

<https://doi.org/10.1038/s44401-025-00058-9>

Modernizing data and data science infrastructure as a strategic asset for cancer center

Check for updates

Yi Guo^{1,2}✉, Thomas J. George^{1,2}, Alison M. Ivey², Jonathan D. Licht^{1,2}, Gigi Lipori³, Aik Choon Tan^{4,5}, Lang Li^{6,7}, Emily C. Webber⁸, Christina M. Scifres⁸, Kun Huang^{8,9,10}, David A. Haggstrom^{8,10}, Kelvin P. Lee^{8,9}, Rachel E. Patzer^{8,10} & Jiang Bian^{8,9,10}✉

Modern cancer centers face increasing data-driven demands from institutional, state, and federal stakeholders, as well as from patients and communities. Data now functions as a strategic asset that supports care delivery, research, education, community engagement, and policy improvement and development. However, many cancer centers continue to operate with fragmented systems, limited interoperability, and underdeveloped analytics capacity, which constrain their ability to meet expectations or drive innovation. This perspective advocates for adopting a Learning Health System - Learning Health Community (LHS-LHC) framework as a strategic approach to modernizing cancer data and data science infrastructure. We outline how cancer centers can develop integrated, interoperable, and inclusive data systems that enable continuous learning, promote equity-driven outcomes, and generate real-world impact. By embedding data and data science as core institutional infrastructure, and by aligning investments across health systems, funders, and community partners, cancer centers can enhance their role as regional anchors for coordinated, patient-centered care and lead a national transition toward data-enabled oncology.

The central role of data and data science in modern cancer centers

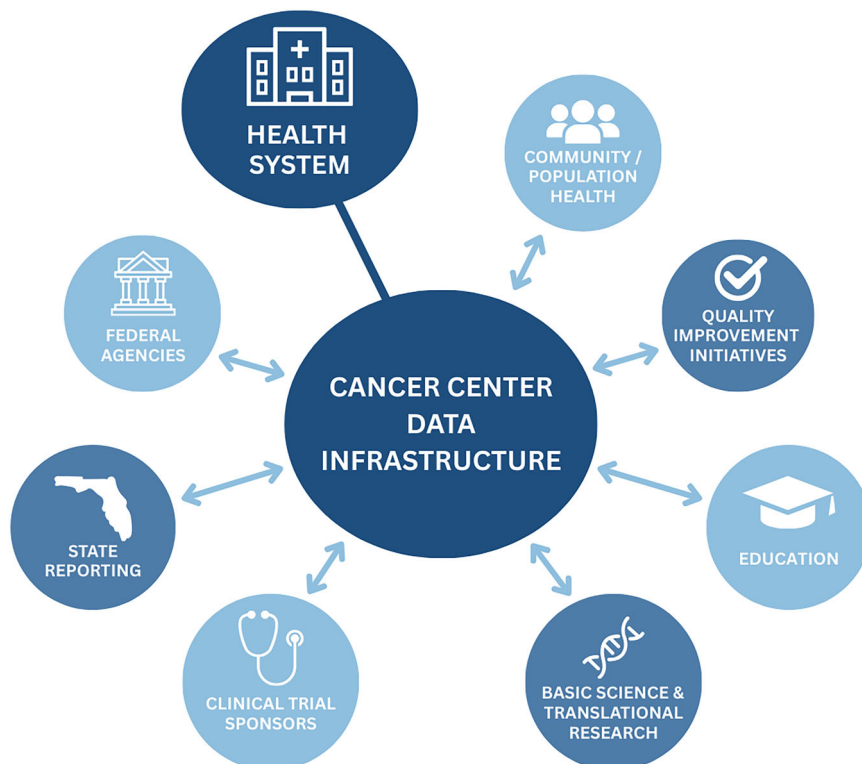
For National Cancer Institute (NCI)-designated cancer centers¹, typically embedded within academic health systems, data serve as the foundation for nearly every dimension of institutional success. Accurate, timely, and integrated data systems are essential for a broad spectrum of core functions, spanning patient care, scientific research, community engagement, education, and policy advocacy. Increasingly, data is recognized not merely as an operational necessity but as a strategic asset through which cancer centers demonstrate value, secure continued investment, and fulfill their mission to reduce the cancer burden across diverse populations. This growing reliance on data also underscores the need to invest in data science expertise, including skilled professionals who can generate insights, drive innovation, and enable data-informed decision-making across the cancer care continuum.

Cancer centers operate within a complex ecosystem of obligations and opportunities (Fig. 1). They must meet a range of reporting requirements from their health systems as well as state and federal stakeholders such as state Departments of Health and the NCI. These obligations are further shaped by demands from clinical trial sponsors, basic and translational research initiatives, population health programs (e.g., Community Outreach and Engagement³), and internal quality improvement efforts. Each domain introduces unique needs for data collection, integration, analysis, and interpretation. Advancing capabilities to generate and harmonize data across silos is not optional, but foundational to both internal planning and external accountability. Moreover, because cancer centers are integrated within larger health systems, their data ecosystems are inherently interdependent with other health system data sources, which makes interoperability essential for seamless data integration and effective collaboration.

¹College of Medicine, University of Florida, Gainesville, FL, USA. ²UF Health Cancer Institute, Gainesville, FL, USA. ³UF Health, Gainesville, FL, USA. ⁴Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA. ⁵University of Utah Health Sciences Center, Salt Lake City, UT, USA. ⁶College of Medicine, The Ohio State University, Columbus, OH, USA. ⁷Ohio State University Comprehensive Cancer Center, Columbus, OH, USA. ⁸Indiana University School of Medicine, Indianapolis, IN, USA. ⁹Indiana University Simon Comprehensive Cancer Center, Indianapolis, IN, USA. ¹⁰Regenstrief Institute, Inc., Indianapolis, IN, USA.

✉e-mail: yiguo@ufl.edu; bianji@regenstrief.org

Fig. 1 | Ecosystem of obligations and opportunities for academic cancer centers.



At the national level, there is a growing emphasis on leveraging data and data science to improve transparency, accountability, and outcomes in cancer care and research^{3,4}. This trend presents a pivotal opportunity for cancer centers to lead the development of Learning Health Systems (LHSs), which support continuous improvement by integrating data-driven practice with evidence generation and clinical care^{5,6}. Cancer centers are increasingly expected to demonstrate measurable impact through discoveries and data that drive paradigm shifts, inform clinical practice, and shape policy^{2,7}. Leveraging these data assets enables cancer centers to improve patient care through earlier detection, personalized treatments, and enhanced survivorship support, and to accelerate the translation of research discoveries into advances in cancer prevention and treatment. As such, cancer centers that can effectively harness and operationalize their data assets will be better positioned to attain or maintain NCI designation, secure research and infrastructure funding, and demonstrate real-world clinical and community impact.

Despite these obligations and opportunities, many cancer centers face persistent gaps in their data and data science infrastructure. Common challenges include fragmented data systems, poor interoperability across clinical and research platforms, underdeveloped analytics capacity, insufficient investment in data governance, and a shortage of skilled professionals who can transform raw data into actionable insight. These challenges are further compounded by the inconsistent integration of diverse real-world data (RWD) sources, including electronic health records (EHRs), cancer registries, patient-reported outcomes (PROs) (e.g., symptoms), patient-generated data (e.g., wearables), patient-derived data (e.g., molecular or imaging data), social determinants of health (SDOH)⁸, and other important data domains. Without robust workflows to incorporate these data, cancer centers are unable to fully leverage them for research, quality improvement, and population health initiatives.

A further barrier is the persistent divide between operations and research. Data science expertise often resides in academic departments, such as biomedical informatics, biostatistics, or data science, and is rarely leveraged to improve institutional operations or health system performance. This siloed structure not only limits the utility of existing data assets but also impairs the cancer center's ability to function as a true LHS. In response,

several leading cancer centers have established or expanded dedicated cancer data science programs to bridge this divide^{9–11}. These initiatives aim to integrate diverse data sources into a cohesive and interoperable infrastructure, promote advanced analytics through cross-disciplinary collaboration, and enable real-time insights across patient care, research, and operational planning.

To fulfill their mission, modern cancer centers must treat data and data science as core institutional infrastructure. This requires not only investment in technology platforms but also support for workforce development and collaborative processes (e.g., between programmers who query EHR data and biostatisticians who conduct statistical analysis). As cancer care becomes more personalized, and therefore more data-intensive, the ability to learn continuously from every patient interaction will determine a center's success, relevance, and long-term impact. In this perspective, we examine the foundational role of data and data science in cancer center performance and accountability, identify persistent structural gaps, and offer recommendations for building data infrastructure that meets the evolving demands of research, clinical operations, and population health within the framework of an LHS.

Obligations, data-driven demands, and structural barriers faced by cancer centers

Modern cancer centers must navigate increasingly complex, data-intensive obligations from institutional, state, and federal stakeholders. Beyond regulatory mandates, data now drive strategic planning, quality improvement, research, education, and community engagement. However, the ability to meet these demands is often hindered by structural barriers (Table 1), including fragmented data systems, limited interoperability, underdeveloped data science capacity, and insufficient investment in governance, workforce, and sustainable infrastructure. These challenges limit the ability of cancer centers to fully harness data and data science in support of their missions.

Institutional obligations and gaps

Cancer centers embedded within academic health systems rely on robust, integrated data and data science infrastructure to meet a broad range of

Table 1 | Key data-driven obligations and structural barriers faced by cancer centers

Domain	Key obligations and data demands	Structural barriers
Institutional	Meet accreditation and reporting requirements (e.g., ASCO, CoC, ACR); support quality improvement, clinical operations, cohort discovery, precision oncology, and AI-ready analytics.	Fragmented data systems (EHR, registries, LIMS, PACS, genomic profiling databases); lack of common data models and standardized terminologies; underdeveloped NLP; inconsistent integration.
State	Contribute to state cancer surveillance; report incidence, mortality, and disparities; coordinate with public health and community providers; strengthen rural and underserved care.	Registry reporting delays and missing data; misclassification of key variables; inconsistent data standards; limited EHR integration; underrepresentation of priority populations.
NCI designation	Define and characterize catchment areas; monitor cancer burden and access; assess outreach and trial participation; demonstrate improvements in early detection and outcomes.	"Invisible denominator" problem; fragmented longitudinal records; limited interoperability across institutions; incomplete capture of population-level data.
Patient and community	Enhance patient experience; support person-centered and coordinated care; identify disparities; guide outreach; collect SDOH and PROs; enable transparency and engagement.	Lack of standardized fields for SDOH; minimal integration of survey and community-level data; limited collection and use of PROs; gaps in transparency, inclusion, and trust.

institutional responsibilities. Accreditation by bodies such as the American Society of Clinical Oncology (ASCO), the Commission on Cancer (CoC), and the American College of Radiology (ACR) requires systematic reporting on care quality, clinical timeliness, and adherence to evidence-based care standards, which rely on timely, accurate data across clinical, administrative, and research domains. Data also support quality improvement efforts such as monitoring treatment delays, evaluating case complexity, benchmarking performance, and optimizing clinical operations (e.g., patient navigation, referral coordination, revenue cycle oversight, and market share analysis). For clinical research, which is central to the mission of NCI-designated cancer centers, curated and integrated data systems enable cohort discovery, automated eligibility screening, dynamic patient-trial matching, longitudinal outcome tracking, and advanced analytics (e.g., predictive modeling for treatment responses). For instance, the ability to link clinical records with molecular profiling data across cancer types, especially for patients with less prevalent malignancies, is essential for precision oncology.

Despite these needs, many cancer centers face persistent fragmentation across core data systems. Platforms beyond the EHR, such as tumor registries, clinical trial management systems, laboratory information management systems (LIMSs), genomic databases, and imaging systems, often operate in silos with limited interoperability. While integrated EHRs have become more common in recent years, they do not fully capture all clinically or operationally relevant information. Most health systems rely on hundreds of specialized, disconnected information systems, which require labor-intensive efforts to curate and use. This challenge is particularly acute in cancer care, where patients often receive different phases of treatment (e.g., surgery, radiation, systemic therapy, survivorship care) across multiple health systems that use disparate EHR platforms. The result is fragmented longitudinal records that complicate care coordination and limit research efforts requiring complete treatment histories.

A more fundamental barrier is the lack of common data models, standardized terminologies, and interoperable interfaces across systems, severely limiting data linkage and aggregation. While many institutions have invested in enterprise data warehouses or data lakes, their maturity and utility vary widely. For example, natural language processing (NLP) tools, including large language model-based tools, are increasingly used to extract critical information from unstructured clinical notes (e.g., treatment data and SDOH), yet implementation of these tools remains inconsistent due to gaps in expertise, infrastructure, and enterprise-level integration.

State obligations and gaps

Cancer centers play critical roles in state-level cancer surveillance and public health planning. State health departments rely on these centers for timely and comprehensive data on cancer burden, including incidence and mortality, across geographic regions and population subgroups. These data inform state-wide cancer control strategies, identify disparities, and guide effective resource allocation. Cancer centers are also expected to coordinate cancer care with community providers and state agencies, including tracking patient referrals into cancer centers, monitoring referral

completions, delivering targeted educational interventions to improve care compliance, and improving communication around clinical trials and specialty services. Such coordination is especially critical in rural or underserved areas, where cancer centers often act as regional hubs for advanced cancer diagnostics, treatment, and support services.

However, state cancer registries face significant limitations in data capacity¹². Data submission lags often exceed 9–18 months due to delays in data abstraction and processing. Key information, such as treatment regimen, screening results, and SDOH, is frequently missing. Additionally, registries commonly encounter data-quality issues such as duplicate records and variable misclassification (e.g., race and ethnicity), partially stemming from inconsistencies in data collection standards¹². Missing data is prevalent, especially for patients with advanced cancer, which reflects the complexity of their care and associated documentation challenges¹³. Moreover, data limitations such as underrepresentation of uninsured and minoritized populations and limited data on structural racism exacerbate disparities by distorting cancer burden estimates and leading to inequitable resource allocation¹⁴. Factors contributing to these limitations include resource constraints, limited integration with EHRs, inconsistent use of standardized terminologies, and reliance on manual data abstraction prone to errors¹². As a result, cancer centers often need to supplement state registry data with internal sources to meet state expectations and deliver actionable insights.

NCI-designation obligations and gaps

NCI-designated cancer centers are expected to define and characterize their catchment areas using demographic, socioeconomic, and epidemiologic data. They need to capture not only who resides in these areas, but also which populations are being screened, which patients are enrolled in clinical trials, which lines of treatment patients receive, and the outcomes they experience. The NCI increasingly emphasizes real-world impact beyond service volume, which requires centers to track metrics such as population reach, outreach effectiveness, and access to clinical trials and related services¹⁵. Data are essential for monitoring trends in cancer burden across the catchment area, including prevalent and high-mortality cancer types and changes in incidence and mortality, to demonstrate measurable improvements in access, early detection, and outcomes, particularly among underserved groups. To meet these expectations, cancer centers must develop integrated data ecosystems that link clinical, research, and community data sources, which enable robust measurement, reporting, and benchmarking through national platforms such as those maintained by the Association of American Cancer Institutes and the NCI.

A critical barrier to characterizing catchment areas lies in establishing reliable denominators. Screening impact cannot be assessed without identifying individuals who are cancer-free and screening guideline-eligible. These data are rarely captured within a single EHR because many healthy individuals have minimal interaction with healthcare systems or receive preventive services elsewhere. Similarly, the cancer-affected population is only partially visible through institutional records, as many patients receive

screening, diagnosis, treatment, or survivorship care across multiple health systems and settings. Without cross-system data interoperability, even basic metrics such as screening rates, treatment penetration, or post-treatment outcomes are difficult to calculate. These gaps produce an “invisible denominator,” obscuring who is reached, who is missing, and whether disparities are narrowing. In practice, this means cancer centers may overestimate performance simply because those not captured in their data systems cannot be counted, which masks unmet need and hinders efforts to demonstrate real improvement in access and equity.

Some EHR systems have introduced cross-institutional data-sharing tools, such as Epic’s Care Everywhere^{16,17}, but these tools primarily support patient-level exchange and are limited in scope. For instance, Care Everywhere is subject to legal, contractual, and technical restrictions that constrain its usability for population health, clinical research, or large-scale data ecosystem development. Broader initiatives such as Health Information Exchanges (HIEs) offer greater potential for data aggregation, yet they vary widely in geographic coverage, data quality, institutional participation, and technical infrastructure^{18,19}. Additional barriers, including residency verification, SDOH collection, record matching across institutions, and disparate data platforms, further complicate efforts. Collectively, these issues contribute to the “invisible denominator” problem, making it difficult to identify populations not currently reached, to target outreach effectively, and to demonstrate progress in reducing cancer disparities—core expectations of NCI designation and renewal.

Patient and community obligations and gaps

Contemporary cancer centers must align their data strategies with the evolving needs of the patients, caregivers, and communities they serve. When used effectively, data can enhance patient experience and support treatment adherence by enabling timely, coordinated, and personalized care. Metrics such as referral completion, time to treatment initiation, and access to navigation services are key indicators of patient-centered care²⁰. At the same time, data must be leveraged to identify and address disparities in access, clinical trial participation, and health outcomes across diverse populations.

From a community perspective, data should guide strategies for outreach, education, and public health partnerships. Community stakeholders, including advocacy groups, faith-based organizations, and local health departments, rely on timely, accessible data to identify unmet needs, tailor interventions, and evaluate engagement efforts. However, current data systems lack standardized fields to capture SDOH that are essential to cancer care, such as transportation challenges or financial hardship. Few systems are equipped to link clinical data with community-level indicators such as cancer screening uptake, early detection rates, or social vulnerability indices. Even when collected, such as through patient or community surveys, these data are rarely integrated systematically with EHRs, cancer registries, or institutional data warehouses, limiting their utility for developing comprehensive, equity-focused strategies.

Meeting patient and community expectations requires not only the collection and integration of diverse data, but also its translation into decision-ready formats. Cancer center leaders, clinicians, and community partners need accessible tools, such as dashboards and interactive visualizations, that transform complex data into interpretable metrics aligned with institutional and community priorities (e.g., referral completion rates, time to treatment initiation, or disparities in screening uptake). Developing and maintaining these tools requires specialized expertise in data visualization and implementation science, ensuring that the data infrastructure supports transparency, accountability, and timely action.

Moreover, patients and communities increasingly expect transparency, accessibility, and agency in how their health data are used. Initiatives such as OpenNotes have shown that providing patients access to their medical records can improve understanding, engagement, and shared decision-making²¹. To support person-centered care, data systems must also be designed to capture the lived experiences of patients through the consistent collection of PROs, particularly regarding symptoms, quality of life, and social needs.

Finally, building culturally competent and inclusive data practices, co-designed with communities and underpinned by strong privacy protections, is essential for establishing trust, improving relevance, and ensuring equitable participation in data-driven initiatives. Without these foundations, cancer centers risk missing critical opportunities to address disparities and fulfill their mission to improve health for all populations they serve.

A learning health system (LHS) and learning health community (LHC) approach for cancer centers

In response to the growing complexity of obligations and persistent data gaps, cancer centers must adopt a more integrative and adaptive approach to data use. The LHS framework provides a strategic model for building data and data science infrastructure that not only meets institutional, state, and federal requirements, but also supports continuous improvement in clinical operations, care delivery, research, education, and meaningful community engagement²². The LHS framework supports a dynamic cycle of data collection, analysis, feedback, and action (i.e., “data-evidence-practice,” see Fig. 2), transforming data from a reporting requirement into a driver of organizational learning and population impact²². A mature LHS should also emphasize sustained engagement with community stakeholders, including patients, caregivers, and community organizations^{23,24}, to collaboratively address unmet social needs and promote health equity. This extension of the LHS to a *Learning Health Community (LHC)* creates an *LHS-LHC dyad*^{25,26}. Realizing the full potential of this model, where best practices are embedded into everyday care delivery²⁷, requires a supporting infrastructure grounded in strong governance, appropriate incentives, and shared values.

From fragmented to integrated, connected, and interoperable data infrastructure

At the core of the LHS-LHC approach is the ability to integrate siloed data systems across the cancer care and research continuum to enable a comprehensive understanding of patient and community needs. By investing in interoperable architectures that link data from EHRs, tumor registries, clinical trial management systems, and other core platforms, cancer centers can begin to address the longstanding data fragmentation challenge. Common data models (e.g., OMOP, PCOR-net), standardized data elements, and consistent terminologies and



Fig. 2 | Learning health system and learning health community framework.

ontologies (e.g., SNOMED CT, LOINC) provide the foundation for harmonizing data across systems. Importantly, achieving interoperability requires engagement with stakeholders upstream of EHR systems. For example, LIMSs must provide results in computable formats rather than static PDF reports, and pathology reports captured via synoptic templates need to retain their structured, computable elements during transmission. Without this upstream standardization, downstream harmonization and integration into EHRs and research platforms remain incomplete. These efforts not only support regulatory compliance and research but also enable real-time analytics to improve patient care, quality improvement, and evidence-based decision-making.

To overcome these limitations, cancer centers should prioritize the systematic mapping of SDOH, PROs, and other patient- or community-reported data elements to reference terminologies such as SNOMED CT and LOINC. Adopting shared standards is essential to making patient- and community-sourced data interoperable, consistently captured across sites, and usable for both research and clinical decision-making. Embedding standardized SDOH and PRO elements directly into workflows lays the foundation for equity-focused analytics and supports the LHS-LHC model at scale.

An integrated data infrastructure also supports the development and deployment of tools for point-of-care decision-making, cohort identification, patient-trial matching, risk stratification, and other critical functions. NLP should be a core component of this infrastructure, enabling the extraction of essential information from unstructured clinical notes (e.g., treatment details, SDOH, and PROs). Only when historically siloed data systems are harmonized at the institutional level can emerging AI technologies be effectively leveraged to identify patterns, detect trends, and perform increasingly complex analytical tasks.

While harmonizing siloed data systems is foundational, integration alone is insufficient to address many high-priority clinical questions in cancer research and care. Critical variables such as cancer recurrence, treatment sequencing, and nuanced imaging interpretations often cannot be derived reliably from structured data or automated pipelines. Instead, these elements require iterative expert curation and sustained interdisciplinary collaboration among oncologists, radiologists, pathologists, and data scientists. Embedding such expert-in-the-loop workflows into the LHS-LHC infrastructure is essential to ensure that curated, high-fidelity variables are available for both research and operational decision-making. Moreover, coupling expert review processes with emerging AI methods (e.g., large language models for clinical abstraction or deep learning for imaging interpretation) offers a path toward scaling these efforts, while preserving the clinical accuracy and contextual nuance that automated methods alone cannot achieve.

One particularly high-impact application of improved data systems is clinical trial accrual, which remains a key metric for maintaining cancer center support grant (CCSG) designation. Cancer centers continue to face challenges in reaching accrual targets, especially for underrepresented populations. By harmonizing EHR, tumor registry, and genomic data, LHS-enabled systems can support more accurate eligibility screening and dynamic trial matching at scale. Commercial platforms (e.g., Paradigm, Triomics) and open-source approaches such as NIH's TrialLLM illustrate the potential of leveraging structured and unstructured data for automated accrual support. Embedding these capabilities into LHS-LHC infrastructures can directly strengthen trial enrollment, expand access for patients, and demonstrate measurable impact for NCI and other stakeholders.

Enhancing responsiveness to state-level needs

The LHS-LHC model can enhance cancer centers' ability to support state-level cancer surveillance, policy development, and care coordination aimed at reducing cancer incidence, improving outcomes, and addressing disparities. By establishing platforms for timely, bidirectional data sharing, cancer centers can assist state health departments more effectively in

monitoring cancer burden across geographic regions and population subgroups. LHS-enabled systems can provide near real-time insights into key metrics, such as screening uptake, referral completion, and treatment initiation, that guide cancer control planning and resource allocation at the state level.

In addition, the LHS-LHC model facilitates care coordination among cancer centers, external providers, and community partners, particularly in rural and underserved areas, through interoperable data systems, shared learning, and sustained stakeholder engagement. Cancer centers can leverage shared dashboards, integrated data platforms, and secure communication tools to track referrals, monitor follow-up, and identify gaps in navigation or specialty care. When effectively implemented, these efforts reinforce the role of cancer centers as regional anchors within broader healthcare delivery ecosystems for equitable, coordinated, and patient-centered care.

Operationalizing expectations for NCI-designated cancer centers

The LHS-LHC approach enables cancer centers to move beyond static data reporting and toward embedded systems that support continuous learning, performance monitoring, and quality improvement. For example, automated analytics pipelines can be developed to monitor screening, diagnostic, and treatment patterns across priority populations and geographic regions, which directly track the metrics emphasized in NCI's CCSG guidelines²⁸.

Importantly, the LHS-LHC model addresses one of the most persistent challenges in catchment area characterization: defining the denominator of individuals affected by or at risk for cancer. By integrating EHR data with claims, census, and community-level data, cancer centers can more accurately estimate the cancer burden, including individuals not yet diagnosed or reached through formal care pathways. This enables more targeted outreach efforts, better resource allocation, and rigorous evaluation of intervention effectiveness, all in alignment with NCI expectations for demonstrating reach, equity, and real-world impact. In addition to improving care delivery and population monitoring, an LHS-LHC infrastructure also strengthens the foundation for research. Standardized, computable data enhance the efficiency of clinical trial recruitment, enable large-scale translational studies, and accelerate discovery science. These downstream benefits create a reinforcing cycle in which research insights inform care, and RWD from care settings generate new avenues for scientific advancement.

While community outreach is a vital component of this strategy, in the context of the U.S. healthcare system, it is unlikely to fully capture the denominator of individuals at risk for cancer. Outreach-based cohorts are often smaller and may not represent the broader catchment population, introducing the potential for selection bias. To mitigate this limitation, cancer centers should integrate outreach data with complementary sources, such as state and regional registries, claims and census data, and HIEs, and apply analytic methods that adjust for underrepresentation. These approaches ensure a more accurate characterization of catchment populations.

Strengthening patient- and community-centered data use

The LHS-LHC approach also helps cancer center to align their data systems with the priorities of patients and communities they serve. To enable timely, coordinated, and personalized care, LHS-informed data infrastructure should integrate SDOH, PROs, and care navigation metrics into routine clinical workflows. This integration supports the identification of barriers to care, disparities in patient experience, and opportunities for continuous improvement in health outcomes. However, PROs, patient surveys, and patient-reported SDOH measures are often incomplete and may not fully represent all patient populations (i.e., selection bias). These limitations highlight the need to supplement such data with additional sources and to apply analytic strategies that adjust for underrepresentation.

From a community perspective, the LHS-LHC model supports co-generation of knowledge by engaging stakeholders and leveraging data and visualization tools, such as community health indicators, to guide meaningful collaboration and strategic planning. For example, cancer centers can

partner with community-based organizations to co-design outreach strategies informed by screening uptake patterns, social vulnerability indices, local health needs, and cultural context. These collaborations enhance the relevance and impact of outreach while strengthening trust and shared accountability between cancer centers and the communities they serve.

Building sustainable capacity for learning and innovation

Central to the LHS-LHC model is the need for sustainable, scalable data and data science infrastructure that is fully embedded within the cancer center's operational fabric. This requires dedicated informatics teams, secure and compliant computing environments, and governance structures that promote transparency, ethical data use, and inclusive stakeholder engagement. Equally important is the development of sustainable cost models to ensure that the personnel, technical, and operational resources needed to build and maintain this infrastructure are realistically planned for and supported over the long term. Data and data science infrastructure must be recognized and resourced as a strategic institutional asset essential to the cancer center's mission in clinical care, research, and community impact.

Ensuring long-term sustainability also requires alignment across institutional priorities, health system investments, and extramural funding sources. Cancer centers should pursue a diversified funding strategy that includes internal institutional investments, state and federal grant support, and partnerships with philanthropic and community-based organizations. Such alignment ensures that data capacity is not only developed and maintained but also continuously enhanced to meet emerging challenges, priorities, and opportunities and accelerate innovation across the cancer care continuum.

Call to action

The transition toward an LHS-LHC model is more than just a hardware and technological upgrade; it is a strategic transformation that requires coordinated action among cancer centers, their affiliated health systems, and external stakeholders. This section outlines key areas where alignment, investment, and sustained commitment are essential to fully realize the promise of an LHS-LHC in advancing cancer care, research, and population health impact.

Aligning cancer centers and health systems

As integral components of larger health systems, cancer centers rely on system-wide technical and governance infrastructures to support the development of a robust LHS-LHC. Health systems must approach cancer data modernization not as a siloed or compliance-driven task, but as a strategic pillar of their enterprise-wide data agenda. This includes prioritizing data interoperability across the enterprise, ensuring timely data delivery across service lines, and aligning institutional goals for data quality, innovation, and equity with cancer center priorities. Cancer centers should be actively represented in shared governance structures that oversee data strategy and investments. These steps are foundational to enabling cancer centers to integrate, analyze, and act on data in ways that support continuous learning and deliver measurable, sustainable outcomes.

State and federal support

State and federal agencies, including public health departments and the NCI, play a critical role in enabling the LHS-LHC transformation. Targeted funding mechanisms are urgently needed to support infrastructure development, integration of diverse data types, and cross-institutional data sharing, activities that are typically underfunded in traditional cancer center operational budgets. In parallel, policies set forth by these agencies should promote the adoption of multi-level metrics (e.g., temporal, descriptive, and outcomes-based statistics across populations and subgroups), support longitudinal data tracking, and incentivize meaningful community engagement. State and federal initiatives should also invest in digital infrastructure, particularly in rural and underserved areas, streamline

duplicative reporting requirements, and create incentives for collaborative data sharing.

These efforts should also be understood in the context of national and industry-led oncology data networks, such as ASCO's CancerLinQ, Flatiron Health, and TriNetX, which already pool large volumes of oncology data across diverse health systems. By strengthening their local data infrastructure, cancer centers not only meet institutional and state obligations but also enhance the quality, representativeness, and analytic power of these larger initiatives. Linking local LHS-LHC efforts with national data collaboratives creates a reinforcing cycle, where improved local data feeds broader networks, and insights generated at scale flow back to improve local care and outcomes. Public-private partnerships and multi-institutional collaboratives can further accelerate innovation and improve scalability and equity across the cancer care continuum.

Building a culture of continuous learning

Adopting the LHS-LHC framework requires not only technological infrastructure but also a cultural and mindset transformation. Cancer centers and their health system partners must cultivate values that prioritize data transparency, inclusive governance, and shared goals and accountability. Administrators, clinicians, researchers, patients, caregivers, and community stakeholders should all be empowered to contribute to the design, implementation, and evaluation of LHS-LHC initiatives. Embedding learning objectives into clinical workflows and research protocols, and fostering bidirectional learning with patients and community partners, are essential steps. By building a culture of continuous learning, cancer centers can ensure that data is not merely collected but actively translated into insights and actions that improve patient care, reduce disparities, and generate generalizable knowledge. Ultimately, this cultural shift supports a more responsive, accountable, and equitable cancer care ecosystem, both within the local communities and across the national landscape.

Building the workforce for an LHS

Achieving the LHS-LHC vision requires not only infrastructure but also a skilled and adaptable workforce. Cancer centers will need to expand capacity in data science, informatics, NLP, and implementation science, while also equipping clinicians, research coordinators, and staff to embed continuous learning into everyday workflows. This includes training in structured data capture, interpretation of real-time analytics, and integration of PROs and SDOH into care delivery. Equally important is fostering new "boundary-spanning" roles that connect clinical teams, data scientists, and community partners to ensure that insights from data translate into improved care delivery and equity. Sustainable staffing models, career development pathways, and institutional incentives will be critical for recruiting, retaining, and supporting the workforce needed to operationalize the LHS-LHC vision. By investing in people as much as in technology, cancer centers can ensure that innovation is scalable, sustainable, and directly tied to improvements in patient outcomes.

Conclusion

Modern cancer centers operate in a rapidly evolving landscape where data is no longer merely an operational necessity but a strategic asset central to care delivery, research, community engagement, and policy influence. The LHS-LHC framework offers a path to meet complex and growing data obligations while simultaneously unlocking the potential of data to drive innovation, advance equity, and support continuous improvement. Realizing this model requires sustained investment in infrastructure, workforce development, and governance, as well as deep alignment among health systems, funders, policymakers, and community partners.

Fragmented and siloed data systems can no longer support the breadth of activities and expectations that define NCI-designated cancer centers. Embracing the LHS-LHC approach enables centers to harmonize data

across domains, promote interdisciplinary collaboration, and cultivate a culture of continuous learning that transforms data into actionable insights. With coordinated action and a shared vision, cancer centers can lead the transformation toward data-enabled oncology, ensuring their efforts yield measurable, lasting impact for the patients and communities they serve.

Data availability

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Received: 14 July 2025; Accepted: 5 December 2025;

Published online: 09 January 2026

References

1. NCI-Designated Cancer Centers. <https://www.cancer.gov/research/infrastructure/cancer-centers> (2012).
2. Pohl, S. A. et al. Evolution of community outreach and engagement at National Cancer Institute-Designated Cancer Centers, an evolving journey. *CA Cancer J. Clin.* **74**, 383–396 (2024).
3. Committee on Improving the Quality of Cancer Care: Addressing The Challenges Of An Aging Population, Board On Health Care Services & Institute Of Medicine. *Delivering High-Quality Cancer Care* (National Academies Press, 2013).
4. National Academies of Sciences, Engineering, and Medicine. *Assessing and advancing progress in the delivery of high-quality cancer care: Proceedings of a Workshop*. (National Academies Press, 2024).
5. Goodby, A. W. & McGinnis, M. Institute of Medicine, Roundtable on Value & Science-Driven Health Care & IOM Roundtable on Evidence-Based Medicine (Series). *Clinical Data as the Basic Staple of Health Learning* (National Academies Press, Washington, 2010).
6. IOM Roundtable on Value & Science-Driven Care & Institute of Medicine. *Integrating Research and Practice* (National Academies Press, 2015).
7. Odedina, F. T., Salinas, M., Albertie, M., Murrell, D. & Munoz, K. Operational strategies for achieving comprehensive cancer center community outreach and engagement objectives: impact and logic models. *Arch. Public Health* **82**, 35 (2024).
8. Braveman, P. & Gottlieb, L. The social determinants of health: it's time to consider the causes of the causes. *Public Health Rep.* **129**(Suppl 2), 19–31 (2014).
9. Cancer Data Sciences Program. <https://www.dficc.harvard.edu/research/research-programs/discipline-based-programs/cancer-data-sciences> (2018).
10. Cancer Data Science. Rogel Cancer Center, University of Michigan <https://www.rogelcancercenter.org/research/shared-resources-and-cores/cancer-data-science> (2015).
11. Institute for Data Science in Oncology. MD Anderson Cancer Center. <https://www.mdanderson.org/research/departments-labs-institutes/institutes/institute-for-data-science-in-oncology.html>.
12. Izquierdo, J. N. & Schoenbach, V. J. The potential and limitations of data from population-based state cancer registries. *Am. J. Public Health* **90**, 695–698 (2000).
13. Yang, D. X. et al. Prevalence of missing data in the National Cancer Database and association with overall survival. *JAMA Netw. Open* **4**, e211793 (2021).
14. Trentham-Dietz, A. et al. Data gaps and opportunities for modeling cancer health equity. *J. Natl. Cancer Inst. Monogr.* **2023**, 246–254 (2023).
15. Goldberg, P. Cancer centers must focus on measurable endpoints in building a diverse workforce, improving health equity. *The Cancer Letter* https://cancerletter.com/health-equity/20231013_3/ (2023).
16. Winden, T. J., Boland, L. L., Frey, N. G., Satterlee, P. A. & Hokanson, J. S. Care everywhere, a point-to-point HIE tool: utilization and impact on patient care in the ED. *Appl. Clin. Inform.* **5**, 388–401 (2014).
17. Ross, M. K. et al. Accuracy of an electronic health record patient linkage module evaluated between neighboring academic health care centers. *Appl. Clin. Inform.* **11**, 725–732 (2020).
18. Dupont, S., Nemeth, J. & Turbow, S. Effects of health information exchanges in the adult inpatient setting: a systematic review. *J. Gen. Intern. Med.* **38**, 1046–1053 (2023).
19. Li, Z., Merrell, M. A., Eberth, J. M., Wu, D. & Hung, P. Successes and barriers of health information exchange participation across hospitals in South Carolina from 2014 to 2020: longitudinal observational study. *JMIR Med. Inform.* **11**, e40959 (2023).
20. Freund, K. M. et al. National Cancer Institute Patient Navigation Research Program: methods, protocol, and measures. *Cancer* **113**, 3391–3399 (2008).
21. Fossa, A. J., Bell, S. K. & DesRoches, C. OpenNotes and shared decision making: a growing practice in clinical transparency and how it can support patient-centered care. *J. Am. Med. Inform. Assoc.* **25**, 1153–1159 (2018).
22. Friedman, C. et al. Toward a science of learning systems: a research agenda for the high-functioning Learning Health System. *J. Am. Med. Inform. Assoc.* **22**, 43–50 (2015).
23. Mullins, C. D., Wingate, L. T., Edwards, H. A., Tofade, T. & Wutoh, A. Transitioning from learning healthcare systems to learning health care communities. *J. Comp. Eff. Res.* **7**, 603–614 (2018).
24. Okun, S. & Goodwin, K. Building a learning health community: by the people, for the people. *Learn. Health Syst.* **1**, e10028 (2017).
25. Palmer, R. C., Ismond, D., Rodriguez, E. J. & Kaufman, J. S. Social determinants of health: future directions for health disparities research. *Am. J. Public Health* **109**, S70–S71 (2019).
26. Palakshappa, D., Miller, D. P. Jr & Rosenthal, G. E. Advancing the learning health system by incorporating social determinants. *Am. J. Manag. Care* **26**, e4–e6 (2020).
27. Davis, F. D., Williams, M. S. & Stamet, R. A. Geisinger's effort to realize its potential as a learning health system: a progress report. *Learn. Health Syst.* **5**, e10221 (2021).
28. PAR-21-321: Cancer Center Support Grants (CCSGs) for NCI-designated Cancer Centers (P30 Clinical Trial Optional). <https://grants.nih.gov/grants/guide/pa-files/PAR-21-321.html>.

Acknowledgements

This work was supported by the Cancer Informatics Shared Resource at the UF Health Cancer Institute. The authors wish to thank the UF Health Cancer Institute Research Development Team for their assistance with graphics.

Author contributions

Y.G. and J.B. conceived the study idea and drafted the manuscript. T.J.G., A.M.I., J.D.L., G.L., A.C.T., L.L., E.C.W., C.M.S., K.H., D.A.H., K.P.L., and R.E.P. provided meaningful inputs, reviewed and edited the manuscript. All authors reviewed and approved the final version of the manuscript for submission.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Yi Guo or Jiang Bian.

Reprints and permissions information is available at

<http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025