Dissecting interRAI Instrument Data Using Visual and Predictive Analytics

Waqar Haque[1][a], Shannon Freeman[2][b] and Piper Jackson[3][c]
[1]*Department of Computer Science, University of Northern British Columbia, Prince George, Canada*
[2]*School of Nursing, University of Northern British Columbia, Prince George, Canada*
[3]*Department of Computing Science, Thompson Rivers University, Kamloops, Canada*

Keywords: Visual Analytics, Predictive Analytics, Geriatrics, Long-term Care, interRAI Assessment Tools.

Abstract: Healthcare data for older adults is often collected through globally standardized instruments and resides in multiple disparate database systems. For gaining insights into this data, an interactive platform has been developed which allows visualization of several actionable key performance indicators along multiple dimensions. The health assessment data was collected from persons receiving community home care services as well as from persons residing in long-term care facilities. The top-level reports provide aggregations across geographical regions at the health service delivery area level with capability to drill down to finer granularity for metrics of interest. By revealing hidden patterns embedded in data, the stakeholders can make informed decisions pertaining to resource allocation and better patient care. The drill-down and drill-through reports include demographics, quality of life, medications, health conditions and disease diagnoses, comorbidities, health service usage, and patient journey across care settings. A predictive model to accurately estimate resource requirements at the time of admission was also developed for data-driven triaging.

1 INTRODUCTION

In recent years, visual analytics has quickly become a staple in various data-rich fields, including health care. This has undeniably led to economic and other benefits by unveiling hidden information and/or patterns in data thus leading to informed administrative decisions, overall better treatment, and more desirable outcomes for patients. interRAI (interRAI, n.d.) is a collaborative global network of researchers and practitioners in over 35 countries who have developed health assessment instruments. These instruments have been mandated for use by several governments, including Canada where the interRAI MDS 2.0 (Morris, et al., 1990) and the interRAI Home Care (Morris, et al., 1997) assessment instruments are widely used for persons receiving long-stay community based home care services and for those receiving support in a long-term care facility. These instruments are based on rigorous research which allows standardized assessment protocols and outcome measures. The data used in our research and the selected key performance indicators (KPIs) are based on these instruments. This research also engaged patient partners for valuable input when selecting relevant metrics and assessing ease of use.

The outcome from this research is twofold. First, the developed platform demonstrates how data visualization can inform clinical and health systems in decision-making for persons receiving care in home and/or long-term care settings. The reports present aggregated results based on analyses conducted on de-identified secondary data. Second, a predictive model has been developed to predict resource requirements when an elderly patient is admitted to a care facility.

2 RELATED WORK

A diverse number of studies where results were presented using visual platforms revealed patterns

[a] https://orcid.org/0000-0002-6921-8097
[b] https://orcid.org/0000-0002-8129-6696
[c] https://orcid.org/0000-0002-7025-5063

Haque, W., Freeman, S. and Jackson, P.
Dissecting interRAI Instrument Data Using Visual and Predictive Analytics.
DOI: 10.5220/0011719100003476
In *Proceedings of the 9th International Conference on Information and Communication Technologies for Ageing Well and e-Health (ICT4AWE 2023)*, pages 110-117
ISBN: 978-989-758-645-3; ISSN: 2184-4984

which were otherwise indiscernible. For instance, the application of visualization techniques to a collection of home care patient datasets led to the discovery of twenty-one significant attribute correlations in admission and discharge data gathered from a total of 988 patients across fifteen distinct home health agencies. Visualizations selected included histograms and heatmaps due to their shared ability to accommodate multiple dimensions and the relative ease with which one can identify patterns embedded therein. Some of the hypotheses generated upon evaluation of these visualizations were later validated via statistical analysis techniques (Monsen, Bae, Zhang, & Radhakrishnan, 2016). While the study was based on a very small number of patients, such a correlation could be used to improve upon various aspects of home health care (such as anticipating a longer episode and/or hospital stay length in patients with urinary incontinence) ultimately improving patient quality of life and reducing health care costs.

Another study found key predictive indicators of transferring persons with dementia into long term care from less dependent forms of care (Cepiou-Martin, Tam-Tham, Patten, Maxwell, & Hogan, 2016). This study involved a meta-analysis of longitudinal data and was primarily conducted via statistical analysis, though tables and funnel plots were used to organize and obtain results. The results reinforce the potential significance of information that has yet to be extracted from otherwise dormant data. The indicators include severity of dementia, exhibition of specific and/or worrisome behaviours, general ability to carry out activities of daily living, and ethnicity.

An ambitious project surrounding a hospital in India was undertaken with the intention of improving effectiveness, efficiency, and cost of care (Menon, Aishwarya, Joykutty, Av, & Av, 2021). The project aimed to replace the hospital's existing visualization tool (Tableau) with a free alternative, constructed using open-source tools, and capable of data pre-processing, data visualization, and predictive modelling via a web portal. Difficulties experienced by the project team include issues with the sample data provided by the hospital, such as missing/ corrupted/invalid attribute values, which had to be dealt with during the data pre-processing stage. Despite difficulties, the tool was found to be "useful for … understanding of trends and patterns" (Menon, Aishwarya, Joykutty, Av, & Av, 2021).

A visualization platform displaying various metrics of interest related to the spread of CoViD-19 in the state of Indiana, USA, saw not only rapid development, but also rapid implementation (Dixon, et al., 2020). Data was retrieved through region-specific health information exchanges and laboratory testing sources and employed automated pre-processing techniques. Figures across both the public-facing and the closed versions were relatively diverse, and included stacked bar charts, line charts, tables, and even geographic heat maps. Figures were also particularly feature-dense, with many offering the ability to "drill down" to finer granularity. Extensive practical usage and feedback suggest that the platform has "provided essential and otherwise unavailable … data to inform key decision makers … and enable a data-driven strategy to a state-wide response" (Dixon, et al., 2020).

A study with a focus on reducing crowding in emergency rooms demonstrated the significance of classification of patients in terms of acuity of care and adjusting and/or prioritizing resources accordingly (Khalifa & Zabani, 2016). In particular, the study selected two key performance indicators: emergency room length of stay (ERLOS) and percentage of patients leaving without treatment. The indicators were chosen for their perceived ability to accurately convey ER efficiency and effectiveness, respectively. The application of descriptive (visual) analytics techniques led to the conception of two procedural adjustments: fast-tracking patients with relatively less serious and/or time-sensitive ailments, and the addition of an internal waiting room to accommodate those who were able to stand, increasing overall bed capacity. Ultimately, the study led to significant reduction in both ERLOS and percentage of patients leaving without treatment, as well as suggestions for additional indicators and procedure changes to further decrease crowding.

3 METHODOLOGY & RESULTS

3.1 Data

The data used in the development of our platform was provided by two provincial health authorities as two separate Microsoft SQL Server (MSSQL) databases. We will henceforth refer to these as databases A and B. The databases contained data collected and stored via Cerner (Cerner, n.d.), DAD (Discharge Abstract Database Metadata (DAD), n.d.), Procura (Procura Home Health Software, n.d.), and Meditech (MEDITECH EMR Software, n.d.) systems. Despite minor differences, the semantics (i.e., the attributes) remained relatively identical across the two databases. No personally identifiable attributes were retained.

A set of SQL views was created to “filter” data without making any permanent alterations to the tables themselves. This pre-processing step involved data cleansing such as the removal of patients under age 50 and patients whose records contained oddities, such as invalid discharge dispositions and inconsistent personal information. Some data types were altered for performance reasons and views were created for easy access to frequently desired record subsets. Similarly, when constructing datasets for chronological metrics in reports, records were restricted to those dated between 2010 and 2019 because the data outside of this interval was relatively sparse and could potentially skew reported numbers.

The databases were stored on a secure server local to the research institution yet on a domain separate from that of the research lab and disconnected from the Internet entirely. The secure server was accessed via a complex authentication protocol. The platform itself was developed on our research lab domain, which included various general-purpose client machines, a file server and a Network Access Storage (NAS). The software used included Microsoft’s web development and business intelligence tool stack consisting of SQL Server (Microsoft, n.d.), SQL Server Reporting Services (SSRS) (Microsoft, n.d.) and SQL Server Data Tools (SSDT) (Microsoft, n.d.), in addition to Visual Studio and the ASP.NET framework (Microsoft, n.d.). QGIS (QGIS, n.d.) was used to develop dynamic maps displaying acute care admissions and assessments by various health regions.

An agile development process was followed using an integrative knowledge translation approach with a patient-oriented research framework in which patient partners and representatives of the health authorities remained engaged at all stages of the research process.

3.2 The Data Visualization Platform

The reports are grouped by care setting, namely Home Care (HC) and Long-Term Care (LTC), with a third group consisting of other reports which fall outside of these categories. The stakeholder-facing landing page provides a synopsis of aggregative statistics, such as the total number of unique patients across HC and LTC including the number of assessments, age group at time of assessment, and the average length of stay (in months) by setting (Figure 1). Additionally, this page allows for navigation to other reports such as geo-mapping and predictive modeling interfaces via a user-friendly menu. Various drill-down and drill-through reports provide information at finer granularity and details, respectively. Tooltips are used throughout the dashboard to popup additional relevant information. For a meaningful analysis, data is normalized, where applicable/possible.

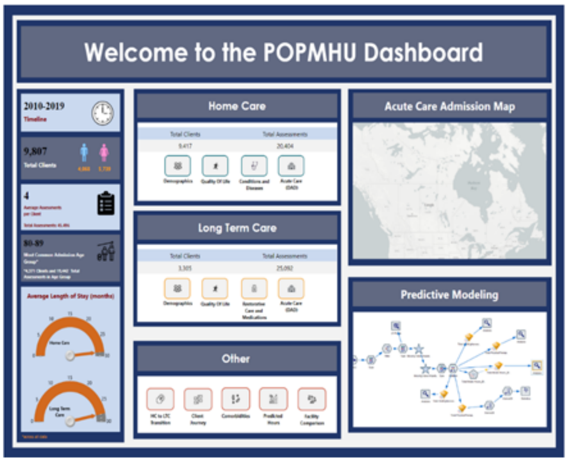

Figure 1: The Landing Page.

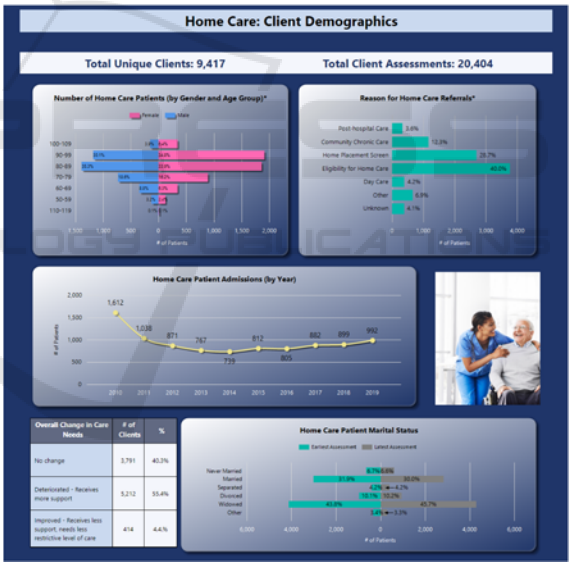

Figure 2: Demographics Report.

3.2.1 Demographics

The demographics report header displays the number of total unique patients and assessments for the selected care type and year interval (Figure 2). A trend chart demonstrates the number of admissions over the selected year/interval. Other charts in this report include patient numbers (and proportions) by gender, age group and marital status (at first and most recent assessment dates), reason for care type referral and assessment, and overall change in care needs.

Tooltips allow viewing of actual patient counts, where applicable.

An interesting trend illustrated in this report is the change in numbers of patients over the years. For instance, a significant (roughly 50%) decrease in HC admissions was observed in Database A between 2010 and 2013, with a gradual crawl upwards thereafter. Database B, however, showed the number of admissions halved from 2010 to 2011, and remained relatively static thereafter. The aggregate marital statuses remain relatively stable suggesting a stable ratio of new and recurring patients. One potentially actionable trend is the skewing of patient counts by gender. In both databases and both modes of care, patient gender ratios skew slightly towards male in patients aged 80-89 years (especially in the younger age ranges), only to skew (often dramatically) towards female thereafter. This is likely a reflection of the differences in general life expectancies between the two populations and may support gender crossover as has been noted in previous studies of the oldest old (Freeman, Hajime, Satoru, & Masahiro, 2009).

3.2.2 Home Care: Client Quality of Life

The Quality of Life reports (Figure 3) have the option to dynamically regenerate the figures using data from a selected year/interval. These reports are broken down into four distinct, horizontally divided groups: Sensory & Cognitive Conditions, Physical Functioning, Pain, and Mood. The selection of metrics therein was based upon data released by the Canadian Institute for Health Information (CIHI), which showed how "long-term care homes are performing on nine [specific] indicators" (McCormick Home, n.d.).

In general, the Sensory & Cognitive Conditions sections include tables displaying patient counts divided according to various levels of visual, auditory, and memory impairment. The Physical Functioning sections display percentages of patients divided according to various metrics representing physical assistance needs and levels of ability to carry out Activities of Daily Living (ADLs). The Pain section touches on patient data regarding pain frequency, presence of pressure ulcers, and frequency of falls. Lastly, the Mood section displays patient data regarding mental health related attributes, such as frequency of feelings of sadness or depression, patient loneliness, and frequency of verbally abusive behavioral symptoms. The charts illustrate that the number of clients with adequate vision was typically greater than the sum of all clients with any sort of visual impairment, regardless of chronology or mode of care. The number of clients with both short term and procedural memory impairments was also often greater than the sum of clients with any single type of memory impairment, or none at all. As expected, the number of clients needing assistance to perform ADLs increases significantly from HC to LTC. In what is likely a testament to quality of care, in Database A the number of clients reporting feelings of sadness or depression seems to decrease from HC to LTC; in Database B, this trend is not seen. Similarly, aggregate reports of pain in any capacity decrease dramatically from HC to LTC, often by more than 15%.

Figure 3: Quality of Life (Home Care).

3.2.3 Conditions and Diseases/Medications

In the Conditions and Diseases report (Figure 4), a stacked bar chart displays the top ten most prevalent diseases. The stacked values represent patient counts, divided according to whether their diagnosis is actively being monitored and/or treated by home care professionals. The other charts on this report display more specific disease/condition-related patient counts for incontinence, nutrition, and certain dermatological issues, each with varying degrees of information. The top three most prevalent diseases across various chronologies and modes of care were Arthritis, Dementia, and Hypertension, though the ordering of these variables varies (e.g., Hypertension is typically the most common disease across all times and care types, though, at a notable number of points in time, Dementia is instead the most common disease in LTC facilities).

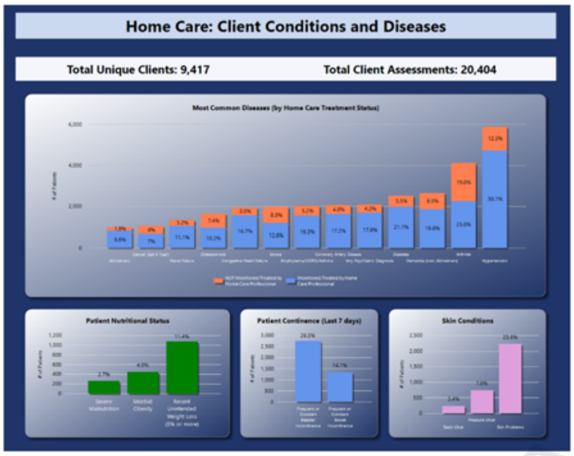

Figure 4: Conditions and Diseases Report.

The Medications report (Figure 5) provide the average number of medications taken for various psychiatric diseases together with frequency of consumption. Two other charts in this report show the average weekly time spent by patients in different types of therapy. These charts are subdivided by various kinds of therapy and/or rehabilitation. Lastly, two bar charts demonstrate both the numbers and proportions of hospital, emergency room, and physician visits.

The average amount of therapy received varied greatly from year to year, with therapies common in one year being sometimes almost non-existent in the next. The average numbers of medications (around nine) do not trend in any particular direction when traversing across age groups. Oddly, in years where average medication counts do not follow this trend, HC clients in older age groups tend to have *fewer* average number of medications than those in younger age groups.

3.2.4 Acute Care (DAD/MEDITECH)

The Acute Care reports (Figure 6) provide visualizations of data collected through two acute/hospital care EHR systems: DAD and Meditech. These reports include trend charts displaying the number of admissions by month, gauges demonstrating the average lengths of stay in both acute and alternative level of cares, and the numbers of admissions for each major clinical category. Other figures found in these reports include

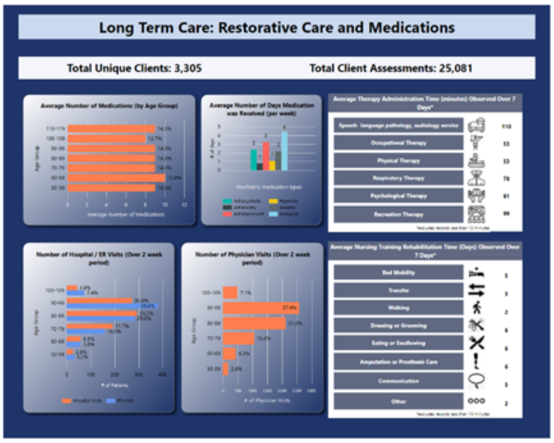

Figure 5: Medications Report.

a bar graph demonstrating the top five facilities by admissions, and a gauge directly comparing the estimated and actual lengths of stay (in days). The bar graph allows one to drill down into a sub-report for information at a finer granularity (Figure 7). Additional figures illustrate the most frequently utilized patient services within the chosen facility, and the number of transfers from other facilities. Other metrics reported here are similar to the parent report, but present values for the selected facility only.

Figure 6: Acute Care.

Perhaps the most intriguing and actionable metric to originate from these reports were the charts demonstrating the numbers of admissions by month. Over the year(s), selected trends that could lead to improved care by better anticipating acute care capacity month by month were demonstrated. For example, regardless of care type, the chart clearly demonstrates that acute care admissions are lowest in December, and seem to peak approximately once per quarter, with differences in peaks and dips consisting

Figure 7: Acute Care drilldown.

of up to hundreds of patients (in HC). In terms of clinical categories at times of admission, diseases of the digestive system seem to be the most common reason.

3.2.5 Other Reports

The dashboard contains several other reports of interest, including a HC to LTC Transition report, a Patient Journey report, and a Comorbidities report. The Transition report (Figure 8) provides a relatively direct comparison between HC and LTC statistics including patient and assessment counts, top three ADLs requiring assistance, and the estimated and actual lengths of stay in acute care. The report also illustrates the average duration of wait times when moving from one form of chronic care to another. An interesting observation revealed that, on average, actual acute care times for patients of either care type are anywhere from 3.5-5 times longer than the estimated times. The report also contains some interesting information regarding transfer times between HC and LTC: although the most common wait time is 30 to 60 days, nearly 10% of clients wait a year or more to transfer.

The Patient Journey report is a chronological representation of events as clients transition from HC to LTC. The statistics on this timeline include the average number and duration of service encounters, ER visits, the most common mode of admission, and the average length of stay in a specific care type. The average duration between the last HC assessment and first LTC assessment was approximately five months. Interestingly, the average number of ER visits seems to drop by 15% between HC and LTC. Similarly, the average number of service encounters among LTC patients is roughly 25% lower than of HC patients.

Lastly, the interactive Comorbidities report (Figure 9) allows a user to select up to three distinct diseases for which patient counts are generated. The counts are broken down by both gender and care type. For instance, one noteworthy comorbidity is osteoporosis and fractures. A large portion of female HC clients have osteoporosis and approximately 10% have experienced a hip fracture while approximately 17% have suffered from another form of fracture (wrist, vertebral, etc).

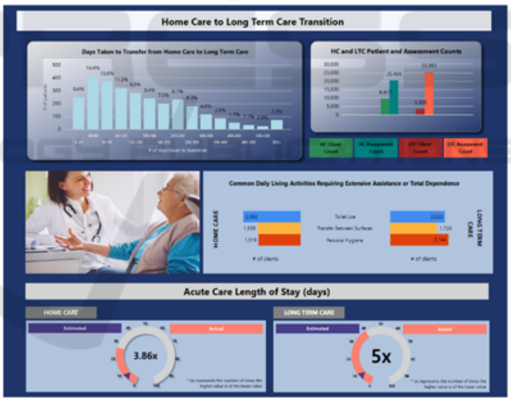

Figure 8: HC to LTC Transition.

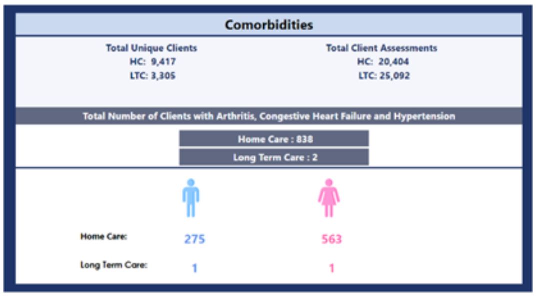

Figure 9: Comorbidities Report.

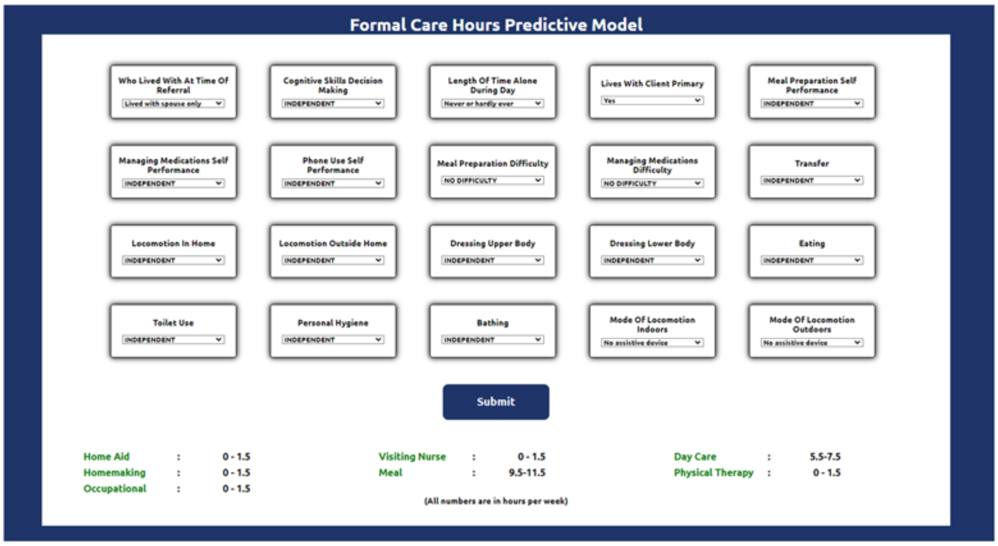

Figure 10: Resource requirements estimated by predictive model.

4 PREDICTIVE MODELING

Estimating the amount of care needed by a patient is a complex problem and usually requires experienced professionals. Given the recent nursing shortages and early retirements, it is challenging to find an experienced nurse to conduct such triaging. Additionally, errors in judgement can result in negative impact on both the patient and the healthcare system. For instance, we noticed variations of over 500% in estimated and actual acute care length of stay. Thus, it is desirable to have a tool which could more precisely estimate the amount of care a patient might need based on their current condition. This could also result in data-driven resource allocation.

To estimate the amount of care a patient might need, we use the fields from MDS (Minimum Data Set) forms such as ADL scores, patient abilities and pre-existing conditions. A predictive model then runs in the cloud and renders results through a simple web interface. The technical details of our model are beyond the scope of this paper. However, the steps involved include data pre-processing, cleansing, performing imputation on categorical variables, identification of features correlated to target variables using statistical tests, and training and testing of the model. To keep the complexity of the model at manageable levels, twenty features (Figure 10) were selected using IBM SPSS modeler (IBM, n.d.) feature selection algorithm. Each feature and its corresponding values were named in accordance with the nomenclature provided by MDS forms. A web form was developed to interact with the model hosted in the cloud. Once the information about a patient's current condition is entered and submitted, the web form sends a request to an API (Application Programming Interface) which processes the information, converts it to an appropriate format and posts to Watson Studio (IBM, n.d.) cloud. The model stored in the cloud processes and returns the predictions to the web form via the same API. The entire process occurs in real-time in a matter of seconds. Figure 10 also illustrates an example where estimated resources (hours per week) are instantly displayed on the bottom of the web form for the selected values of the features.

5 CONCLUSION

Even when data is collected via standardized forms, it poses challenges from quality to comprehension. To extract value from the collected data, it must be pre-processed and converted to a form suitable for visual analytics while maintaining the privacy and confidentiality of patients and facilities. The data obtained from two health authorities was extracted, cleansed, and placed on secure servers. Actionable KPIs were identified for visual analytics of this data. For fair comparison, reports were normalized, where applicable. Displaying data on interactive maps provide a different perspective.

The visual analytics platform enables observation along multiple dimensions which could be used for more informed resource allocation and improved patient care. The reports reveal hidden patterns in the data which could be used for informed decision making and better patient care. For instance, an important observation was the incorrect estimate of length of stay and resources in acute care. This results in unforeseen burden on the healthcare system. To rectify this problem, a predictive model was built to estimate the resources more accurately at the time of admission. This provides a data-driven approach to resource allocation. The quality of life metrics assist in early detection of conditions thus affording the opportunity for addressing situations before they progress to an unmanageable state. The patient journey gives a synopsis of the time and resources that are required by patients as they transition from homecare to long-term care.

ACKNOWLEDGEMENT

This research was funded by a grant from BC Academic Health Science Network (AHSN) under Strategy for Patient-Oriented Research (SPOR) (Methods Cluster) and resulted in training of more than a dozen HQPs. In addition, we acknowledge our patient partners (B. Baker, G. Kramer, I. Muturi, S. Prior, and C. Zannon) who remained engaged throughout this research and provided valuable feedback.

REFERENCES

Cepiou-Martin, M., Tam-Tham, H., Patten, S., Maxwell, C. J., & Hogan, D. B. (2016, Apr.). Predictors of long-term care placement in persons with dementia: A systematic review and meta-analysis. *International Journal of Geriatric Psychiatry, 31*(11), 1151-1171.

Cerner. (n.d.). Retrieved March 21, 2022, from https://www.cerner.com/

Discharge Abstract Database Metadata (DAD). (n.d.). Retrieved March 21, 2022, from https://www.cihi.ca/en/discharge-abstract-database-metadata-dad/

Dixon, B. E., Grannis, S. J., Tachinardi, U., Williams, J. L., McAndrews, C., & Embi, P. J. (2020). Daily Visualization of Statewide COVID-19 Healthcare Data. *2020 Workshop on Visual Analytics in Healthcare*, 1-3.

Freeman, S., Hajime, K., Satoru, E., & Masahiro, K. (2009). Understanding the Oldest Old in Northern Japan: An Overview of the Functional Ability and Characteristics of Centenarians. *Geriatrics & Gerontology International, 10*(1), 78-84.

IBM. (n.d.). *SPSS Modeler*. Retrieved from http://www-01.ibm.com/support/docview.wss?uid=swg27050406

IBM. (n.d.). *Watson Studio*. Retrieved from https://cloud.ibm.com/catalog/services/watson-studio

interRAI. (n.d.). *Speaking One Language for High Quality Care Worldwide*. Retrieved 4 2, 2022, from https://interrai.org/

Khalifa, M., & Zabani, I. (2016). Utilizing health analytics in improving the performance of healthcare services: A case study on a tertiary care hospital. *Journal of Infection and Public Health, 9*(4), 757-765.

McCormick Home. (n.d.). *CIHI Quality Indicators*. Retrieved March 21, 2022, from https://www.mccormickcaregroup.ca/mccormick-home/resident-care/cihi-quality-indicators-nov-2016/

MEDITECH EMR Software. (n.d.). Retrieved March 21, 2022, from https://ehr.meditech.com/

Menon, A., Aishwarya, M. S., Joykutty, A. M., Av, A. Y., & Av, A. Y. (2021). Data Visualization and Predictive Analysis for Smart Healthcare: Tool for a Hospital. *IEEE Region 10 Symposium*, 1-8.

Microsoft. (n.d.). *ASP.NET Overview*. Retrieved May 19, 2022, from https://docs.microsoft.com/en-us/aspnet/overview

Microsoft. (n.d.). *SQL Server 2019*. Retrieved May 19, 2022, from https://www.microsoft.com/en-ca/sql-server/sql-server-2019

Microsoft. (n.d.). *SQL Server Data Tools*. Retrieved May 19, 2022, from https://docs.microsoft.com/en-us/sql/ssdt/sql-server-data-tools?view=sql-server-ver15

Microsoft. (n.d.). *What is SQL Server Reporting Services*. Retrieved May 19, 2022, from https://docs.microsoft.com/en-us/sql/reporting-services/create-deploy-and-manage-mobile-and-paginated-reports?view=sql-server-ver15

Monsen, K. A., Bae, S. -H., Zhang, W., & Radhakrishnan, K. (2016, July). Visual analytics for pattern discovery in home care. *Applied Clinical Informatics, 7*(3), 711-730.

Morris, J. N., Fries, B. E., Steel, K., Ikegami, N., Bernabei, R., Carpenter, G. I., . . . Topinková, E. (1997). Comprehensive Clinical Assessment in Community Setting: Applicability of the MDS-HC. *Journal of the American Geriatrics Society, 45*(8), 1017-1024.

Morris, J. N., Hawes, C., Fries, B. E., Phillips, C. D., Mor, V., Katz, S., . . . Friedlob, A. S. (1990, June). Designing the National Resident Assessment Instrument for Nursing Homes. *The Gerontologist, 30*(3), 293-307.

Procura Home Health Software. (n.d.). Retrieved March 21, 2022, from https://www.alayacare.com/procura

QGIS. (n.d.). Retrieved May 19, 2022, from https://qgis.org/en/site/