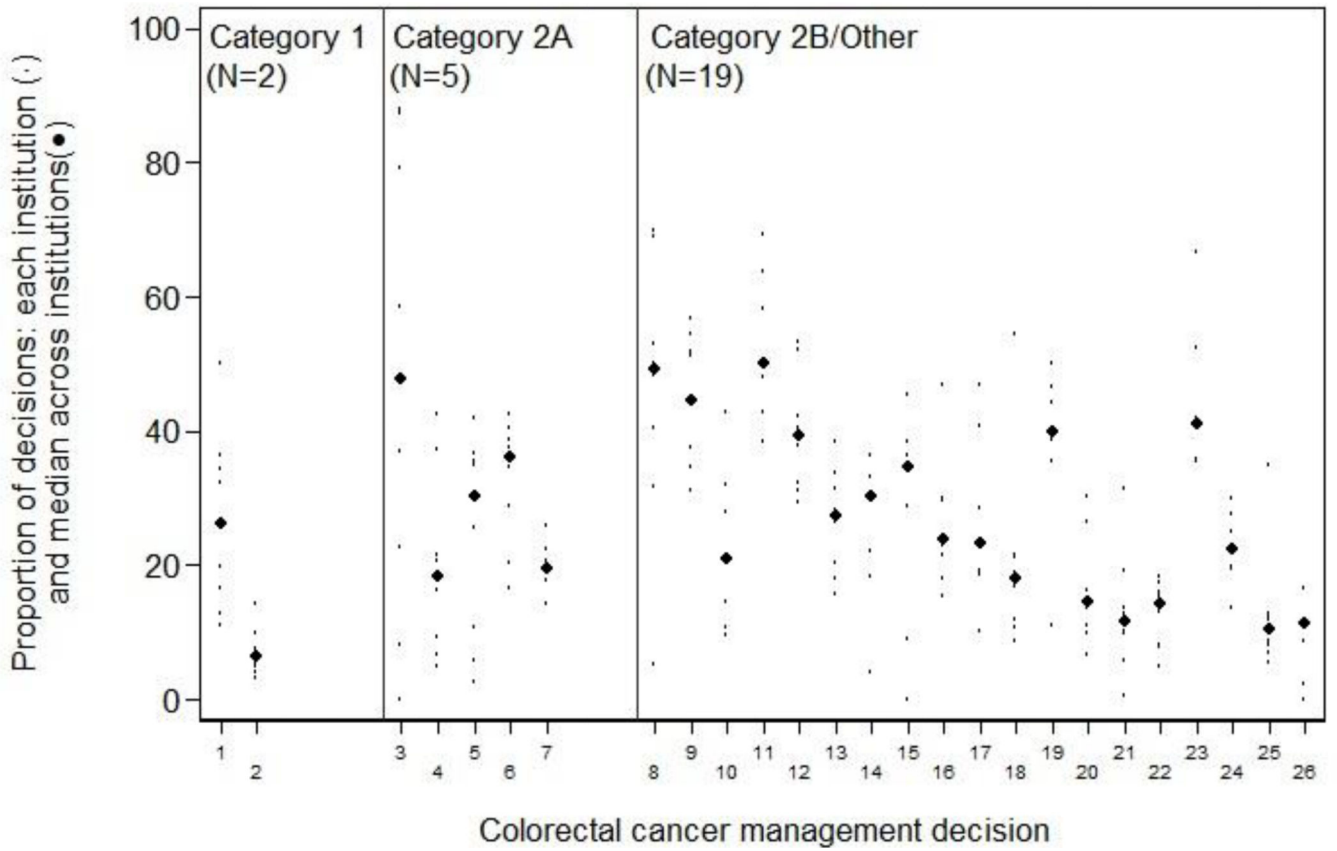


Figure 2. Colorectal cancer: inter-institution variation of management decisions. The small dots represent the institution-specific proportions for institutions contributing at least 10 patients to the decision, and the more prominent dots represent the between-institution medians for each management decision. Management decisions are ordered within NCCN evidence grade category by the median absolute deviation from the median (MAD) with the largest on the left. The numbers on the x-axis correspond to the decision number in Appendix Table 3.

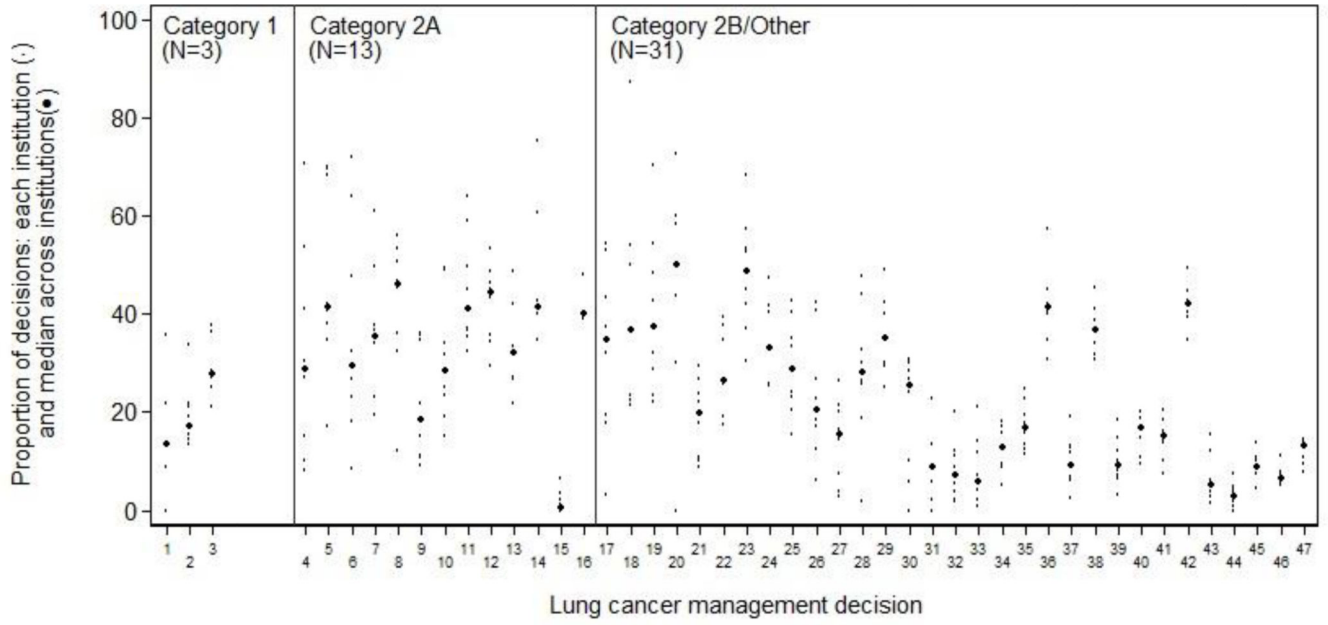


Figure 3. Lung cancer: inter-institution variation of management decisions. The small dots represent the institution-specific proportions for institutions contributing at least 10 patients to the decision, and the more prominent dots represent the between-institution medians for each management decision. Management decisions are ordered within NCCN evidence grade category by the median absolute deviation from the median (MAD) with the largest on the left. The numbers on the x-axis correspond to the decision number in Appendix Table 4.

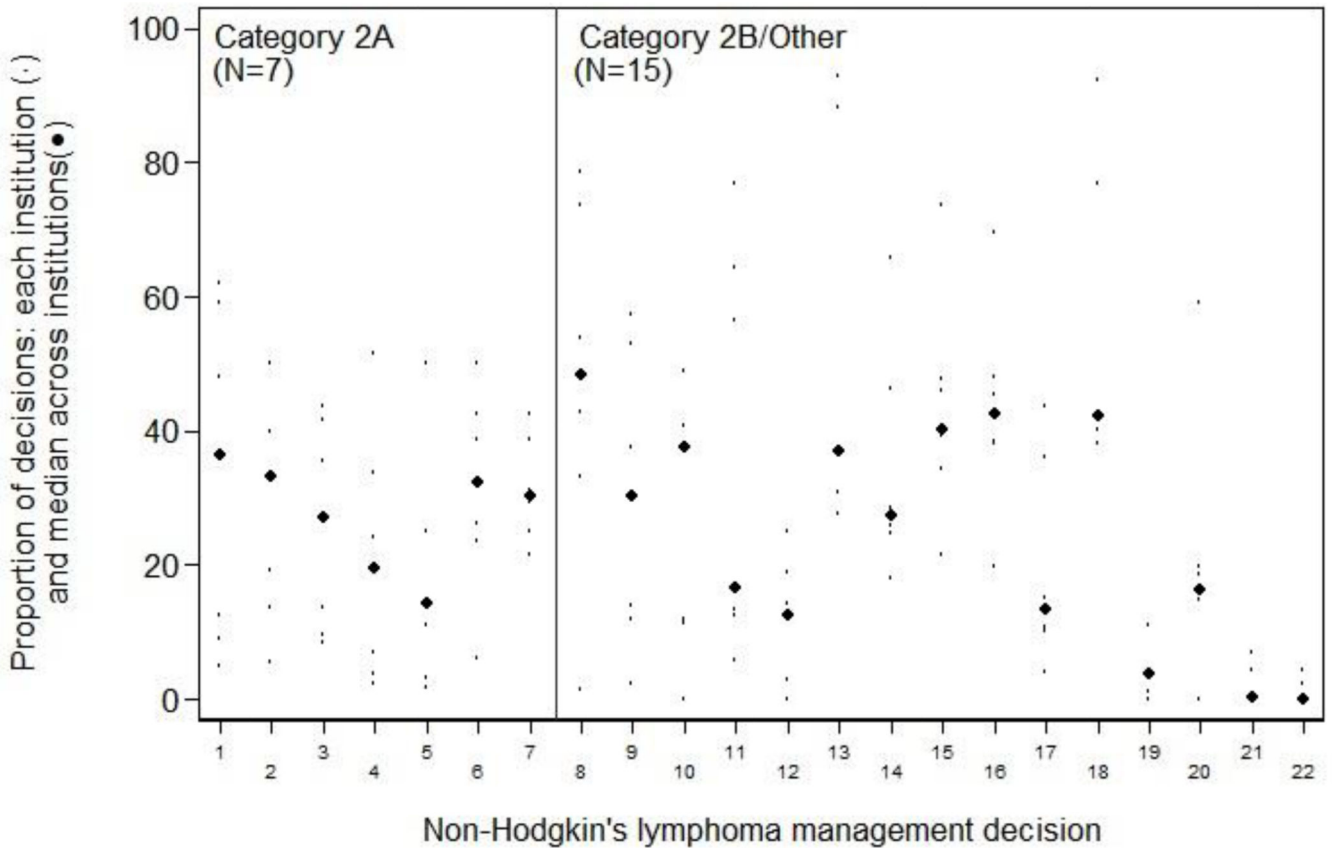


Figure 4. Non-Hodgkin's lymphoma: inter-institution variation of management decisions. The small dots represent the institution-specific proportions for institutions contributing at least 10 patients to the decision, and the more prominent dots represent the between-institution medians for each management decision. Management decisions are ordered within NCCN evidence grade category by the median absolute deviation from the median (MAD) with the largest on the left. The numbers on the x-axis correspond to the decision number in Appendix Table 5.

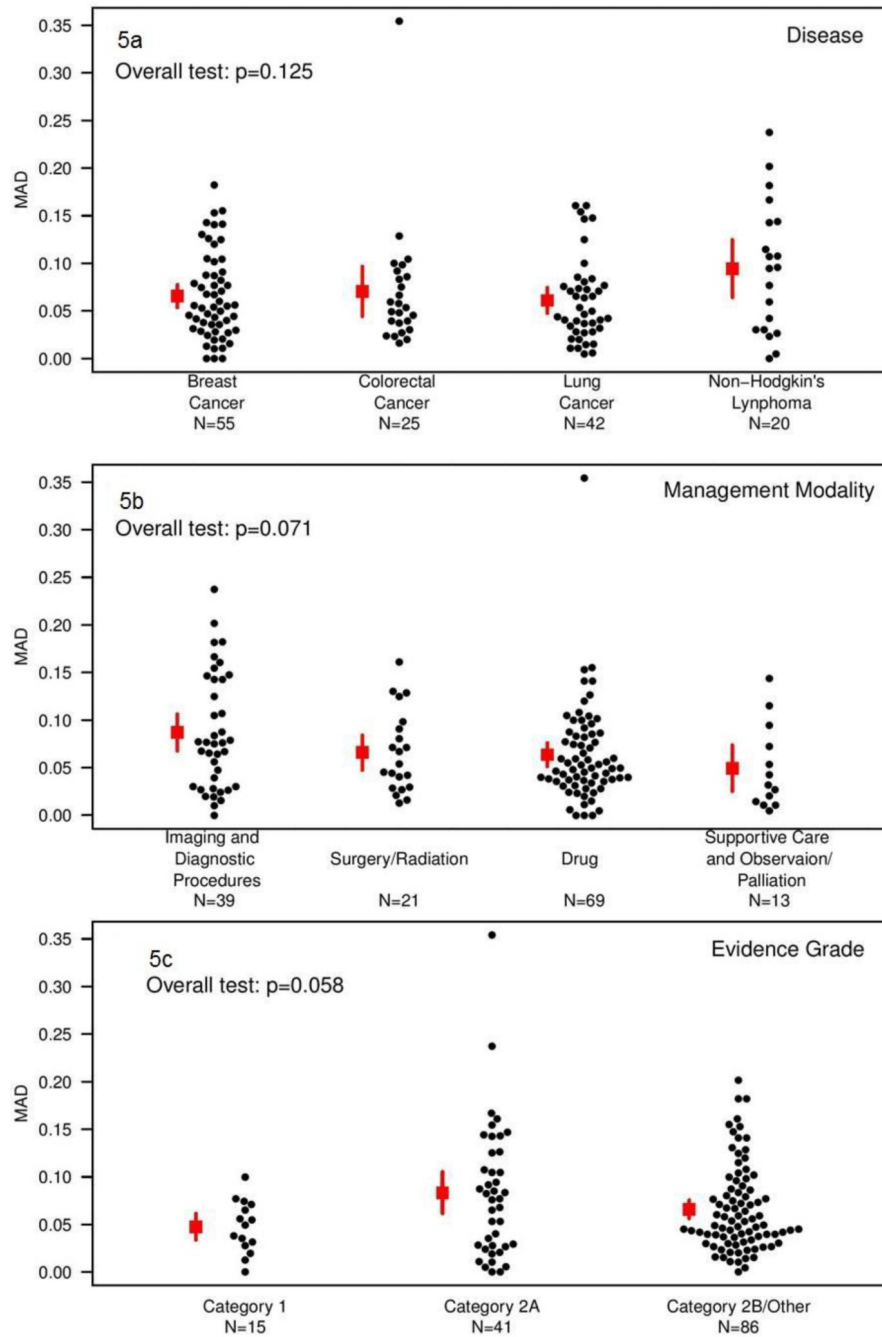


Figure 5. Median absolute deviation from the median by disease (panel a), management modality (panel b), and evidence grade (panel c). Small dots represent the MAD for unique management decisions; squares represent the average MAD and lines represent the corresponding 95% confidence interval.

Top 10 management decisions with the most inter-institutional variation. All statistics are calculated among centers contributing at least 10 patients to the subgroup. The NCCN guideline evidence grade shown in the table is aligned with the endorsed treatment option.

Table 1

Rank across all diseases	Disease	Estimated Annual Population in US	Subgroup	Management Options	NCCN Guideline Evidence Grade	Center Range	Observed Median Absolute Difference	Expected Median Absolute Difference***	Number of Centers
1	Rectal	17,700	Clinical Stage II–III treated with preoperative concurrent chemoradiation with Capecitabine or Infusional 5-FU	Capecitabine	2A	12% – 100%	0.354	0.044	8
				Infusional 5-FU		0% – 88%			
2	NHL	10,100	Follicular lymphoma, all stages, who undergo imaging for staging	No staging PET	2A	5% – 62%	0.237	0.030	7
				Staging PET		38% – 95%			
3	NHL	3,500	Staging of non-gastric MALT Lymphoma, all clinical stages	Bone marrow biopsy	NDR	21% – 99%	0.202	0.059	6
				No bone marrow biopsy		1% – 79%			
4	Breast	13,100	DCIS and Stage I-III without recurrence through 30 months of follow-up	Liver function tests for post-treatment surveillance between months 18 and 30	2A	27% – 83%	0.192	0.044	10
				No liver function tests for post-treatment surveillance between months 18 and 30		17% – 73%			
5	Breast	158,900	Staging of clinical Stage I/II	Chest x-ray	DCC	11% – 83%	0.182	0.018	17
				No chest x-ray		17% – 89%			
6	NHL	10,500	Staging of Follicular Lymphoma, all clinical stages	Bone marrow biopsy	NDR	43% – 97%	0.182	0.028	7
				No bone marrow biopsy		3% – 57%			
7	NHL	1,800	Mantle cell lymphoma, any clinical stage, who undergo imaging for staging	No staging PET	2A	6% – 50%	0.167	0.052	7
				Staging PET		50% – 94%			
8	Lung	19,700	Brain metastases treated initially with radiation therapy	Initial treatment stereotactic radiosurgery	NDR	3% – 55%	0.161	0.032	8
				Initial treatment whole brain irradiation		45% – 97%			
9	Lung	169,700	All Patients	Performance status documented in the chart prior to first treatment	2A	29% – 92%	0.161	0.010	8
				No performance status documented in the chart prior to first treatment		8% – 71%			
10	Breast	6,000	Stage I, HER2-neu positive disease treated with adjuvant or neoadjuvant chemotherapy	Anthracycline-containing regimen	NDR	0% – 55%	0.155	0.068	12
				Non-anthracycline-containing regimen		45% – 100%			

* DCC: depends on clinical characteristics; NDR: no definitive recommendation; NGD: not guideline derived.

Based on 1,000 iterations of simulated datasets from a binomial distribution under the assumption of no inter-institution variation.
**

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Attributes of the decisions included in the analysis of factors associated with inter-institutional variation, stratified by disease (N = 142)

Table 2

Attribute	Overall N (%)	Breast Cancer N=55	Colorectal Cancer N=25	Lung Cancer N=42	Non-Hodgkins Lymphoma N=20
Management Modality					
Imaging and diagnostic procedures	39 (27)	16	4	9	10
Systemic therapy	69 (49)	29	16	18	6
Surgery	12 (8)	4	5	3	0
Radiation therapy	9 (6)	6	0	2	1
Supportive care	3 (2)	0	0	2	1
Observation/palliation	10 (7)	0	0	8	2
Level of Evidence					
Category 1	15 (11)	10	2	3	0
Category 2A	41 (29)	19	5	11	6
Category 2B	4 (3)	3	0	1	0
Depends on clinical characteristics	14 (10)	6	1	6	1
No definitive recommendation	62 (44)	16	14	19	13
Not guideline derived	6 (4)	1	3	2	0