Defining next generation medicine for the patient through translational care: big data scientist training enhancement program (BD-STEP)

To cite this article: Connie S E Lee and Louis D Fiore 2018 *Converg. Sci. Phys. Oncol.* **4** 013001

Convergent Science Physical Oncology

OPEN ACCESS

RECEIVED
1 May 2017

REVISED
18 September 2017

ACCEPTED FOR PUBLICATION
10 October 2017

PUBLISHED
22 December 2017

TOPICAL REVIEW

Defining next generation medicine for the patient through translational care: big data scientist training enhancement program (BD-STEP)

Connie S E Lee[1] and Louis D Fiore[2]

[1] Department of Veterans Affairs, Veterans Health Administration, Washington, DC, United States of America
[2] VA Boston Healthcare System, Boston, MA, United States of America

E-mail: connieselee@gmail.com (C S E Lee)

Keywords: learning healthcare system, big data, precision medicine, oncology, data scientist, translational care, patient centered care

Abstract

A rapid increase in the volume and availability of electronic medical record (EMR) data and its potential to positively impact patient care, led to a collaboration between the Department of Veterans Affairs (VA) and the National Cancer Institute (NCI). The result of the collaboration, the big data scientist training enhancement program (BD-STEP) was launched in 2015 with the objective of enabling a transition from the traditional healthcare model of observe and treat to a desired future state of predict and prevent. The BD-STEP Program brings together expertise of clinicians, researchers and data scientists to use EMR data to directly impact patient decision-making. This program represents a first time collaboration in VA of the Office of Academic Affiliations, Employee Education System, and the Office of Research and Development with the intended outcome of bridging the gap between the administrative, clinical, and research worlds to achieving a common goal—to train the next generation of data scientists to understand data in the context of health care delivery. As an interdisciplinary program, BD-STEP will provide data scientists trained to use EMR data to: (1) accelerate learning (2) inform healthcare systems administrators and; (3) empower clinicians to translate findings to improve patient care. This review will provide an overview of the intersection of precision medicine, data science, and patient engagement, and share a vision for how BD-STEP will use a team science approach to deliver personalized care for patients to reduce cost and improve outcomes.

Introduction

The BD-STEP Program was initially conceptualized in 2012 by VA medical centers and regional offices in Washington, DC and Chicago and was formally established in 2015 as an interagency collaborative effort between the VA and the NCI. The primary goal is to train the next generation of data scientists to understand, interpret, and reuse the large clinical data sets available at the VA to improve patient care outcomes. The Veterans Health Administration is America's largest integrated health care system with over 1700 sites of care, with a mission of honoring America's Veterans by providing exceptional care that improves the health and well-being of the over 8.76 million Veterans served each year [1]. There are opportunities within an integrated system like the VA to reuse clinical data to provide evidence based insights that can then be used to improve clinical care [2]. Specifically in the cancer space where the patient's molecular data is increasingly becoming a more important component of defining treatment plans, the VA provides a unique opportunity to explore a 'big data' world unique to this integrated system that is not easily captured in other healthcare systems. Recent advances in molecular profiling of patients are difficult to test and rapidly deploy in the current translational pipeline, and VA offers a potential solution to this problem; the design of the VA provides ample opportunity to obtain real world evidence within a large healthcare system on the analytical and clinical validity and utility of molecular tests, innovation and advancements that have the potential to improve care. External partnerships with private sector partners (i.e. Flat Iron, Seven Bridges, etc) and across the federal agencies (i.e. Food and Drug Administration, Department of Energy, etc) is demonstrative of the recent interest in using the nation's largest healthcare

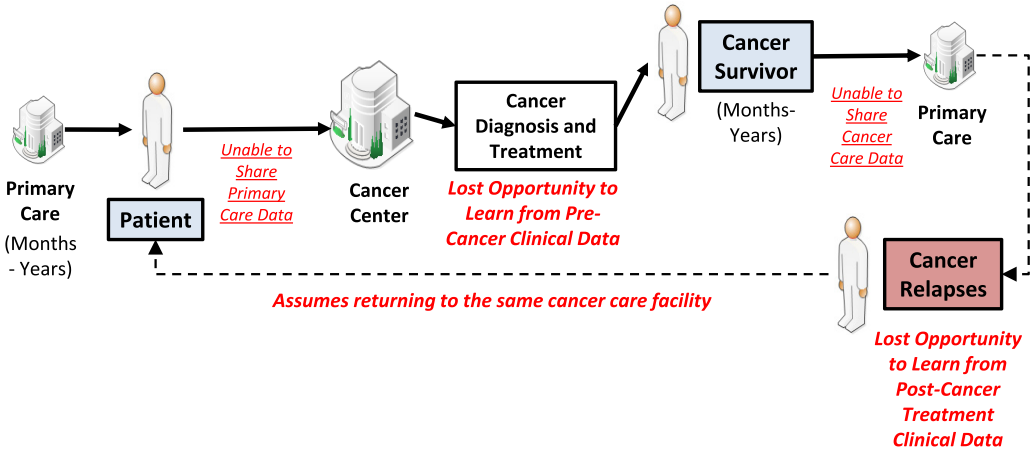

Figure 1. Current cancer patient data paradigm without a learning healthcare system.

Recommendations from the Cancer Moonshot Blue Ribbon Panel Report 2016
A. Establish a network for direct patient involvement
B. Create a clinical trials network devoted exclusively to immunotherapy
C. Develop ways to overcome cancer's resistance to therapy
D. Build a national cancer data ecosystem
E. Intensify research on the major drivers of childhood cancers
F. Minimize cancer treatment's debilitating side effects
G. Expand use of proven cancer prevention and early detection strategies
H. Mine past patient data to predict future patient outcomes
I. Develop a 3-D cancer atlas
J. Develop new cancer technologies

Figure 2. NCI blue ribbon panel report recommendations; highlighted in red are the relevant recommendations that align with the goals of BD-STEP and VA's proposed translational care program specific to cancer.

system to help pave the road ahead for how to truly administer personalized care.

The integration of the EMR across all of the facilities in VA and the low migration rate of VA patients to other healthcare systems makes it possible to follow patients continuously through their life with cancer. This is not typically the case in other healthcare organizations where cancer patients receive their care at a variety of clinical sites that makes it difficult or impossible to aggregate their medical records to obtain a complete patient profile. This limitation is problematic when trying to provide the best care possible for the patient and when trying to gain knowledge by analysis of the aggregated experiences of patients treated by the healthcare system. Figure 1 above depicts the limitations of the current cancer model that is absent of longitudinal data and the capability to share, aggregate and learn from medical record data. The alternative, the learning healthcare system, uses past medical record data to both inform current individual patient care and to create generalizable knowledge that can be shared with other providers [3]. Critical prerequisites to create a learning healthcare system, all present within VA, include availability of complete clinical data, infrastructure to aggregate data and ability to provide an analytic platform and staff trained to do such analyses.

The announcement of the Cancer Moonshot in 2016 was timely and gave traction to the BD-STEP Program. The goal of the Moonshot initiative was 'to dramatically accelerate efforts to prevent, diagnose, and treat cancer—to achieve a decade's worth of progress in 5 years' [4]. A Blue Ribbon Panel of cancer experts was convened to provide recommendations on how to achieve the Vice President's goal. Their report lists 10 recommendations (see figure 2) that include the need to: establish a network for direct patient involvement; to build a national cancer data ecosystem; to mine past patient data to predict future patient outcomes and; to develop new cancer technologies. All of these recommendations were coincidentally aligned closely to the goals of BD-STEP's translational care program activities.

A task force led by former Vice President Joe Biden comprised of federal health agency principals, convened to develop an implementation plan to accomplish these recommendations, proposed five strategic goals: (1) Catalyze New Scientific Breakthroughs; (2) Unleash the Power of Data; (3) Accelerate Bringing New Therapies to Patients; (4) Strengthen Prevention and Diagnosis and; (5) Improve Patient Access and

Care. BD-STEP is in direct in support of strategic goal #2, addressing the need to 'create a knowledgeable, sustainable, and agile biomedical data science workforce' to unleash the power of data [5]. Ultimately, smart collection and use of data can enable the creation of a 'learning health care system for cancer', where as a Nation we learn from the contributed knowledge and experience of every cancer patient [6].

The publicly available report further discusses a multi-pronged approach to addressing the skills and workforce gaps in biomedical data science, and specifically states that support should be included for:

'Interagency fellowship programs to place data science fellows throughout the research and care continuum, building on efforts like VA's Big Data-Scientist Training Enhancement Program (BD-STEP) to expand to other agencies such as NSF, FDA, or CDC. This should include dedicated resources for VA to manage BD-STEP and build data science curricula and community'.

Broad support for the Cancer Moonshot initiative catalyzed growth of BD-STEP over the following year and led to VA's commitment to create a translational care program where the efforts and vision of BD-STEP could be supported. The translational care program proposed to bridge the gap between health care research and operations in a learning healthcare system that enables rapid implementation into the clinical arena of advances in knowledge. Additionally, leadership was looking for ways to establish a learning healthcare system in the Nation's largest integrated health care system beginning with our focus on serving cancer patients. The establishment of a translational care program intended to help VA accomplish this goal.

Learning healthcare system (LHS)

Data collected during routine patient encounters differs from data collected primarily for the purposes of research. Data analysts need to be aware of these differences and cannot assume that clinically derived data provide the full or accurate clinical picture of either individual patients or patient populations. Thus, reuse of EMR data in a learning health system (LHS) requires deep understanding of the purpose of the analysis and the decision making that will result. This requires ongoing collaborations between data analysts and stakeholders who use the results of the analyses with special attention paid to considerations such as public attitudes regarding appropriate use of data [7]. If successfully implemented in oncology, for example, a LHS can provide opportunity to rapidly learn the value/cost tradeoff of tumor mutation detection to advance our understanding of the value of a Precision Oncology Program [8].

Currently in VA, LHS principles are supported and executed through BD-STEP sites. BD-STEP offers an opportunity to share VA clinical data with its data scientist fellows who are able to perform analytics against both the clinical and genomics data to create predictive models to forecast patient outcomes. It's important and essential for VA to lean on external partnerships to maximize the effectiveness of a LHS in VA; these external resources bring expertise and computing capability not currently available within VA to effectively analyze the abundance of data available within the system. The success and outcomes of LHS activities within the BD-STEP network will serve as a use-case for VA to invest in building out the translational care program more broadly across specialty areas outside of oncology.

Precision oncology program (POP)

To date, the benefits of precision medicine are most robust in cancer care. Hence, VA's precision oncology program (POP) was developed within the VA New England Healthcare System and has since been supported by VHA for nationwide deployment and adoption at all of VHA's cancer centers. POP embraces the principles of the rapid learning healthcare system and provides access to modern genomic oncology practice in VA. The primary POP aims include: 1. Determine and disseminate best practices of precision oncology in VA through learning health care system methodologies, 2. Create a cooperative national program to enhance patient and provider engagement and opportunities, 3. Create collaborations between VA, National Cancer Institute, academia, other health care systems and industry to provide cancer patients with state-of-the-art treatments through enhanced clinical care and clinical trial participation, and 4. Monitor implementation of the POP to determine its cost-benefit profile [9]. The expansion of this program nationwide over the past year provides VA an opportunity to discover how to minimize disparities in the delivery of precision oncology across VA. To-date, the program has expanded to over 70 VA sites representing approximately 2000 samples that have been procured as a part of this effort. This innovative interdisciplinary program incorporates LHS principles so that delivery of care is accompanied by analytics that can be applied to decision making for future patients; this makes it a perfect and timely opportunity for BD-STEP trainees to be engaged with as they are able to contribute to the analysis of the data being collected while working collaboratively with the clinicians to help interpret this at the point of care to enhance services and outcomes to the patients they serve.

Big data scientist training enhancement program (BD-STEP)

BD-STEP is a collaboration with VHA's Office of Academic Affiliations, Employee Education System, and the Office of Research and Development in partnership with the National Cancer Institute (NCI) that was created to train a new cohort of talented data

Table 1. Summary of BD-STEP contributors and their mission.

Program	Mission
Office of Academic Affiliations	Develop physician and dentist clinical proficiency, academic scholarship and leadership in emerging areas of health care important to VA; contribute to a talented, well trained recruitment pool for VA; foster an atmosphere of scientific inquiry and excellence within VA.
Employee Education System	Provides dynamic learning to develop a high-performing VHA workforce with the right learning, anytime, anywhere.
Office of Research and Development	To improve Veterans' health and well-being via basic, translational, clinical, health services, and rehabilitative research; to apply scientific knowledge to develop effective individualized care solutions for Veterans; to attract, train, and retain the highest-caliber investigators, and nurture their development as leaders in their fields; and to assure a culture of professionalism, collaboration, accountability, and the highest regard for research volunteers' safety and privacy.
National Cancer Institute, Center for Strategic Scientific Initiatives	Enabling progress in cancer research through advanced technologies, trans-disciplinary programs and resources.

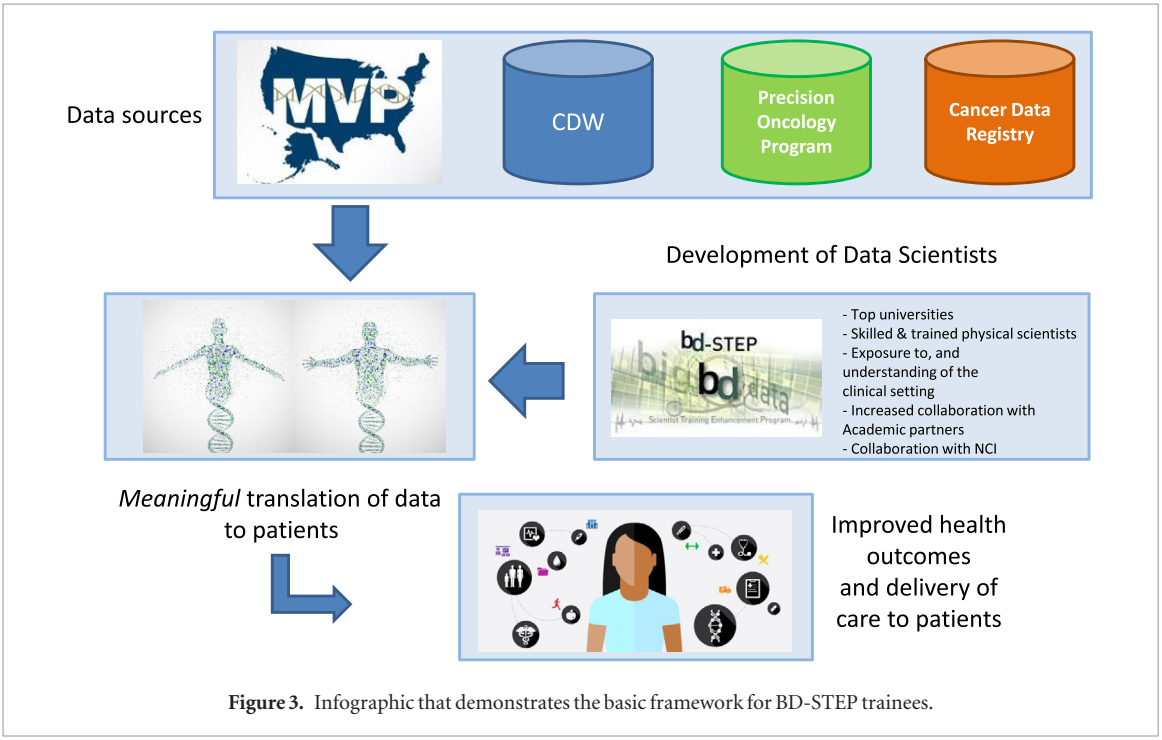

Figure 3. Infographic that demonstrates the basic framework for BD-STEP trainees.

scientists to work collaboratively alongside researchers and clinicians to understand how to harness VA's big data to advance the healthcare of our Veterans (table 1, figure 3). BD-STEP is designed to bring the expertise of highly skilled data scientists-in-training to facilitate the execution of large-scale system changes in clinical care to expeditiously improve patient outcomes. BD-STEP matches VA appointed post-doctoral fellows with VA medical centers across the country to leverage VA data systems, supporting clinically-relevant, year-long training and research opportunities in collaboration with VA clinician scientists. This program also aims to build collaborations and partnerships with academic universities by engaging the post-doctoral fellows' academic preceptors with project development and execution during the fellowship year. As an advanced fellowship program, VA provides stipend support for the appointed fellows of BD-STEP while NCI provides site administrative support to help establish local curriculum and provide travel support for the trainees to attend additional conferences and professional development throughout the course of their year.

BD-STEP's novel approach in leveraging the expertise of each of the discrete VA program offices and NCI's Center for Strategic Scientific Initiatives contributes to the uniqueness of the program curriculum offered to the trainees where they are able to gain the knowledge of experts in each of these areas during their one year curriculum. Each organization contributes a key ingredient and provides valuable input to building a core data scientist workforce that will be trained to understand the critical elements for building LHS capabilities within the healthcare system (figure 4).

Since its creation, BD-STEP has received positive support from the White House Office of Science and Technology Policy for the demonstrated commitment to improving patient outcomes through data science, and for ongoing efforts to improve health outcomes across the country [10]. (see figure 5) Early outcomes

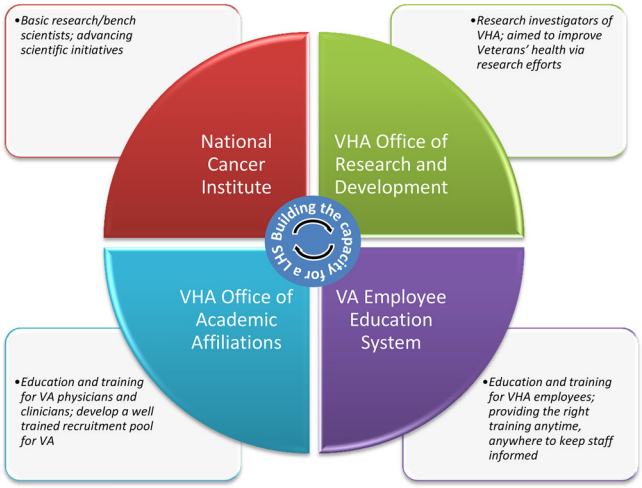

Figure 4. BD-STEP framework: essential Ingredients to establishing a learning healthcare system (LHS).

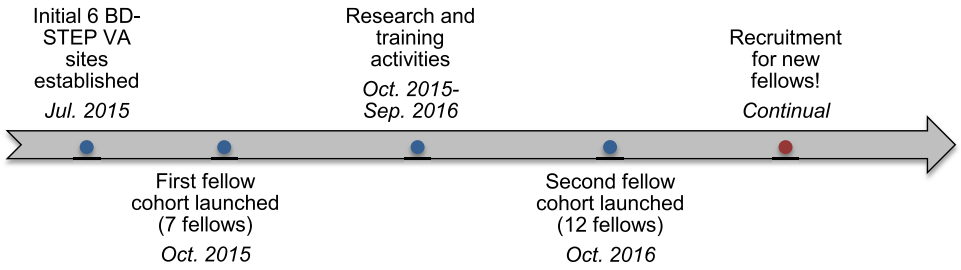

Figure 5. Highlighted BD-STEP activities to-date.

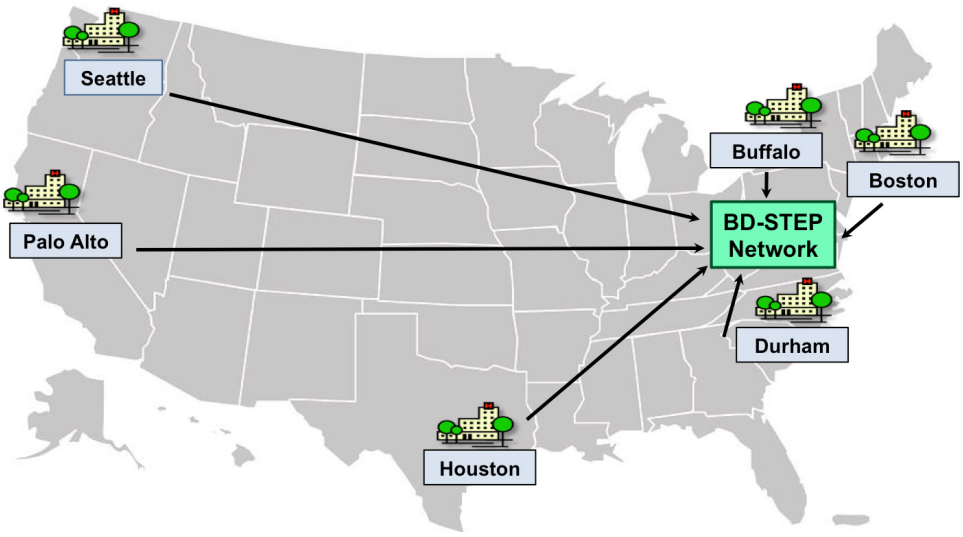

Figure 6. Awarded BD-STEP VA sites for 2015–2016 and 2016–2017 cohorts. Sites are responsible for providing training and support for fellow research projects.

from the first pilot year in Fiscal Year (FY) 2016 (as an unfunded VA program) include the competitive award of BD-STEP site status to six VA Medical Centers with resources and expertise to train data scientists; recruitment and appointment of seven talented BD-STEP trainees from top academic institutions, with backgrounds in physics, computer science, engineering, and epidemiology; and initiation of a diverse set of projects that directly impact cancer research and care. As examples, projects range from analysis of emergency department use by cancer patients to identify gaps in care, to development of predictive models of progression of hepatitis C to liver cancer. Three of the trainees have continued in the program in FY 2017

while the others have assumed permanent faculty, research scientist and postdoc appointments at other academic institutions.

In its first inaugural year as an official VA advanced fellowship program (FY 2017), BD-STEP recruited 13 trainees across the 6 sites that started with the program in October 2016 (figure 6). Their projects range from utilizing basic EHR phenotype data (including imaging data) to accessing the Million Veterans Program genomics data using a series of different analytical tools including predictive modeling and natural language processing.

The promise of this program and demonstrated successes from early outcomes led to the citation of BD-STEP in the Department of Health and Human Services National Institutes of Health, National Cancer Institute's FY 17 budget submission.

'Partnering with the VA to Train Big Data Scientists: The rapidly expanding volume of healthcare data has led to the urgent need to train a diverse pool of scientists with the capability to apply big data to cancer research. To address this need, NCI has partnered with the Veterans Health Administration at the U.S. Department of Veterans Affairs (VA), which is home to the largest integrated medical system in the United States. During FY 2015, this partnership led to the Big Data Scientist Training Enhancement Program (BD-STEP). BD-STEP is training a cadre of diverse junior-level physical scientists and engineers in clinical big data skills. BD-STEP will improve the treatment and care of cancer patients through an expanded ability to manipulate and analyze large-scale patient data sets and to construct new algorithms that advance patient-centered outcomes research'.

As a result of early accomplishments from FY 2016, BD-STEP has secured a partnership with Seven Bridges as a pilot for FY 2017 trainees. Seven Bridges is an organization that accelerates innovation in research and development at the world's largest biopharmaceutical organizations by delivering systems that connect genomic data assets, computational infrastructure, algorithms and teams [11]. Seven Bridges has committed to advising BD-STEP in FY 2017 on the latest science and technology advancements, creating training resources that accommodate trainees from broad backgrounds. Seven Bridges will provide direct mentorship to BD-STEP trainees and has agreed to provide initial sponsorship and mentorship for one postdoctoral trainee to extend biomarker analyses methods to include genomics data with the aim of developing personalized treatments for cancer patients. BD-STEP is committed to furthering public private partnerships such as the one BD-STEP has created with Seven Bridges to enhance industry collaboration and support.

Future goals of BD-STEP include obtaining financial support from VA to invest in this program. This will allow BD-STEP to be sustained and grow. BD-STEP will measure the outcomes and contributions

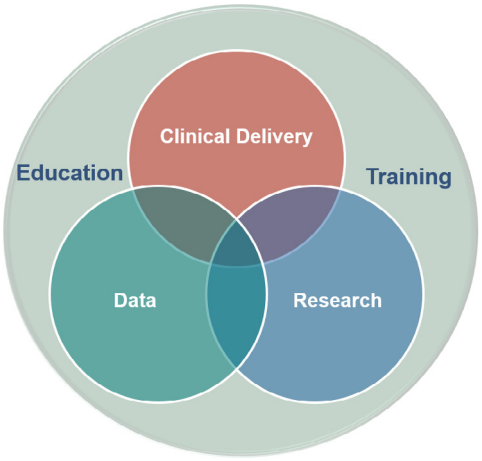

Figure 7. BD-STEP operates at the intersection of clinical delivery, data and research while providing an educational and training opportunity for the trainees, site directors and mentors involved with the program.

towards VA and NCI's strategic goals in the area of furthering cancer research and advancing clinical care in this area (under the advisement of the BD-STEP Advisory Council comprised of leadership from different areas of the organization and NCI) before exploring options to further expand this program into other areas/disciplines for future years.

The program will continue to partner across agencies and academic institutions to build capabilities in new areas. BD-STEP supports the following strategic goals of the Veterans Health Administration, as documented in a FY 2018 VA budget request:

Priority 1: Access: *BD-STEP, in coordination with precision oncology and other precision medicine efforts, provides an unique opportunity to introduce an avenue in which the 'right care can be provided to the right patient at the right time'. Utilizing VA's data infrastructure and their experience with modeling and analytical techniques, data scientists will be afforded the opportunity to work collaboratively with VA clinicians to determine what the best, most effective, and personalized care path is for our Veterans.*

Priority 2: Employee Engagement: *BD-STEP is a program that has been created as a response to meeting the field's unmet need of having data scientist talent explore and partner with VA medical centers. Data scientists currently have no mechanism to expeditiously collaborate and partner with VA. Training fellowships have been limited to those in the medical professions. This program offers an opportunity to expand our capability and lead the health care industry in defining what 'big data' training's potential is and should be about. The success of this program will continue to be dependent on the VAMC's engagement with the trainees.*

Priority 3: High Performance Network: *BD-STEP will build a high performance network of care to best serve Veterans. BD-STEP represents an opportunity for*

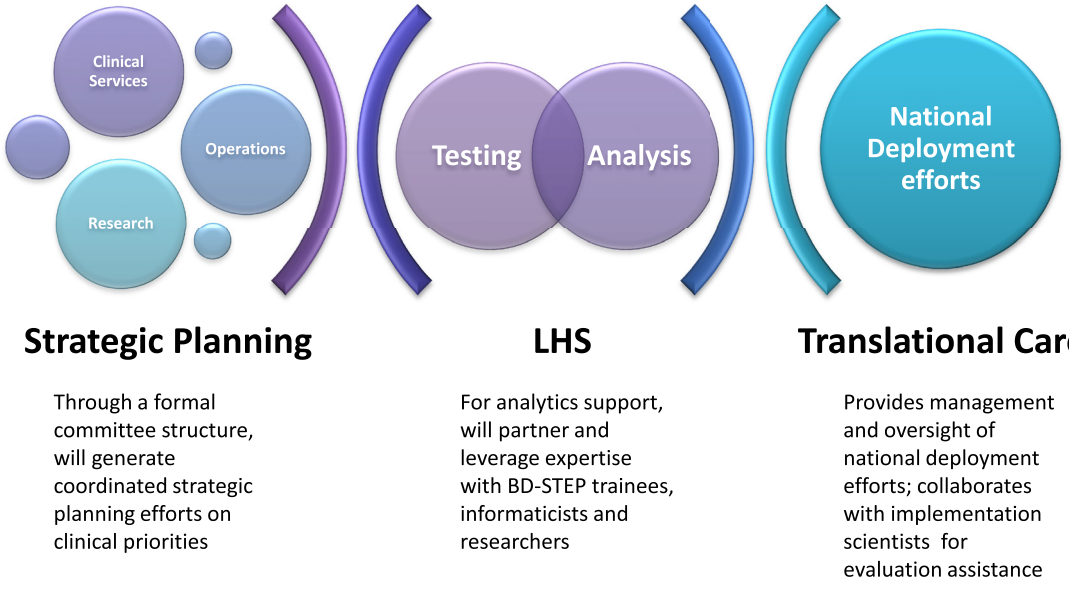

Figure 8. Simplified and coordinated phased approach to national LHS deployment efforts.

the organization to be a leader in the healthcare industry for understanding how big data can be used in healthcare delivery.

Priority 4: Best Practices: *As the program expands, BD-STEP will seek other medical centers who have capacity in precision medicine initiatives and with the infrastructure to support big data scientists, to build on the national 'big data' infrastructure and capacity for sharing educational and training tools throughout the organization.*

Priority 5: Trust in VA Care: *BD-STEP will share results on the successes with collaborating with NCI and other academic universities to provide Veterans with the best care possible.*

Programs such as BD-STEP provide a mechanism that enables the VA to utilize cutting-edge research findings from big data to quickly implement large-scale changes to the administration of clinical care to improve patient outcomes. BD-STEP provides a unique opportunity for data scientists to work in a team science approach in a one year practical experience at a hospital alongside a clinician and a research mentor in the clinic setting to better understand clinical workflow and what clinically relevant questions need to be answered (figure 7). The trainees' time is split between spending time understanding VA's database and infrastructure, and spending time in the clinic to understand workflow and treatment processes.

Understanding how to effectively use 'big data' will impact patient treatment and will result in fewer outpatient visits to the hospital, faster decisions for changing therapies, fewer hospitalizations, and improved, cost-effective health outcomes in VHA and potentially in the nationwide healthcare industry.

The successes of BD-STEP as highlighted within the scope of the former Vice President's Cancer Moonshot Initiative was supported by VA where a need was acknowledged to create a learning healthcare system model that leverages the expertise of clinicians, researchers and data scientists to collaboratively identify an approach to using big data to directly impact point of care clinical care decisions. The establishment of an LHS infrastructure in VA would ensure an oversight and management component for the operational implementation of precision medicine activities, and provides a vision for how to enable a collaborative approach across the silos of clinical delivery, research, and data scientists to effectively deliver a personalized care model for patients. The translational care program will effectively collaborate across the research, clinical and data communities to provide the appropriate training and educational opportunities needed to effectively implement and deploy a learning healthcare system infrastructure throughout the organization.

Patient centered approaches

In an era where consumers are clearly defining and shaping the market for healthcare reform, it is important to provide cancer patients an opportunity to participate in something greater than their individual care treatment plan. Engaging patients in helping to define their treatment plan and educating them on treatment options based on their personal data is a goal of future translational care program efforts. For example, VA has begun to move consented HIPAA-authorized patient data from VA to NCI data repositories for subsequent sharing and analyzing across the research community. This novel approach only requires patients to sign a single Institutional Review Board protocol versus signing multiple papers, and contributes to the greater cancer ecosystem [3]. The patient's contribution towards this effort would

help to enable enhanced data-driven care decisions for themselves as well as other cancer patients within the system.

While more research is needed to determine best practices for engaging patients and to more fully demonstrate possible benefits of the relationship of patient engagement to potential cost savings for the healthcare system, patient engagement has been identified as one proven strategy that achieves the 'triple aim' of improved health outcomes, better patient care, and lower costs. Shared decision-making, active patient engagement at the point of care, and broadening the conversations during patient engagement (to not just focus on the patient's health during the visit, but looking beyond the horizon) lead to better health outcomes at lower costs compared to less engaged patients [12].

What the future holds

Change is needed in the VHA healthcare system to provide a shift in care delivery from the traditional healthcare model of observe and treat to predict and prevent. This shift is necessary given the modernization and advances in emerging technologies used in healthcare. Establishing a learning healthcare system infrastructure for VHA is critical during this time of change in modern day healthcare and will provide guidelines to effect a similar change in the way private sector healthcare is delivered. Figure 8 demonstrates a phased approach to which VA will be able to fully embrace a translational care approach and mindset to leveraging the integrated system structure and way forward with establishing a translational care program.

The possibilities of leveraging big data are endless, but provide us with a unique opportunity to truly think of how to best deliver the next generation of care for our patients. A world in which the data collected within the EMR can be compared against other 'like' patients and leveraged to identify a treatment plan unique to the patient that will most effectively (and economically) treat the disease state is not too far from becoming a reality. Opportunities abound in a system like the VA where, through programs like BD-STEP, a team science approach can be applied to understanding how to maximize the strength of our national treasure. The future of healthcare is dependent on more expeditiously bringing bench side to bed side versus the traditional research model of bench side to bookshelf. This can only be done effectively in a team science approach like the one BD-STEP has been built against, where clinicians, researchers and data scientists work collaboratively together to solve clinically relevant questions.

References to the program throughout the Cancer Moonshot initiative and inclusion in the final task force report demonstrates VA's success in leading the nation in developing the biomedical data science workforce. Thus, BD-STEP now also reflects an important VHA-led initiative that will be expanded across different federal agencies to further support all of the Under Secretary for Health's five priorities, specifically in the areas of best practices and trust in VA Care. By being an exemplar national health care system capable of training the next generation of data scientists to impact Veteran's care, VA will be able to develop, recruit and retain the best and the brightest within the organization.

Author disclosures

The author reports no actual or potential conflicts of interest with regard to this article.

Disclaimer

The contents do not represent the views of the US Department of Veterans Affairs, National Cancer Institute, or the United States Government.

ORCID iDs

Connie S E Lee https://orcid.org/0000-0002-3548-5189

References

[1] www.va.gov/health/ (Accessed: 4 April 2017)
[2] Byrnes N 2014 Data driven healthcare *MIT Technol. Rev. Bus. Rep.* **117** 2
[3] Fiore L D *et al* 2017 Data sharing, clinical trials, and biomarkers in precision oncology: challenges, opportunities, and programs at the Department of Veterans Affairs *Clin. Pharmacol. Ther.* **101** 586–9
[4] The opportunity: ending cancer as we know it (www.medium.com/cancer-moonshot (Accessed: 1 April 2017))
[5] Report of the Cancer Moonshot Taskforce (https://obamawhitehouse.archives.gov/sites/default/files/docs/final_cancer_moonshot_task_force_report_1.pdf (17 October 2016))
[6] Medium.com/cancer-moonshot (Accessed: 1 April 2017)
[7] Deeny S R and Steventon A 2015 Making sense of the shadows: priorities for creating a learning healthcare system based on routinely collected data *BMJ Qual. Saf.* **24** 505–15
[8] Cully M 2015 Advancing precision medicine *in silico Nat. Rev. Drug Discovery* **14** 311
[9] Fiore L D *et al* 2016 Implementation of a precision oncology program as an exemplar of a learning health care system in the VA *Fed. Pract.* **33** S26–30 (www.mdedge.com/fedprac/avaho/article/107141/hospice-palliative-medicine/implementation-precision-oncology-program)
[10] Barbero R and Levy N Increasing the use of data science in US Department of Veterans Affairs Medical Centers (www.whitehouse.gov/blog/2015/10/08/increasing-use-data-science-us-department-veterans-affairs-medical-centers (8 October 2015))
[11] www.sevenbridges.com/ (Accessed: 6 April 2017)
[12] Health policy briefs: patient engagement (www.healthaffairs.org/healthpolicybriefs/brief.php?brief_id=86 (14 February 2013))