Original Paper

Fueling Clinical and Translational Research in Appalachia: Informatics Platform Approach

Alfred A Cecchetti, MScIT, MS, PhD; Niharika Bhardwaj, MBBS, MS; Usha Murughiyan, MBBS; Gouthami Kothakapu, MSCS; Uma Sundaram, MD

Department of Clinical and Translational Science, Joan C. Edwards School of Medicine, Marshall University, Huntington, WV, United States

Corresponding Author:
Alfred A Cecchetti, MScIT, MS, PhD
Department of Clinical and Translational Science
Joan C Edwards School of Medicine
Marshall University
1600 Medical Center Drive
Huntington, WV, 25701
United States
Phone: 1 304 691 1585
Email: cecchetti@marshall.edu

Abstract

Background: The Appalachian population is distinct, not just culturally and geographically but also in its health care needs, facing the most health care disparities in the United States. To meet these unique demands, Appalachian medical centers need an arsenal of analytics and data science tools with the foundation of a centralized data warehouse to transform health care data into actionable clinical interventions. However, this is an especially challenging task given the fragmented state of medical data within Appalachia and the need for integration of other types of data such as environmental, social, and economic with medical data.

Objective: This paper aims to present the structure and process of the development of an integrated platform at a midlevel Appalachian academic medical center along with its initial uses.

Methods: The Appalachian Informatics Platform was developed by the Appalachian Clinical and Translational Science Institute's Division of Clinical Informatics and consists of 4 major components: a centralized clinical data warehouse, modeling (statistical and machine learning), visualization, and model evaluation. Data from different clinical systems, billing systems, and state- or national-level data sets were integrated into a centralized data warehouse. The platform supports research efforts by enabling curation and analysis of data using the different components, as appropriate.

Results: The Appalachian Informatics Platform is functional and has supported several research efforts since its implementation for a variety of purposes, such as increasing knowledge of the pathophysiology of diseases, risk identification, risk prediction, and health care resource utilization research and estimation of the economic impact of diseases.

Conclusions: The platform provides an inexpensive yet seamless way to translate clinical and translational research ideas into clinical applications for regions similar to Appalachia that have limited resources and a largely rural population.

(JMIR Med Inform 2020;8(10):e17962) doi: 10.2196/17962

KEYWORDS

Appalachian region; medical informatics; health care disparities; electronic health records; data warehousing; data mining; data visualization; machine learning; data science

Introduction

Background: Unique Challenges in Appalachia

With regard to health care, Appalachia with its predominantly rural communities is known to have one of the worst outcomes in the United States [1]. This is especially true of southern and central rural Appalachia, which face some of the most severe health disparities in the nation [1]. Over the years, the gap in the overall health between Appalachia and the nation as a whole has continued to grow [2,3]. To close this gap, it is critical to identify the cause of these disparities and direct efforts toward developing necessary interventions to address them.

Such an effort necessitates the adoption of modern technologies such as a centralized research data warehouse to house all data necessary to obtain a comprehensive picture of the health of the Appalachian population before analysis to gain actionable insights can be performed. A centralized data warehouse, once considered strictly a business tool, has evolved into an important instrument for cost containment, tracking of patient outcome, providing clinical decision support at the point of care, improving prognostic accuracy, and facilitating research [4]. Thus, rural academic medical centers have moved toward implementing data warehouse systems that feed analytical systems for research needs [5]. This entails (1) the integration of data from different types of medical settings (ie, multi-institutional) such as hospitals, clinics, and specialty centers; (2) linkage of financial data with clinical data—a well-established practice proven to be pivotal to high-quality care and great economic outcomes [6,7]; and (3) integration of other determinants of health such as environmental [8], social [9], and spiritual factors [10] to create longitudinal health records across the care continuum.

However, there are challenges in creating a multi-institutional data warehouse [11]. The electronic health records (EHRs) do not easily interact with one another due to the use of nonstandard terminologies and difficulty in understanding the flow of information. In addition, significant differences exist between rural and urban health systems [12-16]. Unlike their urban counterparts, health care data in Appalachia are typically fragmented, existing in silos within dissimilar databases, registries, data collections, and departmental systems. With innovations in medical technology, the list of data sources continues to grow, producing unprecedented amounts of data from all aspects of care, including diagnosis, medication, procedures, laboratory test results, imaging data, and patient self-monitoring [17-21]. To complicate matters, the overall health and health behaviors of Appalachians are strongly affected by Appalachia's unique culture, geography, and health system issues [22-24]. Consequently, Appalachian academic medical centers face the complex challenge of collecting, organizing, standardizing, and analyzing these enormous quantities of heterogeneous data originating from a wide variety of sources to address the unmet needs of the population they serve.

Why an Informatics Platform?

Data integration and interoperability have been shown to be key to unlocking these data for data analytics, enabling the development of novel patient management strategies for rural hospitals [25,26] and translational research that leads to new approaches at the bedside for prevention, diagnosis, and treatment of disease, which are essential to improving the health of a population [27-29]. Data analytics, once the domain of the statistician, has now become an equal partner in clinical research and research operations [30,31]. Following the data explosion, data analytics increasingly involves the use of visual analytics tools such as Tableau (Tableau Software Inc) and Power BI (Microsoft Corp) to explore data easily and in a self-service fashion and to clearly and effectively communicate complex ideas [32], especially to those members of the medical community who might not have an intimate understanding of the underlying data. Furthermore, machine learning is gaining importance, especially in the area of predictive analytics, to improve the practice of medicine and to infer potentially innovative risk factors [28,33-35].

However, these applications (eg, data warehouse, data analytics, statistical analysis, machine learning, visual analytics) are generally uncoordinated without any overarching governance. Thus, we developed an informatics platform, that is, a suite of interconnected, coordinated applications hosted within an operational environment [36], called the Appalachian Informatics Platform, in West Virginia—the only state located entirely in Appalachia—that facilitates interoperable access to integrated information, data visualization, and data analytics, thereby functioning as an excellent basis for clinical and translational research to improve health care.

The goal of this study is to describe the structure and process of development of the Appalachian Informatics Platform and demonstrate its value in supporting clinical and translational research.

Methods

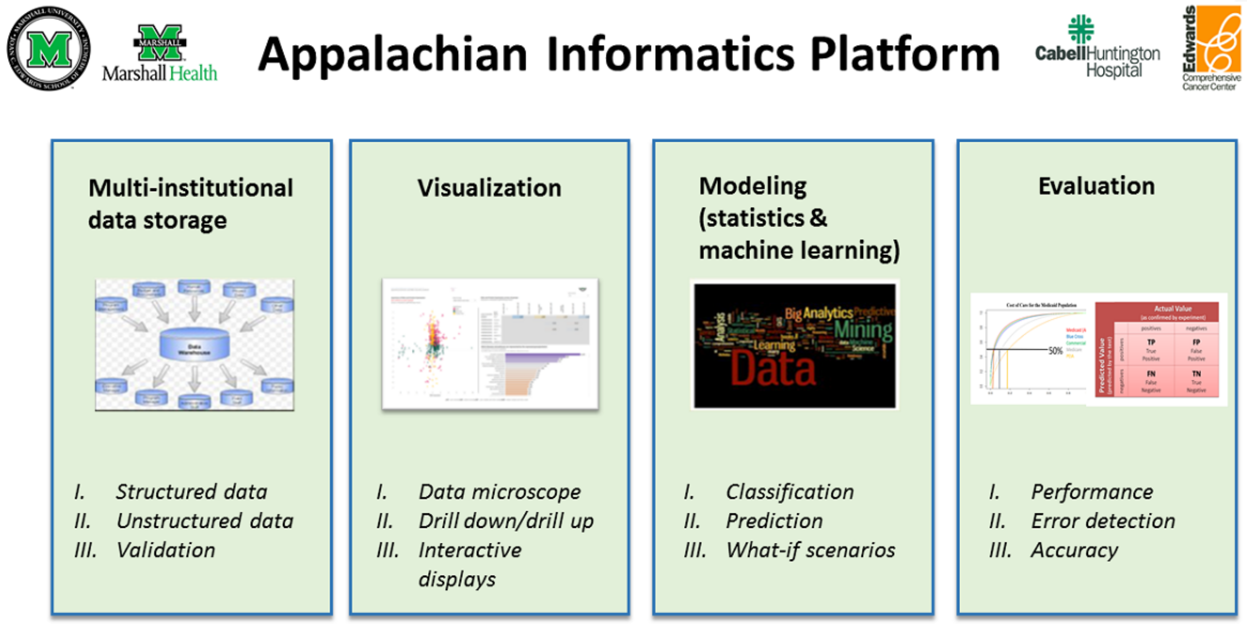
The Appalachian Informatics Platform (Figure 1) is composed of 4 major components: (1) multi-institutional data storage—clinical data warehouse (CDW); (2) modeling (statistical and machine learning); (3) visualization; and (4) evaluation. Each of these components is described in detail in separate sections.

The CDW forms an integral part of the Appalachian Informatics Platform. The Appalachian Informatics Platform, in addition to the CDW, contains embedded data analytics (modeling and evaluation) and interactive visualization tools (eg, Tableau [Tableau Software Inc], Power BI [Microsoft Corp]). Together, these enable the analysis of Appalachian health information to speed up the transition of translational research ideas into clinical practice.

The CDW serves as a secure source of quality data for descriptive, diagnostic, predictive, and prescriptive analytics for research and operational needs. The visual analytics tools enable an initial exploratory analysis of the processed data and the interactive presentation of analytical findings for further analysis and review. Depending on the use case, data can be analyzed using statistical modeling via external (eg, SPSS [IBM Corp], Stata [StataCorp]) or integrated (eg, R [R Foundation for Statistical Computing], Python [Python Software Foundation] in Structured Query Language [SQL]) applications or machine learning modeling. The performance of the resulting models was evaluated using appropriate metrics. Once trained and evaluated, machine learning models can be deployed and stored in the CDW for future use if needed. Furthermore, the stored machine learning models can be continuously evaluated and improved as more data are generated.

XSL•FO
RenderX

Figure 1. Appalachian informatics platform.



The informatics committee governs the access to and utilization of the Appalachian Informatics Platform and ensures adherence to security and privacy rules. In addition, team-building activities are also incorporated into our clinical informatics model to foster the development of an effective clinical informatics team.

Multi-Institutional Data Storage: Appalachian Clinical and Translational Science Institute-Clinical Data Warehouse

The Appalachian Clinical and Translational Science Institute (ACTSI)'s Division of Clinical Informatics solicited buy-in from different entities, namely, Cabell-Huntington Hospital (CHH), Edwards Comprehensive Cancer Institute (ECCC), Marshall Health (MH) practice plan, and Marshall University Joan C Edwards School of Medicine (MU JCESOM), to build the Appalachian Clinical and Translational Science Institute-Clinical Data Warehouse (ACTSI-CDW) in West Virginia. An agreement was created between these entities that provided access to both financial and clinical data.

The multi-institutional CDW contains more than 9 years of billing and clinical data. It comprises relational tables and dimension and fact tables (Online Analytical Processing [OLAP] cube), which enable secure data storage and data access. Designed from the start to facilitate information flow, the CDW can send out a stream of near real-time data that can be used for any authorized research purpose. Documentation includes a data dictionary and flowcharts. Flowcharts follow the patient from admission (or appointment, if outpatient) to discharge (or exit, if outpatient). The data dictionary contains the standardized and source field names, descriptions, and properties along with the associated metadata for the data contained within the data warehouse. For instance, (1) the entry of a patient into any medical service (admission or appointment) was combined with the single term *encounter* and (2) a higher level of precision was introduced by separating patient age into 2 variables, current age or the age when the procedure was performed.

The CDW process is based on an older data warehouse process developed at the University of Pittsburgh [37]. The process is as follows:

1. Data dictionaries are created by recording institutional source field names and field properties and linking them to the standardized CDW names and properties found within the CDW databases. Descriptions of each field (source and CDW) are included.
2. Individual institutional flowcharts show the workflow of the data and the location of the people responsible for the quality of the data, which are also used for quality control purposes.
3. At present, the CDW contains data from 6 institutional software packages hosted in various parts of the country (eg, Cerner data from Kansas City, Missouri; McKesson data from North Druid Hills, Georgia; etc). The data are exported in a standard format (ie, ASCII flat file, XML, etc) and transferred through secure file transfer protocol (eg, Cerberus [Cerberus, LLC]) to the CDW Development server.
4. The data are integrated into the Microsoft SQL databases using Microsoft SQL Server Integration Services (SSIS), a graphical tool that extracts, transforms, and loads (ETL) the data to target schemas that will be used to contain the target data objects: relational tables, dimensions, and cubes. ETL systems enable a smooth migration from one system to another irrespective of the underlying storage system.
5. Conformed dimensions were developed, and patient linkages using various methods (eg, simple heuristics) [38] were also available and made at this time.

XSL•FO
RenderX

6. At present, a transactional grain fact table has been developed, but other fact tables will be created as needed.
7. The CDW contains internal structured billing and EHR data (ie, demographics, encounter details, vitals, medications, procedures, diagnoses, orders, immunizations, laboratory and imaging results, date and time, payee, and provider). It also contains unstructured EHR data (eg, H&P, admission notes, discharge summaries, other clinical notes). These data are received from MH, CHH, and MU JCESOM's ECCC as well as from other outside institutions. In addition, non-EHR data are incorporated using REDCap.
8. Unstructured data are analyzed using text analytics tools, and classification variables based on text mining are incorporated into the CDW.
9. The data structure (OLAP cubes and relational tables), once checked and verified, is transferred from the secure development server to the secure production server for use.
10. Various security measures (eg, IP and password restrictions) are in place to prevent unauthorized use.
11. The CDW structure, which stores multi-institutional medical information, can now provide data for both operational and research analytical model development (statistical or machine learning) using very simple deidentified interfaces (eg, Excel [Microsoft Corp]) or more complex interactive tools (eg, R [R Foundation for Statistical Computing], Tableau [Tableau Software Inc], Power BI [Microsoft Corp], etc). Within the CDW, the data can be manipulated, cleaned, and prepared before the analysis as needed.
12. Structured and unstructured data currently exist within the CDW. Image and BioSample data will soon be incorporated (like the Pittsburgh model), but the full design has not been finalized yet. An *Honest Broker* person assumes control of sample shipping and receiving.
13. Standard Operation Procedures have been developed for administrative and technical areas.
14. The Health Insurance Portability and Accountability Act (HIPAA) guidelines are followed, and protocol to protect patient information has also been implemented.

The CDW is contained within a Microsoft SQL database that can interact with outside objects using other electronic methods such as SignalR, a software library for Microsoft ASP.NET that allows server code to send asynchronous notifications to client-side web applications and SqlDependency, an object that represents a query notification dependency between an application and an instance of SQL server. Objects such as these provide the ability for the data warehouse to interact in real time with the outside regional population using the newest technologies such as Microsoft Machine Learning Server with embedded R or Python procedure coding.

Data Validation

The information derived from multiple data sources can have inconsistencies and missing values because of their heterogeneous nature that needs to be corrected [39-42]. Thus, for each research study, clinical and translational researchers using the data warehouse are required to verify a random sample (calculated on the basis of the size of the study population) of all extracted study data are directly verified at the original data source to ensure data accuracy and validity. Identified errors or omissions are transmitted back to the host systems for correction or inclusion.

Augmenting the CDW Using REDCap

For certain studies, data available in the CDW may not be precise enough or include variables needed to perform this study. For such studies, data can be augmented using data capture tools. One such tool is the Research Electronic Data Capture, or REDCap, a workflow methodology and software solution designed for the rapid development and deployment of electronic data capture tools to support clinical and translational research [43-45].

Our institution has deployed and maintains 2 REDCap servers: secure (located under institutional firewall) and global (outside the firewall). The secure REDCap system is used for storing data considered protected health information (PHI) under HIPAA. The global system, on the other hand, is used to store deidentified or non-PHI data. These data are then transferred to and stored within the multi-institutional data warehouse. This method of augmenting the information pulled from the existing source systems provides research-grade data from outside sources that are normally not contained within a data warehouse.

Visualization

Visualization of information is an excellent method of providing knowledge that can be easily understood by any member of the health care discipline. Within the informatics platform, Tableau provides interactive drill-down and drill-up capabilities for specific projects.

Tableau is a visual analytics tool that provides an interactive method of exploring multidimensional data, optimized from the data warehouse and OLAP data sources. Tableau, using either indexed relational tables or a data cube, can perform associated operations such as slice, dice, roll-up, and drill-down on the data, providing detailed interactive visual overlays that range from the lowest grain of the data to high-level representations of the data. Tableau charts, graphs, filters, and maps can provide visualization of the various subgroups of interest using a storyboard approach that presents a specific question followed by an interactive dashboard that explores that question in detail. The use of visual elements such as logos, pictograms, icons, or pictures into the dashboards, in association with the subgroups, provides easy-to-reference image aids that provide clarity and understanding of complex information. The data warehouse provides the drill-down, drill-up and slice and dice capability, whereas the hub design connects both financial and clinical data to provide a full picture.

The developed interactive dashboards are securely shared with users within a department or a team, as needed, through the use of Tableau Server [46].

Modeling (Statistics and Machine Learning)

The modeling component of the informatics platform supports the construction of tailored regional models (statistical or machine learning) to understand and predict disease and other medical events within this region. EHR is primarily a billing system, research only being a secondary function and, thus, is heterogeneous, incomplete, and noisy [25], leading to

XSL•FO
RenderX

unrepresentative samples, selection bias, and misclassification [47]. During the modeling process, these issues are eliminated or minimized.

To assist in modeling, software packages such as Stata [StataCorp] and SPSS [IBM Corp] and embedded open-source machine learning programs (eg, R [R Foundation for Statistical Computing], Python [Python Software Foundation]) are used. This enables faster and easier development of classification, regression, and clustering algorithms for research use. In addition, we utilize products such as Microsoft's LINQ to electronically gather information and directly incorporate that information into the CDW.

Evaluation

During the modeling process, evaluation of the data set as it relates to the regional population is carried out. Local experts native to this region are asked to evaluate the model from a clinical as well as a financial standpoint. Poverty is endemic within the Appalachian population, and a model that suggests the use of a very expensive medication or procedure over an older but less expensive medication or procedure is unlikely to be used [48]. Thus, the model must take into account whether the patient has the means and access to the recommended medication or procedure [49]. In addition, the willingness of Appalachian medical institutions and health care providers to follow the model's suggestions must also be evaluated.

Once developed, the models were tuned and tested. Location, time of treatment, outside temperature, and other contributory factors available within the CDW were employed to fine-tune the models, as applicable. The performance of the models was measured using the R programming environment using measures such as area under curve, sensitivity, specificity, F_1 score, precision, recall, etc.

Security, Privacy, and the Informatics Committee

Data access and usage are permitted only as described in the mutual agreement between the 3 institutions and are subject to internal security and privacy rules. All data requests must follow the standard operating procedure built on the basis of mutual multi-institutional agreement. Foremost, the researcher must have appropriate credentials and authorization to be able to request for data. If the researcher is authorized to make requests, he or she must obtain the IRB approval for his or her proposed study and submit the IRB proposal and supporting documentation for review by the informatics committee. The informatics committee, independent of the IRB, reviews all requests for data from the data warehouse to ensure compliance with the agreement. If the research project is approved, the research team designated members are scheduled for the deidentified data extraction process.

Team Building

Integral to the informatics platform is team building that builds upon previous work [37]. To facilitate effective team meetings and interprofessional collaboration (local and global) without the need or expense of constant travel, a permanent clinical informatics conference room with a fixed connected computer, an uninterruptable power supply (UPS), a smart board, a camera, and a speaker system, along with a video conferencing system (Zoom) connectivity, was built. This ensures adequate communication among all those involved (ie, team members, users, leadership, etc) and access to resources that would otherwise be unavailable.

Results

Since the implementation of the platform, several studies have been conducted. Each study listed below was approved by the informatics committee, and the deidentified data and platform tools were made available securely to the research team.

To evaluate the functionality and value of this platform, we first analyzed the aggregated data of Medicaid-insured patients across different health systems using the interconnected applications within the platform for population health management. Relevant data were extracted from the CDW, followed by exploratory analysis using a Tableau dashboard. Due to the isolated nature of the study population, regional variables such as distance from the CHH and weather conditions (ie, temperature) were also included. Errors and missing values were identified using the dashboard, and data were subsequently cleaned and prepared. Using these clean data, the regional population was classified into 3 spend categories: *low cost, acute*, and *persistent* subgroups on the basis of the charges accrued. Next, the Charlson Comorbidity Index (CCI) was incorporated into the CDW to predict mortality risk within 1 year of hospitalization for patients with comorbid conditions within each spend category (Table 1) [50,51]. Of these categories, the persistent group had the largest percentage of patients with a high risk of mortality, followed by acute and low cost after excluding the deceased patients (persistent: 898/1247, 72.01%; acute: 2074/6946, 29.86%; low cost: 5130/102,814, 4.99%). The CCI was not very sensitive in predicting the risk of mortality but was very specific and accurate (sensitivity: 896/1512, 59.26%; specificity: 102,905/111,007, 92.7%; accuracy: 103,801/112,519, 92.25%). The effect of distance and weather on the CCI needs further investigation that is being conducted. Adjustments are being made to this standard national index to incorporate other Appalachian characteristics that could improve the sensitivity of this risk scoring system.

This way, the platform has been utilized for a variety of purposes such as increasing knowledge of the pathophysiology of diseases, risk identification, risk prediction, health care resource utilization research, and estimation of the economic impact of diseases to enable data-driven clinical decisions, leading to improved clinical outcomes. Textbox 1 contains a list of studies conducted so far.

XSL•FO
RenderX

Table 1. The 10-year mortality risk predicted using the Charlson Comorbidity Index.

Mortality risk	Deceased, n (%)	Alive, n (%)
High risk	896 (0.80)	8102 (7.20)
Low risk	616 (0.55)	102,905 (91.46)

Textbox 1. Studies conducted using the Appalachian Informatics Platform.

Diagnostic accuracy improvement studies

- Albumin Level as a Risk Marker and Predictor of Peripartum Cardiomyopathy [52]
- Clinical Determinants of Myocardial Injury, Detectable and Serial Troponin Levels Among Patients With Hypertensive Crisis [53]
- Is Fever a Red Flag for Secondary Bacterial Pneumonia During RSV Bronchiolitis [54]
- Metabolic Syndrome: Are Current Colon Cancer Screening Guidelines Enough in a Rural Population? [55]
- Utilization of Appalachian Clinical and Translational Science Institute Data Warehouse to More Accurately Predict Disease Processes Important for Central Appalachia [56]

Resource utilization and financial impact research studies

- Fueling Dementia Research in Appalachia via Appalachian Informatics Platform: A Longitudinal Study [57]
- Hospital Emergency Department Visits For Non-Traumatic Oral Health Conditions [58]

Studies to understand disease pathophysiology

- Serum Calcium Homeostasis and Volume Dynamics in Alzheimer’s Disease and Diabetes Mellitus-2 [59]

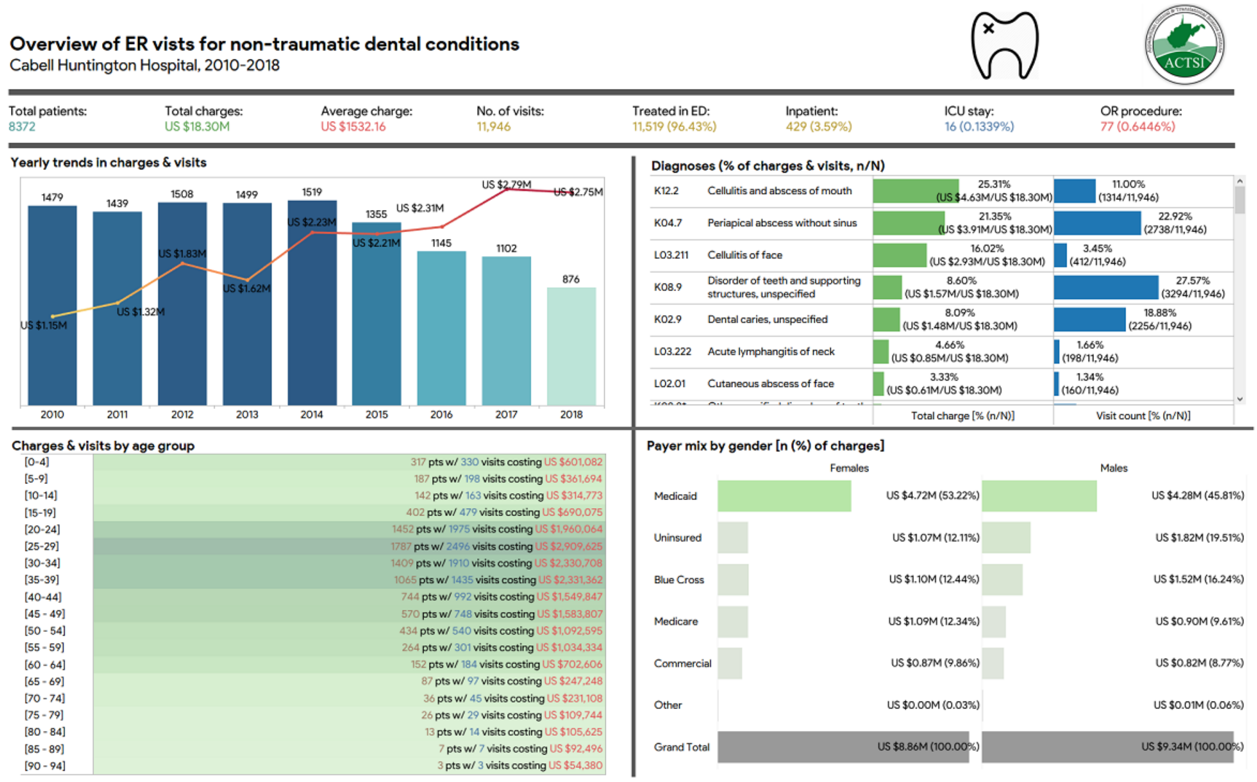
Five studies utilized the platform for risk identification and risk prediction to improve diagnostic accuracy [52-56]. Sundaram et al [56] demonstrated the value of ACTSI-CDW as a primary source to improve the diagnosis of metabolic syndrome, a diagnosis very relevant to the Central Appalachian population. The researchers discovered that utilizing billing codes alone severely underestimated the number of patients with metabolic syndrome by a factor of more than 10 as compared with looking at specific criteria that determine this diagnosis [56]. Another study assessed the relationship between metabolic syndrome and colorectal cancer and found that patients with metabolic syndrome, especially those with insulin resistance, were more likely to have colorectal cancer, indicating the probable need for earlier screening for colorectal cancer in these patients [55]. Elmore et al [54] examined the role of fever in predicting the development of secondary bacterial pneumonia in children with RSV and other viral illnesses. They found that febrile children were 2 to 8 times (RSV, 47/78 vs 27/100; other bronchiolitis, 54/83 vs 7/88) more likely to have secondary bacterial pneumonia compared with afebrile children and, thus, may need to be aggressively evaluated to enable early diagnosis and treatment [54]. Amro et al [52] studied the relationship between hypoalbuminemia and peripartum cardiomyopathy and noted that lower albumin levels were significantly associated with peripartum cardiomyopathy (*P*<.001; odds ratio 0.033, 95% CI 0.034-0.865) and could potentially be used as a risk marker for it. Acosta et al [53] used data from the ACTSI-CDW to identify risk factors (lower BMI, before CHF, and prior use of aspirin) that predict myocardial injury, detectable troponin, and increase in serial troponin levels in patients with hypertensive crisis.

Ferdjallah et al [59] analyzed the data from the ACTSI-CDW to understand how Alzheimer disease and diabetes mellitus affect serum calcium homeostasis and extracellular fluid volume. They observed that acute changes in serum calcium were significantly correlated with changes in extracellular fluid volume in both disease states [59].

The platform has also been applied in 2 studies to assess resource utilization (eg, emergency room, medications, etc) and the financial impact of the disease. For instance, Bhardwaj et al [57] utilized the platform to identify the problems associated with benzodiazepine use in geriatric patients within the health system, such as a higher number of emergency room visits and charges in geriatric patients with dementia plus at least one BZD prescription. In another study [58] that aimed to measure the volume and cost of emergency room use for these conditions and identify the factors that predict such use, the researchers built a dashboard (Figure 2) to easily explore and analyze relevant data on nontraumatic dental conditions that led to emergency room visits and to report the key findings of the study. The authors [58] observed that emergency room visits by uninsured patients were 4 times more likely and those by Medicaid insured 2 times more likely to be for dental problems than Medicare-insured patients.

XSL•FO
RenderX

Figure 2. Tableau dashboard displaying patterns and trends in charges for non-traumatic dental ER visits at Cabell Huntington Hospital between 2010 and 2018. ER: emergency room.



Overview of ER vists for non-traumatic dental conditions

Cabell Huntington Hospital, 2010-2018

Total patients:	Total charges:	Average charge:	No. of visits:	Treated in ED:	Inpatient:	ICU stay:	OR procedure:
8372	US $18.30M	US $1532.16	11,946	11,519 (96.43%)	429 (3.59%)	16 (0.1339%)	77 (0.6446%)

Yearly trends in charges & visits

Diagnoses (% of charges & visits, n/N)

		Total charge [% (n/N)]	Visit count [% (n/N)]
K12.2	Cellulitis and abscess of mouth	25.31% (US $4.63M/US $18.30M)	11.00% (1314/11,946)
K04.7	Periapical abscess without sinus	21.35% (US $3.91M/US $18.30M)	22.92% (2738/11,946)
L03.211	Cellulitis of face	16.02% (US $2.93M/US $18.30M)	3.45% (412/11,946)
K08.9	Disorder of teeth and supporting structures, unspecified	8.60% (US $1.57M/US $18.30M)	27.57% (3294/11,946)
K02.9	Dental caries, unspecified	8.09% (US $1.48M/US $18.30M)	18.88% (2256/11,946)
L03.222	Acute lymphangitis of neck	4.66% (US $0.85M/US $18.30M)	1.66% (198/11,946)
L02.01	Cutaneous abscess of face	3.33% (US $0.61M/US $18.30M)	1.34% (160/11,946)

Charges & visits by age group

[0-4]	317 pts w/ 330 visits costing US $601,082
[5-9]	187 pts w/ 198 visits costing US $361,694
[10-14]	142 pts w/ 163 visits costing US $314,773
[15-19]	402 pts w/ 479 visits costing US $690,075
[20-24]	1452 pts w/ 1975 visits costing US $1,960,064
[25-29]	1787 pts w/ 2496 visits costing US $2,909,625
[30-34]	1409 pts w/ 1910 visits costing US $2,330,708
[35-39]	1065 pts w/ 1435 visits costing US $2,331,362
[40-44]	744 pts w/ 992 visits costing US $1,549,847
[45 - 49]	570 pts w/ 748 visits costing US $1,583,807
[50 - 54]	434 pts w/ 540 visits costing US $1,092,595
[55 - 59]	264 pts w/ 301 visits costing US $1,034,334
[60 - 64]	152 pts w/ 184 visits costing US $702,606
[65 - 69]	87 pts w/ 97 visits costing US $247,248
[70 - 74]	36 pts w/ 45 visits costing US $231,108
[75 - 79]	26 pts w/ 29 visits costing US $109,744
[80 - 84]	13 pts w/ 14 visits costing US $105,625
[85 - 89]	7 pts w/ 7 visits costing US $92,496
[90 - 94]	3 pts w/ 3 visits costing US $54,380

Payer mix by gender [n (%) of charges]

	Females	Males
Medicaid	US $4.72M (53.22%)	US $4.28M (45.81%)
Uninsured	US $1.07M (12.11%)	US $1.82M (19.51%)
Blue Cross	US $1.10M (12.44%)	US $1.52M (16.24%)
Medicare	US $1.09M (12.34%)	US $0.90M (9.61%)
Commercial	US $0.87M (9.86%)	US $0.82M (8.77%)
Other	US $0.00M (0.03%)	US $0.01M (0.06%)
Grand Total	US $8.86M (100.00%)	US $9.34M (100.00%)

Discussion

Utility of the Appalachian Informatics Platform

The Appalachian Informatics Platform has supported several research projects involving the use of different components of the platform, depending on project needs. The studies described reported findings that are seldom reported in this region, enhanced our knowledge of pathophysiology and risk factors, and helped estimate and analyze resource utilization and economic burden of certain diseases within Appalachia using minimal resources (a small IT team and a relatively inexpensive platform).

Before the implementation of the platform, many research studies that followed the patient across multiple care settings or involved analysis of big data were not possible due to the unavailability of technical and economic resources owing to a lack of buy-in from rural health care organizations. As the data existed in silos, there was a lack of standardization and normalization, which resulted in major data inconsistencies. Studies conducted using these disjointed data sets often used unrepresentative small biased samples and had low statistical power and quality.

The introduction of the platform has helped address these issues. It is now easier to pinpoint and correct errors and/or missing values and understand the distribution of data using visual analysis tools. Further, the time needed to conduct these studies from start to finish has been greatly reduced owing to the availability of all applications necessary to complete the study within the platform. This has been specifically useful because many researchers do not have the technical skills needed to perform complex and advanced data analysis, especially on larger data sets.

The paper also revealed that national models do not necessarily perform well when applied to the Appalachian population. The Appalachian Informatics Platform allows for seamless integration of regional variables into the national model, which may improve the performance of these models. For each of the top 10 causes of death in West Virginia in 2017 per the Centers for Disease Control and Prevention [60], a machine learning algorithm was used to predict outcomes on a national level: heart disease [61,62], cancer [63,64], accidents [65,66], respiratory disease [67,68], stroke [69,70], diabetes [71,72], Alzheimer disease [73,74], pneumonia [75,76], kidney disease [77,78], and suicide [79,80]. Each of these cited models could be modified to fit the characteristics of the Appalachian population, especially those characteristics that make this region different in terms of geography, economy, education, and culture from the rest of the United States. The development of these regional models could help rural health general practitioners tackle complex medical conditions without the need for an expensive specialized health care provider nearby [46].

We hope that this paper will help other rural health care organizations, such as ours, that serve underserved populations realize the value and ease of using an informatics platform to conduct research and improving care for their patients despite limited resources.

Ongoing Projects and Future Directions

At present, a model that utilizes embedded data analytics to monitor the side effects of certain types of cancer by ingesting deidentified statements in the regional variety of English

XSL•FO
RenderX

language from patients within this region [81,82] is under development. This model could be used to analyze patient responses at a certain point in time for a cross-sectional study or continuously in real time for a long-term longitudinal study to identify the patients in need of care before their scheduled follow-up visit. The ongoing results from this model would be sent to their health care providers for appropriate actions. In case of an emergency, patient-designated community support networks such as religious or other support groups may be intimated to bring the patient to the emergency department so that the patient can receive timely care.

We plan to expand upon our unified informatics platform to integrate programming applications for the development of state-of-the-art applications targeted specifically toward the unmet health care needs of the Appalachian population.

Conclusions

This paper establishes the value of the Appalachian Informatics Platform in enabling seamless and secure data access, model development through an analytics engine to explore novel and unexpected hypotheses, and simple yet effective communication of all findings via interactive visualization.

The relatively inexpensive nature of such a platform coupled with its demonstrated advantages will hopefully encourage small and midsized rural academic centers, which traditionally have fewer resources than their urban counterparts, to adopt a research informatics platform within their institutions using the template described in this paper as a guide.

Acknowledgments

This work was supported by the National Institutes of Health grants DK-67420, DK-108054, and P20GM121299-01A1 and Veteran’s Administration Merit Review grant BX003443-01 to US and UL1TR00011719 to PK.

Conflicts of Interest

None declared.

References

1. Krometis L, Gohlke J, Kolivras K, Satterwhite E, Marmagas SW, Marr LC. Environmental health disparities in the central Appalachian region of the United States. Rev Environ Health 2017 Sep 26;32(3):253-266. [doi: 10.1515/reveh-2017-0012] [Medline: 28682789]
2. Singh GK, Kogan MD, Slifkin RT. Widening disparities in infant mortality and life expectancy between Appalachia and the rest of the United States, 1990-2013. Health Aff (Millwood) 2017 Aug 1;36(8):1423-1432. [doi: 10.1377/hlthaff.2016.1571] [Medline: 28784735]
3. Marshall J, Thomas L, Lane N. Health disparities in Appalachia. Appalachian Regional Commission. 2017. URL: https://www.arc.gov/wp-content/uploads/2020/06/Health_Disparities_in_Appalachia_August_2017.pdf [accessed 2020-09-25]
4. Foran DJ, Chen W, Chu H, Sadimin E, Loh D, Riedlinger G, et al. Roadmap to a comprehensive clinical data warehouse for precision medicine applications in oncology. Cancer Inform 2017;16:1176935117694349. [doi: 10.1177/1176935117694349] [Medline: 28469389]
5. Kaufman A, Rhyne RL, Anastasoff J, Ronquillo F, Nixon M, Mishra S, et al. Health extension and clinical and translational science: an innovative strategy for community engagement. J Am Board Fam Med 2017 Jan 2;30(1):94-99. [doi: 10.3122/jabfm.2017.01.160119] [Medline: 28062823]
6. Roberts MS, Dreese EM, Hurley N, Zullo N, Peterson M. Blending administrative and clinical needs: the development of a referring physician database and automatic referral letter. Proc Annu Symp Comput Appl Med Care 1991:559-563 [FREE Full text] [Medline: 1807664]
7. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2014;2:3. [doi: 10.1186/2047-2501-2-3] [Medline: 25825667]
8. Ahern MM, Hendryx M. Health disparities and environmental competence: a case study of appalachian coal mining. Environmental Justice 2008 Jun;1(2):81-86. [doi: 10.1089/env.2008.0511]
9. McCulloch B. The relationship of family proximity and social support to the mental health of older rural adults: The Appalachian context. Journal of Aging Studies 1995 Mar;9(1):65-81. [doi: 10.1016/0890-4065(95)90026-8]
10. Simpson MR, King MG. 'God brought all these churches together': issues in developing religion-health partnerships in an Appalachian community. Public Health Nurs 1999 Feb;16(1):41-49. [doi: 10.1046/j.1525-1446.1999.00041.x] [Medline: 10074821]
11. Holve E, Segal C, Lopez MH. Opportunities and challenges for comparative effectiveness research (CER) with electronic clinical data: a perspective from the EDM forum. Med Care 2012 Jul;50 Suppl:S11-S18. [doi: 10.1097/MLR.0b013e318258530f] [Medline: 22692252]
12. Rabinowitz HK, Paynter NP. MSJAMA. The rural vs urban practice decision. J Am Med Assoc 2002 Jan 2;287(1):113. [Medline: 11754723]

XSL•FO
RenderX

13. Anderson AE, Henry KA, Samadder NJ, Merrill RM, Kinney AY. Rural vs urban residence affects risk-appropriate colorectal cancer screening. Clin Gastroenterol Hepatol 2013 May;11(5):526-533 [FREE Full text] [doi: 10.1016/j.cgh.2012.11.025] [Medline: 23220166]
14. Reif S, Whetten K, Ostermann J, Raper JL. Characteristics of HIV-infected adults in the deep south and their utilization of mental health services: a rural vs. urban comparison. AIDS Care 2006;18 Suppl 1:S10-S17. [doi: 10.1080/09540120600838738] [Medline: 16938670]
15. Shubhakaran KP, Khichar RJ. Stroke Management Disparity in Urban Vs Rural Locations. American Academy of Neurology. 2018. URL: https://n.neurology.org/content/stroke-management-disparity-urban-vs-rural-locations [accessed 2020-09-25]
16. Newgard CD, Fu R, Bulger E, Hedges JR, Mann NC, Wright DA, et al. Evaluation of rural vs urban trauma patients served by 9-1-1 emergency medical services. J Am Med Assoc Surg 2017 Jan 1;152(1):11-18 [FREE Full text] [doi: 10.1001/jamasurg.2016.3329] [Medline: 27732713]
17. Chen M, Mao S, Liu Y. Big data: a survey. Mobile Netw Appl 2014 Jan 22;19(2):171-209. [doi: 10.1007/s11036-013-0489-0]
18. Chen M, Hao Y, Hwang K, Wang L, Wang L. Disease prediction by machine learning over big data from healthcare communities. IEEE Access 2017;5:8869-8879. [doi: 10.1109/ACCESS.2017.2694446]
19. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet 2012 May 2;13(6):395-405. [doi: 10.1038/nrg3208] [Medline: 22549152]
20. Wang Y, Kung L, Byrd TA. Big data analytics: understanding its capabilities and potential benefits for healthcare organizations. Technological Forecasting and Social Change 2018 Jan;126:3-13. [doi: 10.1016/j.techfore.2015.12.019]
21. Bhardwaj N, Wodajo B, Spano A, Neal S, Coustasse A. The impact of big data on chronic disease management. Health Care Manag (Frederick) 2018;37(1):90-98. [doi: 10.1097/HCM.0000000000000194] [Medline: 29266087]
22. Elam C. Culture, poverty and education in Appalachian Kentucky. Educ Culture 2002;18(1):10-13.
23. Coyne C, Demian-Popescu C, Friend D. Social and cultural factors influencing health in southern West Virginia: a qualitative study. Prev Chronic Dis 2006 Oct;3(4):A124 [FREE Full text] [Medline: 16978499]
24. Behringer B, Friedell GH. Appalachia: where place matters in health. Prev Chronic Dis 2006 Oct;3(4):A113 [FREE Full text] [Medline: 16978488]
25. Kim J, Ohsfeldt RL, Gamm LD, Radcliff TA, Jiang L. Culture, poverty and education in Appalachian Kentucky hospital characteristics are associated with readiness to attain stage 2 meaningful use of electronic health records. J Rural Health 2017 Jun;33(3):275-283. [doi: 10.1111/jrh.12193] [Medline: 27424940]
26. Mason P, Mayer R, Chien W, Monestime J. Overcoming Barriers to Implementing Electronic Health Records in Rural Primary Care Clinics. The Qualitative Report 2017;22(11):2943-2955 [FREE Full text]
27. Woolf SH. The meaning of translational research and why it matters. J Am Med Assoc 2008 Jan 9;299(2):211-213. [doi: 10.1001/jama.2007.26] [Medline: 18182604]
28. Karstoft K, Galatzer-Levy IR, Statnikov A, Li Z, Shalev AY, Members of Jerusalem Trauma OutreachPrevention Study (J-TOPS) group. Bridging a translational gap: using machine learning to improve the prediction of PTSD. BioMed Central Psychiatry 2015 Mar 16;15:30 [FREE Full text] [doi: 10.1186/s12888-015-0399-8] [Medline: 25886446]
29. Ethier J, Curcin V, Barton A, McGilchrist MM, Bastiaens H, Andreasson A, et al. Clinical data integration model. Core interoperability ontology for research using primary care data. Methods Inf Med 2015;54(1):16-23. [doi: 10.3414/ME13-02-0024] [Medline: 24954896]
30. Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. Health Aff (Millwood) 2014 Jul;33(7):1123-1131. [doi: 10.1377/hlthaff.2014.0041] [Medline: 25006137]
31. Handelsman D. Applying Business Analytics to Optimize Clinical Research Operations. SAS Institute. 2012. URL: https://support.sas.com/resources/papers/proceedings12/171-2012.pdf [accessed 2020-09-28]
32. Simpao AF, Ahumada LM, Gálvez JA, Rehman MA. A review of analytics and clinical informatics in health care. J Med Syst 2014 Apr;38(4):45. [doi: 10.1007/s10916-014-0045-x] [Medline: 24696396]
33. Iwabuchi SJ, Liddle PF, Palaniyappan L. Clinical utility of machine-learning approaches in schizophrenia: improving diagnostic confidence for translational neuroimaging. Front Psychiatry 2013;4:95 [FREE Full text] [doi: 10.3389/fpsyt.2013.00095] [Medline: 24009589]
34. Ainali C. Machine learning for translational medicine. King's College London (University of London). 2013. URL: https://kclpure.kcl.ac.uk/portal/files/31802684/2013_Ainali_Chrysanthi_0829730_ethesis.pdf [accessed 2020-09-28]
35. Jiang M, Chen Y, Liu M, Rosenbloom ST, Mani S, Denny JC, et al. A study of machine-learning-based approaches to extract clinical entities and their assertions from discharge summaries. J Am Med Inform Assoc 2011;18(5):601-606 [FREE Full text] [doi: 10.1136/amiajnl-2011-000163] [Medline: 21508414]
36. Sittig DF, Hazlehurst BL, Brown J, Murphy S, Rosenman M, Tarczy-Hornoch P, et al. A survey of informatics platforms that enable distributed comparative effectiveness research using multi-institutional heterogenous clinical data. Med Care 2012 Jul;50 Suppl:S49-S59 [FREE Full text] [doi: 10.1097/MLR.0b013e318259c02b] [Medline: 22692259]
37. Cecchetti A, Parmanto B, Vecchio M, Ahmad S, Buch S, Zgheib NK, et al. Team building: electronic management-clinical translational research (eM-CTR) systems. Clin Transl Sci 2009 Dec;2(6):449-455 [FREE Full text] [doi: 10.1111/j.1752-8062.2009.00157.x] [Medline: 20443940]

XSL•FO
RenderX

38. Weber SC, Lowe H, Das A, Ferris T. A simple heuristic for blindfolded record linkage. J Am Med Inform Assoc 2012 Jun;19(e1):e157-e161 [FREE Full text] [doi: 10.1136/amiajnl-2011-000329] [Medline: 22298567]
39. Palma G. Electronic Health Records: The Good, the Bad and the Ugly. Becker's Health IT. 2013 Oct 14. URL: https://www.beckershospitalreview.com/healthcare-information-technology/electronic-health-records-the-good-the-bad-and-the-ugly.html [accessed 2020-09-25]
40. Kaufman KR, Hyler SE. Problems with the electronic medical record in clinical psychiatry: a hidden cost. J Psychiatr Pract 2005 May;11(3):200-204. [doi: 10.1097/00131746-200505000-00008] [Medline: 15920394]
41. Shin EY, Ochuko P, Bhatt K, Howard B, McGorisk G, Delaney L, et al. Errors in electronic health record-based data query of statin prescriptions in patients with coronary artery disease in a large, academic, multispecialty clinic practice. J Am Heart Assoc 2018 Apr 12;7(8). [doi: 10.1161/JAHA.117.007762] [Medline: 29650707]
42. Goodloe R, Farber-Eger E, Boston J, Crawford DC, Bush WS. Reducing clinical noise for body mass index measures due to unit and transcription errors in the electronic health record. American Medical Informatics Association Jt Summits Transl Sci Proc 2017 Jul 26;2017:102-111 [FREE Full text] [Medline: 28815116]
43. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap): a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
44. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
45. Harris PA. Research electronic data capture (REDCap) - planning, collecting and managing data for clinical and translational research. BioMed Central Bioinformatics 2012 Jul 31;13(S12). [doi: 10.1186/1471-2105-13-s12-a15]
46. Cecchetti AA. Why Introduce Machine Learning To Rural Health Care? Marshall Journal of Medicine 2018 Apr;4(2) [FREE Full text] [doi: 10.18590/mjm.2018.vol4.iss2.2]
47. McDonald HI, Shaw C, Thomas SL, Mansfield KE, Tomlinson LA, Nitsch D. Methodological challenges when carrying out research on CKD and AKI using routine electronic health records. Kidney Int 2016 Nov;90(5):943-949 [FREE Full text] [doi: 10.1016/j.kint.2016.04.010] [Medline: 27317356]
48. Pierce C, Scherra E. The Challenges of Data Collection in Rural Dwelling Samples. OJRNHC 2004 Dec;4(2):25-30. [doi: 10.14574/ojrnhc.v4i2.197]
49. Verby JE. Patients' and physicians' views of health in a rural area. Acad Med 1989 Nov;64(11):665-666. [doi: 10.1097/00001888-198911000-00009] [Medline: 2803426]
50. Hendryx M, Ahern MM, Nurkiewicz TR. Hospitalization patterns associated with Appalachian coal mining. J Toxicol Environ Health A 2007 Dec;70(24):2064-2070. [doi: 10.1080/15287390701601236] [Medline: 18049995]
51. Ortmeyer CE, Costello J, Morgan WK, Swecker S, Peterson M. The mortality of Appalachian coal miners, 1963 to 1971. Arch Environ Health 1974 Aug;29(2):67-72. [doi: 10.1080/00039896.1974.10666535] [Medline: 4835173]
52. Amro A, Baez GA, Koromia GA, Bhardwaj N, Aguilar R, El-Hamdani M, et al. Albumin level as a risk marker and predictor of peripartum cardiomyopathy. Journal of the American College of Cardiology 2019 Mar;73(9):835 [FREE Full text] [doi: 10.1016/s0735-1097(19)31442-1]
53. Acosta G, Amro A, Aguilar R, Abusnina W, Bhardwaj N, Koromia GA, et al. Clinical determinants of myocardial injury, detectable and serial troponin levels among patients with hypertensive crisis. Cureus 2020 Jan 27;12(1):e6787 [FREE Full text] [doi: 10.7759/cureus.6787] [Medline: 32140347]
54. Elmore D, Yaslam B, Putty K, Magrane T, Abadir A, Bhatt S, et al. Is Fever a Red Flag for Bacterial Pneumonia in Children With Viral Bronchiolitis? Glob Pediatr Health 2019;6:2333794X19868660. [doi: 10.1177/2333794X19868660] [Medline: 31431903]
55. Bhardwaj N, Sundaram S, Carter L, Cecchetti A, Sundaram U. Tu1801 metabolic syndrome: are current colon cancer screening guidelines enough in a rural population? Gastroenterology 2020 May;158(6):S-1167 [FREE Full text] [doi: 10.1016/s0016-5085(20)33589-7]
56. Sundaram S, Bhardwaj N, Gress T, Cecchetti A. Utilization of Appalachian Clinical and Translational Science Institute Data Warehouse to more accurately predict disease processes important for central Appalachia. In: 12th Annual CCTS Spring Conference.: University of Kentucky; 2017 Presented at: CCTS'17; March 30, 2017; Lexington, KY URL: https://www.ccts.uky.edu/media/913
57. Bhardwaj N, Cecchetti AA, Murughiyan U, Neitch S. Analysis of Benzodiazepine prescription practices in elderly Appalachians with dementia via the Appalachian informatics platform: longitudinal study. J Med Internet Res Med Inform 2020 Aug 4;8(8):e18389 [FREE Full text] [doi: 10.2196/18389] [Medline: 32749226]
58. Khanna R, Gress T, Cecchetti A. Hospital Emergency Department Visits For Non-Traumatic Oral Health Conditions. National Oral Health Conference. 2018. URL: http://www.nationaloralhealthconference.com/pdfs/2018-poster-abstracts.pdf [accessed 2020-09-29]
59. Ferdjallah M, Driscoll H. Serum Calcium Homeostasis and Volume Dynamics in Alzheimer’s Disease and Diabetes Mellitus-2. The Health Science Center 32nd Annual Research Day at Marshall University. 2020. URL: https://jcesom.marshall.edu/media/58548/9110_researchsyllabus_2020.pdf [accessed 2020-09-29]

XSL•FO
RenderX

60. Stats of the State of West Virginia. Centers for Disease Control and Prevention. 2017. URL: https://www.cdc.gov/nchs/pressroom/states/westvirginia/westvirginia.htm [accessed 2020-09-25]
61. Patil P, Kinariwala S. Automated Diagnosis of Heart Disease using Random Forest Algorithm. GitHub. 2017. URL: https://github.com/mbbrigitte/Predicting_heart_disease_UCI/blob/master/heartdisease_UCI.Rmd [accessed 2020-09-25]
62. Sreejith S, Rahul S, Jisha R. A real time patient monitoring system for heart disease prediction using random forest algorithm. In: Advances in Signal Processing and Intelligent Recognition Systems. Cham, UK: Springer; Dec 25, 2015:485-500.
63. Tanaka T, Voigt MD. Decision tree analysis to stratify risk of de novo non-melanoma skin cancer following liver transplantation. J Cancer Res Clin Oncol 2018 Mar;144(3):607-615. [doi: 10.1007/s00432-018-2589-5] [Medline: 29362916]
64. Paxton R, Zhang L, Wei C, Price D, Zhang F, Courneya KS, et al. An exploratory decision tree analysis to predict physical activity compliance rates in breast cancer survivors. Ethn Health 2019 Oct;24(7):754-766. [doi: 10.1080/13557858.2017.1378805] [Medline: 28922931]
65. Moreno HA. Predicting car accidents in Barcelona using a Random Forest model. Universitat Politècnica de Catalunya. 2017 Jan. URL: http://hdl.handle.net/2117/100298 [accessed 2020-09-25]
66. Xiaohui J. Forecast model of road traffic accidents based on LS-SVM with grey correlation analysis. Appl Res Comput 2016;3:038 [FREE Full text]
67. Khatri KL, Tamil LS. Early detection of peak demand days of chronic respiratory diseases emergency department visits using artificial neural networks. IEEE J Biomed Health Inform 2018 Jan;22(1):285-290. [doi: 10.1109/jbhi.2017.2698418]
68. Oijuela-Canon AD, Gomez-Cajas DF, Sepulveda-Sepulveda A. Respiratory Diseases Discrimination Based on Acoustic Lung Signals and Neural Networks. In: 20th Symposium on Signal Processing, Images and Computer Vision. 2015 Presented at: STSIVA'15; September 2-4, 2015; Bogota, Colombia. [doi: 10.1109/stsiva.2015.7330461]
69. McKinley R, Häni L, Gralla J, El-Koussy M, Bauer S, Arnold M, et al. Fully automated stroke tissue estimation using random forest classifiers (FASTER). J Cereb Blood Flow Metab 2017 Aug;37(8):2728-2741 [FREE Full text] [doi: 10.1177/0271678X16674221] [Medline: 27798267]
70. Chen L, Bentley P, Rueckert D. A novel framework for sub-acute stroke lesion segmentation based on random forest. In: Ischemic Stroke Lesion Segmentation. 2015 Presented at: ISLES'15; October 5, 2015; Munich, Germany p. 17-20 URL: http://www.isles-challenge.org/ISLES2015/pdf/20150930_ISLES2015_Proceedings.pdf#page=17
71. Xu W, Zhang J, Zhang Q, Wei X. Risk Prediction of Type II Diabetes Based on Random Forest Model. In: Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics. 2017 Presented at: AEEICB'17; February 27-28, 2017; Chennai, India. [doi: 10.1109/aeeicb.2017.7972337]
72. Shukla N, Arora M. International Journal of Computer Sciences and Engineering 2016;4(7):101-104.
73. Basaia S, Agosta F, Wagner L, Canu E, Magnani G, Santangelo R, Alzheimer's Disease Neuroimaging Initiative. Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. Neuroimage Clin 2019;21:101645 [FREE Full text] [doi: 10.1016/j.nicl.2018.101645] [Medline: 30584016]
74. Lu D, Popuri K, Ding GW, Balachandar R, Beg MF, Alzheimer’s Disease Neuroimaging Initiative. Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structural mr and FDG-PET images. Sci Rep 2018 Apr 9;8(1):5697 [FREE Full text] [doi: 10.1038/s41598-018-22871-z] [Medline: 29632364]
75. Wiemken TL, Furmanek SP, Mattingly WA, Guinn BE, Cavallazzi R, Fernandez-Botran R, et al. Predicting 30-day mortality in hospitalized patients with community-acquired pneumonia using statistical and machine learning approaches. The University of Louisville Journal of Respiratory Infections 2017;1(3) [FREE Full text] [doi: 10.18297/jri/vol1/iss3/10/]
76. Menéndez Villanueva R. [The diagnostic evaluation of rapid sputum technics for Pneumococcus in community-acquired pneumonia. The usefulness of Bayes theorem for clinical application]. Arch Bronconeumol 1995;31(7):317-322. [Medline: 8777525]
77. B.V R, Sriraam N, Geetha M. Classification of non-chronic and chronic kidney disease using SVM neural networks. 2017 Dec 31;7(1.3):191-194 [FREE Full text] [doi: 10.14419/ijet.v7i1.3.10669]
78. Annapoorani J, Gnanaselvam C. Enhancing prediction accuracy of chronic kidney disease using neural networks. Automation and Autonomous System 2018;10(1):10-15. [doi: 10.36039/AA012018003]
79. Ayat S, Farahani HA, Aghamohamadi M, Alian M, Aghamohamadi S, Kazemi Z. A comparison of artificial neural networks learning algorithms in predicting tendency for suicide. 2012 Jul 26;23(5):1381-1386 [FREE Full text] [doi: 10.1007/s00521-012-1086-z]
80. Bhat H, Goldman-Mellor S. Predicting adolescent suicide attempts with neural networks. arXiv 2017 Dec 01:- epub ahead of print [FREE Full text]
81. Wolfram W, Christian D. Appalachian speech. Center for Applied Linguistics. 1976. URL: http://files.eric.ed.gov/fulltext/ED130511.pdf [accessed 2020-09-25]
82. Luhman R. Appalachian english stereotypes: language attitudes in Kentucky. Lang Soc 2008 Dec 18;19(3):331-348. [doi: 10.1017/s0047404500014548]

Abbreviations

ACTSI: Appalachian Clinical and Translational Science Institute

XSL•FO
RenderX

CCI: Charlson Comorbidity Index
CDW: clinical data warehouse
CHH: Cabell-Huntington Hospital
ECCC: Edwards Comprehensive Cancer Institute
EHR: electronic health record
ETL: extract, transform, and load
HIPAA: Health Insurance Portability and Accountability Act
MH: Marshall Health
MU JCESOM: Marshall University Joan C Edwards School of Medicine
OLAP: Online Analytical Processing
PHI: protected health information
SQL: structured query language

Edited by G Eysenbach; submitted 24.01.20; peer-reviewed by T Muto, B Fan, N Mohammad Gholi Mezerji; comments to author 12.05.20; revised version received 24.07.20; accepted 27.07.20; published 14.10.20

Please cite as:
Cecchetti AA, Bhardwaj N, Murughiyan U, Kothakapu G, Sundaram U
Fueling Clinical and Translational Research in Appalachia: Informatics Platform Approach
JMIR Med Inform 2020;8(10):e17962
URL: http://medinform.jmir.org/2020/10/e17962/
doi: 10.2196/17962
PMID: 33052114

©Alfred A Cecchetti, Niharika Bhardwaj, Usha Murughiyan, Gouthami Kothakapu, Uma Sundaram. Originally published in JMIR Medical Informatics (http://medinform.jmir.org), 14.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on http://medinform.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX