The Community Analytics Academy Pilot: A Community-Academic Partnership for Building Community Health Care Analytic Capacity

Rhonda BeLue[1] [iD]**, Lorinette Wirth[1], Amanda Steormer[2], and Suzanne Alexander[1]**

Abstract

The ability to analyze data to identify best practices is key to improving quality of care for community-based health care organizations (CBOs). Leading commercial statistical software remains too costly for many CBOs operating in underserved communities. The St Louis Integrated Health Network (IHN) collaborates with CBOs to increase access to health care. IHN and a local university developed the Community Analytics Academy (CAA), a training collaborative designed to meet the need for data-informed decision making among CBOs. Establishing analytics training collaboratives for CBOs empowers organizations to respond to the ever-growing amounts of health care data and the need for data-driven decision making.

Keywords

safety-net data managers, community capacity, analytic capacity, community-based organization

Introduction

Empowering CBO Data Managers

Health care is moving toward a population health management framework, where reimbursement is tied to outcomes and value-based payment initiatives are becoming more common. The ability to analyze available data to identify treatments, practices, and processes that optimize benefit to an organization's population is paramount to providing cost-effective, high-quality care.[1] The need to manage larger amounts of data, to share and coordinate data between organizations, and to conduct more complex data analysis is increasing.

Many community-based health care organizations (CBOs) experience challenges to training staff in data analysis skills, date presentation, and interpretation.[2] The volume of data, challenges of sharing and coordination, and complexity of predictive analysis require increasingly advanced software with specialized training. Unfortunately, leading commercial statistical software remains too costly for many not-for-profits (NFPs) and CBOs operating in underserved communities.

Community Capacity Building

This encompasses the expansion of skills and expertise by an organization to address the issues they face, such as analytic and data processing capacity.[3] To successfully reduce health disparities and increase organizational sustainability, CBOs need to build capacity by acquiring new skill sets and resources.[4] While the concept of capacity building among CBOs is not new, the ability of a CBO to make data-driven, evidence-based decisions to address health equity issues and improve patient and client services is becoming increasingly critical in the era of increasing data availability and complexity.[5,6]

[1]University of Saint Louis, St Louis, MO, USA
[2]Integrated Health Network, St Louis, MO, USA

Corresponding Author:
Rhonda BeLue, Department of Health Management and Policy, College for Public Health and Social Justice, University of Saint Louis, 3545 Lafayette Avenue, St Louis, MO 63104, USA.
Email: rhonda.belue@slu.edu

Addressing Access Through Education: Need and Desire

For data analytics to be effective, health care organizations need access to the right tools and the right training. Community-based, NFP health care organizations that care for underserved communities and serve as a region's health care safety net often lack the financial resources to procure tools and training; hence, there is a demonstrated need. The St Louis Integrated Health Network (IHN) has created a collaborative model among CBO leaders, who routinely share concerns and challenges impacting the ability to model and forecast advances in care management; through these meetings, a desire to use data analytics in modeling has been emphasized. Health care safety net organizations employ masters-level individuals who are highly skilled in their field, but a lack of access to technological resources limits their ability to optimize their organization's data analytics.[7] To address this issue, the Community Analytics Academy was developed as a collaboration between regional safety net health care organizations and university partners, with a goal of training local safety net data analysts in a free software platform with advanced analytic capacity, to facilitate the growing need for data informed decision making.

The Partnership

A local university departments of health policy and data science collaborated with the St Louis IHN. The IHN networks community health centers, hospital systems, academic medical schools, public health departments, and other safety net institutions to increase access to high-quality, affordable healthcare for all residents of Metropolitan St Louis.[7]

Methods: Program Overview

All program components were developed using community-based participatory research principles.[8] Participants, course content, and schedule were all informed by the CBO pilot course participants.

Who Are the Targeted Participants and What Are Their Specific Needs?

Participants were data managers and data analysts at NFP health care and public health CBOs, including federally qualified health centers, local health departments, and other health intermediary organizations that serve underserved communities. Participants demonstrated a varied degree of proficiency with statistical analysis and represented a spectrum of experience, longevity, and education levels (see "Experience" in the Appendix A). It was important to respond to each individual's comfort level with the software chosen, and to provide an inclusive environment that not only facilitated but motivated continued learning as well as to encourage program alumni to provide mentoring to future program participants.

Why R?

The traditional statistical software packages such as SPSS, SAS, and Stata are too costly for many NFPs to purchase[9]; however, today's data cleaning and analysis tasks are too difficult or impossible for Microsoft Excel to perform efficiently. R is optimal because it is free, open-source, and robust—since its inception in 1993, a strong, participatory community of scientists and scholars have continually refined it into a tool that is easily adapted to an organization, if the staff has the knowledge and skill to properly utilize it.[10]

How Were Participants Recruited?

Recruitment for the CAA pilot R Workshop occurred in a mostly organic way via 3 routes: (1) IHN membership organizations were invited to select staff to attend; (2) The IHN Director of Evaluation, Quality & Learning shared the opportunity with an informal network of peers employed at local NFP public health/community based organizations, who have collaborated over the years to optimize their data collection, analysis, and evaluation processes; and (3) IHN shared a workshop flyer at various advisory meetings within the region's safety-net community. All participants who wanted to attend were accommodated and accepted into the program.

How Was the Educational Activity Implemented?

IHN and St Louis University stakeholders met regularly over the course of 4 months to develop a training implementation plan. An online survey was first developed and distributed to gauge interest in learning R. IHN acted on behalf of the partnership, leveraging established relationships to distribute the survey as broadly as possible through their existing relationships within the safety net (see Appendix A, "Survey Results"). Responses to the survey indicated a very wide range of experience with analytical software. Furthermore, the survey queried which analytic procedures were needed in day-to-day work and should be taught in the workshops. The content of the workshops was driven by the participants.

Program Participation

Per respondent preferences, sessions were scheduled for Saturday mornings. A total of 12 participants registered to attend one or all the sessions, at a charge of $20.00 per

Table 1. Chronological Order of Each Workshop.

Q&A about previous session	10 minutes
Group review/with multiple-choice test projected on screen; participants volunteered answers which were discussed for clarification and synthesis	10 minutes
Lecture segment	60 minutes
Break	5 minutes
Q&A for lecture segment	5 minutes
Activity 1—Participants worked to apply a concept and produce output, with instructor assistance as needed	30 minutes
Q&A for Activity 1	15 minutes
Activity 2 (same format as Activity 1)	30 minutes
Final Q&A	10 minutes
Brief review of topics covered from first workshop to present	2.5 minutes
Assignment of at-home exercise to be completed prior to the next session	2.5 minutes
Session feedback	
Total	4 hours

session for instructor compensation. Participants could choose their level of attendance; slides and supporting information was made available to participants who missed sessions for free. The session was taught by a masters-prepared PhD candidate in data analytics with significant training in R, STATA, and advanced analytic procedures, as well as previous experience in conducting research and data analysis related to quality and access to care in safety net health care facilities. Note that the IHN and partnering CBOs advertised the workshop to their staff and stakeholders whom they thought would benefit from the workshop.

What Are the Specific Objectives of the Education/Training? The training was designed to meet participant preferences, including

- how to download free R software and additional plugins,
- how to import and clean data, and
- how to conduct both basic and multivariate statistical analyses.

Three-hour sessions were held weekly for 4 weeks. Participants were provided exercises to walk through with the instructor for each session's designated topics of interest, with all materials available prior to the start of the sessions. All sessions included a break, question and answer sessions and time for students to receive help from the instructor. All students brought their own personal laptop.

Prior to the Workshops

Participants were sent instructions via email for downloading materials and software prior to their first session. Instructions were given for using Anaconda to download R, and the open source platform Jupyter Notebook for file and code sharing. Participants brought their own laptops in

order to utilize their own organizational data during in-class tutorials, which also allowed them to save any code they would build during class for future use. All workshop materials were made available via an online portal designed by the instructor.

Workshop Structure

An adapted flipped learning model was used to limit in-class lectures, in favor of using the time for exercises with instructor presence and assistance. While the schedule for each workshop had been planned prior to the first session, participant feedback was solicited at the end of each workshop and future course content was adjusted accordingly (Table 1). With only 4 days to deliver the content, the program prioritized introducing attendees to the tools available for working with R. Becoming aware of the extensive R-related help resources available would thus foster self-efficacy in troubleshooting challenges with the application of the tools. There was a strong focus on understanding the strengths and weaknesses of the R software, the importance and depth of the R community, and the overall functionality of the software in addition to the exploration and application of R tools. Each workshop was developed so that new lessons built upon content introduced in previous sessions, allowing an opportunity to review and reinforce prior knowledge as well as exploring new concepts.

Lesson Content

In alignment with the flipped learning model, the workshop was designed to limit total lecturing time to 1 hour at most, with 2 hours of questions and answers (Q&A) and hands-on activities. All workshop materials can be found at https://osf.io/bdzfm/.

A brief syllabus is included in Appendix B. As needed to improve comprehension and retention, the lecture segment

would be divided with material pertinent to Activity 2 presented after Activity 1 was completed. The instructor was also able to tailor subsequent classes to fit the needs of the participants as indicated in session feedback responses.

Next Steps

As evidenced by positive evaluations and continued interest, The CAA pilot was a success in terms of participant satisfaction and desire for additional training using the format. There remain, however, opportunities and challenges to its sustainability. While the pay-per-session model covered the instructor's actual time spent in sessions, lesson preparation in advance of the program required a significant investment. Funding sources including grants are currently being sought. Additionally, researcher partners are investigating the creation of additional R modules, as well as plans for developing pedagogy for other freeware.

While upskilling existing data managers and analytics in CBOs is a necessary first step, building analytics capacity throughout CBOs, including data awareness, use, and integrity is integral to achieving the CBOs' strategic goals and objectives. We suggest both increasing the skill of CBO staff, whose main task is data analysis and management, and increasing the analytic capability of end users and staff who make data-informed decisions. The latter would involve training of non-analytics focused staff in best practices for data use and integration in daily decision making.

Measures for increasing the sustainability, reach, and effectiveness of the R Workshop are based in part on recommendations derived from organizer and participant experience through feedback and formal evaluations collected after each session and are currently under consideration for implementation. These measures largely refine workshop content and CBO community support for continued training. Additional feedback should be solicited from CBO leadership to ensure the skills learned are retained and to record distribution to nonworkshop attendees within the organization.

Workshop Structure

- Offer the introductory workshop annually, with additional single-session workshops for special or advanced topics, and offer various levels to accommodate beginning and more advanced participants with face-to-face troubleshooting.
- Explore additional freeware options that support SQL (structured query language), data visualization, and other emerging needs in health analytics.
- Expand workshop participants to both data manager and end-user groups so that increased analytic capacity can be disseminated more efficiently throughout participating organizations.

CBO Community Support

As this project was a pilot study, further actions are needed to extend the work and assure integration into CBOs:

- Set up coalition of CBO data managers to help guide the CAA, maintain and lead ongoing trainings, and stay abreast of CBO data analysis needs.
- Continue to document and manualize training so that it can be replicated in other communities.
- Develop ongoing evaluation, including continued use of R in the workplace.
- As recommended by participants: "Allow time during trainings for us to bring specific questions from our CBOs to the class so we can get started on our own projects with assistance".
- Develop additional metrics for assessing uptake of newly acquired skills and appropriate timing for booster and additional sessions. Metrics will include organizational readiness to advance organizational capacity for advanced analytics and data-based decision making.
- Work directly with CBO leadership to assure ongoing support for ongoing analytics training.
- Work directly with CBO leadership to assess post-workshop retention and application of skills learned to organization mission.

In addition, we recognize that to fully understand the impact of the program, longitudinal assessment is necessary to determine how skills acquired in the workshop are practiced in the workplace. For the next iteration of the workshop, we plan to incorporate a one and three-month post-assessment to understand how participants used their new skills in the workplace.

Conclusion

Having audience-appropriate technical training sessions is important and not to be underestimated. It is rare to have this level of training focused on the needs and use cases relevant to CBOs, and the evidence of need and feasibility of such program implementation indicate it is warranted. This is an important gap to fill as CBOs fill an important gap in our communities, especially for serving underserved populations—they too need to be equipped to optimally use advanced tools to mitigate disparities. Initial needs assessment identified that there is a need and desire for such training. Survey results showed that most organization wanted to expand into more advanced statistics application in the future but lack training in these areas. Feedback from sessions showed that the content of this training was relevant to participants' job and would be beneficial. This project highlights the need for such training, shows that these programs are feasible, and provides a model for reproducibility across the nation.

Appendix A

Survey Results

A total of 19 CBO data analysts/managers were queried from 14 CBOs.

Experience

- Six data analysts had at least 1 year of experience using statistical software that requires writing code
- Ten had previously only used point-and-click statistical software
- Three had no statistical software experience

Analytical Approaches Desired

- Seven analysts responded that their organization does not currently do any advanced statistical analysis, but would like to expand in the future
- Twelve analysts indicated that their produce statistical analyses such as descriptive statistics (means and frequencies and odds rations) or epidemiologic measures, such as prevalence and incidence

Experience With R

- Nine respondents indicated that they had no experience with R
- Eight respondents had experience, but were uncomfortable using it
- Three respondents had experience, but were only somewhat comfortable using it

Barriers to Learning

- Fifteen respondents said that cost would be a major factor
- Several reported scheduling issues

Appendix B

Workshop Curricula

Workshop 1: Integrating R in Your Organization

- Exploring the R environment
  - Reference guides, troubleshooting sources, and commands lists
- Compatibility with Excel, Access, SQL
- Basics of importing data into R from Excel and SQL (Since only one attendee worked with SQL at their organization, individual instruction was provided to that participant only)
- Data preparation
  - Data structures, data frames, and recoding variables
  - Subsetting, sorting, and researching potential issues

After a brief introduction to each concept with examples of methods for solving potential problems, attendees were prompted to seek solutions to provided scenarios, using a curated list of resources like RBloggers and StackOverflow. By applying the flipped learning model, participants were able to remain engaged during the workshop, while practicing problem-solving techniques with real-world applications. Attendees were given a take-home assignment for further practice of the day's concepts, along with information on a data manipulation package in R called, "Dplyr."

Workshop 2: Using dplyr for Data Analysis

- Review, Q&A for Workshop 1
- Data manipulation through dplyr
- Building reports
- Linear and logistic regressions
- Working WITH Packages in R
- Finding and installing R packages
- Using the "survey" package to work with weighted data

Workshop 3: Data Analysis Through dplyr (Cont'd). While Workshop 3 was originally intended to focus on propensity score matching and weighting, along with an introduction to graphics in R, the session was adjusted to address the needs of the attendees. Feedback provided at the end of Workshop 2 expressed a desire for continued review of previous topics and holding off on providing new information. The introduction of propensity scoring technique was to provide context for conducting evaluations; however, a discussion with attendees revealed that there was limited familiarity with these statistical techniques. Participants and the instructor agreed that these methods would be best introduced in a separate workshop, such as "Evaluating Community Interventions in R."

Workshop 4: Base Graphics and ggplot2. The final workshop was used to introduce base graphics in R and the functionality of the R package, ggplot2, for visualizing data analysis results. After discussing materials covered over the 4 weeks, attendees reflected and discussed perceived opportunities for utilizing R in their daily workflows. All participants received a certificate of completion.

Declaration of Conflicting Interests

ORCID iD

Rhonda BeLue ![iD] https://orcid.org/0000-0001-6163-9543

References

1. Liberato SC, Brimblecombe J, Ritchie J, Ferguson M, Coveney J. Measuring capacity building in communities: a review of the literature. *BMC Public Health*. 2011;11:850. doi:10.1186/1471-2458-11-850

2. Griffith DM, Allen JO, DeLoney EH, et al. Community-based organizational capacity building as a strategy to reduce racial health disparities. *J Prim Prev*. 2010;31:31-39. doi:10.1007/s10935-010-0202-z

3. Cottrell E, Darney BG, Marino M, et al. Study protocol: a mixed-methods study of women's healthcare in the safety net after Affordable Care Act implementation—EVERYWOMAN. *Health Res Policy Syst*. 2019;17:58. doi:10.1186/s12961-019-0445-y

4. Humphries DL, Carroll-Scott A, Mitchell L, et al. Assessing research activity and capacity of community-based organizations: development and pilot testing of an instrument. *Prog Community Health Partnersh*. 2014;8:421-432. doi:10.1353/cpr.2014.0067

5. Bekemeier B, Park S, Backonja U, Ornelas I, Turner AM. Data, capacity-building, and training needs to address rural health inequities in the Northwest United States: a qualitative study. *J Am Med Inform Assoc*. 2019;26:825-834. doi:10.1093/jamia/ocz037

6. Community Science. Not as easy as it seems: the challenges and opportunities to build community capacity to use data for decisions and solutions. Accessed May 20, 2020. https://www.communityscience.com/news-detail.php?news=195

7. St. Louis Integrated Health Network. IHN's History. Accessed May 20, 2020. http://stlouisihn.org/our-history

8. Komaie G, Goodman M, McCall A, et al. Training community members in public health research: development and implementation of a community participatory research pilot project. *Health Equity*. 2018;2:282-287. doi:10.1089/heq.2018.0043

9. Hauge Ø, Ayala C, Conradi R. Adoption of open source software in software-intensive organizations—a systematic literature review. *Inf Softw Technol*. 2018;52:1133-1154. doi:10.1016/j.infsof.2010.05.008

10. Hartzband D, Jacobs F. Deployment of analytics into the healthcare safety net: lessons learned. *Online J Public Health Inform*. 2016;8:e203.