

Capturing Demographic, Health-Related, and Psychosocial Variables in a Standardized Manner: Towards Improving Cancer Screening Adherence

Yannan Lin, MD, MPH^{1,2}, Ruiwen Ding, BS^{1,2}, Ashley E. Prosper, MD², Denise R. Aberle, MD^{1,2}, Alex A.T. Bui, PhD^{1,2}, William Hsu, PhD^{1,2}

¹Department of Bioengineering, University of California, Los Angeles, CA, USA; ²Medical & Imaging Informatics Group, Department of Radiological Sciences, University of California, Los Angeles, CA, USA.

Abstract

Determining factors influencing patient participation in and adherence to cancer screening recommendations is key to successful cancer screening programs. However, the collection of variables necessary to anticipate patient behavior in cancer screening has not been systematically examined. Using lung cancer screening as a representative example, we conducted an exploratory analysis to characterize the current representations of 18 demographic, health-related, and psychosocial variables collected as part of a conceptual model to understand factors for lung cancer screening participation and adherence. Our analysis revealed a lack of standardization in controlled terminologies and common data elements for these variables. For example, only eight (44%) demographic and health-related variables were recorded consistently in the electronic health record. Multiple survey instruments could collect the remaining variables but were highly inconsistent in how variables were represented. This analysis suggests opportunities to establish standardized data formats for psychological, cognitive, social, and environmental variables to improve data collection.

Introduction

Anticipating patient behavior and providing specific interventions are important components of successful cancer screening programs. If the benefits of cancer screening are to be achieved (i.e., improved early detection rates and reduced cancer-specific mortality), participation in and adherence to recommended actions are surely critical. But the collection of variables needed to understand what drives patient participation in and adherence to cancer screening is hugely inconsistent. One set of variables that are associated with breast, cervical, and colorectal cancer screening participation are social determinants of health (SDH),¹ which are “conditions in the places where people live, learn, work, and play that affect a wide range of health and quality of life risks and outcomes.”² When incorporating SDH into research, a major challenge is that *there is presently a lack of consensus on standards for representing or capturing SDH in electronic health records (EHRs).*³ Besides SDH-related variables, barriers to cancer screening involve factors at several levels.⁴⁻⁷ For example, at the patient level are psychological barriers such as denial, fear, and stigmatization; lack of education about cancer and cancer screening; lack of access to health care; and the quality of patient-provider communication. At the provider level, there may be limited knowledge – and outright skepticism – regarding screening guidelines and benefits. More generally, systemic barriers include lack of insurance coverage, access to care, and repeated healthcare visits.

These issues are paramount in emergent areas, such as lung cancer screening. Lung cancer remains the leading cause of cancer-related death in the United States (US).⁸ Clinical trials have shown that screening with low-dose computed tomography (LDCT) is effective in reducing lung cancer death rate by up to 20% or more.^{9, 10} The US Preventive Services Task Force recently recommended that individuals who are 50 to 80 years of age with a minimum 20 pack-year smoking who currently smoke, or have quit within the past 15 years, receive annual screening with LDCT.¹¹ A national coverage decision from the Centers for Medicare and Medicaid Services has covered lung screening since 2015¹² with a recent expansion in coverage February 2022.¹³ Although both Medicare and, in some states, Medicaid cover annual screening LDCTs, an analysis by the American Lung Association through 2021 reveals low screening rates across all states among eligible individuals, ranging from 1-18%.¹⁴ Moreover, our recent meta-analysis showed that patient adherence to baseline lung cancer screening recommendations was only 57-65% across clinical lung cancer screening programs in the US.¹⁵ Given the relative nascence of lung cancer screening, little is known about why screen-eligible smokers decide (not) to undergo screening. To identify factors that are associated with screening behavior in lung cancer, Carter-Harris et al. developed a conceptual model for lung cancer screening participation and adherence.¹⁶ This model proposes that multiple factors can influence lung cancer screening participation, including: psychological variables; demographic and health status characteristics; cognitive variables; receiving a healthcare provider recommendation; social and environmental variables; lung cancer screening health beliefs; and the shared decision-making process between an individual and their health care provider (**Figure 1**). Although the conceptual model

provides a blueprint of what variables should be considered, it does not specify how to measure and encode these variables to facilitate data sharing and semantic interoperability. Markedly, the current state of data capture for the enumerated variables (e.g., cancer fatalism, smoking-related stigma, lung cancer worry, fear, etc.) is not well-characterized, and it is not clear how to best collect this information and from what data source.

Our long-term objective is to improve the overall participation and adherence rate to lung cancer screening among eligible patients by enabling individualized interventions that encourage screening participation. In this work, we focus on capturing the antecedents from the Carter-Harris conceptual model in a consistent and standardized manner. Antecedents are the circumstances that exist before a behavior related to cancer screening. In the Carter-Harris model, antecedents, a combination of SDH, and psychological and cognitive variables, are precursors to the stage of adoption for lung cancer screening, the shared decision-making process, and the subsequent outcomes concerning lung cancer screening behavior (**Figure 1**). Markedly, in prior studies, such antecedents have been shown to correlate with patient participation in lung, breast, cervical, or colorectal cancer screening programs.¹⁶ Our goal is to examine the current state and gaps in standardized collection of SDH data, using cancer screening as a driving application. We investigate whether data standards exist for demographic, health-related, and psychosocial variables and their level of completeness in the EHR. To our knowledge, no analysis has characterized the current representations of variables affecting lung cancer screening participation – or more generally – and adherence across the EHR and existing medical ontologies.

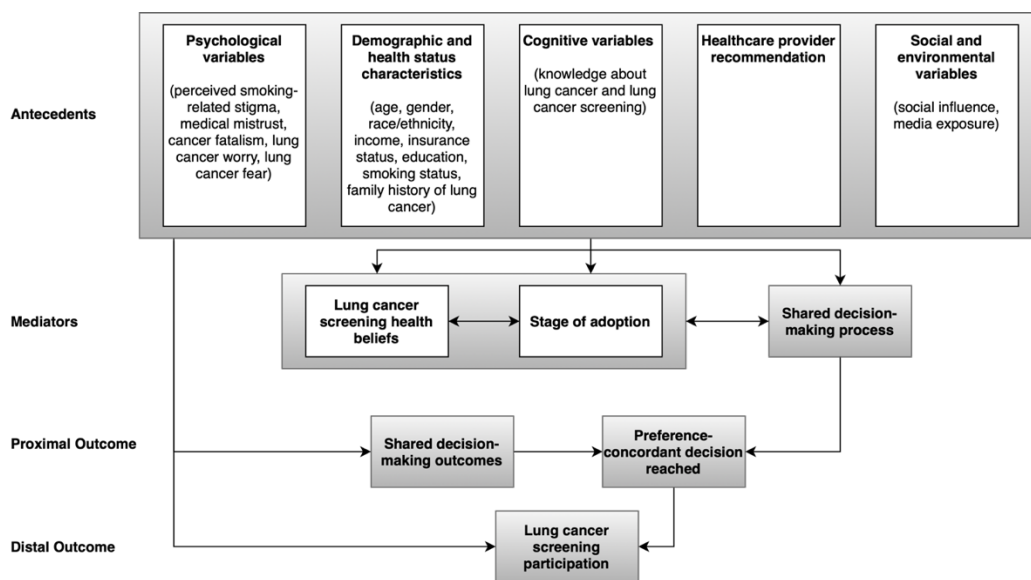


Figure 1. The Carter-Harris conceptual model for lung cancer screening participation.

Methods

Defining and mapping variables

Carter-Harris et al. grouped antecedents into five categories (**Figure 1**): 1) psychological variables, 2) demographic and health status characteristics, 3) cognitive variables, 4) healthcare provider recommendation and 5) social and environmental variables. Each category comprises of a set of variables, such as social influence and media exposure. Among the 18 antecedents, seven are broadly considered SDH, including gender, race/ethnicity, income, education, smoking-related stigma, social influence, media exposure.¹⁷ For each variable, we defined a data element with permissible values, mapping it to EHR data relevant to the 18 antecedents, published ontologies, controlled vocabularies (see *Identifying relevant data elements in the EHR and existing ontologies* section below), and/or survey instruments whenever possible. Next, we identified potential data sources for each data element. Information sources were represented in various formats, including the EHR and questions in a survey instrument, such as Shen’s scale for measuring fatalism¹⁸. When more than one source of representation was available for a specific variable, we listed the most used representation(s) reported in systematic reviews on measures/survey instruments of these variables (e.g., a systematic review on measuring medical mistrust¹⁹). Considering data elements that are not currently collected

in a standardized manner, we identified existing techniques in literature that have been used to collect this information (e.g., survey instruments).

Identifying relevant data elements in the EHR, existing ontologies, and literature

1. ***Search strategy.*** The search terms were the exact expression of antecedents from the Carter-Harris conceptual model. Specifically, we independently searched representations for each antecedent variable across potential data sources. For example, ‘medical mistrust’ was used as the search term (or keyword) for the medical mistrust variable mentioned in the antecedents. We did not add synonyms of the antecedents in the search terms. However, we also used a more general term that was not specific to lung cancer screening for certain variables. For example, we searched cancer “worry” and cancer “fear” in addition to more specific lung cancer worries and fears.

2. ***Searching the EHR.*** Using our institution’s EHR (Epic Systems, Verona, WI) as a representative example, we investigated which variables were presently captured, whether captured data were collected in a consistent and standardized format; and if the variables were not available, alternative sources that could be used. This process was conducted in three ways: 1) examining data elements that are displayed in the EHR user interface, 2) using the ‘Search’ bar with keywords, and 3) consulting with clinicians (D.R.A and A.E.P) on unstructured fields that may contain relevant information. In addition to examining data elements that are explicitly captured in the EHR, we also examined data elements collected as part of a questionnaire administered to the patient before his/her LDCT screening exam; a digitized copy of the questionnaire is retained within the EHR. This questionnaire collects data variables required by the PLCO_{M2012} 6-year lung cancer risk model²⁰ in a standardized manner (i.e., all multiple-choice questions, no free-text questions, see Supplemental Materials for a list of variables included in this questionnaire). The EHR search was conducted by Y.L who has 2 years’ experience of extracting patient information from our institution’s EHR.

3. ***Searching other catalogs and resources.*** Alongside searching the EHR, we queried BioPortal²¹, a comprehensive repository of biomedical ontologies and terminologies, and used ‘Class Search’ to examine existing medical ontologies for the antecedents. In addition to BioPortal, search results from three vocabulary systems/toolkit were summarized, NIH Common Data Elements (provides access to structured data elements that have been recommended or required to use in research by NIH institutes or centers or other organizations)²², NIH RADx-UP Common Data Elements (captures a variety of variables such as sociodemographics, housing, insurance, medical history, health status, tobacco use, medical trust, etc.)²³, and PhenX Toolkit (covers SDH variables, tobacco use, etc.)²⁴. Finally, PubMed and Google Scholar were used to identify measures not captured in the EHR. Two authors (Y.L and R.D) searched BioPortal, the NIH Common Data Elements, NIH RADx-UP Common Data Elements and PhenX Toolkit for relevant concepts (end date of search: Mar 8, 2022). Discrepancies in search results were resolved through a consensus discussion.

Data quality assessment

We focused on one dimension in data quality assessment – data completeness – where we characterized the current level of coverage for antecedents by reporting what percentage of variables could be represented using existing standardized data elements. Specifically, we identified what percentage of data elements can be populated using information that is readily collected in the EHR, NIH Common Data Elements, NIH RADx-UP Common Data Elements and PhenX Toolkit because these ontologies were likely to capture representations for a large number of antecedent variables given their broad coverage of data elements in demographics, health-related, and psychosocial variables. We also assessed the percent of antecedents captured in survey instruments from literature (unstandardized data).

Results

Table 1 (a simplified version, see Supplemental Materials for the full version) summarizes possible data sources for antecedents from the Carter-Harris conceptual model.

Few psychological, cognitive, social, and environmental variables are standardized in the EHR and existing ontologies

Demographic and health status characteristics (age, gender, race/ethnicity, income, insurance status, education, smoking status and family history of lung cancer) were well-standardized in current medical vocabularies. All variables were captured in the EHR system at our institution in a normalized manner (i.e., each stored as a variable with standardized values in the EHR database). Seven of eight variables could be obtained from all three medical ontologies (the NIH Common Data Elements, NIH RADx-UP Common Data Elements, and PhenX Toolkit). Our institution’s EHR implementation lacks a structured field that indicates whether a screen-eligible patient received a

recommendation for lung cancer screening from a healthcare provider. Still, healthcare provider recommendations for lung cancer screening among high-risk individuals were documented in physicians' notes or the "Indication" section of a screening CT interpretation.

The remaining three categories of antecedents were largely unstandardized: we found standardized mappings for a few psychological, cognitive, social, and environmental antecedents from our EHR system, BioPortal searches, or the three medical ontologies (the NIH Common Data Elements, NIH RADx-UP Common Data Elements, and PhenX Toolkit), and most of these mappings were proxies. For example, there existed ontologies for stigma (as a proxy for smoking-related stigma) and fatalism (as a proxy for cancer fatalism) in BioPortal. Therefore, we attempted to map the measurements of these variables with survey instruments developed in the literature. For the five psychological variables (i.e., perceived smoking-related stigma, medical mistrust, cancer fatalism, lung cancer worries, and lung cancer fear), we found at least one instrument that had been used to measure these variables. The instruments were either a direct mapping of the antecedent developed from a screening or non-screening cohort or a proxy instrument used in other domains (such as cancer screening and COVID-19) that could potentially be used in cancer screening. For cognitive variables, the Lung Cancer Awareness Measure³⁸ could be used to assess patients' knowledge about lung cancer. Given that no instruments had been developed to measure patients' knowledge about lung cancer screening, we listed a proxy instrument, a modified version of an instrument developed to assess patients' knowledge about colorectal cancer screening³⁹. Similarly, no instruments existed for measuring social influence among participants in lung cancer screening. We adapted instrument originating from breast cancer screening.⁴¹ A proxy instrument⁴⁴ for measuring general and health-specific media exposure was included for media exposure because no studies had investigated the effect of media exposure on lung cancer screening behavior.

In total, among 18 antecedents, nine (50%) variables were captured in the EHR system at our institution (**Table 2**). Two standardized medical vocabulary repositories captured up to half of the antecedents: eight (44%) variables found in the NIH Common Data Elements repository and nine (50%) variables were indexed in the NIH RADx-UP Common Data Elements repository. The PhenX Toolkit had representations for eight (44%) antecedents. Although the EHR and three medical vocabularies captured 44-50% of the antecedents, most variables were from the demographic and health status characteristics category. Survey instruments from the literature provided measures for the nine (50%) psychological, cognitive, social, and environmental variables, six of which (including proxies) were indexed in BioPortal. Using a combination of EHR and survey instruments from the literature, all 18 antecedents were captured. Yet 22% of these antecedents (including three survey instruments that were not included in BioPortal and the healthcare provider recommendation variable, which was documented in free text in the EHR) *lacked a standardized data format and varied in semantics and permissible values, depending on survey instrument of EHR implementation.*

Table 1 (Simplified Version). Potential data sources of antecedents in the Carter-Harris conceptual model for lung cancer screening participation.

Variable	Definition	Common data sources	Dimensions	No. Items	Scales/Values
Perceived smoking-related stigma	A social process by which exclusion, rejection, blame or devaluation occurs ²⁵	Proxy: stigma. BioPortal: two original ontologies in psychology and nursing practice.			
		Proxy: 'covid_iso_chal' in NIH RADx-UP CDE.			
		Stuber et al. ²⁶ 2009	Devaluation	2	Four-point Likert scale
			The respondents' perceptions that they are the subject of differential treatment due to smoking	3	Dichotomous
		Internalized Stigma of Smoking Inventory ²⁷ ($\alpha=0.80, 0.81, \text{ and } 0.70$ for self stigma, felt stigma, and discrimination experiences, respectively) 2015	Self stigma	3	Four-point Likert scale
			Felt stigma	3	
			Discrimination experiences	2	

Table 1 Continued (Simplified Version). Potential data sources of antecedents in the Carter-Harris conceptual model for lung cancer screening participation.

Variable	Definition	Common data sources	Dimensions	No. Items	Scales/Values
Medical mistrust	Distrust of medical personnel and organizations ²⁸	Proxy: 'trust_doc_2' in NIH RADx-UP CDE. BioPortal: none exists.			
		Medical Mistrust Index ²⁹ ($\alpha=0.76$) 2009	NA	7	Four-point Likert scale
		Group-Based Medical Mistrust Scale ³⁰ ($\alpha=0.83$) 2004	Suspicion	6	Five-point Likert scale
			Group disparity	3	
	Lack of support	3			
Other instruments mentioned in a systematic review ¹⁹					
Cancer fatalism	The belief that death is inevitable when cancer is present ³¹	Proxy: fatalism. BioPortal: two original ontologies in psychology and consumer health.			
		Shen et al. ¹⁸ (applicable across a wider range of health conditions and with a broader set of culture) (overall $\alpha=0.88$, $\alpha=0.86$, 0.80 , 0.82 for predetermination, luck, and pessimism, respectively) 2009	Predetermination	10	Five-point Likert scale
			Luck	4	
			Pessimism	6	
Other instruments mentioned in a systematic review ³²					
Lung cancer worry	Concerns about developing cancer or cancer recurrence, and the impact of these concerns on daily functioning, among individuals at risk for hereditary cancer ³³	NIH CDE, BioPortal: one original ontology in LOINC.			
		Proxy: cancer worry. BioPortal: one original ontology. Cancer Worry Scale ³⁴ ($\alpha=0.87$) 2014	NA	8	Four-point Likert scale
		Proxy instrument: breast cancer worry ³⁵ ($\alpha=0.85$) 2012	NA	2	Categorical
Lung cancer fear	The threat of what a lung cancer diagnosis may mean to the individual ^{36,37}	Proxy: cancer fear. BioPortal: three original ontologies in primary care and clinical terms.			
		Psychological Consequences Questionnaire ³⁶ 2008	NA	3	Five-point scale
Age	Age	EHR, NIH CDE, NIH RADx-UP CDE, PhenX Toolkit, BioPortal: >10 original ontologies.	NA	NA	Continuous or categorical
Gender	Gender	EHR, NIH CDE, NIH RADx-UP CDE, PhenX Toolkit, BioPortal: >10 original ontologies.	NA	NA	Dichotomous
	Proxy: sex	EHR, NIH CDE, NIH RADx-UP CDE, PhenX Toolkit, BioPortal: >10 original ontologies.	NA	NA	Dichotomous
Race/ethnicity	Race/ethnicity	EHR, NIH CDE, NIH RADx-UP CDE, PhenX Toolkit, BioPortal: >10 original ontologies.	NA	NA	Categorical
Income	Income: ontology-specific definitions	EHR, NIH CDE, NIH RADx-UP CDE, PhenX Toolkit, BioPortal: >10 original ontologies.	NA	NA	Continuous or categorical
	Proxy: zip code (map family income)	EHR, NIH CDE, NIH RADx-UP CDE, PhenX Toolkit, BioPortal: >10 original ontologies.	NA	NA	Continuous or categorical

Table 1 Continued (Simplified Version). Potential data sources of antecedents in the Carter-Harris conceptual model for lung cancer screening participation.

Variable	Definition	Common data sources	Dimensions	No. Items	Scales/Values
Insurance status	Insurance status	EHR, NIH CDE, NIH RADx-UP CDE, PhenX Toolkit, BioPortal: >10 original ontologies.	NA	NA	Categorical
Education	The highest level of education	EHR (source: UCLA-specific questionnaire), NIH CDE, NIH RADx-UP CDE, PhenX Toolkit, BioPortal: >10 original ontologies.	NA	NA	Categorical
Smoking status	Smoking status	EHR, NIH CDE, NIH RADx-UP CDE, PhenX Toolkit, BioPortal: >10 original ontologies.	NA	NA	Categorical
Family history of lung cancer	A reported family history of lung cancer in one or more family members	EHR, BioPortal: one original ontology.	NA	NA	Dichotomous
Knowledge about lung cancer	Awareness of symptoms and risk factors of lung cancer ³⁸	BioPortal: none exists.			
		Lung Cancer Awareness Measure ³⁸ (overall $\alpha=0.88$, $\alpha=0.91$ and 0.74 for the warning signs and risk factors subscales) 2012	Socio-demographical characteristics	6	Dichotomous or categorical
			Knowledge of warning signs for lung cancer	14	Continuous
Knowledge about lung cancer screening	Knowledge about lung cancer screening guidelines and frequency	BioPortal: none exists.			
		Proxy: adapted from colorectal cancer screening ³⁹	Screening participation	1 (2 follow-up questions)	Dichotomous (follow-up questions: free text)
			Screening frequency	1	Free text
Healthcare provider recommendation	Documented recommendations of getting a screening LDCT from healthcare providers	BioPortal: none exists.			
		EHR (free text)	NA	NA	Free text
Social influence	The influence of family and friends on an individual's behavior ⁴⁰	BioPortal: one original ontology.			
		Proxy: adapted from breast cancer screening ⁴¹ ($\alpha=0.93$)	NA	7	Five-point scale
Media exposure	The potential influence of commercial, print, and social media on cancer screening participation ^{42,43}	BioPortal: one original ontology in psychology.			
		Proxy: 'Media Use During COVID-19' in PhenX Toolkit.			
		Proxy: media exposure ⁴⁴ ($\alpha=0.74$ and 0.65 for general and health-specific media exposure, respectively) 2014	General media exposure	2	Continuous
		Health-specific media exposure	3	Categorical	

α : Cronbach's alpha, a measure of internal consistency or reliability for a survey/questionnaire.

EHR: electronic health record, NIH: National Institutes of Health, CDE: common data elements, LOINC: Logical Observation Identifier Names and Codes.

Table 2. Summary of representations of antecedents in the Carter-Harris conceptual model.

No	Variable	EHR	NIH Common Data Elements	NIH RADx- UP Common Data Elements	Phenx Toolkit	Instruments from literature
1	Perceived smoking-related stigma			√*		√
2	Medical mistrust			√*		√
3	Cancer fatalism					√
4	Lung cancer worry		√			√
5	Lung cancer fear					√
6	Age	√	√	√	√	
7	Gender (or sex)	√	√	√	√	
8	Race/ethnicity	√	√	√	√	
9	Income ^a	√	√	√	√	
10	Insurance status	√	√	√	√	
11	Education	√	√	√	√	
12	Smoking status	√	√	√	√	
13	Family history of lung cancer	√				
14	Knowledge about lung cancer					√
15	Knowledge about lung cancer screening					√*
16	Healthcare provider recommendation	√**				
17	Social influence					√*
18	Media exposure				√*	√*
Percent captured (%)		50 (9/18)	44 (8/18)	50 (9/18)	44 (8/18)	50 (9/18)

^a Include family income mapped by zip code.

* Need to adapt from other domains, such as COVID-19, breast and colorectal cancer screening.

** Unstandardized. E.g., free text.

EHR: electronic health record, NIH: National Institutes of Health.

Discussion

Using lung cancer screening as an example, our study highlights the lack of consistent and standardized representations for variables that are needed to understand the drivers of patient participation in and adherence to cancer screening. In this exploratory analysis, we mapped antecedents from the Carter-Harris conceptual model to existing standardized medical vocabularies and EHR data, identifying gaps in data elements that needed to be collected from additional sources (i.e., survey instruments). Our analysis suggests that many common antecedents, including psychological, cognitive, social, and environmental variables, have yet to be standardized and consistently represented. While a previous study revealed more than 1,000 clinical codes in common medical vocabularies (i.e., LOINC, SNOMED CT, ICD-10-CM, and CPT) that could potentially be used to document SDH-related clinical activities,⁴⁵ our study showed that a number of antecedents are missing from these codes.

As there is a lack of standardized representations from either the EHR or common data vocabularies, researchers face the challenge of selecting the most appropriate survey instrument to address a specific antecedent. For example, a systematic review on medical mistrust measures revealed at least 12 measures or scales for assessing medical mistrust across a wide variety of health topics, including cancer screening, and observed varied conceptualizations of the term ‘medical mistrust’.¹⁹ We must understand how medical mistrust and other antecedents should be conceptualized in the context of cancer screening before suggesting standardized representations.

Facilitating clinical research in cancer screening by capture of psychological, cognitive, social, and environmental variables is critical to ensuring the completeness and consistency of collected data. Improving the systematic collection of these antecedents in a standardized manner will aid in the identification of factors that predict whether a patient will be adherent to screening follow-up recommendations; this information can be used to tailor interventions to patients to encourage their adherence. While a number of toolkits and resources that include SDH data elements exist, these disparate efforts combined with a lack of awareness among investigators result in poor adoption and inconsistent use of standards. One promising initiative, the Gravity Project, has been developing consensus-driven standards to promote interoperability of available SDH data in the EHRs.⁴⁶ A centralized clearinghouse for SDH resources and data collection instruments could aid in. While several groups have proposed various ontologies to represent different aspects of SDH^{47, 48}, much work needs to be done to broaden the coverage of existing ontologies. As demonstrated in our work, common data elements need to be developed around specific use cases such as cancer screening. Societies and professional organizations should promote the development of these common data elements and serve as resources for their respective communities on how to utilize these resources. For example, societies that run national registries (e.g., National Lung Screening Registry) could promote the use of standardized SDH specific to screening to ensure interoperability of collected data across sites. When these standardized variables are readily available for use in clinical research, researchers can verify them, allowing more opportunities to refine and improve our knowledge in predicting patient participation in and adherence to cancer screening.

This work has several notable limitations. A single author (Y.L.) conducted the searches in our institution's EHR. Additional raters for this task may minimize errors in the searches and increase the reliability of this study. This study is limited to determining the completeness of obtaining antecedent information from EHR and other data sources. We did not assess data quality in other dimensions, such as data consistency, accuracy, timelessness, and validity. We limited the variables to the 18 antecedents from the Carter-Harris conceptual model. However, additional barriers to cancer screening are unaddressed by this model, including the patient's lack of access to health care, ongoing skepticism about screening benefits, insufficient time for providers to discuss cancer screening, and a provider's knowledge deficits about screening guidelines^{49, 50}. We did not perform a comprehensive analysis of mapping quality between antecedents and possible data sources. Although this study examined antecedents specific to lung cancer screening, our approach could generalize to any domain, i.e., mapping standardized representations to data elements in a conceptual model that can later be incorporated into analyses to inform clinical decisions.

Conclusion

A deep understanding of disparities in cancer screening can facilitate interventions to improve patient participation in and adherence to cancer screening programs. Current EHR systems and standardized medical vocabularies (i.e., NIH Common Data Elements and NIH RADx-UP Common Data Elements, etc.) cannot comprehensively represent variables that capture patients' beliefs about smoking, cancer and cancer screening, social, and environmental factors in a standardized manner. Systematic collection of this information could help researchers understand why screen-eligible patients decide (not) to undergo screening and why screening patients (do not) adhere to screening guidelines. While there exist survey instruments in the literature for measuring psychological, cognitive, social, and environmental variables, a lack of consistent representations of these variables impedes reliable and reproducible research. To systematically collect psychological, mental, social, and environmental variables that influence participation in and adherence to cancer screening recommendations, we need to be attuned to how these variables are conceptualized, determine standardized representations through systematic reviews, make the variables available in common clinical data sources (such as EHR), and encourage researchers to verify and improve the standardization in clinical research.

Supplemental Materials

The full version of Table 1 and a list of variables captured by the UCLA questionnaire can be found here: https://github.com/allyn1982/AMIA_2022_Student_Paper.

Acknowledgment

This work was supported by the National Institutes of Health under awards R01 CA210360 and R01 CA226079.

References

1. Kurani S, McCoy R, Lampman M, Doubeni C, Finney Rutten L, Inselman J et al. Association of Neighborhood Measures of Social Determinants of Health With Breast, Cervical, and Colorectal Cancer Screening Rates in the US Midwest. *JAMA Network Open*. 2020;3(3):e200618.

2. Social Determinants of Health | CDC [Internet]. Cdc.gov. 2022 [cited 22 February 2022]. Available from: <https://www.cdc.gov/socialdeterminants/index.htm>
3. Cantor M, Thorpe L. Integrating Data On Social Determinants Of Health Into Electronic Health Records. *Health Affairs*. 2018;37(4):585-590.
4. James T, Greiner K, Ellerbeck E, Feng C, Ahluwalia J. Disparities in colorectal cancer screening: a guideline-based analysis of adherence. *Ethn Dis*. 2006;16(1):228-233.
5. Schueler K, Chu P, Smith-Bindman R. Factors Associated with Mammography Utilization: A Systematic Quantitative Review of the Literature. *Journal of Women's Health*. 2008;17(9):1477-1498.
6. Fox S, Heritage J, Stockdale S, Asch S, Duan N, Reise S. Cancer screening adherence: Does physician–patient communication matter?. *Patient Education and Counseling*. 2009;75(2):178-184.
7. Peterson E, Ostroff J, DuHamel K, D'Agostino T, Hernandez M, Canzona M et al. Impact of provider-patient communication on cancer screening adherence: A systematic review. *Preventive Medicine*. 2016;93:96-105.
8. Siegel R, Miller K, Fuchs H, Jemal A. Cancer statistics, 2022. *CA: A Cancer Journal for Clinicians*. 2022;72(1):7-33.
9. de Koning H, van der Aalst C, de Jong P, Scholten E, Nackaerts K, Heuvelmans M et al. Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. *New England Journal of Medicine*. 2020;382(6):503-513.
10. National Lung Screening Trial Research Team, Aberle D, Adams A, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365(5):395-409.
11. Recommendation: Lung Cancer: Screening | United States Preventive Services Taskforce [Internet]. [Uspreventiveservicestaskforce.org](https://www.uspreventiveservicestaskforce.org). 2022 [cited 22 February 2022]. Available from: <https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/lung-cancer-screening#fullrecommendationstart>
12. NCA - Screening for Lung Cancer with Low Dose Computed Tomography (LDCT) (CAG-00439N) - Decision Memo [Internet]. [Cms.gov](https://www.cms.gov). 2022 [cited 22 February 2022]. Available from: <https://www.cms.gov/medicare-coverage-database/view/ncacal-decision-memo.aspx?proposed=N&NCAId=274>
13. NCA - Screening for Lung Cancer with Low Dose Computed Tomography (LDCT) (CAG-00439R) - Decision Memo [Internet]. [Cms.gov](https://www.cms.gov). 2022 [cited 8 March 2022]. Available from: <https://www.cms.gov/medicare-coverage-database/view/ncacal-decision-memo.aspx?proposed=N&ncaid=304>
14. State of Lung Cancer | California. [Lung.org](https://www.lung.org). 2022 [cited 22 February 2022]. Available from: <https://www.lung.org/research/state-of-lung-cancer/states/california>.
15. Lin Y, Fu M, Ding R, Inoue K, Jeon C, Hsu W et al. Patient Adherence to Lung CT Screening Reporting & Data System–Recommended Screening Intervals in the United States: A Systematic Review and Meta-Analysis. *Journal of Thoracic Oncology*. 2022;17(1):38-55.
16. Carter-Harris L, Davis L, Rawl S. Lung Cancer Screening Participation: Developing a Conceptual Model to Guide Research. *Research and Theory for Nursing Practice*. 2016;30(4):333-352.
17. Islam M. Social Determinants of Health and Related Inequalities: Confusion and Implications. *Frontiers in Public Health*. 2019;7.
18. Shen L, Condit C, Wright L. The psychometric property and validation of a fatalism scale. *Psychology & Health*. 2009;24(5):597-613.
19. Williamson L, Bigman C. A systematic review of medical mistrust measures. *Patient Education and Counseling*. 2018;101(10):1786-1794.
20. Tammemägi M, Katki H, Hocking W, Church T, Caporaso N, Kvale P et al. Selection Criteria for Lung-Cancer Screening. *New England Journal of Medicine*. 2013;368(8):728-736.
21. Noy N, Shah N, Whetzel P, Dai B, Dorf M, Griffith N et al. BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Research*. 2009;37(Web Server):W170-W173.
22. NIH Common Data Elements (CDE) Repository [Internet]. [Cde.nlm.nih.gov](https://cde.nlm.nih.gov/home). 2022 [cited 22 February 2022]. Available from: <https://cde.nlm.nih.gov/home>
23. NIH RADx-UP Common Data Elements - RADx-UP [Internet]. [RADx-UP](https://radx-up.org/learning-resources/cdes/). 2022 [cited 22 February 2022]. Available from: <https://radx-up.org/learning-resources/cdes/>
24. Hamilton C, Strader L, Pratt J, Maiese D, Hendershot T, Kwok R et al. The PhenX Toolkit: Get the Most From Your Measures. *American Journal of Epidemiology*. 2011;174(3):253-260.
25. Thornicroft G, Brohan E, Rose D, Sartorius N, Leese M. Global pattern of experienced and anticipated discrimination against people with schizophrenia: a cross-sectional survey. *The Lancet*. 2009;373(9661):408-415.
26. Stuber J, Galea S, Link B. Stigma and Smoking: The Consequences of Our Good Intentions. *Social Service Review*. 2009;83(4):585-609.

27. Brown-Johnson C, Cataldo J, Orozco N, Lisha N, Hickman N, Prochaska J. Validity and reliability of the internalized stigma of smoking inventory: An exploration of shame, isolation, and discrimination in smokers with mental health diagnoses. *The American Journal on Addictions*. 2015;24(5):410-418.
28. Omodei M, McLennan J. Conceptualizing and Measuring Global Interpersonal Mistrust-Trust. *The Journal of Social Psychology*. 2000;140(3):279-294.
29. LaVeist T, Isaac L, Williams K. Mistrust of Health Care Organizations Is Associated with Underutilization of Health Services. *Health Services Research*. 2009;44(6):2093-2105.
30. Thompson H, Valdimarsdottir H, Winkel G, Jandorf L, Redd W. The Group-Based Medical Mistrust Scale: psychometric properties and association with breast cancer screening. *Preventive Medicine*. 2004;38(2):209-218.
31. Powe B, Finnie R. Cancer Fatalism. *Cancer Nursing*. 2003;26(6):454-467.
32. Valenti G, Faraci P. Instruments measuring fatalism: A systematic review. *Psychological Assessment*. 2022;34(2):159-175.
33. National Cancer Institute Thesaurus - Cancer Worry Scale - Classes | NCBO BioPortal [Internet]. Bioportal.bioontology.org. 2022 [cited 22 February 2022]. Available from: <https://bioportal.bioontology.org/ontologies/NCIT?p=classes&conceptid=http%3A%2F%2Fncicb.nci.nih.gov%2Fxml%2Fowl%2FEVVS%2FThesaurus.owl%23C176441#details>
34. Custers J, van den Berg S, van Laarhoven H, Bleiker E, Gielissen M, Prins J. The Cancer Worry Scale. *Cancer Nursing*. 2014;37(1):E44-E50.
35. Zhang L, Chiarelli A, Glendon G, Mirea L, Knight J, Andrulis I et al. Worry Is Good for Breast Cancer Screening: A Study of Female Relatives from the Ontario Site of the Breast Cancer Family Registry. *Journal of Cancer Epidemiology*. 2012;2012:1-9.
36. Byrne M, Weissfeld J, Roberts M. Anxiety, Fear of Cancer, and Perceived Risk of Cancer following Lung Cancer Screening. *Medical Decision Making*. 2008;28(6):917-925.
37. Patel D, Akporobaro A, Chinyanganya N, Hackshaw A, Seale C, Spiro S et al. Attitudes to participation in a lung cancer screening trial: a qualitative study. *Thorax*. 2011;67(5):418-425.
38. Simon A, Juszczyk D, Smyth N, Power E, Hiom S, Peake M et al. Knowledge of lung cancer symptoms and risk factors in the UK: development of a measure and results from a population-based survey. *Thorax*. 2012;67(5):426-432.
39. Tessaro I, Mangone C, Parkar I, Pawar V. Knowledge, barriers, and predictors of colorectal cancer screening in an Appalachian church population. *Prev Chronic Dis*. 2006;3(4):A123.
40. Allen J, Sorensen G, Stoddard A, Peterson K, Colditz G. The relationship between social network characteristics and breast cancer screening practices among employed women. *Annals of Behavioral Medicine*. 1999;21(3):193-200.
41. Steadman L, Rutter D. Belief importance and the theory of planned behaviour: Comparing modal and ranked modal beliefs in predicting attendance at breast screening. *British Journal of Health Psychology*. 2004;9(4):447-463.
42. Anderson J, Mullins R, Siahpush M, Spittal M, Wakefield M. Mass media campaign improves cervical screening across all socio-economic groups. *Health Education Research*. 2009;24(5):867-875.
43. Morrell S, Perez D, Hardy M, Cotter T, Bishop J. Outcomes from a mass media campaign to promote cervical screening in NSW, Australia. *Journal of Epidemiology & Community Health*. 2009;64(9):777-783.
44. Jung M, Chan C, Viswanath K. Moderating Effects of Media Exposure on Associations between Socioeconomic Position and Cancer Worry. *Asian Pacific Journal of Cancer Prevention*. 2014;15(14):5845-5851.
45. Arons A, DeSilvey S, Fichtenberg C, Gottlieb L. Documenting social determinants of health-related clinical activities using standardized medical vocabularies. *JAMIA Open*. 2018;2(1):81-88.
46. The Gravity Project - Gravity Project - Confluence. Confluence.hl7.org. <https://confluence.hl7.org/display/GRAV/The+Gravity+Project>. Published 2022. Accessed July 10, 2022.
47. Jani A, Liyanage H, Okusi C, Sherlock J, Hoang U, Ferreira F, Yonova I, de Lusignan S. Using an Ontology to Facilitate More Accurate Coding of Social Prescriptions Addressing Social Determinants of Health: Feasibility Study. *J Med Internet Res*. 2020;22(12):e23721.
48. Brenas JH, Shin EK, Shaban-Nejad A. Adverse Childhood Experiences Ontology for Mental Health Surveillance, Research, and Evaluation: Advanced Knowledge Representation and Semantic Web Techniques. *JMIR Ment Health*. 2019;6(5):e13498.
49. Bromley E, May F, Federer L, Spiegel B, van Oijen M. Explaining persistent under-use of colonoscopic cancer screening in African Americans: A systematic review. *Preventive Medicine*. 2015;71:40-48.
50. Schueler K, Chu P, Smith-Bindman R. Factors Associated with Mammography Utilization: A Systematic Quantitative Review of the Literature. *Journal of Women's Health*. 2008;17(9):1477-1498.