HHS Public Access

Author manuscript

J Registry Manag. Author manuscript; available in PMC 2015 July 06.

Published in final edited form as:

J Registry Manag. 2012 ; 39(3): 121–132.

Evaluation of Primary/Preferred Language Data Collection

Linh M. Duong, MPH[a], **Simple D. Singh, MD, MPH**[a], **Natasha Buchanan, PhD**[b], **Joan L Phillips, CTR**[a], and **Ken Cerlach, MPH, CTR**[a]

[a]Cancer Surveillance Branch, Division of Cancer Prevention and Control, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, GA

[b]Epidemiology and Applied Research Branch, Division of Cancer Prevention and Control, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, GA

Abstract

A literature review was conducted to identify peer-reviewed articles related to primary/preferred language and interpreter-use data collection practices in hospitals, clinics, and outpatient settings to assess its completeness and quality. In January 2011, Embase (Ovid), MEDLINE (Ovid), PubMed, and Web of Science databases were searched for eligible studies. Primary and secondary inclusion criteria were applied to selected eligible articles. This extensive literature search yielded 768 articles after duplicates were removed. After primary and secondary inclusion criteria were applied, 28 eligible articles remained for data abstraction. All 28 articles in this review reported collecting primary/preferred language data, but only 18% (5/28) collected information on interpreter use. This review revealed that there remains variability in the way that primary/preferred language and interpreter use data are collected; all studies used various methodologies for evaluating and abstracting these data. Likewise, the sources from which the data were abstracted differed.

Keywords

data collection; interpreter use; medical records; primary language; preferred language

Introduction

Vulnerable populations – including racial/ethnic minorities, older adults, and those with low income–are at risk for poorer health and adverse health communication outcomes when they have low health literacy,[1-3] ″the inability to obtain, process, and understand health information to make appropriate decisions.″[1] Studies have demonstrated that low literacy and low health literacy are associated with impaired patient-provider communication, patient

Address correspondence to Linh M. Duong, MPH, Centers for Disease Control and Prevention, 4770 Buford Highway NE, Mail Stop F-69, Atlanta, GA 30341. Telephone: (770) 488-3122. Fax: (770) 488-4759. lduong@cdc.gov..

These findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

non-adherence, increased hospitalization, and poorer health.[4-7] Similarly, research examining the effects of patient-provider language discordance on the quality of care found that language barriers are associated with less health education, worse interpersonal care, and lower patient satisfaction.[8] Access to a translator may facilitate transmission of health education, but having an interpreter present does not serve as a substitute for language concordance between patient and provider.[8] To accurately quantify health disparities due to low health literacy level, a standardized measure would be helpful for primary/preferred language data collection practices. The need for a standardized measure becomes increasingly important as the United States grows more linguistically diverse.

According to a 2010 Census Bureau report, the United States is becoming more linguistically diverse.[9] The number of people 5 years of age and older who speak a language other than English at home has more than doubled in the last 3 decades, a growth rate that is 4 times greater than that of the overall US population.[9] Within this time frame, the number of speakers of non-English languages grew by 140% while the overall US population grew by 34%,[9] highlighting the importance for culturally diverse health care in medical facilities and practices.

To address the health-care needs of this growing population in the United States, a better understanding is needed on how information on language spoken or primary/preferred language is being collected and utilized currently. In the past, data collection practices in health care facilities on language and communication have been limited by systems which were incomplete and used incongruously.[7-10] However, with the growing use of electronic health records (EHR), there may be the potential to track and maintain information which is currently difficult to collect in a standardized manner. In addition, the use of an electronic health system has the potential to improve the quality and completeness of these data collection methods. Recently, standards have been developed for certification of EHR technology by the US Department of Health and Human Services (DHHS); standardization will promote the systematic collection of health data to inform care.[11,12]

While much remains to be done in the development of a standardized set of criteria to evaluate language as related to disease outcomes, the implementation of EHR is a step toward building a foundation for primary/preferred language data collection. Our literature review evaluates current practices on primary/preferred language data collection in hospital medical records and includes an assessment of data collection on interpreter use.

Methods

The objectives of this literature review were to address the following questions: (1) Are primary/preferred language data being collected? (2) If so, how is primary/preferred language information being captured and collected? (3) What sources are collecting primary/preferred language data? (4) Is interpreter-use data being collected? (5) If so, how is interpreter-use information being captured and collected? (6) Where is interpreter-use data being collected?

This report defines completeness and interpreter use in the following manner: Completeness of primary language data is defined as data in which the majority (greater than 80%) of the language data is provided (ie, is not missing or unknown). Interpreter use data is defined broadly in this report as any data which records use of interpreter services, data which reports patient lack of English-proficiency requiring the aid of an interpreter to understand the physician, and if data was recorded on language-appropriate written material being provided to the patient.

Data Sources

A literature review was conducted to identify peer-reviewed articles related to primary/preferred language collection practices in hospitals, clinics, and outpatient settings as well as an assessment of data collection on interpreter use. Initially, our literature review (primary search) focused on studies related to cancer/neoplasms and primary/preferred language data collection practices. To increase our sample size, the search was subsequently expanded (secondary search) to include studies of all other diseases and primary/preferred language data collection practices. The primary search was conducted on January 6, 2011 using Embase (Ovid) and MEDLINE (Ovid) for the years 1988-2010. The secondary search was conducted on January 27, 2011 using PubMed and Web of Science to find eligible studies and restricted the search to articles published in the last 5 years. Search terms used in both searches can be found in Table 1. We excluded the following key words: programming language, ontology, language publication, language restriction, language articles, and natural language processing as these terms were not relevant to identifying spoken or vernacular languages of study subjects and as a result did not meet inclusion criteria. The results of both searches were then combined into a single library in EndNote and checked for duplicates, which were subsequently removed.

Study Selection and Data Abstraction

We reviewed titles, abstracts, and full texts of the identified citations and selected eligible articles based on prespecified criteria described below. The selection of eligible articles for data abstraction was based on a 2-phased approach. In the preliminary phase, one coauthor reviewed eligible articles that included primary/preferred language as an outcome variable of interest in either the title or abstract, and determined whether the quality of the data was assessed. Full texts of eligible articles were obtained for the second phase review. For the second phase, studies were selected if they included primary/preferred language data collection in medical records or used other data sources (survey, interview, US Census data, Medicaid data, health-plan data, etc) to obtain primary/preferred language information. Initially, 2 of the coauthors conducted the secondary phase independently. After each reviewer compiled a list of eligible articles, the 2 lists were evaluated by all authors. Then, a final list of eligible articles was compiled for data abstraction and determination of topic area for the tables. The following information was collected from each eligible article: first author name and publication year, methodology used (data source and primary/preferred language variable), and key findings of the study.

Results

This literature search (Figure 1) yielded 768 articles after duplicates were removed. After primary and secondary inclusion criteria were applied, there were 28 eligible articles which remained for data abstraction. These 28 articles were then divided between Table 2 and Table 3 according to data sources from which the primary/preferred language information was obtained. All 28 articles in this literature review reported collecting primary/preferred language data.

Table 2 includes 10 articles that are related to primary/preferred language data collected from medical records. Availability of information on primary/preferred language was quite high. Approximately 60% (6/10) of studies reported information on primary/preferred language.[13-18] Furthermore, among these studies, data for primary/preferred language had a high level of completeness ranging from 82% to 96%.[13,18] McClure et al reported that overall information on primary/preferred language was available in medical records for 86.4% of study participants.[13] Of 27 facilities that did not have primary/preferred language data available to abstract electronically, 81.6% had that data in the medical records, most often in the admission records.[13] Polednak found that about 8.4% (64/765) of cases had an unknown preferred language.[14] Similarly, in another study he noted that records for all but 9.7% of 992 Hispanics or Asian patients had information on language preference and that this information was missing more often in Asian than in Hispanic-American patient records.[15] Additionally, a third study by Polednak found that primary/preferred language was not recorded in 8.4% of records abstracted.[16] Solberg et al reported that primary/preferred language data was missing for 6,972 (3.7%) of cases in their study sample.[17] In another study, Solberg et al found that primary/preferred language data was 96% complete.[18] In regards to the collection of interpreter use information, as presented in Table 2, 67% (2/3) reported data completeness information (64% and 93%; respectively)[15-19] and 33% (1/3) reported where interpreter use data were collected (consent forms and nurses notes).[13]

Table 3 includes 18 articles that used other data sources (eg, surveys, interviews, US Census data, Medicaid data, health-plan data) to obtain primary/preferred language information. Only 6% (1/18) provided data on the availability of primary/preferred language and this study did not specify a percentage missing; rather, researchers just stated that preferred language was usually recorded.[20] There were 11% (2/18) of studies that collected interpreter use information; however, neither study reported the completeness of data on the use of an interpreter.[21,22]

Overall, 25% (7/28) of studies[13-18,20] reported completeness of information on primary/preferred language while 18% (5/28) of studies[13,15,19,21,22] reported the completeness of data on the use of an interpreter (Table 2, Table 3).

A number of studies shown in Table 3 combined primary/preferred language variables obtained from surveys with data from medical records.[20,22-29] Others combined primary/preferred language data obtained from interviews with data from medical records,[21,30-34] or from the US Census,[35] or from Medicaid data,[36] or from the health plan.[37] In a study by

Polednak, primary/preferred language data (having a follow up physician with a Spanish-language practice) was obtained from a physician profile survey.[28] However, data on primary/preferred language of each Hispanic patient were not available in this study.[28]

Collection of preferred language data in medical records, surveys, interviews, US Census data, Medicaid data, or in health plan data was not limited to a single disease (Table 2, Table 3). In fact, it appears that physicians from a range of specialties recorded information on primary/preferred language including physicians who treat cancer,[13-16,20,22-25,27-32,34-36,38] stroke,[33] mental health disorders,[16] sexually transmitted diseases,[39] diabetes,[14,26,37,40] nutrition,[38] and tobacco use.[17-19] Thus, primary/preferred language affects many diverse specialties. This is further supported by a study by Hasnain-Wynia et al in which 20 practices nationwide, each with 5 or fewer physicians, were interviewed, found that primary/preferred language data was collected across several disciplines.[21] These practices represented a diverse set of specialties including internal medicine, obstetrics and gynecology, family medicine, pediatrics, geriatrics, and pulmonary medicine.[21] Results of this study found that of the 20 practices interviewed, 9 reported collecting demographic data (eg, race, ethnicity, and/or primary language).[21]

Primary/preferred language variables differed for all studies (Table 2, Table 3). Some studies had a single primary/preferred language variable.[13-19,25,29-31,33,36,38] Other studies had 2 or more primary/preferred language variables.[20,21,24,26,34,35,37,39] And, others still developed their own intrinsic scale used to measure primary/preferred language[22,32] or used a scale already developed.[23] Examples of some of the single primary/preferred language variables used include the following variable categories: English or non-English,[13] patient′s preferred language,[19] preferred language (English, Other, and Unknown),[14] preferred/primary language (English, Spanish, Bilingual, Asian, Inconsistent, and Other, Unknown),[38] English or Other (Spanish, bilingual, inconsistent),[15] primary language (Spanish/bilingual vs English),[16] preferred language (English, Other, or No Data),[17] language type (English-speaking or Non-English-speaking),[40] and preferred language (English, Half-English/Half Not English, Not English).[31] Likewise, studies with 2 or more primary/preferred language variables also had wide-ranging categories. Examples include a study by Gindi et al which had 2 primary/preferred language variables: language spoken (English-speaking or Spanish-speaking) and language status (Latino English-Proficient, Latino Spanish-Speaking, or Non-Latino)[39] and a study by Hawley et al which had the following primary/preferred language variables: race/ethnicity (Latina-Spanish speaking, Latirta-English speaking, African American, Caucasian) and health literacy (low, moderate, high).[24] Studies that developed their own primary/preferred language variable include a study by John et al that created an acculturation index based on language usage and generational status[32] and a study by Johnson-Kozlow which developed an acculturation scale with scores that were composed of 7 status variables that measured English language use and proficiency, nativity, citizenship, and years living in the United States.[22] Lastly, Hamilton et al used a scale already developed called the Short Acculturation Scale for Hispanics to assess primary/preferred language.[23]

Discussion

Our literature review indicates that although the completeness of primary/preferred language data collection is high, 96% completeness for primary/preferred language data collection in 1 study,[18] there remains variability in the way this information is collected. For example, investigators used different protocols for evaluating the collection of primary/preferred language and the sources used to collect this information also varied. In addition to using hospital medical records to obtain primary/preferred language information,[13-19,38-40] investigators used surveys,[20,22-29] interviews,[21,30-34] US Census data,[35] Medicaid data,[36] or health plan data.[37] This information was then combined with the medical records to assess disparities in health outcomes based on primary/preferred language concordance. These findings show that primary/preferred language data collection occurs in multiple ways within various settings. Moreover, the collection of data from various sources reduces the ability to make comparisons across studies as well as limits the possibility of aggregating primary/preferred language data study results to obtain a global review of health outcomes. Similarly, the lack of a common definition and standard codes may impede research efforts. Therefore, a standardization of primary/preferred language collection practices may be warranted.

Furthermore, the collection of interpreter service data differed between studies. Of the studies that reported collecting primary/preferred language data, only 18% (5/28) collected information on interpreter use.[13,15,19,21,22] These included a study by McClure et al that stated interpreter use information was found in consent forms and nurses' notes[13] while Polednak reported that information on interpreter use was missing for 36.1% of 653 probable non-English-preferring patients.[15] And, even when interpreter use data are collected, it is unclear how this information is used to improve services for these patients.

Our findings regarding the variability in the collection of primary/preferred language data is similar to what was found in a 2007 Joint Commission on the Accreditation of Healthcare Organizations (JCAHO) report.[10] In 2006, JCAHO required the maintenance of records on patients "language and communication needs."[41] These standards are intended to support the provision of care, treatment, and services in a manner that is conducive to cultural, language, literacy, and learning needs of individuals [41] For example, these provisions include standards for respecting values and beliefs of the patient, appropriate communication, including interpreter and translation services, effective communication throughout organization, ensuring that orientation and ongoing staff education is appropriate to the needs of patient population, and the collection of data, documentation of needs and access to data.[41] However, a review in 2007 of these records reveals that there still remains much work in improving this system as noted by an evaluation of 60 representative US hospitals.[10] In this review, JCAHO found that systems for collection of required data on language and communication were "underdeveloped" and used "inconsistently."[10] However, this may change as a result of implementation of the certification criteria for EHR technology.

In 2010, the DHHS issued a final rule to complete the adoption of an initial set of standards, implementation specifications, and certification criteria for EHR technology.[11,12] Stage 1

criteria for EHR certification states the minimum elements required in support of meaningful use by eligible professionals, eligible hospitals, and/or critical access hospitals under the Medicare and Medicaid EHR Incentive Programs.[11,12] Specifically, the collection of demographic data includes a record of preferred language and demonstration of meaningful use of this technology.[11,12] However, a data collection field in the EHR does not necessarily indicate that the data will be collected by physicians. Providing an appropriate incentive may be needed to assist physicians in collecting these data.

In addition, a report by the Institute of Medicine (IOM) stated that data on a person's language and communication needs should be a part of any minimum data set related to health care delivery and quality improvement.[6] The IOM subcommittee for this report recommended identifying spoken language need in a stepwise approach: first by determining how well the individual believes he/she speaks English and second by asking what language he/she needs for a health-related encounter.[6] This will allow for improved quality of services in subsequent encounters, in analysis of health disparities, and in system-level planning (determining the needs for interpreters and matching patients to language-concordant providers).[6] A study by Karliner et al, which adds support to this recommendation, found that a screening question asking how well a patient speaks English followed by language preference for medical care was most inclusive and accurate for identifying patients likely to benefit from language assistance.[7] Certification of EHR did not include recommendations from this IOM report.[11,12]

Policies and initiatives to strengthen health literacy across the nation have also begun to take root. For example, Healthy People 2020 has begun incorporating objectives to improve health literacy and provider communication. Specifically, these 2 objectives seek to "improve the health literacy of the population (HC/HIT-1)" and to "increase the proportion of persons who report that their health care providers have satisfactory communication skills (HC/HIT-2)".[42] The overall goals are to improve health outcomes, health care quality, and achieve health equity through the use of health communication strategies and health information technology.[42]

Similarly, recent federal policy initiatives have begun to bring health literacy to the forefront of the health care discussion, including the Affordable Care Act of 2010, the DHHS' National Action Plan to Improve Health Literacy, and the Plain Writing Act of 2010.[3] The Affordable Care Act addresses health literacy by integrating training on health literacy for health professionals (section 5301) and requiring that health plans and insurers provide consumers with a summary of health information, benefits, and coverage options that is clear and consistent and that can be compared to other plans (section 2715).[3] The National Action Plan to Improve Health Literacy provides a consolidated structure with which to unite health literacy goals and strategies for the nation.[3] And, the Plain Writing Act of 2010 specifies that federal agency documents must be written clearly so the public will be able to understand them.[3]

As uniform data collection becomes the norm, the strengthening of health information technology across the nation may have direct implications for medical records. Medical records at the hospital level may benefit from the development of standardized protocols for

primary/preferred language data collection practices. These protocols may improve the ability of hospitals to help address patients' linguistic needs and support studies related to reducing health disparities. The standardization of primary/preferred language data would allow an accurate assessment of differences across hospitals, clinics, and outpatient settings.

The ultimate goal of primary/preferred language concordance between patient and provider is providing access to health services that are required for effective treatment, especially for patients with complex illnesses such as cancer. Such patients commonly require access to multiple specialists, effective coordination of care, accurate information about disease and treatment options, and timely attention to symptoms.[43]

The need for high-quality data that are complete and accurate is not unique to primary/preferred language data collection.[44] In fact, obtaining data on variables such as race/ethnicity,[44-46] socioeconomic status,[47] and stage at diagnosis[48,49] also have been difficult. As a result, in addition to the standardization of data collection practices, continuous quality-control activities are also needed to identify and correct errors and to ensure uniformity and accuracy of the data collected.

Our findings should be considered in light of several limitations. Although we employed a thorough and extensive search strategy and literature review, some studies may not have been identified and included in this review. In particular, since we focused on peer-reviewed publications, we did not examine unpublished documents or reports on this topic. In addition, due to variability in the manner that the primary/preferred language data were collected across the studies included in this review, we were not able to aggregate studies for meta-analysis. In spite of these limitations, this literature review is among the first assessments, to our knowledge, to examine primary/preferred language and interpreter use data collection practices in hospital medical records, to explore the completeness of these data, and to identify areas in need of improvement.

Conclusions

As the United States moves toward improving the health literacy of its population by strengthening provider communication through health information technology and by passing federal initiatives and policies to support these goals, a more uniform protocol may emerge for collecting information on primary/preferred language. This is especially important in light of the Healthy People 2020 objectives to improve population outcomes related to health-care quality and health equity through the use of health information technology. The development of a standardized protocol to collect data on primary/preferred language may improve research methods used to analyze health disparities related to language spoken and utilization of interpreter services which has the potential to impact disease outcomes. The ability to describe areas in which resources are lacking for vulnerable populations unable to access health care due to patient-provider discordance in language may aid in creating public health interventions targeted at improving and increasing needed resources at facilities which serve a diverse population.

Acknowledgements

The authors would like to acknowledge Onnalee Gomez and Katherine Tucker, CDC Library personnel, who conducted the literature review search and consolidation of articles for this paper. In addition, we would like to thank Qiang Ling, Carissa Holmes, Mridhula (Maya) Kumar, Megan Crawley, and Kate Allen for their comments and suggestions in the development of this paper.

References

1. Garcia S, Hahn E, Jacobs EA. Addressing low literacy and health literacy in clinical oncology practice. J Supportive Oncol. 2010; 8(2):64–69.
2. Zun L, Sadoun T, Downey L. English-language competency of self-declared English-speaking Hispanic patients using written tests of health literacy. J Natl Med Assoc. 2006; 98(6):912–917. [PubMed: 16775913]
3. Koh H, Berwick D, Clancy C, et al. New federal policy initiatives to boost health literacy can help the nation move beyond the cycle of costly 'crisis care'. Health Aff. (Millwood). 2012; 31(2):434–443. [PubMed: 22262723]
4. Berkman, ND.; DeWalt, DA.; Pignone, MP., et al. Literacy and Health Outcomes. Evidence Report/Technology Assessment no. 87. Agency for Healthcare Research and Quality; Rockville, MD: 2004. AHRQ publication 04-E007-2
5. Hahn EA, Cella D. Health outcomes assessment in vulnerable populations: measurement challenges and recommendations. Arch Phys Med Rehabil. 2003; 84(suppl 2):S35–S42. [PubMed: 12692770]
6. IOM (Institute of Medicine). Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement. The National Academies Press; Washington, DC: 2009.
7. Karliner LS, Napoles-Springer AM, Schillinger D, Bibbins-Domingo K, Perez-Stable EJ. Identification of limited English proficient patients in clinical care. J Gen Intern Med. 2008; 23(10): 1555–1560. [PubMed: 18618200]
8. Ngo-Metzger Q, Sorkin DH, Phillips RS, et al. Providing high-quality care for limited English proficient patients: the importance of language concordance and interpreter use. J Gen Intern Med. 2007; 22(suppl 2):324–330. [PubMed: 17957419]
9. US Census Bureau. [Accessed July 13, 2010] New Census Bureau Report Analyzes Nation's Linguistic Diversity. Available at: http://www.census.gov/newsroom/releases/archives/american_community_survey_acs/cb10-cn58.html
10. Wilson-Stronks, A.; Calvez, E. [Accessed December 20, 2011] Hospitals, Language, and Culture: A Snapshot of the Nation: The Joint Commission and The California Endowment. 2007. Available at: http://www.jointcommission.org/assets/1/6/hlc_paper.pdf
11. The Office of the National Coordinator for Health Information Technology. Department of Health and Human Services. Federal Register, Health Information Technology. [Accessed 31 August 2011] Initial Set of Standards, Implementation Specifications, and Certification Criteria for Electronic Health Record Technology; Final Rule. Available at: http://edocket.access.gpo.gov/2010/pdf/2010-17210.pdf
12. The Office of the National Coordinator for Health Information Technology. Department of Health and Human Services. [Accessed 31 August 2011] Standards & Certification Criteria; Final Rule. Available at: http://healthit.hhs.gov/portal/server.pt/community/healthit_hhs_gov__standards_ifr/1195
13. McClure LA, Claser SL, Shema SJ, et al. Availability and accuracy of medical record information on language usage of cancer patients from a multi-ethnic population. J Immigrant Minority Health. 2010; 12(4):480–488.
14. Polednak AP. Prevalence and predictors of comorbid diabetes among newly diagnosed Hispanic cancer patients in Connecticut. Cancer Detect Prev. 2007; 31(6):453–456. [PubMed: 18061370]
15. Polednak AP. Obtaining information on language preference among newly diagnosed Hispanic and Asian American cancer patients in Connecticut. J Registry Manage. 2009; 36(3):77–82.
16. Polednak AP. Comorbid mental disorders in hospital records of Hispanic patients diagnosed with cancer in Connecticut. J Registry Manage. 2009; 36(4):111–116.

17. Solberg LI, Flottemesch TJ, Foldes SS, Molitor BA, Walker PF, Crain AL. Tobacco-use prevalence in special populations taking advantage of electronic medical records. Am J Prev Med. 2008; 35(6S):S501–S507. [PubMed: 19012845]
18. Solberg LI, Parker ED, Foldes SS, Walker PF. Disparities in tobacco cessation medication orders and fills among special populations. Nicotine Tob Res. 2010; 12(2):144–151. [PubMed: 20018945]
19. Parker ED, Solberg LI, Foldes SS, Walker PF. A surveillance source of tobacco use differences among immigrant populations. Nicotine Tob Res. 2010; 12(3):309–314. [PubMed: 20083645]
20. Polednak AP. Collecting information on race, Hispanic ethnicity, and birthplace of cancer patients: policies and practices in Connecticut hospitals. Ethn Dis. 2005; 15(1):90–96. [PubMed: 15720054]
21. Hasnain-Wynia R, Van Dyke K, Youdelman M, et al. Barriers to collecting patient race, ethnicity, and primary language data in physician practices: an exploratory study. J Natl Med Assoc. 2010; 102(9):769–775. [PubMed: 20922920]
22. Johnson-Kozlow M. Colorectal cancer screening of Californian adults of Mexican origin as a function of acculturation. J Immigrant Minority Health. 2010; 12(4):454–461.
23. Hamilton AS, Hofer TP, Hawley ST, et al. Latinas and breast cancer outcomes: population-based sampling, ethnic identity, and acculturation assessment. Cancer Epidemiol Biomarkers Prev. 2009; 18(7):2022–2029. [PubMed: 19549806]
24. Hawley ST, Janz NK, Hamilton A, et al. Latina patient perspectives about informed treatment decision making for breast cancer. Patient Educ Couns. 2008; 73(2):363–370. [PubMed: 18786799]
25. Kaplan CP, Napoles AM, Hwang ES, et al. Selection of treatment among Latina and non-Latina white women with ductal carcinoma in situ. J Womens Health. 2011; 20(2):215–223.
26. Karter AJ, Ferrara A, Darbinian JA, Ackerson LM, Selby JV. Self-monitoring of blood glucose: language and financial barriers in a managed care population with diabetes. Diabetes Care. 2000; 23(4):477–483. [PubMed: 10857938]
27. Napoles-Springer AM, Ortiz C, O'Brien H, Diaz-Mendez M, Perez-Stable EJ. Use of cancer support groups among Latina breast cancer survivors. J Cancer Survivorship. 2007; 1(3):193–204.
28. Polednak AP. Identifying newly diagnosed Hispanic cancer patients who use a physician with a Spanish-language practice, for studies of quality of cancer treatment. Cancer Detect Prev. 2007; 31(3):185–190. [PubMed: 17706369]
29. Yoon J, Malin JL, Tao ML, et al. Symptoms after breast cancer treatment: are they influenced by patient characteristics? Breast Cancer Res Treat. 2008; 108(2):153–165. [PubMed: 17492377]
30. Gomez SL, Glaser SL. Quality of cancer registry birthplace data for Hispanics living in the United States. Cancer Causes Control. 2005; 16(6):713–723. [PubMed: 16049810]
31. Gomez SL, Glaser SL, Kelsey JL, Lee MM. Bias in completeness of birthplace data for Asian groups in a population-based cancer registry (United States). Cancer Causes Control. 2004; 15(3): 243–253. [PubMed: 15090719]
32. John EM, Phipps AI, Davis A, Koo J. Migration history, acculturation, and breast cancer risk in Hispanic women. Cancer Epidemiol Biomarkers Prev. 2005; 14(12):2905–2913. [PubMed: 16365008]
33. Smith MA, Lisabeth LD, Bonikowski F, Morgenstern LB. The role of ethnicity, sex, and language on delay to hospital arrival for acute ischemic stroke. Stroke. 2010; 41(5):905–909. [PubMed: 20339124]
34. Sweeney C, Edwards SL, Baumgartner KB, et al. Recruiting Hispanic women for a population-based study: validity of surname search and characteristics of nonparticipants. Am J Epidemiol. 2007; 166(10):1210–1219. [PubMed: 17827445]
35. Kouri EM, He Y, Winer EP, Keating NL. Influence of birthplace on breast cancer diagnosis and treatment for Hispanic women. Breast Cancer Res Treat. 2010; 121(3):743–751. [PubMed: 19949856]
36. Ramsey SD, Zeliadt SB, Richardson LC, et al. Discontinuation of radiation treatment among Medicaid-enrolled women with local and regional stage breast cancer. Breast J. 2010; 16(1):20–27. [PubMed: 19929888]

37. Traylor AH, Schmittdiel JA, Uratsu CS, Mangione CM, Subramanian U. The predictors of patient-physician race and ethnic concordance: a medical facility fixed-effects approach. Health Serv Res. 2010; 45(3):792–805. [PubMed: 20337734]

38. Polednak AP. Indicators of nutritional screening in hospital records of newly diagnosed Hispanic and Asian-American adult cancer patients in Connecticut. Nutrition. 2008; 24(10):1053–1056. [PubMed: 18562169]

39. Gindi RM, Erbelding EJ, Page KR. Sexually transmitted infection prevalence and behavioral risk factors among Latino and non-Latino patients attending the Baltimore City STD clinics. Sex Transm Dis. 2010; 37(3):191–196. [PubMed: 19910863]

40. Tocher TM, Larson E. Quality of diabetes care for non-English-speaking patients. A comparative study. West J Med. 1998; 168(6):504–511. [PubMed: 9655991]

41. JCAHO. Joint commission 2006 Requirements Related to the Provision of Culturally and Linguistically Appropriate Health Care. Joint Commission on Accreditation of Healthcare Organizations; 2006.

42. Healthy People. [Accessed March 6, 2012] Healthy People 2020 Topics and Objectives, Health Communication and Technology. Available at: http://healthypeople.gov/2020/topicsobjectives2020/overview.aspx?topicld=18

43. Ayanian JZ, Zaslavsky AM, Guadagnoli E, et al. Patients' perceptions of quality of care for colorectal cancer by race, ethnicity, and language. J Clin Oncol. 2005; 23(27):6576–6586. [PubMed: 16116149]

44. Agency for Healthcare Research and Quality, Race, Ethnicity, and Language Data. [Accessed September 1, 2011] Standardization for Health Care Quality Improvement: Improving Data Collection Across the Health Care System. Available at: http://www.ahrq.gov/research/iomracereport/reldata5.htm

45. Ford ME, Kelly PA. Conceptualizing and categorizing race and ethnicity in health services research. Health Serv Res. 2005; 40(5 Pt 2):1658–1675. [PubMed: 16179001]

46. Wallman KK, Evinger S, Schechter S. Measuring our nation's diversity: developing a common language for data on race/ethnicity. Am J Public Health. 2000; 90(11):1704–1708. [PubMed: 11076235]

47. Krieger N. Overcoming the absence of socioeconomic data in medical records: validation and application of a census-based methodology. Am J Public Health. 1992; 82(5):703–710. [PubMed: 1566949]

48. Duong LM, Ajani UA, Wilson RJ. Comparative evaluation of uterine cancer staging data using two different staging systems, 2001-2005. J Registry Manage. 2009; 36(4):125–129.

49. Klassen AC, Curriero F, Kulldorff M, Alberg AJ, Platz EA, Neloms ST. Missing stage and grade in Maryland prostate cancer surveillance data, 1992-1997. Am J Prev Med. 2006; 30(suppl 2):S77–87. [PubMed: 16458794]

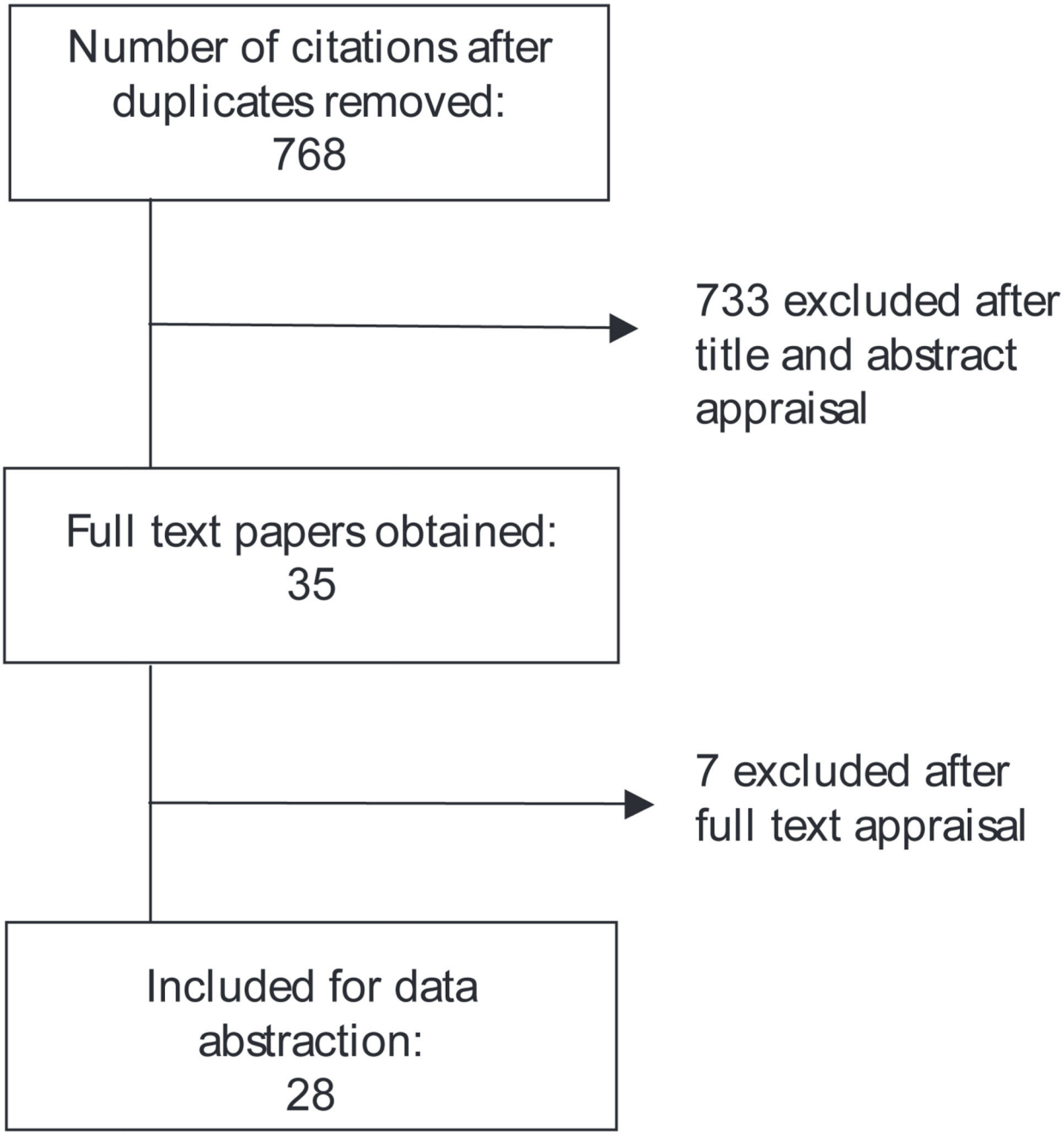

Figure 1. Results of Literature Search

Table 1

Databases and Search Terms Used for Literature Review[†]

Database	*Search terms*
Primary search—cancer/neoplasm only	
Searched on January 6, 2011 **Embase (Ovid)** Limits: published between 1988 and 2010 **MEDLINE (Ovid)** Limits: published between 1988 and 2010	Language.hw,kw,sh,ti. **AND** (registry or registries).hw,kw,sh,ti. **AND** Cancer.hw,kw,sh,ti. or Neoplasm"*".hw,kw,sh,ti. hw = heading word kw = key word sh = subject headings ti = title
Secondary search—no focus on a particular disease	
Searched on January 27,2011 **PubMed** Limits: published in the last 5 years	Registry or registries or registries [MeSH][‡] or "Electronic Health Records" [MeSH] or "electronic medical record"*"[§][‖] or "hospital record"*" or "reporting hospital"*" **AND** Language[MeSH] or "language data" or "primary language" or "preferred language" or "native tongue" or "native language" or "language spoken" or "language proficiency" or "preferred language" or "primary spoken language" or "language proficiency" or "language proficient" or "Spanish speaking" or "native speaker" or "non-English" or "non-native speaker" or "language codes" or "linguistically" or "language barrier"*" or "languages spoken" or "language concordance" or "language concordant" or "translation services" or "collection of language" or "language data" or "linguistic" or "patient language" or "language of patient" or multilingualism[MeSH] or bilingual or bilingualism[MeSH] or multilingualism
Secondary search—no focus on a particular disease	
Searched on January 27, 2011 **Web of Science** Limits: published in the last 5 years	Registry or registries or "Electronic Health Records" or "electronic medical record"*" or "hospital record"*" or "reporting hospital" or "reporting hospitals" **AND** "Language"

Database	*Search terms*
	or “language data” or “primary language” or “preferred language” or “native tongue” or “native language” or “language spoken” or “language proficiency” or “preferred language” or “primary spoken language” or “language proficiency” or “language proficient” or “Spanish speaking” or “native speaker” or “non-English” or “non-native speaker” or “language codes” or “linguistically” or “language barrier″*″” or “languages spoken” or “language concordance” or “language concordant” or “translation services” or “collection of language” or “language data” or “linguistic” or “patient language” or “language of patient” or “multilingualism” or “bilingual” or “bilingualism” or “multilingualism”

† Excludes the following search terms: programming languages, ontology, language of publication, language restriction, language articles, and natural language processing as these terms were not relevant to identifying spoken or vernacular languages of study subjects and as a result did not meet inclusion criteria.

‡ MeSH; Medical Subject Headings used by National Library of Medicine for indexing, cataloging, and searching for biomedical and health-related information and documents.

§ Quotation marks surround words that were searched as a phrase.

// =wildcard truncation for plurals.

″*″ =wildcard truncation for plurals.

Table 2
Studies Related to Language Data Collection in Medical Records

First author and publication year	*Methodology*	*Key findings*
Gindi et al, 2010[39]	Data Source: EMR abstraction from Baltimore City public STD clinic. Primary/preferred language variable: Language spoken (English-speaking or Spanish-speaking) Language status (Latino English-Proficient, Latino Spanish-Speaking, or Non-Latino)	2% (39,728) patients were Latinos. More than half of Latino patients were Spanish Speaking (60%). This differed by gender and age group.
McClure et al, 2010[13]	Data Source: MR from Greater Bay Area Cancer Registry and California Cancer Registry. Primary/preferred language variable: English or non-English	Overall, information on spoken language was available in MR for 86.4% of study participants. For 27 facilities for which language data was not abstracted electronically: 81.6% had language information in MR, most often in admission records. Information on interpreter use was collected in consent forms and nurses' notes. Significant differences by race, year of diagnosis, and advanced stage were found.
Parker et al, 2010[19]	Data Source: EMR Minnesota multispecialty care delivery organization. Primary/preferred language variable: Patient's Preferred Language	Language interpreter data exist for 93% of this sample, and country of origin data exist for 53%. Total number of patients is not provided for language interpreter data or country of origin data.
Polednak, 2007a[14]	Data Source: MR from Connecticut population-based registry. Primary/preferred language variable: Preferred Language (English, Other, Unknown)	Prevalence of comorbid diabetes was 25.1 % (192/765). About 15.1 % of 166 preferred English vs 30.3% of 535 who preferred a non-English language (predominantly Spanish). About 8.4% (64/765) cases had an unknown preferred language.
Polednak, 2008[38]	Data Source: MR from Connecticut population-based registry. Primary/preferred language variable: Preferred/primary language (English, Spanish, Bilingual, Asian, Inconsistent, and Other, Unknown)	Recent weight loss was mentioned for only 21.5% and was less frequent (12.7%) among 237 preferring English vs 28.2% of 418 preferring Spanish and 28.6% preferring an Asian language; the association with language category persisted when other variables were considered. No indication of completeness on language data provided.
Polednak, 2009a[15]	Data Source: MR from Connecticut population-based registry. Primary/preferred language variable: English or Other (Spanish, bilingual, inconsistent)	Only 9.7% of 992 Hispanic or Asian patients had no information on language preference. Information on use of an interpreter was missing for 36.1% of 653 probable non-English-preferring patients. Missing information was more frequent in Asian than Hispanic American patients.
Polednak, 2009b[16]	Data Source: MR from Connecticut population-based registry. Primary/preferred language variable: Primary language (Spanish/bilingual vs English)	Prevalence of a comorbid mental disorder declined with age but did not differ by primary language (Spanish/bilingual vs English). Primary/preferred language was not recorded in 8.4% of records abstracted. Total number of patients is not provided for primary language.
Solberg et al, 2008[17]	Data Source: EMR from Minnesota HealthRartners Medical Group (HPMG) multi-specialty care delivery organization. Primary/preferred language variable: Preferred language (English, Other, or No Data)	Overall, 19.7% with recorded status were tobacco users as were 8.5% of those whose preferred language was other than English. Language data was missing for 6,972 (3.7%). Underreporting of tobacco status appears to be correlated with the absence of other data such as insurance information (84.4%), ethnicity (80%), and preferred language (73.6%).
Solberg et al, 2010[18]	Data Source: EMR from Minnesota HealthPartners Medical Group (HPMG) multi-specialty care delivery organization. Primary/preferred language variable: Language preference (English, Spanish, or Other)	Groups receiving fewer [tobacco cessation prescription] orders than their comparison groups included those with non-English preference. The same groups were less likely to fill that prescription, except patients with non-

First author and publication year	*Methodology*	*Key findings*
		English preference or Medicaid. Language data was 96% complete. Total number of patients is not provided for primary language.
Tocher et al, 1998[40]	Data Source: EMR from clinical and administrative databases at the University of Washington Medical Center and Haborview Medical Center. Primary/preferred language variable: Language type (English-speaking or Non-English-speaking)	At these institutions, the quality of diabetes care for non-English speaking patients appears as good as, if not better than, for English-speaking patients. Physicians may be achieving these results through more frequent visits and laboratory testing. No indication of completeness on primary language.

EMR=electronic medical record. MR=medical records. STD=sexually transmitted disease.

Table 3

Use of Other Data Sources (Survey, Interview, US Census, Medicaid, Health Plan) to Get Language Info

First author and publication year	*Methodology*	*Key findings*
Gomez et al, 2004[31]	Data Source: MR from Greater Bay Area Cancer Registry. Language data obtained from interview. Primary/preferred language variable: Preferred Language (English, Half English/Half Not English, Not English)	Among US-born Asians, those misclassified as foreign-born were more likely than those correctly classified to prefer a non-English primary language. Asian subgroups varied by preferred language. The multiple-race Asian group was most likely to prefer to use English (79%), followed by Japanese (86%), other Asian (61%), and Filipinos (56%), while the majority of Vietnamese (62%) preferred not to use English.
Gomez et al, 2005[30]	Data Source: Greater Bay Area Cancer Registry. Language data obtained from interview. Primary/preferred language variable: Preferred language (English, Half English/Half Not English, Not English, or Not Asked/Refused)	About 40% preferred to use English as a primary language, and 30% preferred another language. Patients who preferred speaking a language other than English were half as likely to have unrecorded birthplace, although the magnitude of this association was diminished somewhat in the adjusted model.
Hamilton etal, 2009[23]	Data Source: Los Angeles Cancer Surveillance Program. Language data obtained from survey. Primary/preferred language variable: Short Acculturation Scale for Hispanics 5-point scale) - Only English, English better than Spanish, both equally, Spanish better than English, Only Spanish). Respondents answered the following questions using 5-point scale: (1) What language(s) do you read and speak? (2) What language(s) do you usually speak at home? (3) In what language do you usually think? (4) What language do you usually speak with your friends?	Greater than 50% of the self-identified Latinas indicated that they preferred to speak Spanish over English. The Short Acculturation Scale for Hispanics results suggests that those strongly preferring Spanish reported the lowest levels of education, being born in the United States, and having either parent born in the United States.
Hasnain-Wynia et al, 2010[21]	Data Source: Data collected from 20 practices nationwide and were from medical practices with 5 or fewer physicians. Language data obtained from interview. Primary/preferred language variable: Preferred or primary language Use of interpreter	Of the 20 practices surveyed, 9 reported collecting either race, ethnicity, or primary language; 3 collected race/ethnicity and primary language data; 5, only race/ethnicity; and 1, only primary language. Only 1 practice feature facilitated demographic data collection: use of EMR system (7 of 10 practices with an EMR collected data). When patient information on language is collected, it is rarely used to schedule interpreters or to guide the translation of patient materials, even when these services are offered by the practice.
Hawley et al, 2008[24]	Data Source: MR from Los Angeles metropolitan SEER registries data. Language data obtained from survey and merged to SEER data. Primary/preferred language variable: Race/ethnicity (Latina-Spanish speaking, Latina-English speaking, African-American, Caucasian) Health literacy (low, moderate, high) Translation (did not need, family or friend, doctor or staff)	The analytic sample included 877 women: 24.5% Latina-Spanish speaking (Latina-SP), 20.5% Latina-English speaking, 24% African-American and 26.6% Caucasian. Approximately 28% of women in each ethnic group reported a surgeon-based, 36% a shared, and 36% a patient-based surgery decision. Spanish preferent Latina women had the greatest odds of high decision dissatisfaction and regret controlling for other factors. Low health literacy was independently associated with dissatisfaction and regret and slightly attenuated associations between Latina-SP ethnicity and decision outcomes.
John et al, 2005[32]	Data Source: Greater Bay Area Cancer Registry. Language data obtained from interview. Primary/preferred language variable: Acculturation index based on language usage	Among long-term foreign-born residents, breast cancer risk was lower among Hispanics who moved to the United States at age >20 years and those who spoke mostly Spanish.

First author and publication year	*Methodology*	*Key findings*
	and generational status	
Johnson-Kozlow, 2010[22]	Data Source: 2005 California Health Interview Survey (CHIS). Language data obtained from survey. Primary/preferred language variable: Acculturation was score composed of seven status variables that measure English language use and proficiency, nativity and citizenship, and years lived in the US	Approximately 18% said that only English was spoken at home; 5% said they had difficulty understanding their doctor at their last doctor visit. Of those 82% said they had difficulty understanding the doctor due to language and 66% said they needed another person to help them understand the doctor.
Kaplan etal, 2011[25]	Data Source: Eight California Cancer Registry regions and linked to survey data about patient treatment decision making. Language data obtained from survey. Primary/preferred language variable: Ethnicity language group (White women, English-speaking Latinas, or Spanish-speaking Latinas)	English-speaking Latinas (ESL) were more likely to receive radiation than their Spanish-speaking or white counterparts, controlling for demographic and other factors. A greater proportion of white women had a college education compared to ESL and Spanish-speaking Latinas (SSL) women. The majority of white and ESL women were privately insured, but this was not true for SSL women. A larger proportion of white and ESL women reported having a relative with a history of breast cancer compared with SSL women.
Karter et al, 2000[26]	Data Source: Kaiser Permanente Northern California Region health survey. Language data obtained from survey. Primary/preferred language variable: Language measure (prefer to communicate in non-English language) English language difficulty (Yes or No?)	Among Hispanics and Asian/Pacific Islanders, 26 and 30%, respectively, were identified as having difficulties communicating in English or as preferring languages other than English. However, only 1% of non-Hispanic Caucasian and African-American members with diabetes had language difficulties. In most cases, those patients with language difficulties were less likely to practice self-monitoring of blood glucose (SMBG) at recommended levels compared with subjects who were fluent in English.
Kouri et al, 2010[35]	Data Source: SEER population-based data. Language data obtained from US Census. Primary/preferred language variable: Race, Ethnicity, Birthplace	Foreign-born Hispanic women in the United States have a lower probability of being diagnosed at earlier stages of breast cancer and, for women with early-stage disease, of receiving radiation following breast conserving surgery compared to US-born Hispanics and whites. Adjusted rates of stage at breast cancer diagnosis included an adjustment for Spanish language proficiency. Adjusted rates of breast-conserving surgery (BCS) without radiation, BCS with radiation and mastectomy included an adjustment for Spanish language proficiency.
Napoles-Springer et al, 2007[27]	Data Source: Population-based SEER registry. Language data obtained from telephone survey. Primary/preferred language variable: Ethnicity Language of Interview (English or Spanish)	Results suggest that families play an important role in promoting use of support groups among Latina breast cancer survivors, and that spirituality may offer an alternative source of support. More effort should be directed toward providing culturally and linguistically appropriate support services to breast cancer survivors, and increasing awareness of these services among oncologists, patients, and family members.
Polednak, 2005[20]	Data Source: 30 acute care hospitals were surveyed. Language data obtained from survey. Primary/preferred language variable: Race, Ethnicity, Birthplace	At least one staff member at 86% of 28 responding hospitals reported a hospital policy to ask patients about their race, vs 25% for ethnicity and 57% for birthplace, and patient self-reports were reportedly used to obtain race in 100% of hospitals vs 54% for ethnicity. Ethnicity was rarely recorded on any specific type of document, although preferred language was usually recorded.
Polednak, 2007b[28]	Data Source: MR from Connecticut population-based registry. Language data obtained from Physician Profile Survey (PPS).	Having a Follow-up physician (FUP) with a Spanish-language practice (SLP) was statistically significantly associated with receipt of radiotherapy for breast cancer but not for

First author and publication year	*Methodology*	*Key findings*
	Primary/preferred language variable: Data on primary/preferred language of each Hispanic patient was not available in this study.	prostate cancer. This methodology should be explored in states with larger Hispanic populations, and future efforts should include efforts to obtain data on other cancer treatments (eg, chemotherapy and hormone therapy).
Ramsey et al, 2010[36]	Data Source: Washington State Cancer Registry (WSCR) and Medicaid enrollment and claims records. Language data obtained from Medicaid (per Scott Ramsey via email on 08/31/2011). Primary/preferred language variable: Primary language (English or Other)	Factors associated with not receiving radiation included in situ disease and non-English as a primary language.
Smith etal, 2010[33]	Data Source: Brain Attack Surveillance in Corpus Christi (BASIC) - stroke surveillance study. Language data obtained from interview. Primary/preferred language variable: Language (self-reported language fluency and dichotomized as "Spanish" or "English"; English speakers included subjects fluent in both languages)	Mexican Americans were less likely than non-Hispanic whites to arrive by emergency medical services (odds ratio, 0.6; 95% CI, 0.4, 0.8). Men were more likely than women to present to the hospital within 3 hours (odds ratio, 0.7; 95% CI, 0.5, 0.9); language was not associated with study outcomes.
Sweeney et al, 2007[34]	Data Source: Utah Cancer Registry and New Mexico Tumor Registry. Language data obtained from interview. Primary/preferred language variable: Surname, Ethnicity	Hispanics who were correctly classified differed from those who were misclassified, reporting lower language acculturation and education attainment. The authors conclude that a surname search efficiently identifies Hispanics, although individuals identified using this method are not completely representative. Recruitment of Hispanic cases and controls does not appear to be affected by selection bias related to community characteristics.
Traylor et al, 2010[37]	Data Source: Kaiser Permanente's Northern California Diabetes Registry of 2005. Language data obtained from health plan, at plan level separate from the registry, and available through automated clinical data (per co-author Julie Schmittdiel via email on 09/07/2011 and 09/08/2011). Primary/preferred language variable: Race/Ethnicity/Ratient language	Patients who chose their physicians were more likely to have a same race/ethnicity physician with OR of 2.2 (95% CI 1.74-2.82) for African American patients, 1.71 (95% CI 1.44-2.04) for Hispanic patients, 1.11 (95% CI 1.04-1.18) for white patients, and 1.38 (95% CI 1.23, 1.55) for Asian patients. Limited English language was a strong predictor of concordance for Hispanic patients (OR 4.81; 95% CI 4.2-5.51) and Asian patients (OR 9.8; 95% CI 7.7, 12.6)
Yoon et al, 2008[29]	Data Source: MR from Los Angeles County SEER Registry. Language data obtained from survey. Primary/preferred language variable: Race/Ethnicity (Black, Hispanic English speaker, Hispanic Spanish speaker, Other, White)	Multi-variate analysis controlling for patient characteristics and treatment showed that older, black, Hispanic Spanish-speaking, widowed or never married, and working women were less likely to report severe symptoms than other women.