A DATA INFRASTRUCTURE FOR GLOBAL PERIOPERATIVE OUTCOMES

UNDERSTANDING HOW LANGUAGE BARRIERS AFFECT DATA COLLECTION IN EMERGENCY SITUATIONS

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Systems and Information Engineering

By
Bhavana Channavajjala

October 31, 2019

Technical Project Team Members

Rex Focht
Luke McPhillips
Sarah Winston Nathan
Nathan Ohene
Victoria Rho
Angela Yi

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

Signed: ______________________ Date: 10/31/2019

Approved: Catherine D. Baritaud Date: Dec. 5 2019
Catherine D. Baritaud, STS Division, Department of Engineering and Society

Approved: ______________________ Date: 12/11/2019
Donald Brown, Department of Systems and Information Engineering

As found by Petroze et al. (2013), operative diseases comprise approximately 11% of the global burden of disease (p. 457). Namely, diseases requiring surgical care account for a sizable portion of all communicable and non-communicable diseases throughout the world (Petroze et al, 2013). In high income countries, surgical care is highly available and is supported with proper resources and real-time data collection to ensure patient safety. However, in many low and middle income countries (LMICs), surgical resources and data collection methods are outdated. According to a 2017 study by Sileshi et al. perioperative data, or data collected during a patient's surgical procedure is often handwritten by nurses and anesthesiologists and is inconveniently stored, preventing comprehensive research studies on the data (Sileshi et al., 2017, p. 251). A study conducted by Abahuje et al. (2019) in Rwanda studied the effect of an added Acute Care Surgical unit on the operations of the hospital, but was limited in its scope by illegible and unavailable data. Though unsuccessful, the study highlighted the need for accessible data and robust emergency care services, which are often overlooked in LMICs (Abahuje et al., 2019). Thus, enabling more robust data collection practices in underdeveloped countries could enhance data collection and support progress in medicine.

The technical portion of this project will explore how manually collected surgical data can be digitized using image processing techniques. The digitized data will then be maintained in a relational database and made available for further analysis. The technical project considers the scarcity of medical data infrastructure in Rwanda, while the STS portion of the paper will explore language barriers to data collection in countries like Rwanda. With a special focus on data collection in emergency situations, the STS paper will analyze how a lack of common understanding leads to a dearth of data in urgent circumstances. The findings from the STS topic will take a practical, socially-oriented outlook on data collection, which will add more context to the discoveries in the technical project.

Figure 1 below indicates intended timelines for both the STS and technical projects. As shown in the figure, image processing algorithms for the technical portion of the project will be developed through early December of 2019, and tested using statistical and user-centric methods until February 2020. The backend database containing the digitized data will also be implemented by early February 2020. Holistic system testing will occur from late February 2020 until early March 2020, after which the results of the project will be finalized and detailed in a conference paper. Meanwhile, as the Gantt Chart shows, preliminary research on data collection in emergencies in LMICs will be conducted until late December 2019. Between November 2019

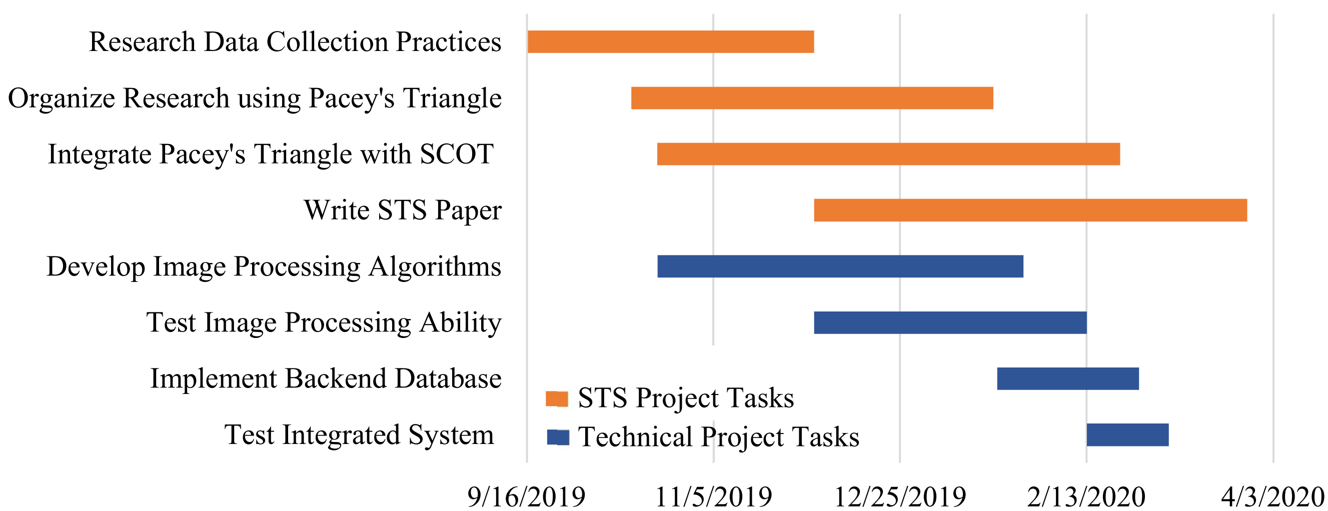

Figure 1: Gantt Chart of STS and Technical Projects: Visualization of project progress from Fall 2019 through mid-April 2020 (Created by Channavajjala, B., 2019).

and late January 2020, Pacey's Triangle, a framework developed by author Arnold Pacey, will be used organize and interpret the research so that the impact of language on the cultural, organizational and technical contexts of emergency data collection can be better understood (Pacey, 1983). The Social Construction of Technology theory (SCOT), as defined by Deborah

Johnson, will simultaneously leverage the wider contexts identified by Pacey's Triangle to identify how relevant social groups use language to interact with data collection technology (Johnson, 2005, p.1792). The STS analysis will be completed by late March 2020 and the final report will be written by April 2020. The completion dates are subject to change as the topics are studied further.

A DATA INFRASTRUCTURE FOR GLOBAL PERIOPERATIVE OUTCOMES

The technical portion of this paper discusses digitizing handwritten preoperative, intraoperative and postoperative records and storing them in a central database. Many patient records are handwritten in low and middle income countries, discouraging researchers from using collected data. Information on patient vitals, surgical procedures and drugs administered, can determine the chances of mortality for a given patient. However, important patient data is not stored centrally or made accessible, leading to a lag in research studies (M. Durieux, personal communication, September 13, 2019). A study to enable electronic collection of perioperative data was conducted by Sileshi et al. (2017) on behalf of Vanderbilt University in a tertiary hospital in Kenya. By supplanting paper records with a real-time, electronic perioperative data collection system, researchers found they could effectively implement a more modern data infrastructure in a middle income country. The infrastructure accounted for spotty Internet connection in the area and lack of centralized servers, which are problems the technical project could address as well (p. 252). However, Sileshi et al. (2017) acknowledged that the prior training of the clinical staff in Kenya enabled the quick adoption of a new record-keeping system, meaning that the collection system could not be easily implemented in hospitals in other LMICs (p. 256). This technical project seeks to minimize the disruption of workflow for hospital staff and instead focuses on digitizing previously collected data for further analysis. In this way, data will be readily available for research studies.

Perioperative mortality rate (POMR) is a metric often used to quantify the number of mortalities from surgeries compared to the number of surgeries conducted in a given hospital. Ng-Kamstra et al. (2018) maintain that POMR is used throughout hospitals in Africa, but that its use is not standardized, thus leading to incoherent results within hospitals (n.p.). However, POMR is a strong indicator of surgical capabilities in a given hospital and is also used by agencies such as the World Health Organization to determine if funding should be given to medical facilities (Khan, Penoff, Pirrotta, & Hosang, 2018, p. 2). In 2012, researchers Vaid, Bell, Grim and Ahuja retroactively analyzed preoperative data of patients from the American College of Surgeons, National Surgery Quality Improvement Program database and developed a score to predict perioperative mortality for patients. Their study used patient demographics, physical characteristics and disease history to calculate the score, and ultimately found that preoperative markers were successful and efficient in predicting mortality (Vaid, Bell, Grim, & Ahuja, 2012, p. 12). Thus, centrally collected electronic data could support better predictions of mortality, which could result in more funding for new medical technology. By creating digital copies of previously recorded data, hospital administrators can calculate metrics to assess current hospital functionality and can measure progress compared to previous years.

The objective of the research work is to implement a system for creating electronic records from manually collected patient records. As shown in Figure 2 included on page 6, the process will begin by uploading a scanned document with patient information to a database located at the University of Virginia. Using image classification, optical mark recognition, and graph conversion methods, the team seeks to implement a software program which can successfully read and extract data from the various fields of the records and store the data in a backend relational database. The program will be able to recognize doctor handwriting, analyze graphical displays of patient vitals, and collect data on the type of surgical procedure as it relates to the patient's condition. The data collected will be limited to the data available on the provided flowsheet. Figure 2 on page 6 then shows that data within the database will be filtered for any extraneous noise and will be made available to facilities in the corresponding hospital for analysis. By way of this process, the group aims to create a smoothly functioning system to translate medical records into data at a low error rate.

In order to implement image classification and optical character recognition, an open source program or existing method will have to be implemented. The types of writing that have to be recognized include: sketches of graphs, demographic information contained in check boxes, medications administered to the patient during surgery and any doctor notes specific to a patient's case. The team aims to find previous research which coincides with the objectives of the project and leverage software development methods to create image processing algorithms. The team will also use a relational database management system to create a backend system to contain both raw and processed data. This database will have to be made available to doctors at corresponding hospitals.

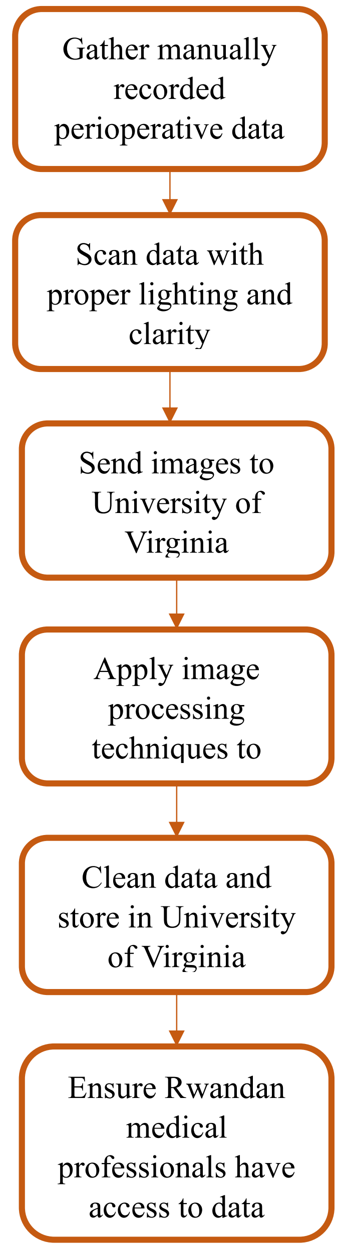

Figure 2: Process Flow of Data Digitization: High-level depiction of the process of digitizing data obtained from University Teaching Hospital in Kigali, Rwanda (Created by Channavajjala, B., 2019).

This project will be conducted under the guidance of Professor Donald Brown and Professor Benjamin Lobo of the Systems and Information Engineering department at the University of Virginia. It will also be conducted in conjunction with fourth year undergraduates in the Systems and Information Engineering department, Rex Focht, Luke McPhillips, Sarah Winston Nathan, Nathan Ohene, Victoria Rho, and Angela Yi. Working with Dr. Marcel Durieux, an anesthesiologist at the University of Virginia Health System, and Dr. Christian Ndaribitse, an anesthesiology postgraduate at the University of Rwanda in Kigali, the team will create a plan to implement the technology at the University Teaching Hospital of Kigali. The team hopes to digitize information contained in the records at the hospital and store the electronic data such that doctors in Rwanda can also access patient records. The project will be conducted with fabricated data and will not require Institutional Review Board approval.

A comprehensive study of the history of medical data digitization, by author Jeff Hecht, stated that the origins of the idea came from doctors who were often burdened with manually entering patient data. The doctors' tendency to scribble notes coupled with emerging medical technology such as magnetic resonance imagining led to a massive amount of electronic data, with minimal infrastructure and maintenance. Hecht asserts that this problem spurred a revolution in medical recordkeeping across the United States (Hecht, 2019). A similar problem of overworking doctors is seen in countries such as Rwanda, and by implementing technology that requires doctors to simply take a picture of handwritten notes, data can be properly and efficiently handled. Hecht does mention pitfalls to digitization such as violation of patient privacy and computer error which can lead to fatal errors in the medical field. However, Hecht details how electronic health records systems have been iterated on to eventually create the system and capabilities seen today. Machine learning algorithms and neural networks can now aid doctors in diagnoses while taking patient history into full consideration (Hecht, 2019). Thus, when implementing the digitization technology in Rwanda, the team can apply lessons learned from the digital revolution in America to understand the impact and potential of electronic medical data collection in low and middle income countries.

UNDERSTANDING HOW LANGUAGE BARRIERS AFFECT DATA COLLECTION IN EMERGENCY SITUATIONS

A study conducted on public hospital access in sub-Saharan Africa by Marsh and Rouhani (2018) showed that approximately 29% of the population lives more than two hours away from the closest hospital (p. 240). The sheer lack of access to medical facilities calls for a better emergency transportation system, as well as a medical system that is more understanding of emergencies. Khan, Penoff, Pirrotta, and Hosang (2018) explained in their research that low and middle income countries (LMICs) have opted to fund more primary care facilities than emergency facilities to increase access to care. However, emergency surgeries are more common than elective surgeries in underdeveloped nations, making an emergency care network imperative. Khan and her team found that trauma was the leading cause for hospital admission in the population of Fort Liberte, Haiti, supporting the idea of greater access to emergency services in LMICs (Khan et al. 2018).

The African Federation of Emergency Medicine is currently seeking to establish more urgent care facilities throughout the continent, and is especially focused on Out-of-hospital emergency care (OHEC), or pre-hospital care to reduce the seriousness of illness or injury. The efforts have been widespread, but have low yield as explained by Kironji et al. (2018) in their study on barriers to out-of-hospital care. The team identified cultural, financial and technical barriers and blamed the unstandardized practices of OHEC facilities for their lack of impact (Kironji et al., 2018). While better training programs for staff and strict government regulations could improve adoption of emergency care facilities, standardized data collection could also have a profound impact on the operations of a critical care center. Scott et al. (2017) explored real-time data collection in a trauma care center in Rwanda and specifically examined data regarding hospital performance. The team found that consistently relaying feedback to hospital staff improved the quality of work at the hospital and motivated the staff to perform better (p. 1380). The research conducted in Rwanda demonstrates that data collection can positively shift the operating standards of a hospital and supports the idea that data collection in the hospitals of underdeveloped countries should be more ubiquitous.

Data collection infrastructure in hospitals requires financial investment and interest from stakeholders, both of which can be difficult to acquire because the value of data is seldom recognized in LMICs (Akhlaq, McKinstry, Muhammad, and Sheikh, 2016, p. 1315). However, Sileshi et al. (2017) demonstrated that electronic data collection can be cost-effectively implemented in hospitals in their study in Kenya. A more significant barrier standing in the way of digitizing medical data is communication. As studied by Tiffin, George, and LeFevre in 2019, the lack of communication between hospital staff and more importantly, between patients and doctors, leads to decreased implementation of data collection technology. Especially in countries with large, diverse populations, conversations can be marred by language differences, resulting in low-quality treatments for patients (p. 3). Countries such as India have a multitude of primary languages, as shown in Figure 3 on page 10. Although close to 43% of the country claims Hindi as their primary language, the sheer volume of languages claimed by the other 57% of the country shows that it is unfair to assume that rural populations can understand Hindi (Office of the Registrar General in India, 2011, Statement - 4, p.15). Thus, local tongues and communication styles in countries such as India must be understood in order to facilitate the smooth transition to new technology. In this way, language differences pose a threat to mutual understanding between medical professionals and patients in low and middle income countries.

In order to understand how language poses a problem, Hunter-Adams and Rother (2017) studied how migrant populations in Cape Town, South Africa struggled to receive proper maternal care because of their inability to speak any of South Africa's many languages. Women in the study discussed how language was used discriminatorily in healthcare settings and how they were often forced into interventions they did not feel they needed (p. 5). Hunter-Adams and Rother even recounted one woman's observation that emergency services were only available to local citizens but not to migrants like her. The researchers noted that most of the women felt hesitant, skeptical and untrusting of medical services because of the communication gap between patients and doctors (Hunter-Adams and Rother, 2017, p. 4). Consequently, language differences could be the reason why emergency services are not quickly appreciated in LMICs. In 2015, Calvello, Skog, Tenner, and Wallis examined how a three-delay model, which was once used to evaluate maternal mortality rates, could be used to understand emergency services in low-resource countries. The model identifies the willingness of patients to seek care as a factor affecting the underappreciation of critical care

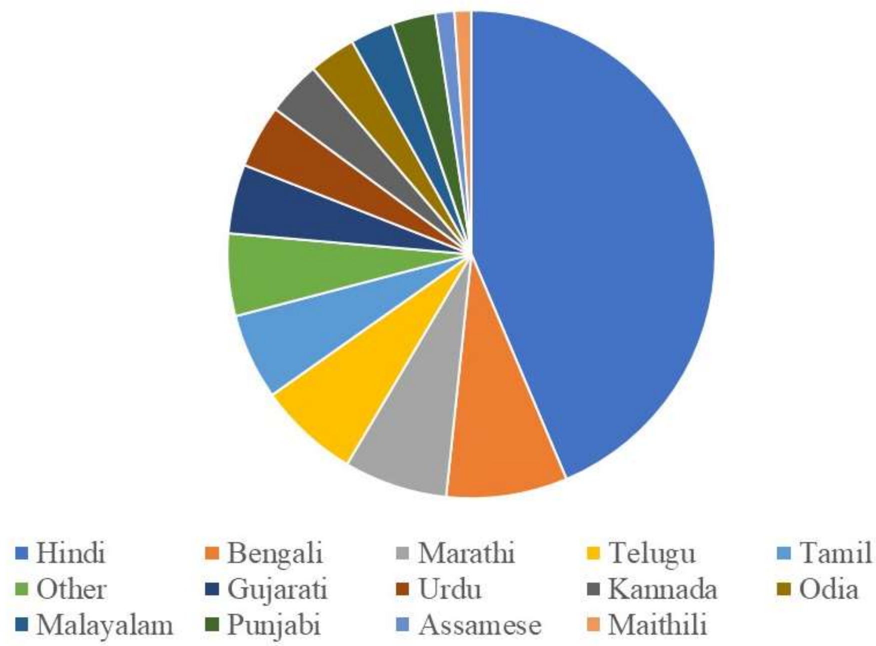

Figure 3: Distribution of Mother Tongues in India: Visual representation of most languages and their popularity throughout India (Adapted by Channavajjala, B., from data provided by Office of Registrar General, India, 2019).

facilities (Calvello, Skog, Tenner, and Wallis, 2015, p. 418). It then follows that language and discriminatory practices may be deterring patients from seeking urgent care to begin with, which leads to a dearth of electronic data. Ultimately, language could be barring the collection of data in emergency settings, resulting in poor quality of treatment and low adoption of emergency care services in low and middle income countries.

The STS Project will examine how language discourages proper data collection in urgent medical situations. As shown in Figure 4, the analysis will use Pacey's Triangle as a framework for examining the various contexts in which emergency data collection exists and how language barriers penetrate each context. The cultural context of emergency data collection encompasses the social norms, religious beliefs and educational values, which determine how medical practices are adopted. Language is the primary vehicle for disseminating these cultural elements and is therefore a prominent barrier in understanding how different populations view modern medicine.

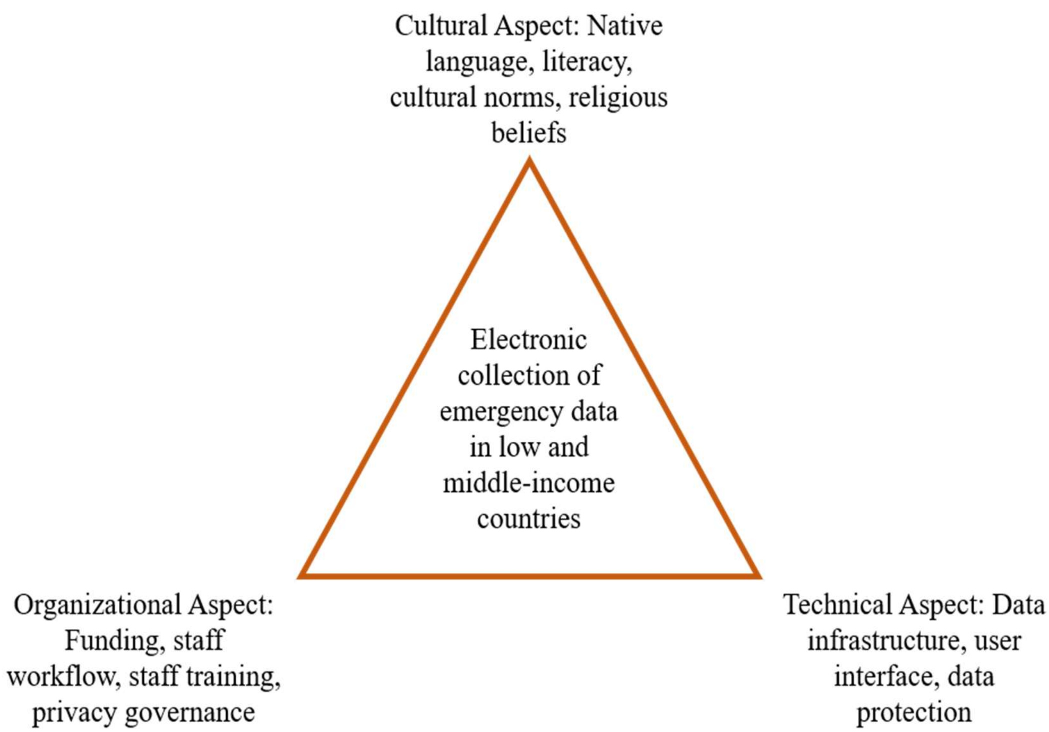

Figure 4: Contexts of Emergency Data Collection: Adapting Pacey's Triangle to the cultural, organizational and technical aspects of emergency data collection in low- and middle-income countries (Adapted by Channavajjala, B. from Pacey A., 2019)

Hospital funding, medical staff training, emergency procedures and hospital policies are fundamental to the organizational context of emergency data collection, as Figure 4 on page 11 depicts. While many LMICs have a national language which is used for all medical purposes, interactions between emergency staff and families of patients requires knowledge of other tongues spoken in the country. Proper use of language is also required to clearly articulate medical procedures and associated risks to patients and their families. This requires proper staff training so that medical professionals can communicate appropriately in urgent situations. In this way, language barriers also exist in the organizational context of emergency data collection.

The technical elements of data collection include: surveys or questionnaires used to collect data, infrastructure used to maintain data, and security measures used to protect data. Questionnaires can pose a prominent communication barrier in emergency situations, as lack of understanding can lead to incomplete or biased data as demonstrated in the study by Hunter-Adams and Rother in South Africa (Hunter-Adams and Rother, 2017). Additionally, the patient's inability to understand the data privacy policy of medical institutions can lead to distrust, as observed by Calvello, Skog, Tenner, and Wallis in their study to identify delays to emergency care. Thus, Pacey's Triangle elucidates how language barriers are present in all contexts of emergency data collection and how they can inhibit ethical and efficient data collection processes.

The analysis conducted using Pacey's Triangle fits into Deborah Johnson's perspective of the Social Construction of Technology (SCOT) theory which discusses the wider contexts in which a technology exists. SCOT supports the idea that emergency data collection is pertinent in different situations and that data collection impacts a number of relevant social groups (Johnson, 2005, p. 1792). Based on the contexts identified using Pacey's Triangle, the ideas of SCOT can identify how various social groups interact to shape the linguistic component of emergency data collection. Understanding the social groups in each context of Pacey's Triangle will clarify who affects the establishment of language barriers, and in turn, who can work to dissolve those barriers. Together, Pacey's Triangle and SCOT can be used to analyze how language impacts the many facets of data collection operations and how specific social groups directly influence the communication barriers.

A plethora of resources exist to explain the barriers to adoption of emergency care services in LMICs but there are not sufficient studies discussing how language threatens digital data collection. Therefore, the STS project seeks to acknowledge the relative nascence of electronic data collection in LMICs and detail how a mutual understanding of languages can expedite the implementation of better emergency care services.

WORKS CITED

Abahuje, E., Sibomana, I., Rwagahirima, E., Urimubabo, C., Munyaneza, R., & Rickard, J. (2019). Development of an acute care surgery service in Rwanda. *Traume Surgery & Acute Care Open, 4*(1). :10.1136/tsaco-2019-000332

Akhlaq, A., McKinstry, B., Muhammad, K. B., & Sheikh, A. (2016), Barriers and facilitators to health information exchange in low- and middle-income country settings: A systematic review. *Health Policy and Planning, 31*(9), 1310-1325. doi: 10.1093/heapol/czw056

Calvello, E. J., Skog, A. P., Tenner, A. G., & Wallis, L, A. (2015). Applying the lessons of maternal mortality reduction to global emergency health. *Bulletin of the World Health Organization, 93*(6), 417-423. doi: 10.2471/BLT.14.146571

Channavajjala, B. (2019). *Contexts of emergency data collection*. [Figure 4]. *Prospectus* (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia, Charlottesville, VA.

Channavajjala, B. (2019). *Distribution of mother tongues in India*. [Figure 3]. *Prospectus* (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia, Charlottesville, VA.

Channavajjala, B. (2019). *Gantt chart of STS and technical projects*. [Figure 1]. *Prospectus* (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia, Charlottesville, VA.

Channavajjala, B. (2019). *Process flow of data digitization*. [Figure 2]. *Prospectus* (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia, Charlottesville, VA.

Hecht, J. (2019). The future of electronic health records. *Nature, 573*. doi: 10.1038/d41586-019-02876-y

Hunter-Adams, J. & Rother, H. (2017). A qualitative study of language barriers between South African health care providers and cross-border migrants. *BMC Health Services Research, 17*(1). doi: 10.1186/s12913-017-2042-5

Johnson, D. G. (2005). Social construction of technology. In *Encyclopedia of Science, Technology, and Ethics* (pp. 1791-1795). Detroit, MI; Macmillan Reference USA.

Khan, A., Penoff, B. T., Pirrotta, E. A., & Hosang, R. (2018). Shifting the paradigm of emergency care in developing countries. *Cureus, 10*(2). 10.7759/cureus.2219

Kironji, A. G., Hodkinson, P., de Ramirez, S. S., Anest, T., Wallis, L., Razzak, J.,…Hansoti, B. (2018). Identifying barriers for out of hospital emergency care in low and low-middle income countries: A systematic review. *BMC Health Service Research, 18*. https://doi.org/10.1186/s12913-018-3091-0

Marsh, R. H., & Rouhani, S. A. (2018). Gaps in physical access to emergency care in sub-saharan Africa. *The Lancet Global Health, 6*(3), 240-241. https://doi.org/10.1016/S2214-109X(18)30026-3

Ng-Kamstra, J.S., Arya, S., Greenberg, S. L. M., Kotagal, M., Arsenault, C., Ljungman, D.,…Shrime, M. G. (2018). Perioperative mortality rates in low-income and middle-income countries: A systematic review and meta-analysis. *BMJ Global Health*, *3*(3). 10.1136/bmjgh-2018-000810

Office of the Registrar General, India. (2011). *Census of India 2011: Language*. (Paper 1 of 2018). Retrieved from http://censusindia.gov.in/2011Census/C-16_25062018_NEW.pdf

Pacey, A. (1983). *The Culture of technology*. Cambridge, MA: MIT Press.

Petroze, R. T., Groen, R.S., Niyonkuru, F., Mallory M., Ntaganda, E., Joharifard, S.,…Calland, J. F. (2013). Estimating operative disease prevalence in a low-income country: Results of a nationwide population survey in Rwanda. *Surgery, 153*(4), 457-464. 10.1016/j.surg.2012.10.001

Scott, J. W., Nyinawankusi, J. D., Enumah, S., Maine, R., Uwitonze, E., Hu, Y.,…Jayaraman, S. (2017). Improving prehospital trauma care in Rwanda through continuous quality improvement: An interrupted time series analysis. *Injury, 48*(7), 1376-1381. doi: 10.1016/j.injury.2017.03.050.

Sileshi, B., Newton, M.W., Kiptanui, J., Shotwell, M. S., Wanderer, J.P., Mungai, N.,…Mcevoy, M.D. (2017). Monitoring anesthesia care delivery and perioperative mortality in Kenya utilizing a provider-driven novel data collection tool. *Anesthesiology, 127*(2), 250-271. 10.1097/ALN.0000000000001713

Tiffin, N., George, A., & LeFevre, A. E. (2019). How to use relevant data for maximal benefit with minimal risk: Digital health data governance to protect vulnerable populations in low-income and middle-income countries. *BMJ Global Health, 4*(2). doi:10.1136/ bmjgh-2019-001395

Vaid, S., Bell, T., Grim, R., & Ahuja, V. (2012). Predicting risk of death in general surgery patients on the basis of preoperative variables using American College of Surgeons National Surgical Quality Improvement Program data. *The Permanente Journal*, *16*(4), 10–17. doi:10.7812/tpp/12-019