Title
Short Term Surgical Mission Evaluation: A Medical Record Analysis

Author
Johnson, Barbara Daugherty

Publication Date
2011

Short Term Surgical Mission Evaluation:
A Medical Record Analysis

by

Barbara D. Johnson

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Nursing

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

[redacted]

Dedication and Acknowledgments

Dedicated to the children and families of Faces of Hope

And

Peter, Megan and Timothy Johnson

There are so many people I would like to thank for their help and support.

Erika Froelicher, my academic advisor, mentor and guide, without whom I could not have completed this work.

The qualifying and dissertation committee members: your time and expertise made this possible.

Erika Froelicher, RN, MPH, Ph.D, FAAN

Nancy A. Stotts, RN, CNS, EdD, FAAN

Sally H. Rankin, RN, Ph.D, FAAN

Janet Myers, Ph.D, MPH

Research assistant: many thanks to Timothy Johnson.

Statistical support: many thanks to Steven Paul, Ph.D.

Graphic support: many thank to Sherri Goodman.

Financial support: many thanks to The Graduate Dean's Health Science Fellowship.

Lastly, for their faith in me and encouragement:

Stefan F. Klein, MD

And many dear friends and family.

Abstract

Background: Oral clefts occur in approximately one out of 700 births worldwide and can affect the lip, the palate, or both. In developing countries, lack of awareness of cleft treatment options, scarcity of surgical services, and poverty create barriers to corrective surgeries. To reduce global cleft care disparities, many nongovernmental organizations have developed short-term missions that provide surgical treatment. Few of these organizations have evaluated the impacts of their programs for the children and their families.

Aim: To evaluate the activities of the surgical outreach organization Faces of Hope (FOH) for the years 2005 through 2009.

Conceptual Framework: The Center for Disease Control Framework for Program Development in Public Health was used to initiate the evaluation of FOH. Research questions evaluated patient eligibility to receive services, and the outputs and outcomes of the services provided.

Design: A cross sectional design utilizing five consecutive years of data collected for clinical purposes.

Methods: Medical record review data provided the independent, dependent and covariate variables.

Sample: A sample of 706 medical records documented the clinical encounters of Guatemalan families seeking surgical services from the organization FOH for 2005 to 2009.

Results: There is no evidence that surgical repair of primary orofacial clefts occurred sooner over the years of FOH activities. Surgical complications compared by years suggest that a statistically significant difference exists between surgical complications in 2006 and 2009. There were no complications in 2009 compared to 2006 in which there was an 11.3% complication rate. Anesthetic complications did not have an association by year. Diagnosis, illness and wt/age- in- month Z scores were statistically significant predictors of complications.

Eligibility (receiving surgical treatment) was predicted by previous surgery, illness and diagnosis. Pre-operative vital signs and blood hemoglobin level did not predict eligibility. Returning in subsequent years to seek additional services was predicted by diagnosis, wt/age-in-month Z scores and documented futures needs.

Significance: Evaluation of short-term surgical missions is crucial as these organizations continue to participate in the provision of surgical care for children who may otherwise not receive treatment.

Approved:

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1: Background and Significance

Introduction

Oral clefts occur in approximately one out of 700 births worldwide and can affect the lip, the palate, or both. In developing countries scarcity of surgical services, lack of awareness of cleft treatment options, and poverty create barriers to corrective surgical procedures. In an effort to reduce global cleft care disparities, many nongovernmental organizations have developed short-term missions that provide surgical treatment. Few of these organizations have evaluated the impacts of their programs for the children and their families.

For children born with oral clefts in developing countries, scarcity of surgical intervention means that they must live with their condition, enduring both physical and social challenges. These include difficulties with eating, speech and hearing, social awkwardness and ridicule (Damiano et al., 2006; Noor & Musa, 2006; Patel & Ross, 2002; Turner et al., 1997).

As part of the armamentarium addressing global healthcare inequities, by one estimate, 543 volunteer medical organizations exist in the United States that serve developing countries (cite). By the year 2004, an estimated 6,000 international short-term medical and surgical missions had been conducted by these organizations, with a projected $250 million in expenditures (Maki et al., 2008). Varying degrees of local partnerships and multidisciplinary care is provided. For some organizations this is a major focus, for others primary care of oral clefts is emphasized, with a focus on maximizing the surgical volume of the missions. These international efforts must meet the challenge to provide care with the greatest beneficial impacts and with the aims of reducing health disparities between populations. There is also a responsibility on the behalf of the volunteer providers to assure that the healthcare provided is at an acceptable quality compared to that of wealthier nations (Walsh, 2004).

The literature provides very few examples of research specific to volunteer surgical missions serving poor countries. Maki et al. (2008) remarked about the paucity of literature on the subject despite the cost, number, and scope of short-term missions. Additionally, for the many organizations involved in surgical projects for children with cleft lip and palate in developing countries, few have reported the long-term outcomes (Canady, et al 2006; Uetani et al., 2004). The International Task Force on Volunteer Cleft Missions reported that research is not usually undertaken during cleft missions and underscored its importance, recommending that all cleft missions include research as part of organizational aims. Therefore, there is a need to critically evaluate the value of missions

(Yeow et al., 2002).

While acknowledging the need for outside volunteer assistance due to an increasing global surgical burden of disease, made worse by a lack of trained healthcare personnel in developing countries, program evaluation and monitoring are not routine (McQueen et al., 2009). Outcome indicators and measures of effectiveness are not collected, and systems to monitor surgical missions for quality assurance are not in place (Banatavalan & Zwi, 2000; Yeow et al., 2002).

Criticism of short-term medical missions has resulted from these shortcomings. Much debate exists in the literature, with some references to the missions as surgical safaris, surgical tourism and medical colonialism (Dupuis, 2004; Maki et al., 2009; Mulliken, 2004; Zbar, Rai & Dingman, 2000). According to the critics this is said to occur when inexperienced, unqualified or untrained individuals engage in surgeries they would not perform in their home countries; when surgeries are performed on children who would be considered poor surgical candidates in wealthy countries; when follow up care in not available; or when missions divert patients or resources from local providers.

Walsh (2004) questioned the effects of short-term humanitarian interventions, and if interventions conducted within the context of Western values were sustainable globally. She also raised the issue of whether providing highly technical care, without ongoing resources to support that level of care, meets the goals of reducing disparities and serving poorer nations. These concerns about potential paternalism and sustainability were considered by Bunyavanich & Walkup (2001), who stated that healthcare providers can broaden their understanding by listening to the poor people of the world to uncover their priorities. Uniform indices to compare quality of performance and equity of care, and the creation of an evidence base to guide the development of standards and outcome measures, would do much to address these criticisms (Banatalva & Zwi, 2000; McQueen et al., 2009).

Preliminary steps have been taken to measure quality and evaluate outcomes of short-term surgical missions (STSM). Operation *Smile International* implemented an electronic medical record system in 2005 to help monitor a number of quality indicators during surgical missions that will enable audit and reviews (McQueen et al., 2009). These efforts are aligned with the World Health Organization's proposals for global strategies to reduce the health-care burden of craniofacial anomalies (Mossey et al., 2004). The proposals included the adoption of guidelines for the maintenance and analysis of clinical records of cleft care. Establishing databases to record and document clinical activities and consideration of ways to create and disseminate evidence concerning humanitarian intervention is needed to prevent unnecessary duplication of effort and uninformed decisions (Banatavalan & Zwi, 2000; Mills, et al., 2005; Pezella, 2006; Smoot, 1998). Surgical safety is a substantial global-health concern. Surgery occurs at a tremendous volume globally with documented risks (Dupuis, 2002; McQueen et al., 2009; Smoot et al., 1992). In developing countries a death rate of 5-10% has been estimated for major surgery. Nosocomial (hospital acquired) infection rates were purported to be higher for short-term medical missions with rates at two or three times that of the United States (McQueen et al.; Smoot et al.). Developing methods to evaluate volunteer surgical missions is necessary for quality measurement in global care. Review of the medical records of volunteer surgical efforts and analysis of data is necessary to advance this agenda.

Research Aims

To evaluate the activities of Faces of Hope Surgical Mission in Guatemala for children both seeking and receiving surgical services from the organization for the years 2005-2009.

Specific research aims:

1. Describe the demographic and clinical characteristics of the population and compare these features over the time course of 2005-2009.
2. Determine if the demographic and clinical profiles provide evidence for earlier treatment for primary orofacial clefts over time (2005-2009).
3. Describe clinical and quality indicators (complication rates, procedure times), and benchmark with the American College of Surgeons National Quality Improvement Program Pediatric Phase 1 standards.
4. Determine the independent predictors (out of RQ 1-2) of eligibility for receiving surgical treatment.
5. For a subset of subjects who have data available, determine the predictors (out of RQ 1-3) of receiving additional procedures, in subsequent years, from Faces of Hope volunteers.

Chapter 2 will discuss the current literature on the subject of short-term surgical missions (STSM). Chapter 3 will describe the conceptual framework that is used to guide the evaluation the activities of FOH, which is based on the Centers for Disease Control Framework for Evaluation in Public Health. Medical record abstraction methodology as it relates to the research will be detailed in Chapter 4, and the research results presented in Chapter 5. The concluding chapter (6) includes discussion and recommendations for further study.

Chapter 2: Literature Review

Introduction

The purpose of this chapter is to review and critique the published literature focused on evaluation of Short-term Surgical Missions (STSM) operating in developing countries. This is undertaken in an effort to uncover the state of the science of STSM research vis a' vis the robust debate on the subject of them. The literature provides very few examples of research specific to STSM serving poor countries. Maki et al. (2008) commented about the paucity of research on the subject despite the cost, number, and scope of short-term missions.

Literature Search

Review of the literature began with a search of Pub Med, Cinhal, and Web of Science databases. The key terms used were: cleft lip, cleft palate, program evaluation, short-term medical missions and charitable surgical missions. Selection criteria were research studies focused on program evaluation of STSM that provide surgical services in developing countries for children with orofacial clefts. Investigations of cleft care provided in developing countries were also included. The term "paucity" has been used to characterize the published research on this subject (Maki, et al., 2008; Fischer, et al., 2001). Evaluation research addressing STSM for patients with cleft lip and palate appears to be in a nascent phase, though debate on the subject is robust. This search produced 59 publications using key search terms. Of those 16 met selection criteria and were included in this review.

Access to Surgical Care

Postponed treatment was identified as the most important unresolved issue regarding cleft treatment in developing countries (Uentani, et al., 2005). Age guidelines vary in developed countries; however primary lip repair usually occurs at approximately three months and palate repair at 15-18 months (Bearn, et al., 2001; Shaw, et al., 2001). Delayed closure of cleft lip or palate can lead to less desirable outcomes, e.g., impaired speech development, hearing loss, nutritional deficiencies with subsequent developmental delays, and stigmatization (Bermudez et al., 2010; Eberlin, et al., 2008; Shaw, et. al, 2001; Schwarz & Khadka, 2004). The longer a child has to wait until surgical treatment is available, the more exposure to ridicule or isolation may be experienced.

Lack of local providers and affordable services

Lack of local providers and high cost of services can inhibit access to surgical care for impoverished people in developing countries (Donkor et al., 2007; Moore & Fernandes, 2006, Schwarz & Khadka, 2004; Uetani et al., 2005). Delayed treatment due to the cost, paired with necessary travel and accommodations, influence the age of primary repair of cleft lip and /or palate (Donkor et al., 2007; Moore & Fernandes, 2006; Schwarz & Khadka, 2004; Uetani et al., 2005). Necessary time away from work for family members to access distant services, and the time required for treatment and recovery, compound expenses and delays in seeking treatment.

Describing and categorizing reasons for late presentation for cleft treatment was accomplished in a cross sectional survey conducted at a referral center in central Nepal (Schwarz & Khadka, 2004). Either the father or mother of a child who presented to the center for care was interviewed (n=53). Children arriving for initial treatment who were over three months of age met the study definition of late presentation. Seventy-nine percent of subjects met this criterion. The authors indicated that all cleft patients who arrived at the center during the study period were included, though details of how consent to participate and inclusion criteria was not described. Testing anticipation may be a concern that could potentially bias participant responses. The survey instrument and preparation of the interviewers was not detailed, precluding assessment of potential reliability and validity issues.

The results underscored that many barriers to accessing care exist. The most common reasons cited by the authors were a lack of knowledge of availability of repair (31%), lack of accessibility of care (29%) and lack of finances (24%).
A multidirectional approach to improving access was deemed necessary, and itinerant services were posited as a viable way in which obstacles to oral cleft care could be overcome (Schwarz & Khadka, 2004).

Outcomes of STSM

Since early repair is medically indicated, studies have been completed measuring the age of primary repair of children with oral clefts and surgical volume by cleft type in developing countries (Donkor et al., 2007; Moore & Fernandes, 2006; Uetani et al., 2005). Medical record abstraction methodology was used to evaluate the outputs and outcomes of surgical services provided by nongovernmental organizations in Vietnam, Ghana and East Timor (Donkor et al, 2007; Moore & Fernandes, 2006; Uetani et al., 2005). In Vietnam (n=790) the medical record audit indicated that the median patient age for lip and palate repair was lowered over the ten-year period of the organizations activities. The median patient age for lip repair was lowered from 14 years in 1993 to 1.3 years in 2003. For children with cleft palate the median age of repair was lowered form 13.5 years in 1993 to 5 years in 2003. Age ranges for the children seeking both cleft lip and palate repair were also reduced in 2003 compared to 1993
(Uetani et al., 2005).

Five years of operating room records were analyzed in Ghana (n=344), and a fourfold increase in the number of individuals with clefts seeking and receiving surgery was reported (Donkor et al., 2007). The increase is surgical volume for cleft lip and palate repair was posited by the authors to have resulted from greater community awareness of available surgical services, and removal of financial barriers. There was a sharp increase in surgical volume between 2003 and 2004, years in which funding for local surgeons was provided by a US NGO and the additional presence of a visiting volunteer surgical team. The researchers suggested that increasing volume each year was a proxy measure for satisfaction with outcomes, that volume growth associated with children, who returned for additional treatment, as in the case of children with both cleft lip and palate, was a demonstration of satisfaction with initial care (Donkor et al., 2007). The removal of financial barriers, such as providing financial support for local providers and creating partnerships with STSM was proposed as a strategy which could lead to increased access to surgical services for patients in developing countries (Donkor et al., 2007). The results were not reported but no statistical analysis were mentioned, thus it is unclear if the differences in volume of the surgical procedures reported were statistically significant. In the study procedure frequencies were totaled for the five year study period. Breakdown of this data by individual year would have enabled the calculation of percent increase in the numbers by year, with the potential to identify annual trends. Interrater reliabilities or validity measures for the record auditing processes were not described in the reports. Procedures for handling missing data were also not identified.

Causation beliefs and expectations

Beliefs about causation of oral clefts and the pre-operative expectations of parents of children awaiting surgery from STSM in India, Egypt and Brazil have been studied. Cultural differences among the groups were reported (El-Shazly et al., 2010; Reeve et al, 2004; Weatherly-White et al., 2005). The studies used qualitative methodology and open-ended response interview guides. Beliefs about causality fell into three primary categories: contagion, God's will, and personal conduct. Contagion was thought to occur by ingestion of harmful substances or in some instances witchcraft (El-Shazly et al., 2010). In Brazil parents felt implicated in the contagion for allowing "exposure" (Reeve et al., 2004). For families assigning the cleft to the will of God, concern was expressed that repairing it would be interfering. Others were resigned towards their child's condition based this belief. In all of the countries, blame was often placed on the parents for sins committed either in their present lives or a past one (El-Shazley et al.,2010 Reeve, et al., 2004; Weatherly-White et al., 2005). In rural India it was reported that the mother's behavior was the most influential in creating the cleft condition (El Shazly DATE et al.). The parents all expressed their hope that their child's life would improve after surgery, though priorities differed among the groups. An increased educational opportunity for the children was the hope of all of the parents.

The pre-operative perceptions and expectations of patients and families were also assessed (Reeve et al., 2004; Weatherly-White et al., 2005). In the study conducted in Brazil, the potential for improved speech was assigned the greatest importance by patients and families (Reeve et al., 2004). This was viewed as critical to facilitate performance in school and the potential for obtaining satisfying work. Furthermore, they also was also reported that appearance was significant though not the primary concern.

In rural India, appearance was perceived to be the most significant concern of the parents (Weatherly-White et al., 2005). Improvements in speech and social acceptance were anticipated, but to a lesser extent. Marriage prospects were seen as crucial, and the surgery was thought to influence chances of having a beneficial marriage. This was particularly relevant to girls as it was believed they may have no marriage prospects without lip repair leading to economic burdens on the family. Educational opportunities were also impacted by the child's appearance. Parents of both boys and girls children stated that their children had been refused admission to school (Weatherly-White et al., 2005).

For each of the studies (El-Shazly et al., 2010; Reeve et al, 2004; Weatherly-White et al., 2005), the internal validity would have been strengthened if theoretical frameworks or specific qualitative methodologies were identified. Domain specifications for the interview questions were unclear; more detail would have made a clearer case for content validity. Reliability measures were not reported by study authors. Women respondents from rural India were described as being extremely deferential and hesitant to answer the questions posed them (El-Shazley et al., 2010). This potentially affected the accuracy or depth of their responses. Testing anticipation was possible as the parents were waiting to have their children selected for surgery, and the researchers were involved with the surgical missions providing the care, potentially biasing their responses (El-Shazley et al, 2010; Weatherly-White et al., 2005).

These reports suggest that expectations may be defined by culture and can have an impact on outcomes. For example in the study in rural India (Weatherly-White et al., 2005) where appearance was the primary concern, the immediate surgical effect of closing the cleft may fulfill the expectation of improved appearance. In this situation positive long-term outcomes may be perceived due to achievement of pre-operative expectations. In contrast, for the patients or family members in Brazil who considered improved speech the highest priority (Reeve et al., 2004), outcomes may not as closely match expectations. Post-operative speech therapy is usually necessary after repair of a cleft palate (Patel & Ross, 2002; Reeve et al., 2004; Sharp et al., 2008). If this therapy is not available expected speech outcomes may not be attained and satisfaction with care diminished. It becomes critical for visiting surgical teams to assess pre-operative expectations so that evaluation of outcomes and impact includes the degree to which the pre-operative expectation of the children and their families are met. Evaluations of STSM which focus only on the surgical services provided may have limited benefit in particular cultural contexts.

Causation beliefs inform surgical experience and need to be considered to optimize access to care. For families that consider God's will to be the causative factor, fatalistic aspects of this perception may affect or delay seeking care and surgical outcomes viewed as predestined. Family beliefs may preclude them from seeking the series of interventions necessary for optional functional benefit, if available, as many cleft patients require more than one surgery, as well as dental care and speech therapy. This is especially true for patient with both cleft lip and palate. When immediate corrective and social results are not experienced the belief that the condition was pre-destined may be re-enforced and follow-up care not obtained.

The perceptions of traditional healers regarding causation and treatment of cleft lip and palate was studied in South Africa (Dagher & Ross, 2004). In developing countries traditional healers are often more accessible to patients and many rely on them. It is ironic, in light of this, that they are rarely included as members of cleft care teams (Dagher & Ross, 2004). Qualitative methodology was used to investigate how a sample of African traditional healers approached the treatment of cleft lip and palate. Standardized, semi-structured interviewers were conducted (n=15). Snowball sampling was undertaken and selection criteria included a minimum of 1 year in practice. The interview questions were pilot tested on a small subsample of traditional healers who met the selection criteria but were not study participants (Dagher & Ross, 2004).

Concerns about testing were addressed by the researchers who chose to conduct the interviews in the settings where the study participants typically practiced. A rigorous approach to qualitative methodology was described, though a specific method, e.g. phenomenology or grounded theory, was not identified. An "immersion in the data" was undertaken to establish themes (Dagher and Ross, 2004). Subsequent analysis of the themes was completed including independent expert review. Descriptive statistics were obtained for demographic information. It was puzzling that the study authors referred to the small sample size (n=15) as a potential study weakness in view of the qualitative approach undertaken to answer the research questions.

The study supported previous investigations into the role of cultural belief systems and how they may affect both perceptions of orofacial cleft etiology and access to surgical treatment (Patel & Ross, 2002; Reeve et al., 2004; Sharp et al., 2008). Trepidation toward the recommendation of surgical intervention was described by the traditional practitioners, and often was related to the belief that the patient's ancestors had caused the cleft. The reasons for such ancestral actions were viewed by some of the healers as a punishment, and for some an outward sign that the child was bestowed with special powers. Regardless of their particular viewpoint, the healers considered that changing the cleft with surgery might be more harmful than the cleft itself (Dagher & Ross, 2004). However, most of them ultimately recommended surgery to their patients, if available, to prevent adverse psychological consequences.

The process of referral for surgical services demonstrated the belief of study participants that collaboration with Western providers was in the ultimate best interest of the patients. Most expressed a willingness to continue the effort towards building strong partnerships to ensure patients were treated in a holistic manner. The study participants defined satisfactory healing as involving the ultimate social and psychological reintegration of the patient to their community, a viewpoint which extends beyond the mere recovery from symptoms (Dagher & Ross, 2004).

The studies described are important in examining cultural beliefs and expectations regarding oral cleft conditions and how these beliefs can affect treatment. Patients, caregivers and traditional providers defined the socio-cultural dimensions relating to access and utilization of cleft care services. With few published guidelines regarding the evaluation of volunteer programs, data provided from multiple perspectives will be needed to develop them.

Program Evaluation and Short-term Surgical Missions

Cleft care is an ongoing process. Both psychological and functional measures have been used to examine patient and caregiver perspectives regarding their care and experiences. Researchers have theorized that ongoing functional limitations and psychosocial concerns may be perceived by patients and their parents (Damiano, et al., 2006; Hunt, et al., 2005; Noor & Musa, 2006; Turner et al., 1997).

Satisfaction with treatment

Satisfaction with treatment has been measured using the Cleft Evaluation Profile (CEP). The CEP was developed by the Royal College of Surgeons Cleft Lip and Palate Audit Groups as a measure of satisfaction with cleft related features (Noor & Musa, 2006, Turner et al., 2007). It consists of an eight-item list, which includes speech, hearing, lips, nose, teeth, bite, breathing and profile. Content validity for the CEP was established by both domain specification and face validity (Noor & Musa, 2006; Turner et al., 1997).

In a study conducted in Malaysia by Noor & Musa (2006) the CEP was used to compare parent and child satisfaction with surgical outcomes. A weighted kappa statistic was used to determine agreement between parent-child pairs. Wilcoxon signed rank sum tests measured differences in parent-child scores. Fair to moderate agreement between parent/child pairs for the CEP items was reported (n=120). For their research the study authors reported that the use of self-reported measures of cleft care outcomes could be confounded by cultural values, with answers reflecting perceptions of social desirability rather than individual opinions regarding outcome. It becomes important for visiting teams to develop strategies to evaluate satisfaction with cleft care that are culturally relevant.

The concept of satisfaction was used in all of the studies to frame outcomes. The concept fit the research problems because it incorporated the multiplicity of concerns faced by cleft patients and their families. The theoretical links of the concept were articulated (Noor & Musa, 2006). A comparison group of children without oral clefts was utilized in two of the studies (Damiano et al., 2006; Hunt, et al., 2005). This strengthened the internal validity of the research, and increased causal inference regarding the psychosocial effects experienced by children with clefts. Random sampling was not undertaken in any of the projects so generalizations of results are limited, however given the small sample of children having the procedure, the data does approximate the population that was treated.

In an initial effort to measure satisfaction with cleft treatment outcomes for children in the Philippines (n=44), an exploratory study was conducted to develop a model to locate post-op patients following surgery by STSM and to collect primary outcome data (Sharp et al., 2008). Structured interviews with open ended questions and scored questionnaires were used. Improvements were reported in speech, eating, socialization and physical appearance. Reduced teasing and other social benefits were also measured. Ability to make friends was described as “a little better” (Sharp et al., 2008). The study also concluded that it was feasible to conduct research to measure outcomes of surgery delivered by medical missions, and was the first published attempt to collect data in this area. Reporter bias was minimized in the research by using a local team to conduct the study that had no involvement with the medical missions and was unaware of the presurgical expectations of the families. The study was limited to only primary benefit measurement, and did not attempt to measure long-term impacts or outcomes of the surgery for the patients and their families.

Measuring patient and parental satisfaction with clinical outcome is necessary to ensure that humanitarian care provided around the world incorporates the perceptions of care recipients in determining quality, as occurs in wealthy nations (Smoot, et al., 1992).. Any suggestion by mission team members that patients in developing countries should be grateful for the free medical service provided is an attitude that can hinder discharge care planning and preclude assessment of patient satisfaction (Smoot, et al., 1992).

Practice profiles and quality indicators

Developing quality indicators and practice profiles for the evaluation of services provided by STSM is an important next step to address the aforementioned criticisms of them(Fischer, et al., 2001; Bermudez, et al., 2010, McQueen, et al., 2009). Evaluation research regarding the anesthetic practices of volunteer providers abroad was assessed using a self-report quality survey (Fischer, et al., 2001). Physicians and nurses participating in Operation Smile surgical missions completed the survey at multiple sites over an 18-month period (n=6,037). The authors reported that the study demonstrated the feasibility of obtaining data to evaluate anesthesia delivery abroad, and served as a call to action. The investigators reported that to their knowledge this study was the first attempt to develop quality profiles in order to evaluate the anesthetic outcomes of STSM, a statement which is unsettling in light of the volume of surgeries performed by them.

The tools used to evaluate anesthetic practices were developed by Operation Smile anesthesia providers in an effort to establish face validity for the survey. Adverse events, e.g. laryngospasm, upper airway obstruction, bronchospasm or arrhythmia, were included as measures of quality (Fischer, et al., 2001). These events were chosen as the outcome indicators to be examined based on the literature and personal experience. After completion, survey reports were presented for six age categories and were profiled based on age strata. The surveys were verified by comparing the self-report to the documentation in the patient chart. The data extractor also imported missing items from the medical record to the survey instrument. The investigators did not report reliability measures taken regarding this process and offered no comments on reviewer training.

The study was important in profiling anesthetic complications and how the age of the patient relates to them. For example, airway complication while the children were in the operating room were the most frequent challenges and children less than 5 years old were reported have a relative risk of airway complication 3.4 times that of children greater than 5 years. Inadvertent extubation was strongly correlated with younger age. Statistical values and significance data were not reported. This risk of largospasm was 10 times higher for patients who lived with a smoker (Fischer, et al., 2001). This is an important factor to consider in developing countries where many families cook inside, over open fires.

Self-reporting by providers regarding the complications experienced by patients under their care is a weakness of the study design. The potential for reporting bias to influence the data collected must be considered. Interpersonal aspects of STSM team formation have been posited to influence the occurrence of compilations and the resolution of them (Fischer, et al., 2001). The potential of the interpersonal dimensions of volunteer teams to influence the validity of self-reporting for this project was not explored.

A longitudinal investigation by Bermudez et al (2010) of cleft surgical outcomes was undertaken to assess the quality of services provided by Operation Smile surgeons (Bermudez, et al., 2010). For this multicenter pilot project photographs of patients receiving treatment were taken pre-operatively and immediately post-operatively. At 6-12 month intervals a follow-up photo was taken and the images were consolidated along with clinical data, and entered into a database (Bermudez, et al., 2010). The photographs and accompanying data were sent to a group of independent experts for evaluation of outcomes (n=562). The quality indicators assessed related to facial appearance, and included the shape and symmetry of the nose and lip. The presence of post-operative oral or nasal fistulas following cleft palate repair was also documented.

Mono-method bias was a concern with this study as photographs were used as the measure of quality outcomes. The addition of clinical data regarding complication, e.g. drawing of fistula locations and incorporation of physician notes, helped to minimize this threat to validity. It was mentioned that speech outcome data was also collected in this effort, but no details or analysis were reported. Reliability issues associated with multiple individual taking the photographs were explored by study authors who proposed that in the future images will be standardized, and obtained by individuals trained in the standardization process.

After the post-operative images and clinical data were analyzed by blinded independent experts, feedback reports were sent to 134 of the volunteer surgeons. This auditing process was a particular strength of the study design as the nature of STSM often precludes follow-up by surgeon providers, who may be unaware of the outcomes of their efforts.

Selection bias is a concern for both of the projects. Bermudez, et al., (2010), reported that not all sites returned collected images to Operation Smile investigators. Of the sites that did provide images of patients pre-op and immediately post-op (n=1,917), only 36.6% (n=703) had obtained 6 month-1 year follow-up images. Of the initial 4,100 photographed subjects, only 562 were entered into the database and analyzed. The investigators of anesthetic outcome (Fischer, et al., 2001) stated that cases were dropped for analysis if key data was missing from the medical record. This included the type of service provided, the type of anesthetic administered or the patient's age. However, it was not indicated if the missing data which precluded inclusion was associated with a particular provider or site. For both of the projects specific details concerning which sites returned or had missing data and methods to address this potential threat to reliability and validity were not provided. The patients who were able to return for post-operative follow-up photographs may have only included families with the means to travel for follow up, families with significant outcome concerns, or families living in locations in which the political and social climate did not inhibit travel for follow-up.

In 2005 an electronic medical record (EMR) was instituted by Operation Smile and reported to be the first known attempt by an international surgical organization to institute one (McQueen, et al., 2009). Implementation of the EMR afforded monitoring of outcome indicators for quality assurance reviews. Review of existing medical record data was completed (n=8,000). Program development, priority of surgery and complications were analyzed. Volume data (the number of operations) by procedure type (the specific surgery performed) was described and related to the age of the patient when service was received. Later presentations for surgical services were documented. Age/weight ratios were measured and compared to the (US) CDC growth charts. Complication rates were reported to be similar to those in the US and Europe (McQueen, et al., 2009).

Using electronic technology for medical record data facilitates research efforts for STSM that are aligned with established quality measures. Creating electronic databases allows researchers easy access to data. This research allowed comparison of quality indicators such as anesthetic or surgical complications to standard indices in the US and United Kingdom. It also developed quality measures specific to surgical missions, such as prioritization of care strategies and resource allocation guidelines. The authors acknowledged that surgical priorities may differ from standards in the U. S. and Europe (McQueen, et al., 2009). By offering unique guidelines for surgical resource allocation amid the challenges encountered in addressing the surgical burden of disease worldwide, the study serves as an anchor point for research related to the ongoing response of the medical outreach community to disparities in surgical care.

Drawing from 20 years of experiences Eberlin, et al. (2008) recommended guidelines for quality assurance and bridge building to ensure continuity for surgical outreach programs. A set of quality indicators were proposed that were based on the expertise of members of the organization Medical Missions for Children (MMFC). The study authors described a typical MMFC mission and presented the quality guidelines they developed within this context. Three phases of MMCF surgical mission were identified: the pre-operative phase, the peri-operative phase and post-operative care. Careful detail was described for each of these phases and specific goals were developed for them. At each phase interface with the local community was emphasized. Safety and preparation tactics were also incorporated into the goals. Parity with US quality standards was stressed in goal setting. The authors reported historical outcome data but offered little analysis.

Evaluation models for medical and surgical missions

Models developed to monitor humanitarian activities in the developing world through self evaluation and benchmarking have provided mechanisms to review and audit practices. The value of these activities cannot be understated, as current debate demands that the medical outreach community develop and maintain methods for accountability. Evaluation of planning and development, outcomes and impacts can be facilitated by implementing evaluation models.

A standardization tool for short-term medical missions was developed to evaluate patient safety and quality measures (Maki, et al., 2008). The authors decried the paucity of literature on the subject and emphasized that establishing the means to evaluate short-term medical missions is crucial for quality. Quality was defined using the definition developed by the Institute of Medicine, which established criterion validity for study concepts. The study was executed in 3 phases, which consisted of a base-needs analysis, surveys for self-evaluation by mission personnel and field testing of the surveys (Maki, et al., 2008).

A focused approach to developing the survey instrument was undertaken with investigators conducting in depth interviews of mission personnel and recipients of care in an attempt to uncover important factors in the evaluation of quality. Three missions operating in Latin America were selected to participate in survey development. Criteria for inclusion were identified and the authors reported that heterogeneity of mission size, goals and affiliations were sought. Six major and 30 minor factors were identified as important in evaluating mission quality. Major factors included: cost, efficiency, impact, preparedness, education and sustainability. The constructs defining these factors were presented. For example, impact was defined as a measure of the quality and effectiveness of the collective medical interventions as perceived by patients and providers. A matrix detailing the relationship between the major and minor factors was created, outlining a framework to establish content validity for the surveys.

The six major and 30 minor factors were incorporated into 5 distinct surveys for mission stakeholders, including the host country/local providers, mission personnel, patients, mission administrators and mission directors. Having specific surveys for the type of mission participants was designed to offer a 360 degree viewpoint for evaluation purposes (Maki, et al., 2008). This design feature strengthened the internal validity of the evaluation model as a measure of overall quality. The surveys took approximately 10 minutes to complete, with the exception of the one developed for mission directors, which required approximately 45 minutes (Maki, et al., 2008). Details regarding why the survey for mission directors was so much longer and how this variance may impact the 360 degree vantage point were not offered.

During the pilot phase five missions participated in the survey. Inclusion criteria for pilot participation was documented and designed to reflect heterogeneity of types of surgeries performed by the medical missions operating around the globe. The investigators attempted to secure a sample of 10% of patients in each study site, although inclusion criteria for optimal model usage were vague. All mission personnel completed the questionnaire (Maki, et al., 2008). The surveys were a combination of forced response, Likert scale and open-ended questions. The surveys for patients were given orally in their spoken language. Each question was assigned a maximum value of five and each response was given a corresponding point value. The open-ended questions were not scored numerically, and served solely for informational purposes (Maki, et al., 2008). Completed surveys were entered into a database. Feedback reports were compiled and a percentile evaluation of the six major factors was presented.

Responses could be compared to those of other missions. The model provided a mechanism for survey results to be compared with other missions. The snap shot provided for in the evaluation model can be immediately useful. For example if the result of a mission survey indicate areas of weakness in comparison with percentiles provided in the feedback report, plans for growth in those quality factors can begin.

The study investigators reported the surveys were the first objective measure of short-term medical mission quality that can be standardized across missions (Maki, et al. 2008). They also serve to stimulate dialogue and focus program development. The use of self-report survey to measure quality is a design weakness. This was mitigated somewhat by the 360 degrees of evaluation accomplished by the having surveys completed by all mission stakeholders. Validity of the surveys for specific types of missions, e.g. surgical vs. primary care, needs to be established. Further testing of the survey instrument is an important and necessary step to advance them as a measure of quality to evaluate short-term medical missions.

A philosophical model was structured on three levels was developed by Zbar, Rai, & Dingman (2000). The three phases of the model are observation, integration and independence. Study authors questioned current models in which visiting surgical teams provide free surgical services, and suggest that this is based on the premise that surgery alone can achieve quality outcomes for children with cleft lip and/or palate (Zbar, Rai, & Dingman, 2000). Research has demonstrated that cleft care is ongoing and requires a multidisciplinary team approach (Bearn, et al., 2001; Shaw, et al., 2001). Functional and psychosocial outcomes are supported when follow-up services are provided and necessary treatment is available. Without provisions for aftercare, STSM were characterized as aesthetic rather than functional undertakings (Zbar, Rai & Dingman, 2000).

Observation is phase I of the model, and requires an interested local host. During this phase surgeries performed by visiting teams must meet medical standards for peri-operative care used in the developing word. Sensitivity to local culture is underscored. A diagram of this phase is presented and the terms misdirection and stagnation are introduced (Zbar, Rai & Dingman, 2000). Misdirection was defined as a failure to meet the goals of active participation by host provider, to maintain quality standards, or to progress towards phase II. Medical colonialism was defined as the profiting of guest providers from the engagement in a manner unavailable in their countries of origin. This included the independent performance of procedures that the provider is unqualified to perform, or lack of privileges to perform in their home countries.

Phase II of the model is integration. At this level the host health care providers play key roles in the provision of care. Collegial interaction and academic exchanges are emerging (Zbar, Rai & Dingman, 2000). A prerequisite at this phase is the establishment of a cleft board, which consists of host and guest providers whose function is to analyze cleft cases and care provided. Misdirection and stagnation are also possible at this phase. Termination is incorporated into the model at each phase if the pitfalls of misdirection or stagnation cannot be overcome.

Phase III, the final phase, is independence. This occurs when the host providers maintain the site during absence of visiting teams and offer ongoing clinical services. Financial hurdles at this phase are a concern. As guest teams begin to withdraw so often does the funding they provide.

The authors propose that the model outlines a safe and effective mechanism for the provision of orofacial cleft services in developing countries. The model remains untested and reliability and validity have not been identified or addressed. Strategies for overcoming lack of progress among the phases were not presented. The potential for stagnation and medical colonialism described in the model offers a chilling portrait of the potential perils of humanitarian efforts.

Summary

Summarizing the published literature on the subject of STSM uncovered themes: access, outputs, outcomes and evaluation. Access to care is inhibited by lack of local providers and services, lack of knowledge about treatment options, poverty and personal and cultural attitudes and beliefs. Outputs of STSM were reported as the volume and type of surgeries performed, annual trends in volume, practice profiles and quality. Outcomes of STSM were operationalized by analysis of the age of primary repair of the lip or palate documented over a time course, and by measuring patient and parent satisfaction with treatment. Models to evaluate STSM have also been developed, which underscore the value of partnerships with local providers and systems.

STSM are an important vehicle for addressing the large numbers of untreated clefts in developing countries (Bermundez, et al., 2010). Future research aimed at evaluation of the activities of STSM is necessary to inform the debate surrounding them. A scientific approach is necessary to measure the stability and quality of present models, and to predict the future utility of them, as the surgical outreach community responds to global disparities in surgical services for children born with orofacial clefts in developing countries.

Chapter 3 will present the CDC Framework for Evaluation in Public Health (2007), which was used to develop an organizing framework for program evaluation (Figure 3.1) that was utilized in meeting the research aims detailed in Chapter 1. Creation of the organizing framework enabled the development of the research design and methodology, in an effort to scientifically approach STSM evaluation.

Chapter 3: An organizing framework based on the U.S. Center for Disease Control and Prevention Framework for Evaluation in Public Health

Introduction

The purpose of this paper is to explore how the first four steps of the U.S. Center for Disease Control Framework were used to guide the development of a study to evaluate a short-term surgical mission (STSM). I have volunteered with a STSM called Faces of Hope (FOH) for five years, in both the capacity of clinical nurse in surgical services, and as the Charge Nurse for the mission. This paper will describe how an evaluation process for Faces of Hope was initiated using the CDC Framework as a guide.

CDC Framework for Evaluation in Public Health

The CDC Framework for Program Evaluation in Public Health (CDC Framework) was developed in response to a lack of conceptual models for the evaluation of public health programs (CDC, 1999). Evaluation is described as the systematic investigation of the merit, worth, or significance of an object (CDC, 1999). Acknowledging that public health is expanding its targets beyond communicable disease to include chronic diseases, violence and the social contexts, which influence health disparities, the process of developing a framework was started in the hopes of combining evaluation with program management (CDC, 1999). Integrating the CDC Framework in evaluation design was proposed as a method to foster program improvement and detect program effects or outcomes.

Initial development of the model began in 1997 when the Evaluation Working Group (EWG) was commission by CDC executives to create a framework that summarized and organized the basic elements of program evaluation (CDC, 1999). The EWG consisted of health care representatives from throughout the CDC in collaboration with state and national public health officials. Additional contributors came from 8 selected reference groups; made of evaluation experts, public health program managers and officials, community based researches and non-governmental health organization directors.

The structure of the CDC Framework is comprised of the following steps: 1) engage stakeholders, 2) describe the program, 3)focus the evaluation design, 4) gather credible evidence, 5) justify conclusions & 6) ensure use and share lessons learned. Additionally, 30 standards for assessing the quality of evaluation activities were identified and organized into four groups: 1) *utility*-confirming that the information needs of the users of the evaluation are met, 2) *feasibility*- establishing that the evaluation is viable, practical and designed to produce valuable findings, 3) *propriety*-ensuring that the evaluation is ethical and undertaken with regard towards the interests of those involved and effected, & 4) *accuracy*- confirming that the findings can be considered correct.

Validity

A workshop was sponsored by the EWG to develop the content and establish the validity of the CDC Framework. Ninety health experts from the eight reference groups participated. Additionally interviews were conducted with 250 public health stakeholders. Ultimately an estimated 10,000 professionals were involved in the development of the CDC Framework, by participating in information sharing sessions through the CDC's Public Health Training Network(CDC, 1999).

The CDC Framework was tested in two studies. In a study by Logan, Boutotte, Wilce, & Etkind (2003), it was used to develop a self-evaluation tool for tuberculosis contact investigation programs. For the study researchers used the first three steps of the CDC Framework: 1) engage stakeholders, 2) describe the program, & 3) focus the evaluation design (Logan, Boutotte, Wilce, & Etkind, 2003). The study authors concluded that the CDC Framework provided a useful methodology for beginning the evaluation process of TB contact investigation programs. It was reported that by engaging stakeholders to assist in planning the evaluation, as detailed in the CDC Framework, the self evaluation tool developed was based on a variety of value systems, which supported its validity (Logan, Boutotte, Wilce, & Etkind, 2003). The use of logic models in describing the program, also suggested when using the CDC Framework, was credited with documenting the programs effect on outcomes.

To address the fragmentation of monitoring and evaluation efforts across the multiple agencies responding to HIV/AIDS, the CDC Framework was utilized in concert with the UNAIDS' National AIDS Programs Guide to Monitoring and Evaluation to develop a CDC Public Health Questions approach to unify HIV/AIDS monitoring and evaluation (Rugg, Carael, Boerma & Novak, 2004). The CDC Framework was used to establish an evaluation model in which problem identification, understanding potential responses, monitoring and evaluating national programs and determining collective effectiveness were organized in a step wise hierarchical fashion. The model developed was structured so that program inputs, activities, outputs and outcomes could be identified and measured (Rugg, Carael, Boerma & Novak, 2004).

Step 1: Engage stakeholders

Engaging stakeholders marks the beginning of the evaluation cycle (CDC, 1997). Stakeholders include the people and organizations that are invested in the knowledge to be gained from the evaluation and how that information will be utilized. The CDC Framework categorizes stakeholders as: a) those involved in program operations, b) those served or affected by the program, and c) primary users of the evaluation. The CDC Framework acknowledges that variations in stakeholder involvement exist, and depend upon the situations surrounding individual programs. For example, STSM operating in developing countries may be constrained in engaging the children or families who received surgical services; however when stakeholders are not engaged important element of the program may be missed.

Engaging surgical outreach programs with the research community can be a starting point for the development of evaluation studies Many STSM are operated by private organizations and NGOs with limited, if any, relationships with academic or research institutions. Trust must first be established between researchers and mission volunteers. The perception that the evaluation activities are an effort to judge the program must be avoided. Against the current backdrop of often rancorous debate on the subject of STSM, the process of forging trust can be challenging. However overcoming this challenge is necessary to ensure that the values of a particular stakeholder or groups of stakeholders are not over-, or under-emphasized (CDC, 1997). A solid engagement in the evaluation process is required for mission directors to open up their records or supply data for research designed to evaluate the activities of their organizations. For the volunteer directors of STSM their experiences with the research process may be limited. Knowledge of the ethics and Institutional Review Boards, which drive protocols, may be undeveloped. If necessary, STSM personnel must be willing to expand their knowledge of the research process and participate in the protection of human subjects as they begin to engage with the research community.

Step 2: program description

Faces of Hope (FOH) are a private volunteer organization which was established in 2005 and provides surgical and dental services in Antigua, Guatemala. Headquartered in Oakland, CA, FOH was started by Rotary District 5170 in collaboration with medical professionals and community volunteers. An international team of physicians, dentists, nurses and support volunteers convenes annually for a 10-day period, and surgical repair of oral clefts and dental care are the primary activities.

Mission

a) To provide life changing plastic surgeries, particularly cleft lip and palate surgery, for those with the most need and least access.

b) To build the capacity of Guatemalan health care programs. The three key elements of the programs are: a) surgical and dental team, b) patient and family care team, & c) partnership building team. These elements are combined to provide surgical and dental care to children in Guatemala with limited or no access to services, while building the capacity of existing programs in the country.

Resources

FOH is funded by private donations, district Rotary funds and matching grants with Rotary clubs in Guatemala. The process of establishing matching grants between Rotary clubs in the United States and Guatemala solidifies partnerships and engages stakeholders in both countries. Additional resources and support have been secured by working with organizations operating in Guatemala, such as Partners for Surgery, and with Peace Corps volunteers who provide patient referral and fieldwork prior to the arrival of the teams. Surgical and dental services are delivered in Antigua, Guatemala at the Obras Socialies Del Santo Hermano Pedro (Obras). This facility is operated by Franciscan monks and nuns, and serves as a hospital, an orphanage, and home for mentally and/or physically disabled Guatemalans. Twenty-four hour nursing care is provided. Four operating theatres at Obras were equipped modernized with financial assistance from Partners for Surgery and private donations. Faces of Hope medical personnel work with the medical and nursing staff at Obras in the provision of surgical care.

Stage of Development

The CDC Framework recognized that programs mature over time. Three stages of development are defined within the CDC Framework and include planning, implementation, and effects (CDC, 1997). FOH has been in operation since 2005 and is in the implementation stage of development. The goals of evaluation activities at this stage include the characterization of the real activities (outputs and outcomes) of the program. These activities are then compared with the programs aims and missions (access). Data regarding outputs and outcomes is collected in an effort to describe operations and relationships. Improvements are initiated if necessary. Plan revisions may also occur, based on the findings at this stage.

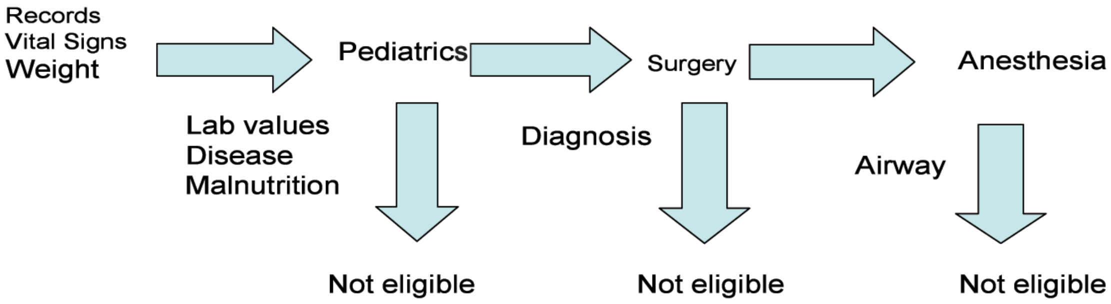
Context

FOH operates within the context of humanitarian assistance. The organization endorses the premise that lack of access to surgical services for children with orofacial clefts in developing countries leads to ongoing functional and social limitations. This includes concerns with speech, hearing, appearance and stigmatization. FOH endeavors to meet the challenge of addressing global disparities in surgical care by offering surgical and dental assistance for children born in Guatemala with orofacial clefts.

Logic Modeling

The CDC Framework describes the used of logic models to map the sequence of events that are designed to bring about change (CDC, 1997). In the study conducted by Logan, Boutotte, Wilce, & Etkind (2003) this process was credited with documenting the TB contact investigation programs effect on outcomes. By employing a logic modeling technique the relationship between patients accessing services from FOH and eligibility to receive those services was mapped. The model charts patient flow at the opening clinic for Faces of Hope and the interfaces in which eligibility to receive services may be denied.

Figure 3.0. Model of patient flow at opening clinic.



Steps 3 focus evaluation design; Step 4 gather credible evidence

Focusing the evaluation design was achieved by creating an organizing framework (Figure 3.1) which was adapted from the Public Health Questions Approach to Unifying HIV/AIDS Monitoring and Evaluation model, described previously, and developed by Rugg, Carael, Boerma & Novak (2004).

Figure 3.1. Organizing Framework for *Faces of Hope* Evaluation.

Data applied to the components of the organizing framework can be used to begin the process of program evaluation. An example would be to use data abstracted from medical records kept by FOH to answer research questions framed within the components of the model; assess, outputs and outcomes.

Medical record data is considered to be credible and has been referred to as the 'gold standard' (Aaronson & Burman, 1994; Banks, 1998; Ballard, 2003; Zhan, et al., 2007). Clinical indicators contained in medical records have been used to establish baseline performance (Ballard, 2003; Bermundez, et al., 2010; Shaw, et al, 2001). Opportunities to compare individual organization quality indicators with national and international trends are also afforded (Ballard, 2003). Medical record abstraction methodology is described in detail in chapter 4. Examples of how medical record data can be used within the organizing framework for FOH evaluation is detailed below.

Access

Anthropometric measures such age, weight, hemoglobin and hematocrit (H&H) are routinely collected and recorded in the medical records of STSM (Eberlin, et al., 2008). FOH medical records contain this data. Weight/age ratios when compared with World Health Organization (WHO) growth standards can provide a glimpse of nutritional and developmental status. Children with low weight for their age or abnormal H&H values are often denied access to surgical services provided by STSM (Eberlin, et al., 2008). Unfortunately, for families in countries with political and economic instabilities, significant risks may have been undertaken to travel for services, and at considerable expenditure. When their children are deemed ineligible due to nutritional or developmental deficiencies, they are faced with the prospect of a potentially dangerous and costly return home. The possibility of having them return for subsequent missions may be diminished.

Medical record data used in evaluation processes over time can provide information about changes and trends in anthropometric measures. If children are trending downward on measures such as weight- for- age and Hgb/Hct values, characteristics of the political and economic climate need to be considered and used as a guide to planning and resource allocation (Schwartz & Khadka, 2004). Evaluating the predictors of receiving surgical services can provide information to design and implement procedures to maximize eligibility. Additional fieldwork prior to surgical teams arriving may be indicated, with vitamin or nutritional supplements provided to families in anticipation of travel to seek services. If necessary, assessment of educational needs or concerns regarding feeding can be undertaken and any needed supplies (syringes, specialized nipples) distributed before surgical teams arrive. Parasite screening and treatment can be undertaken to improve anemia. These possibilities underscore the importance of STSM forging strong partnerships between local organizations and care workers in the field (Bermundez, et al., 2010; Moore & Fernandez, 2006; Sharp, et al., 2008; Zbar and Rai, 2008). Without these relationships the ability to implement adaptive strategies to foster eligibility to receive surgical services may be constrained.

Outputs

Cross-sectional data from FOH clinical records can be used to measure and describe relationships within the organizing framework. Comparing data abstracted over a time course builds in additional advantages, including the evaluation of changes over time.

Quality

An example is to measure complications, such as wound dehiscence and infections, and benchmark with standard indices. The Surgical Quality Improvement program in the United States uses routine medical record review to measure and compare quality indicators for surgical procedures (Kaafarani & Rosen, 2009). Taken at a single point the data has value in evaluating the activities of a given year. This can trigger further analysis including the complications associated with individual providers, or per diagnoses. Appropriate feedback can be given based on these results. Detecting providers with low quality outcomes and harmful or unnecessary procedures as soon as possible is important to maintain and build quality for humanitarian volunteer-based surgical models (Bermudez, et al., 2009).

Analysis of FOH medical record data over time may be used to identify trends. For instance, are complication and infection rates staying the same, or are they changing over the years? Do they go up or down? Evaluating the activities of organizations over a time course of 5 years establishes a track record. It has been proposed by Bermudez, et al. (2010) that detection of treatment modalities for cleft care that may be unnecessary or cause harm can be accomplished within five years, if significant differences in intercenter variation on outcome measures are demonstrated.

Volume

Surgical volume measures can be considered similarly. Volume data taken as a single data point is an important descriptor of the service provided in a particular year. For private organizations such as FOH, where a large percentage of annual funds come from individual or group donations, volume data per annum provides donors with information regarding outputs and utilization of their contributions. Measuring surgical volume over successive years allows trending. Program stability can also be assessed. In Ghana, researchers suggested that increasing volume each year was a proxy measure for satisfaction with outcomes (Donkor et al., 2007). It was posited that volume growth associated with children who returned for additional treatment, as in the case of children with both cleft lip and palate, was a demonstration of satisfaction with initial care. This underscores the importance of organizing frameworks, which reflect the fluidity of the evaluation process.

Resource planning for the next year also begins with output data. Describing the relationships between additional surgical needs and specific variables, such as pre-operative diagnosis, and integrating this information with volume data informs materials management activities in which provisions are made available to meet anticipated demands. For example, if specific diagnoses are found to significantly predict additional surgical needs, frequency data regarding diagnoses can be used to foresee probable case mix and case load for future missions.

Outcomes

Medical record data obtained from FOH over its 5-year history can be used to analyze age and clinical profiles over the time course and may provide evidence for earlier treatment. If a statistically significant difference in mean age is measured over time, specifically if the mean age for primary cleft repair is lowered, the value of FOH in providing access for early surgical treatment of orofacial clefts is supported.

Conclusion

The CDC Framework proved a useful guide for developing an organizing framework (access, output, and outcomes) to initiate evaluation of FOH (Figure 3.1). Engaging the FOH Mission Director was a crucial first step in this process. Incorporating the perspectives of investigators from the University of California, San

Francisco (UCSF) with mission personnel was necessary to initiate an evaluation process which reflected shared values. In this way trust was established which may support inclusion of evaluation results in ongoing FOH activities. Trust and engagement between investigators and FOH administrators was fostered by partnering to discuss the FOH goals and activities that define the program Vis a Vis the current themes and gaps identified in the literature. This discussion served to motivate FOH administrators to work with investigators from UCSF to advance the science of evaluating the STSM model in addressing disparities in surgical care. The result was a willingness of the FOH Mission Director to actively participate in the evaluation. Ultimately FOH administrators agreed to share 5 years of medical record data with investigators from UCSF after Institutional Review Board approval was secured.

By utilizing a logic model (Figure 3.0), as described in the CDC Framework, the sequence of events involved in providing access to surgical services from FOH teams was mapped. This process identified eligibility as a concept in which to operationalize and measure access. Steps 3 & 4 of the CDC Framework (focus study design and gather credible evidence) resulted in the development of an organizing framework to evaluate FOH activities (Figure 3.1), as well as a method to obtain data (medical record abstraction). These results are aligned with published data in which medical records have been utilized to study the impacts of STSM in addressing case backloads resulting from cleft care disparities and to evaluate services provided (Donkor et al., 2007; Moore & Fernandes, 2006; Uetani et al., 2005).

Chapter 4 describes the method of medical record abstraction and it's application to this study. The results obtained by implementing the research design and methodology are detailed in Chapter 5. Discussion of the results and their value in evaluating FOH are described in Chapter 6.

Chapter 4: Methodology

Methods

This chapter will be organized in the following manner. After description of the study design, sample and measurements, there will be a thorough discussion about medical record abstraction to detail the theoretical underpinnings of the methodology. A discussion about how specific details of the methodology apply to this study will be included in the discussion of data abstraction, management and analysis.

Study design

A cross sectional design was used to evaluate the activities of Faces of Hope (FOH) for the years 2005-2009, through analysis of data abstracted from their medical records. The FOH records documented the clinical activities of FOH in Antigua, Guatemala. Yearly cross sectional data from FOH clinical records was used to measure and describe relationships. Comparing the data abstracted over the years allowed evaluation over time. The Framework for Program Evaluation in Public Health, developed by the Center for Disease Control (CDC) in 2007, guided the development of a conceptual framework that was used to organize the study. The framework, entitled the Organizing Framework for *Faces of Hope* Evaluation, is described in Chapter 3 (Figure 3.1). The framework provided an anchor for the medical record data, enhancing the construct validity of the study design.

Sample

The sample size is fixed. There are 706 written medical records, which contain demographic and clinical information that describes patient screening and selection, operative and post-operative care, and discharge. Data were abstracted from January through April 2011 by the principal investigator. Data from the written records were entered into SPSS 17 software.

Power analysis was completed using G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007). Table 4.0 shows the power analysis and demonstrates that there is adequate power to answer the research aims.

Table 4.0 Power Analysis Calculation Methods

Outcome	n	alpha	power	method	effect	Effect direction/size
Total volume* (R.A. #1)	65	.05, 2 tail	0.8	X2 df=10	0.5	large **
	181	.05, 2 tail	0.8	X2 df=10	0.3	medium**
Volume by diagnosis*: (R.A. # 1)	65	.05, 2 tail	0.8	X2 df=10	0.5	large **
Mean age (R.A. #2)	200	.05, 2 tail	0.8	ANOVA	0.25	medium**
Eligibility (R.A. # 4)	466	.05, 2 tail	0.8	L.R.	OR=.75	
Subsequent tx (R.A. # 5)	242	.05, 2 tail	0.8	L.R.	OR=1.5	

L.R. = logistic regression   R.A.= Research aim (see Chapter 1).

* To evaluate this variable both the total volume of all surgical procedures combined, and the volume by specific pre-operative diagnosis was used as a measure of research aim #1.

Ethics

The University of California Committee on Human research approved this study and granted a waiver of informed consent. This ethical issue was considered by previous research involving medical records, which concluded that the use of existing medical records for observational research posed no risk to physical harm and minimal risk to privacy (McCarthy, et al., 1999). Retrospective medical record research has been exempted by current federal policy from Institutional Review Boards when no patient identifiers are included (McCarthy, et al., 1999).
The records reviewed for this research were de-identified by a research assistant working for the primary investigator to protect subject privacy.

Measures:

Patient Demographics:

- Gender: boys and girls.
- Location of current residence:. Clinical measurements
- Age in weeks and age in months: measured using date of birth data in the clinical record.
- Weight in kilograms: obtained during pre-operative evaluation by team RNs.
- Hemoglobin (Hgb): the oxygen-carrying pigment of the red blood cells and is reported in grams per 100ml (Tilian, S., Conover, M., & Tilkian, A, 1979). This

test is part of the pre-operative screening process and is completed at the laboratory associated with the host hospital.

- Hematocrit (Hct): is the volume of packed red blood cells found in 100ml of blood (Tilian, S., Conover, M., & Tilkian, A, 1979), and is part of the pre-operative screening process. The test is completed in association with the Hgb analysis at the host hospital.
- Weight/age ratio: this is the weight in kilograms divided by the age in months.
- Weight/age in months Z scores: calculated for the sample in SPSS 17.
- Oxygen saturation at pre-operative clinic: measured by the team RNs at the pre-operative screening process as percentage of Hgb oxygen saturation, measured on a digit, using a portable oximeter device.
- Body temperature at pre-operative clinic in degrees Celsius. This measurement is obtained by the team RNs at the pre-operative screening clinic using a tympanic membrane thermometer.
- Pre-operative diagnosis: determined at the pre-operative screening clinic by team surgeons.
- Illness/Co morbidities: determined at the pre-operative screening clinic by team physicians. This includes upper respiratory infections, infectious

disease such as scabies, growth and development delays, gastro-intestinal illnesses, cardiac anomalies or syndromes.

- Surgical history: obtained at the pre-operative screening clinic by team surgeons.
- Type of surgical service provided: documented in the medical record by team surgeon.
- Providers: surgeons and anesthesiologists.
- American Society of Anesthesiologist (ASA) score: ASA physical status classification system was developed and is used for pre-operative assessments. A person with an ASA of 1 is classified as a normal healthy individual. A person with an ASA of 2 is classified as having mild systemic disease. This system increases incrementally to ultimately include an ASA of 6, which is a person who has been diagnosed with brain death.
- World Health Organization (WHO) weight- for- age- growth standard Z scores. These international standards were developed by the WHO to reflect both acute and chronic malnutrition. The WHO Z scores give an indication of the overall level of malnutrition.

Clinical quality indicators

- Length of procedure in minutes

- Anesthetic complications: airway issues.
- Surgical complications: wound dehiscence, infection, hemorrhage.
- American College of Surgeons National Surgical Quality Improvement Program Pediatric: Phase I: developed to identify outcomes of children’s surgical care to target for quality improvement (Raval, et al., 2011).

Table 4.1 displays the reference variables abstracted from the medical record and how they were coded and categorized.

Table 4.1 Variables and Coding

Variable		**Code**	**Categories**
Diagnosis	1=cleft lip 2=cleft palate 3=cleft lip & palate 4=bilat. cleft lip 5=bilat. cleft palate 6=fistula 7=other	Diagnosis 3 group	1=cleft lip 2=cleft palate 3=cleft lip and palate
Diagnosis	1=cleft lip 2=cleft palate 3=cleft lip & palate 4=bilat. cleft lip 5=bilat. cleft palate 6=fistula 7=other	Diagnosis 5 group	1=cleft lip 2=cleft palate 3=cleft lip & palate 6=fistula 7=other
Diagnosis	1=cleft lip 2=cleft palate 3=cleft lip & palate	Diagnosis 6 group	1=cleft lip 2=cleft palate 3=cleft lip & palate

	4=bilat. cleft lip 5=bilat. cleft palate 6=fistula 7=other		4=bilat. Cleft lip 6=fistula 7=other
Illness	1=URI 2=GI 3=growth & dev. 4=scabies 5=other	Illness 3 group	1=URI 3=G&D 5=other
Previous surgery	0=none 1=cleft lip 2=cleft palate 3=cleft lip & palate 4=fistula 5=lip revision 6=other	Prev. surg. 6 group	0=none 1=cleft lip 2=cleft palate 3=cleft lip & palate 6=other
post-op complications	1=hemorrhage 2=dehiscence 3=infection 4=other	Any complications	1=no 2=yes
anesthesia complications	1=respiratory 2=cardiac 3=inadvertent extubation 4=difficult IV 5=allergic reaction 6=aspiration	any anesthetic comp	1=no 2=yes
complications		anesthetic & surg.	1=no 2=yes
future needs	0=none	future needs 4 group	1=none 2=surgical

1=surgical 2=dental 3=follow-up 4=exam in 1 year 5=speech therapy 6=not documented 7=other		2=follow up 3=other

Medical Record Abstraction

The use of medical records in evaluating short-term surgical missions (STSM) minimizes language issues when anthropometric or frequency data are the variables to be abstracted (Moore & Fernandes, 2008). Difficulties associated with loss to follow up due to the transitory aspects of the missions are also mitigated (Moore & Fernandes, 2008). Families often travel great distances and return home after receiving treatment. The costs associated with further travel for follow-up can be prohibitive (Schwarz & Khadka, 2004). Transience associated with the poverty these families experience may also exacerbate the problem of locating them post-operatively. Initial efforts in STSM outcome research have underscored this issue. Dr. J. Canady, MD, a leading scientist in oral cleft research, remarked that while trying to locate families for an outcome study of STSM in the Philippines, that the families were more transient than expected. Locating them was not trivial. Without involving local guides, it would have been extremely difficult (personal communication, October 14, 2008). In light of these difficulties, medical records can provide valuable data to advance outcomes and evaluation research of STSMs.

Medical chart abstraction is one of the most common methods for measuring processes of providing clinical care, including that of competence in the clinical setting and activities of organizations (Ballard, 2003; Peabody, et al., 2000 Zhan, et al., 2007). The Joint Commission (JCAHO) Core Measures program requires medical record data for four topic areas as a measure of organizational quality (Ballard, 2003). Medical record data are used in error reporting systems and to monitor patient safety. It is frequently considered the 'gold standard' (Aaronson & Burman, 1994; Banks, 1998; Ballard, 2003; Zhan, et al., 2007). Clinical indicators contained in medical records have been used to establish baseline performance in the identification and dissemination of best practices (Ballard, 2003; Bermundez, et al., 2010; Shaw, et al, 2001). Opportunities to compare individual organization quality indicators with national and international trends are also afforded (Ballard, 2003).

Advantages and limitations of medical record abstraction (MRA)

MRA has the advantages of accessibility and low cost. This is a particular strength when further contact with subjects is precluded. A richness of clinical data exist in medical charts, which is often standardized in content and format (Luck et al., 2000; Pan, et al., 2005; Wu, 1997). Medical records are universal and abstraction is a routine and efficient method of data collection (Luck, et al., 2000; Pan, et al., 2005; Wu, 1997).

There are important limitations of this method to consider. The accuracy and completeness of the medical record must be confirmed if it will be used in analysis and interpretation. This information is often lacking (Pan et al., 2005). Higher methodological standards for chart review have been called for due to their frequent use in clinical and epidemiological research (Gilbert, et al., 1996). Methods for medical record abstractions must be detailed and clear to ensure quality. This approach is also limited by ambiguous research questions (Allison, et al, 2000; Banks 1998). Aggregate concepts need to be well defined to confirm alignment with the data contained in the medical record (Allison, et al., Huff, 1997).

Chart review methodology is limited by both the documentation itself and by the reviewer (Wu & Ashton, 1997). The document may be incomplete and necessary data unrecoverable. The data may be flawed by false or inaccurate information (Wu & Ashton, 1997).

Secondary data

Measurement errors occur in health records. They can be both systematic and random (Aaronson & Burman, 1994). Error can occur at many places. Data obtained from charts is not primary data, but has been filtered, often by many providers. It is removed once again by the abstraction process (Aaronson & Burman, 1994; Banks, 1998). Because chart audits are retrospective, clarification of confusing or missing information is limited (Banks, 1998). Threats to quality exist during original collection, documentation, extraction and interpretation.

Investigations of health records have uncovered quality issues associated with data accuracy (Pan, et al., 2005). Medical records are potentially subject to recording bias. Some practitioners fail to record all of their activities, while others may document well but are poor clinicians (Aaronson & Burman, 1994; Luck et al., 2000). Quality may be over or underestimated by chart review. This challenges a bias long held by some investigators that the information in medical records is correct (Aaronson & Burman, 1994).

Reliability

Information regarding the reproducibility of data obtained from chart review is often lacking (Gilbert et al., 1996; Huff, 1997). Taking measures of reproducibility and discrimination will establish the reliability of the chart review processes. This can be assessed with interrater agreement measures and calculation of error rates (Aaronson & Burman, 1994, Allison, et al., 2000, Grunfeld et al, 2006). Abstraction protocols which include these procedures will address measurement error during data extraction and interpretation (Allison, et al., 2000).

Interrater reliability (IRR) can be measured by having different individuals abstract the same record, or records from the same pool, and the results compared. Percentages of agreement and Kappa statistics may then be obtained (Allison, et al., 2000, Banks, 1998; Eder, Fullerton, Benroth & Lindsay, 2005; Grunfeld, et al, 2006; Huff, 1997). Desired percentages of agreement are set a priori. For important variables a 95% agreement rate is often the criterion (Allison, et al., 2000, Banks, 1998; Eder, Fullerton, Benroth & Lindsay, 2005; Grunfeld, et al, 2006; Huff, 1997). Intrarater reliability is assessed by an individual abstracting the same chart on two occasions and the percentage of agreement calculated. This procedure can be helpful in training phases of chart abstraction studies (Allison, et al, 2000, Banks, 1998).

Reliability has been demonstrated to be adversely affected when the chart reviewers were required to make clinical interpretations regarding the data (Banks, 1998; Eder, Fullerton, Benroth & Lindsay, 2005). Strategies which facilitate standard interpretations, such as algorithms, can promote reliability (Banks, 1998). Data variance is increased if information is obtained from a large number of providers or from different locations in the record. Format differences in charts and the period of time in which the data are collected also impacts inter rater reliability (Banks, 1998; Eder, Fullerton, Benroth & Lindsay, 2005).

Assessing reliability of abstraction protocols can also be accomplished through calculation of error rates. Transcribing hard data facts such as anthropometric measures have the lowest rate of error (Wu & Ashton, 1997). Error rates are measured by having a second individual re-examine the chart to determine if hard data was transcribed correctly. Levels of accuracy are determined by measuring mean error rates for individual indicators (Pan, et al., 2005; Wu & Ashton, 1997). A priori target accuracy of 95% is ideal (Pan, et al., 2005).

The type of data collected affects its reliability during abstraction. Physical assessment finding which are not ambiguous are more likely to be recorded accurately (Aaronson & Burman, 1994). Good inter and intra rater agreement has been measured with variables such as diagnosis, test results, and chief complaints. Less agreement was measured with patient histories (Aaronson & Burman, 1994).

Demographic, outcome and discharge data were reported as most reliable in a study by Grunfeld, et al. (2006). Diagnostic and surgical information were determined the least reliable, as they were subject to a lot of interpretation by data abstractors (Grunfeld, et al., 2006). Standardized charts increase the reliability of the abstraction process. Single abstracted data elements were more reliable than aggregated ones (Grunfeld, et al., 2006; Huff, 1997).

Reliability of medical record abstraction methodology is based on certain assumptions which are often unmet. These include that the data will be present and consistent throughout the record. The ability for the information to be interpretable by abstractors is also assumed (Eder, Fullerton, Benroth & Lindsay, 2005). Errors can occur due to missing data or inability to locate necessary information in the record (Gilbert, et al., 1996). Entries which are illegible, vague or incomplete challenge the precision of the record. Coding inconsistencies and transcription errors also reduce reliability (Eder, Fullerton, Benroth & Lindsay, 2005; Gilbert, et al., 1996).

Factors influencing data quality have been identified (Allison, et al., 2000; Grunfeld, et al., 2006; Huff, 1997). Information related to surgical procedures poses reliability concerns, however error is reduced with abstracter training. Surgical nurses have been considered gold standard in chart abstractions measuring surgical quality (Aaronson & Burman, 1994; Grunfeld, et al., 2006; Hanauer, et al., 2009).

Validity

Unreliable data subsequently compromises validity (Huff, 1977). Validity issues with medical record data begin with the information provided by the patients themselves. Uncertainties in patient responses of up to 8% (N=104) of questions posed to them by health care providers have been reported (Nagurmey, et al., 2005). These same subjects changed their responses when re-interviewed with a subset of similar questions up to 3% of the times (Nagurmey, et al, 2005). Patient cooperation and competence are important factors impacting data validity, which relies on accurate patient input (Aaronson et al., 1994, Nagurmey et al., 2005).

Content validity of medical record abstraction methodology is supported by large and more detailed the abstraction tools. A representative abstraction tool increases the likelihood that using the medical record as a measurement tool will address issues of content validity (Allison, et al., 2000). Construct validity in enhanced when an underlying theory or conceptual framework is used to anchor health record data in answering research questions (Allison, et al., 2000; Banks, 1998).

Establishing criterion validity is helpful in improving the quality of medical record abstraction (Allison, et al., 2000). In this process expert clinical judgment is used as the gold standard, and is developed for each data element to be extracted. Comparison between the information obtained by individual chart abstractors to this gold standard is undertaken, and agreement measured (Allison, et al., 2000).

For much of the data found in health records, the health provider is the measuring instrument. Validity is compromised by provider characteristics, and can result in systematic measurement error (Aaronson & Burman, 1994). False positives were declared when actions recorded in the chart were not reported by standardized patients. False negatives occur when information or interventions are omitted from the chart (Aaronson & Burman, 1994; Dressehaus, Luck & Peabody, 2002; Nagurmey et al., 2005). When false positives occur the medical record may overestimate the quality of care. When false negatives occur, the quality of care may be underestimated (Dressehaus, Luck & Peabody, 2002).

Sensitivity and Specificity

Research comparing the information obtained from medical records with that provided by standardized patients after a medical encounter underscored validity concerns (Luck, et al., 2000). The sensitivity and specificity of the record was reported to have varied with data domains, and researchers concluded that medical records may lack both sensitivity and specificity in certain clinical interactions (Luck, et al., 2000). Medical records lack sensitivity due to false negatives. They lack specificity due to false positives (Aaronson & Burman, 1994; Dressehuas, Luck & Peabody, 2002; Luck, et al., 2000). The sensitivity of the medical record was found to be increased when objective data was recorded and diminished when subjective services, such as patient education and counseling, were documented (Eder, Fullerton, Benroth & Lindsay, 2005). Research by Luck, et al. (2000) suggests that the medical record is more specific than it is sensitive.

Rigor

Though challenges to the quality of clinical data exist, medical record abstraction provides a starting point to begin program evaluation activities. Accountability is crucial for service organizations. Definitive guideline for quality assurance in the surgical outreach community are limited (Eberlin, et al., 2008), and their development can be informed by data abstracted from medical records. Through rigorous protocol, measurement error and biases associated with this methodology can be anticipated, acknowledged and/ or measured.

Missing data

Missing data contributed to lower than planned accuracy targets in research aimed at measuring quality using chart abstraction (Pan, et al., 2005). In the study by Nagurmey, et al.(2005), 85% of patients (N=104) indicated the location of their chest pain, but this was described in their charts only 44% of the time.

Procedures for missing data include data imputation, and case wise and list wise deletions of variables or subjects. The approach is dependent on the specific research question and associated statistical analysis. Missing anthropometric data can be imputed by calculating means, such as age in months. Case wise deletions can be implemented for logistic regression analysis and list wise deletions for frequency variables.

Bias

Potential bias exists for the study. Recording bias may be present in the medical record. Another possible source of bias is that the data abstractor is not blinded to study hypothesis. Because the principal investigator has a dual role in FOH, as a provider of clinical services and a researcher, a potential for bias exists.

Quality

The review process can compound issues with the reliability and validity of the abstracted data if abstractors receive insufficient training or clinical judgment is required to complete the abstraction. The overall quality of medical record methodology depends on complete and accurate documentation as well as reviewer consistency, impartiality and skill in extracting the information from the record (Wu & Ashton, 1997).

Heath care organizations around the world are using clinical quality indicators to identify improvement opportunities, measure interventions and link quality with cost measures (Ballard, 2003). Medical records are a source of data that can be aligned with national and international trends. Evaluations using medical records serve as valuable adjuncts to other quality control measures (Pan, et al., 2005). Health care quality has been defined as the degree to which health care services increase the likelihood of desired outcomes, and are consistent with current knowledge (Grunfeld, et al., 2006). Quality of care studies relying on chart review can not address the degree to which the care received is in line with the non medical needs of the patient (Steffen, 1988). Failure to meet non-medical goals of patients represents substandard care, and underscores the importance of ongoing evaluation research which includes other sources of data to supplement that from medical records (Steffen, 1998; Wu & Ashton, 1997).

Data abstraction, management and analysis

The records from FOH are standardized and specific. This may increase the precision of the abstraction process (Grunfeld, et al., 2006; Huff, 1997). The data extracted were single elements, which have also been reported to be more dependable during abstraction than aggregated ones (Grunfeld, et al., 2006; Huff, 1997).

All of the data contained in the written records were abstracted and entered into SPSS 17 software. The data abstracted were aligned with that which has been reported to have greater reliability during the abstraction process. These include diagnosis, test results, chief complaints, demographic, outcome and discharge data (Aaronson & Burman, 1994; Grunfeld, et al., 2006). The data were used to create two demographic variables, 15 clinical measurement variables and four clinical and quality indicators (see figure 4.0). Prior to abstraction the medical records were de-identified by a research assistant working for the principal investigator.

Having the abstraction for this research completed by a surgical Registered Nurse helped support data reliability, as surgical nurses have been considered the gold standard in chart abstractions measuring surgical quality (Aaronson & Burman, 1994; Grunfeld, et al., 2006; Hanauer, et al., 2009). Issues with inter-rater reliability that result from multiple data abstractors was also mitigated.

Construct validity for this project is supported by utilizing the CDC Public Health Framework for Program Evaluation. This allowed the medical record data to be anchored to a framework to support program evaluation activities. Content validity is fostered by the research questions that benchmark medical record data with established quality and growth indices.

After the data was abstracted, frequencies were run for all of the variables and outliers were double checked with the written records to confirm accuracy. Intra-rater reliability was then checked. This was accomplished by selecting a random sample of 8% of all entries and re-verifying the data with the written record. An error rate of .005 was calculated, an acceptable level of error.

Missing or incomplete data from the records was handled by case-wise deletion during analysis. After deletion, sample sizes were still large enough to meet power analysis estimates for all variables except hemoglobin. Families accessing services from Faces of Hope are required to pay for their laboratory studies. Because of this expense most families wait until they have been selected for surgery before obtaining lab studies, decreasing the number of records, which contained this data. Hemoglobin as a predictor of eligibility was therefore run as a separate logistic regression, and lacked sufficient power to rule out the possibility of a type II error.

1) Describe the demographic, clinical and anthropometric characteristics of the population and compare these features over the time course of 2005-2009.

Frequency histograms and distributions for clinical and demographic variables were completed. Frequency statistics were compared between the years by Chi-square testing. Weight/age data were compared to WHO child growth standards for analysis of growth and development. Univariate logistic regression explored gender as a predictor of pre-operative diagnosis; the data is reported as Odds Ratio with 95% CI.

2) Determine if the demographic and clinical profiles provide evidence for earlier treatment for primary repair or orofacial clefts over time (2005-2009).

To determine if the demographic and clinical profiles provide evidence for earlier treatment across 2005-2009, children with a diagnosis of primary cleft lip, palate or cleft lip and palate, had their age data analyzed. This analysis included: age range, mean age and SD. Estimation of differences between age statistics for the years 2005-2009 was achieved utilizing ANOVA.

3) Describe clinical and quality indicators (complication rates, procedure times), and benchmark with American College of Surgeons National Quality Improvement Program Pediatric Phase 1 standards.

Time data abstracted from the medical records was used to maximize output and evaluate case flow: mean operative times per pre-op diagnosis and by provider, with range and SD. Clinical quality indicators abstracted from the medical records were benchmarked with the National Surgical Quality Standards, phase I, for pediatrics (Raval, et al., 2011).

4) Determine the independent predictors (out of RA 1-2) of eligibility for receiving surgical treatment.

To explore independent predictors of receiving surgical treatment, multivariable logistic regression analysis was used to estimate a set of predictors and their effects on eligibility (receiving services). The data is reported as Odds Ratios with a 95% CI.

5) For a subset of subjects who have data available, determine the predictors (out of RA 1-3) of receiving additional procedures, in subsequent years, from Faces of Hope volunteers?

The relationships between diagnosis, provider, complications, illness/co-morbidities, demographic data and receiving additional procedures in subsequent years was explored with multivariable logistic regression analysis. The data is reported as Odds Ratios with a 95% CI.

Significance

While acknowledging the need for international volunteer assistance due to an increasing global surgical burden of disease, made worse by a lack of trained healthcare personnel in developing countries, program evaluation and monitoring are not routine (McQueen et al., 2009). Criticism of short-term medical missions has resulted from these shortcomings (Dupuis, 2004; Maki et al., 2009; Mulliken, 2004;

Zbar, Rai & Dingman, 2000). Uniform indices to compare quality of performance and equity of care, and the creation of an evidence base to guide the development of standards and outcome measures, would do much to address these criticisms, (Banatavalan & Zwi, 2000; McQueen et al., 2009). Developing methods to evaluate volunteer surgical missions is necessary for quality measurement in global care. Review of the medical records of volunteer surgical efforts and analysis of abstracted data is necessary to advance this agenda.

Chapter 5: Results

Sample

The sample for the study is comprised of 706 medical records that document the clinical encounters of Guatemalan families seeking surgical services from the organization Faces of Hope for the years 2005 thru 2009. The geographic areas served by Faces of Hope (FOH) mission in Antigua, Guatemala, are displayed in Figure 5.0., and show the representativeness of the sample. The data were abstracted from January thru April 2011 from the medical records. The results for the five specific research aims to evaluate the activities of FOH in Antigua, Guatemala, are given next.

Results to answer research Aim #1:

Describe the demographic and clinical characteristics of the population and compare these features over the time course of 2005-2009.

The frequency of patients coming to Antigua seeking services from different geographic regions is represented in the yellow circles. The district (department) of Sacatepequez is the department most represented by families seeking services. This department includes Antigua, so the travel distance is reduced. The two departments which followed in frequency were Alta Vera Paz and Quiche.

Figure 5.0. Frequency of patient visits by geographical region.

Appendix A, Figure 1.0, displays the average age-in-weeks of the patients seeking surgical care. The mean total age in weeks across all five years was 317.48 (SD± 434.62). The ANOVA analysis comparing the average age- in- weeks did not change statistically significantly over the 5 year time course ; [F(4,309)=.259, p=.974].

Appendix A, Figure 2.0, displays the average weight in kilograms of patients seeking surgical services. The mean weight was 18.00 (SD±15.78) across the five years. The ANOVA analysis comparing the average weight-in-kilograms over the 5 year time course did not change significantly;[F (4,306)=.124, p=.974].

Appendix A, Figure 3.0, displays weight to age-in-month ratios, calculated as an initial indicator of nutritional status. The mean weight to age ratio in months was 0.52 (SD±0.47). The ANOVA analysis comparing the average weight/ age-month ratio over the 5 year time course did not change significantly; F(4,305,)=1.34, p=.254.]

Figure 5.1 displays the frequency by type of surgery performed each year. A Chi square analysis showed that no statistically significant association was found between the frequencies of surgical treatment and year.

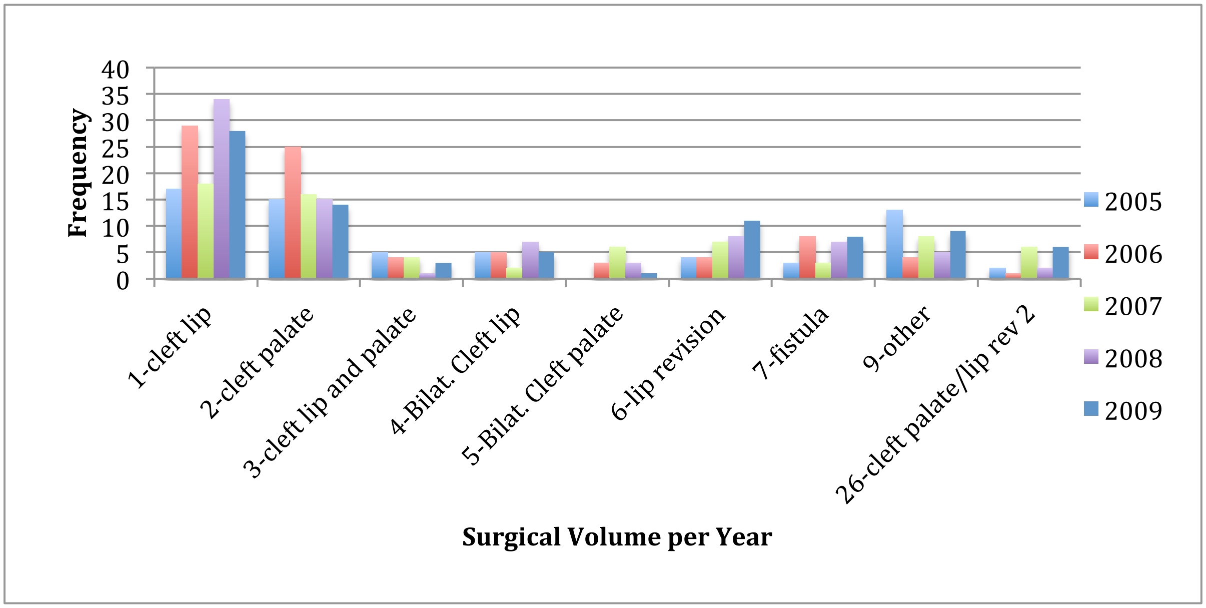


Figure 5.1. Surgical volume per year: treatment provided by diagnosis (n=393).

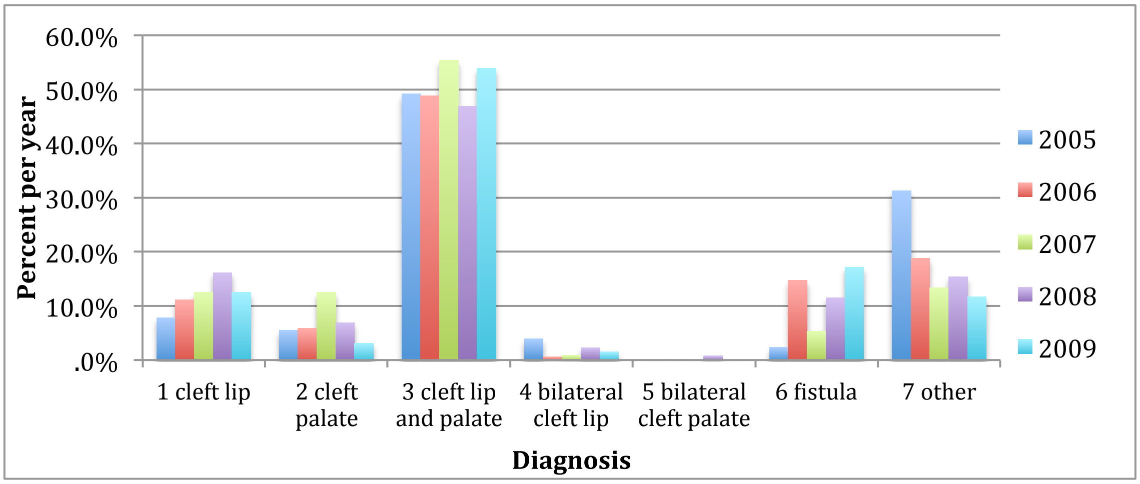


Figure 5.2. Annual pre-operative diagnostic categories in percent for the years 2005-2009 (n=541).

Figure 5.2 depicts the annual pre-operative diagnostic categories in percent, for patients seeking services for the years 2005 through 2009. A Chi-square test for independence (alpha .05) indicated a statistically significant association between diagnosis and year (Fig. 5.6), Chi-square (24, n=541) =60.41, p=.000. The number of patients presenting for surgery for lip revision and fistula repair increased from 2005 to 2009. Conversely, patients needing non-cleft related surgeries decreased from 2005 to 2009. However cell sizes violated the assumption of the test. Ten cells had a count less than 5 (29%). Diagnostic variables were collapsed into 3 groups with cleft related diagnosis; cleft lip, cleft palate or cleft lip and palate. Chi-square test for independence (alpha .05) was not significant for an association between diagnosis and year for orofacial clefts; Chi-square (8, n=401) =11.63, p=.169.

Figure 5.3 represents the frequency of diagnosis by gender. A Chi-square test for independence (alpha .05) indicated a statistically significant associated between gender and diagnosis; Chi-square (2,n=400)=8.09, p=.018. A post hoc comparison determined that children presenting with cleft lip and palate were more likely to be boys; Chi-square (1, n=130)= 8.09, p=.004.

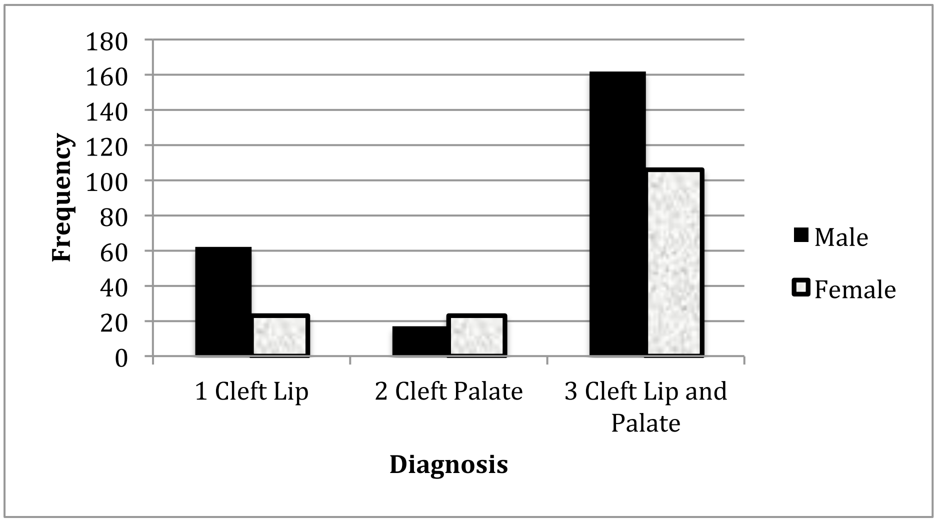


Figure 5.3. Histogram of diagnosis by gender, n=400

Neither age, weight, weight/age ratio, volume of procedures or diagnostic categories changed over the time course from 2005 thru 2009. The only statistically significant finding was that a higher proportion of boys than girls presented with a diagnosis of both cleft lip and palate.

Results to answer research Aim # 2:

Determine if the demographic and clinical profiles provide evidence for earlier treatment of primary orofacial clefts over time (2005-2009).

Table 5.0 displays the mean age in weeks for children (<780 weeks) with a primary diagnosis of an orofacial cleft, for the years 2005 through 2009. The average age for all years was 166.77(SD±185.68). The youngest mean age was found in 2008 of 146 weeks (SD±168.62), and the oldest 183.12(SD±194.28) in 2007. Therefore there was no evidence that children received earlier treatment for primary orofacial cleft over time. The ANOVA analysis comparing the mean ages in weeks per year for patients with primary cleft lip, cleft palate or cleft lip and palate were not statistically significantly different across the five years; [F(4, n=367)=.731, p=.572].

Table 5.0. Mean age in weeks for children <780 weeks.

Year	n	Mean age	S. D
2005	81	177.26	173.21
2006	86	148.57	168.93
2007	64	183.12	194.28
2008	68	146.00	168.62
2009	69	182.50	225.03
Total	368	166.77	185.68

Table 5.1 shows the mean age in weeks for all patients with primary cleft related diagnosis. The ANOVA analysis comparing the mean age in weeks for patients by diagnostic group was not significant; [F (2,398) =1.11, p=.330].

Table 5.1 mean age in weeks for primary cleft related diagnosis

Independent variable	n	Mean	Std. Deviation
Cleft Lip	90	267.23	394.77
Cleft Palate	40	258.28	293.05
Cleft Lip and Palate	269	217.22	259.40
Total	399	232.61	298.45

Table 5.2 shows the mean age in weeks based on primary diagnosis for patients <780 weeks (15 years). The ANOVA analysis comparing the mean age in weeks for

patients < 780 weeks old by diagnostic group was not significantly different by diagnostic category;[F(2,367)=.242, p=.785].

Table 5.2: Mean age in weeks for sample <780 weeks (15 years) per diagnosis.

Independent variables	n	Mean	S D
Cleft Lip	80	155.95	206.67
Cleft Palate	36	180.36	177.49
Cleft Lip and Palate	252	168.30	180.34
Total	368	166.77	185.68

The box plot depicted in figure 5.4 shows the mean and age ranges for patients with a primary diagnosis of orofacial cleft for the years 2005-2009 and demonstrates that there were no changes in age profiles across the years.

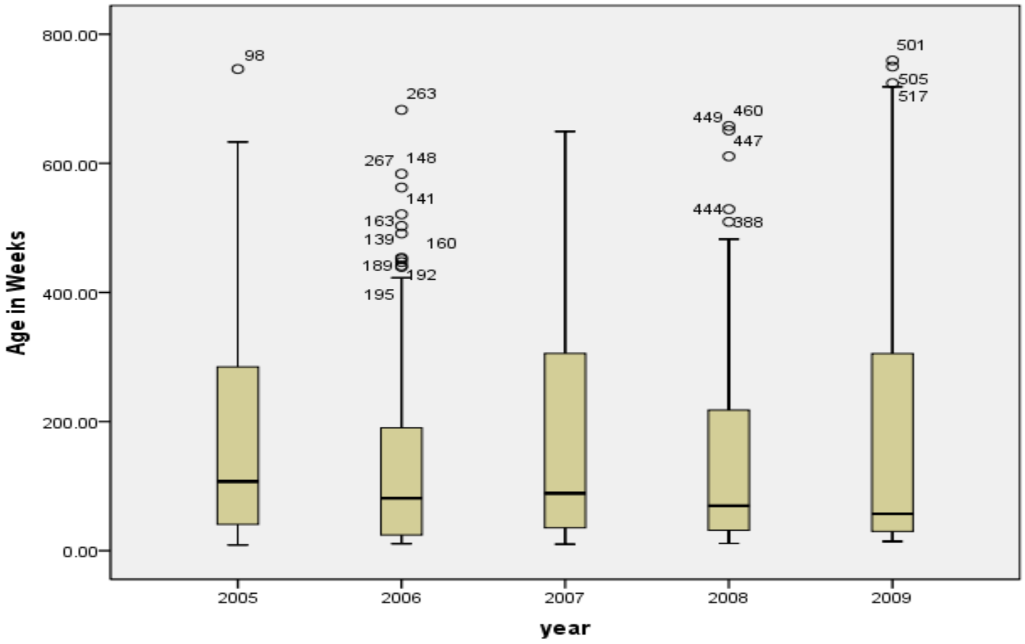


Figure 5.4. Box plot depicting the age ranges for patients with a primary diagnosis of cleft lip, cleft palate and cleft lip and palate for the years 2005 through 2009.

Results to answer research Aim # 3:

Describe clinical and quality indicators (complication rates, procedure times), and benchmark with American College of Surgeons National Quality Improvement Program Pediatric Phase 1 standards.

The percentage of surgical complications per year is depicted in figure 5.5. A comparison of surgical complications across the years shows a statistically significantly higher proportion of complications occurred in 2006 compared to 2009 [Chi-square (1, n=146) =9.708, p=.002]. The Fisher's exact test result was also significant, p=.002.

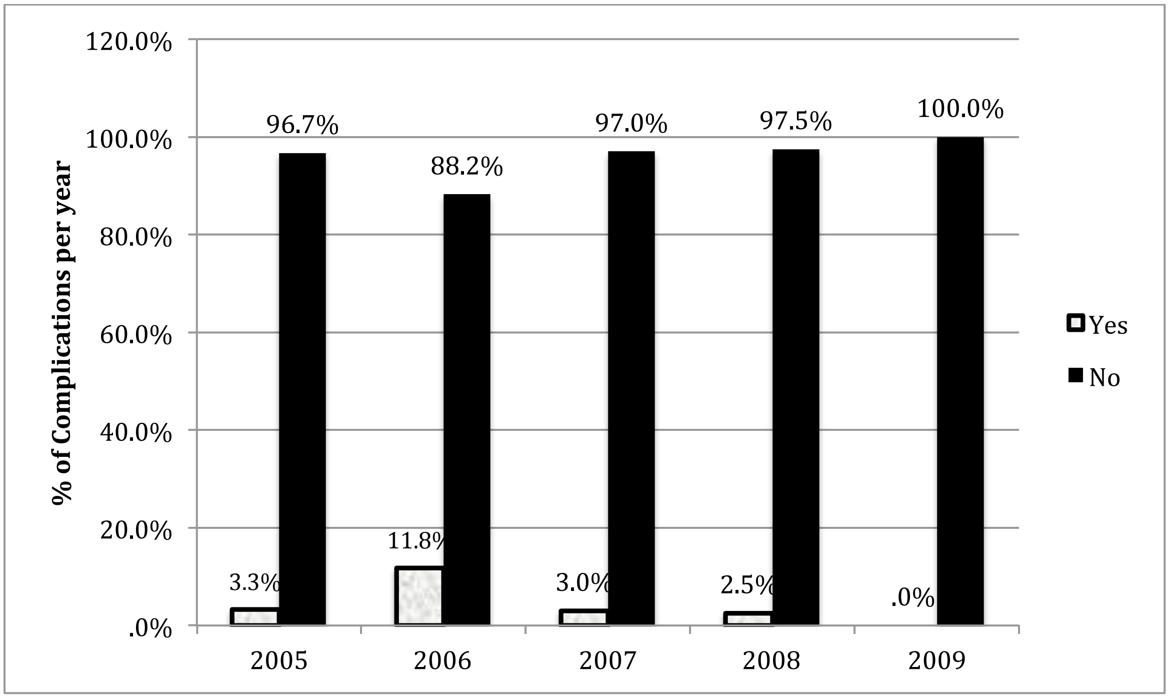


Figure 5.5. Histogram of surgical complications per year (n=354).

For children with surgical complications in 2006 that were under 5 years of age, n=6, their age-in-months and weight in kilograms were plotted against WHO international growth standards for weight-for-age (Fig.5.6 and 5.7). In 2006 there was statistically significantly higher frequency of complications on post-hoc analysis compared to 2009. Among boys (n=3), two had documented wound dehiscence and one had an anterior fistula. Among girls (n=3) two had documented wound dehiscence and one had an anterior fistula. The two patients over 5 years of age were compared to WHO growth charts, which revealed that the boy had a -2Z score (less than 3rd percentile), and the girl a -1Z score (below the 15th percentile) . Both of the children over 5 years of age had post-operative hemorrhages. All of the children (n=8) had diagnosis of both cleft lip and palate, with 4 of the cases requiring a bilateral cleft lip and palate repair.

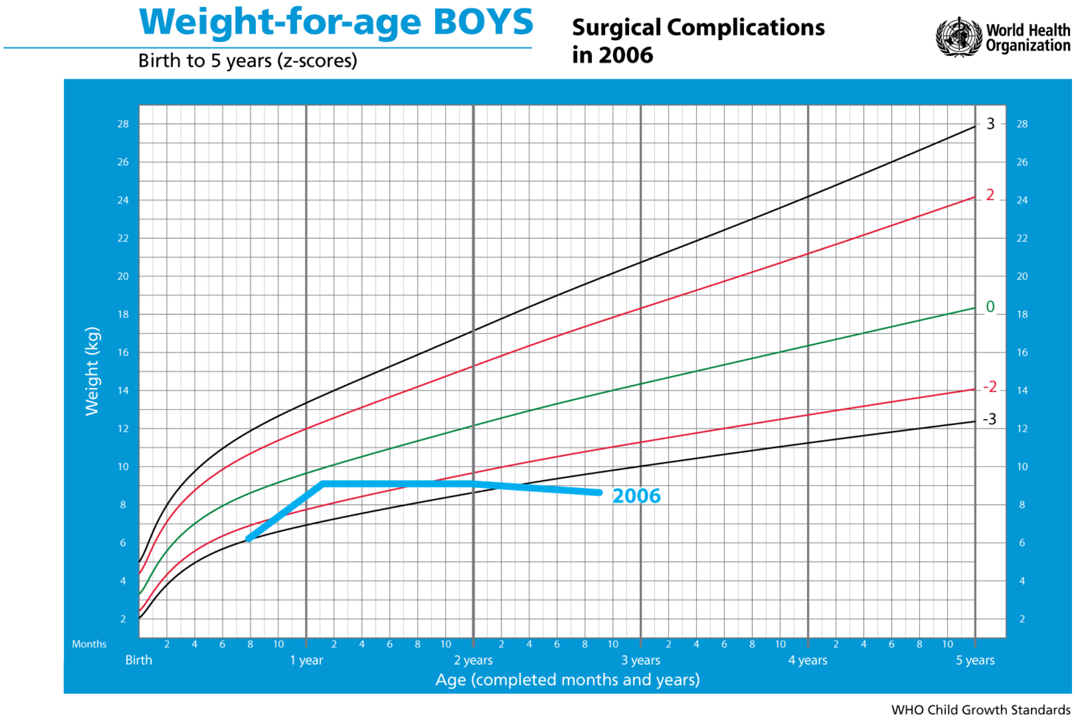


Figure 5.6. WHO weight-for-age growth reference for boys under 5 years of age with surgical complications in 2006.

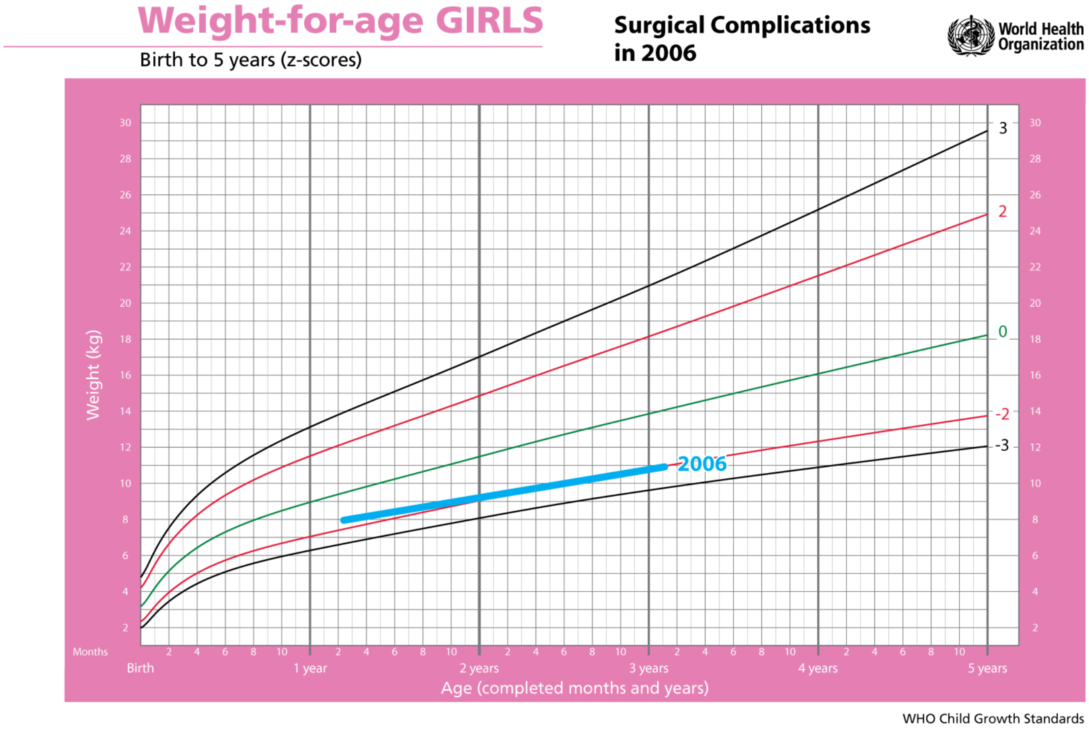


Figure 5.7. WHO weight-for-age growth reference for girls under 5 years with surgical complications in 2006.

Figure 5.8 displays the percentage of surgical complication per surgeon in 2006, as a description of provider quality outputs. Small cell sizes precluded an analysis between surgeon and surgical complications.

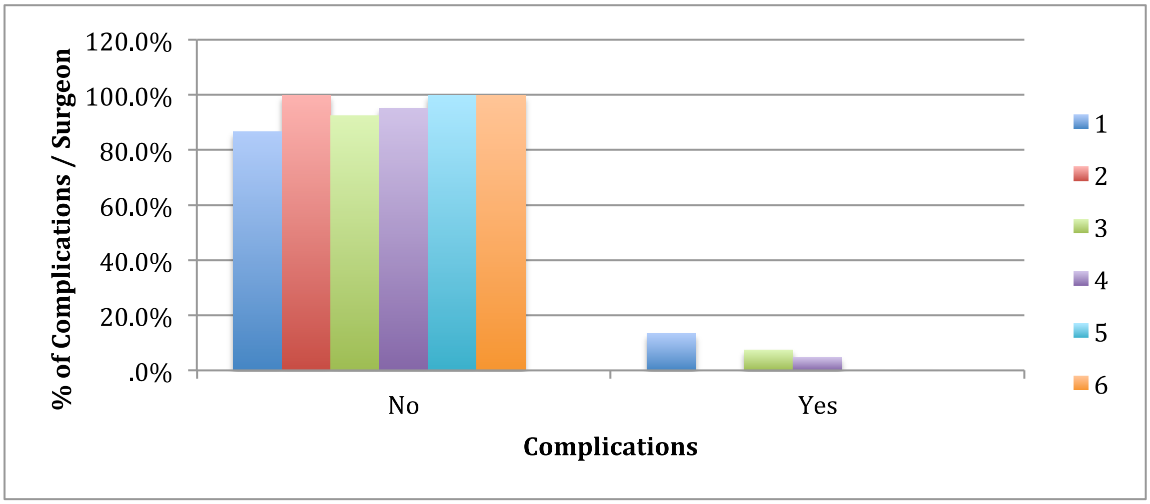


Figure 5.8. Percentage of surgical complications per surgeon in 2006 (n=8).

Table 5.3 shows the frequency of anesthetic complications for the years 2005 through 2009. There was not a statistically significant association between the frequency of anesthetic complications and year.

Table 5.3. Anesthetic complications 2005-2009.

Variable	n	%
None	348	91.1
Respiratory complication	33	8.6
Inadvertent extubation	1	0.3
Total	382	100

For children with reported anesthetic complications that were under 5 years of age, n=31, their age-in-months and weight in kilograms were plotted against WHO international growth standards for weight-for-age. The three other documented anesthetic complications were all respiratory. Two of the patients were over 10 years of age, and could not be referenced with WHO standards (one female with cleft lip, one male with cleft lip and palate). The third patient was a male with a cleft palate; whose weight for age Z score was -2 when plotted again WHO international growth standards.

In this sample (n=370) two statistically predictors of having an anesthetic compilation were found: namely, illness/co-morbidities and wt/age-in-month Z scores. Patients with an upper respiratory infection were 3.8 times more likely to have an anesthetic complication compared to patients without an illness or co-morbidity (CI: 1.50, 9.7;p=0.005). Those with wt/age-in-months Z score above the

.50 cut value were 1.5 times less likely to have a documented anesthetic complication.

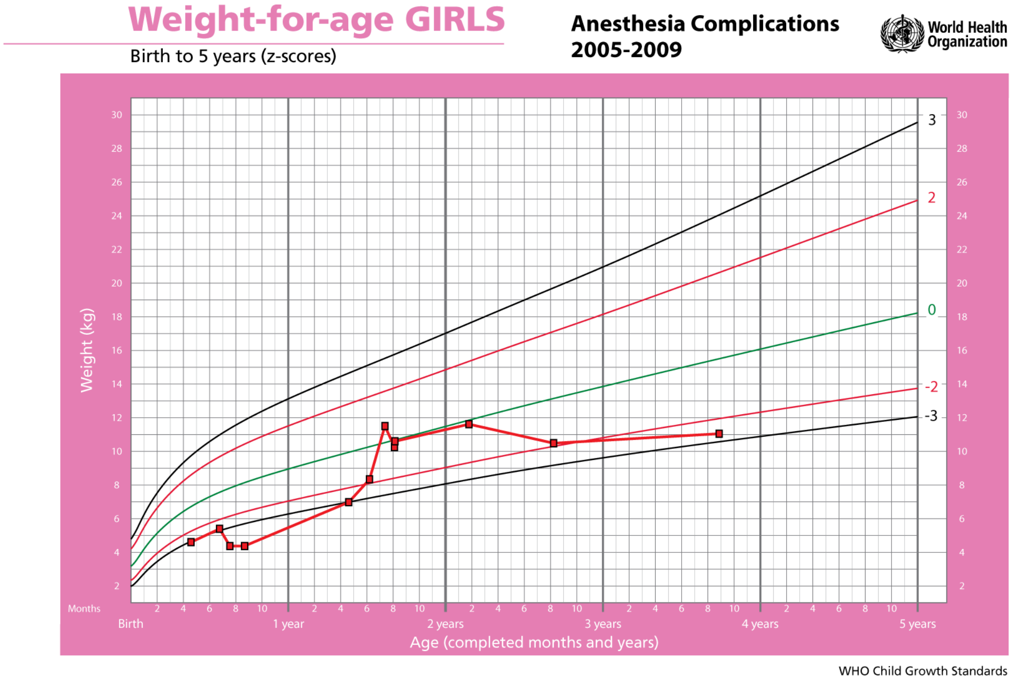


Figure 5.9. Girls under 5 years with anesthesia complications (2005-2009) weight for-age Z scores.

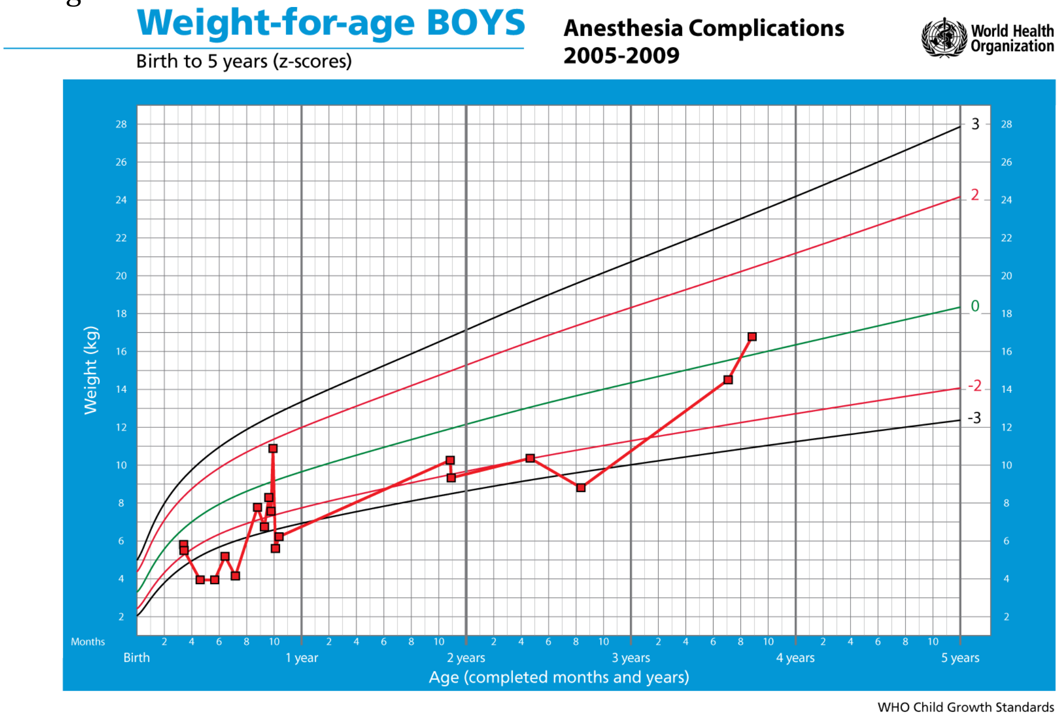


Figure 5.10. Boys under 5 years with anesthesia complications (2005-2009) weight-for-age Z scores.

Table 5.4. Logistic regression of two predictors of anesthesia complications.

Variable	OR	95% C.I.		p	n
		Lower	Upper		
None	1			.044	313
Upper Respiratory Infections	3.80	1.50	9.70	.005	30
Growth & development	.000	.00	.	.999	4
Other	1.86	.51	6.80	.346	23
Weight age ratio z-score	1.47	1.05	2.05	.025	370
Constant	.08			.000	

American Society of Anesthesiologists (ASA) physical status classification system was developed and is used for pre-operative assessments. A person with an ASA of 1 is classified as a normal healthy individual. A person with an ASA of 2 is classified as having mild systemic disease. For this sample ASA classification, when documented, ranged from 1 to 2, and was a significant predictor of anesthetic complications(see table 5.5). Compared to patients with an ASA of 1, those with ASA of 2 were 3.25 times more likely to have an anesthetic complication (OR 3.25: CI 1.4, 7.5, p=.006).

Table 5.5. Logistic regression of ASA as a predictor of anesthetic complications.

n=330	df	OR	95% C.I.		p
			Lower	Upper	
ASA	1	3.25	1.41	7.47	.006
Constant	1	.073			.000

Three predictors of peri-operative complications, including anesthetic and surgical complications, were analyzed (see tables 5.6 and 5.7). Combining surgical and anesthetic complications re-confirmed illness and wt/age-in-month Z scores as statistically significant predictors of peri-operative complications. For patients with an upper respiratory infection, their odds of having a complication was increased

2.65 times compared to patients without illness/co morbidities (OR 2.64, CI 1.0, 6.56, p=.04). Those with wt/age- in- month Z scores above the .50 cut value were 1.5 times less likely to have a documented peri-operative complication (OR 1.45, CI 1.0,2.0, p=.04).

Diagnosis emerged as an additional statistically significant predictor when both surgical and anesthetic complications were combined. Patients with a diagnosis of cleft palate were 6.47 times more likely to have a peri-operative complication compared to patients with cleft lip (CI 1.7, 24.3, p=.01). Patients with both cleft lip and palate were 3 times more likely to have a complication (CI 1.08, 8.38, p=.04), than patients with cleft lip only.

Table 5 6. Frequencies of diagnostic and illness variables

Independent variables		**%**
Diag5grp	Cleft lip	70
	Cleft palate	28
	Cleft lip and palate	207
	Fistula	46
	Other	20
Illness3grp	None	314
	Upper respiratory infection	30
	Growth and development	4
Wt/age-in-month z score	Other	23 371

Table 5.8 (below) depicts FOH complication percentages benchmarked to the American College of Surgeons National Surgical Quality Improvement Program Pediatric (NSQIPP): phase 1(Raval, et al., 2011.)

Table 5.7. Logistic regression of three predictors of peri-operative complications (surgical and anesthetic).

Independent variables	OR	95% C.I.		p
		Lower	Upper	
Diag5grp				.086
Diag5grp(1)	6.47	1.71	24.34	.006
Diag5grp(2)	3.00	1.08	8.38	.036
Diag5grp(3)	2.22	.51	9.67	.289
Diag5grp(4)	.000	.00	.	.998
Wtagemonz	1.45	1.01	2.10	.044
Illness3grp				.222
Illness3grp(1)	2.64	1.06	6.56	.036
Illness3grp(2)	.000	.00	.	.999
Illness3grp(3)	1.06	.29	3.88	.934
Constant	.05			.000

Table 5.8. Percentage of complications National Surgical Quality Improvement Program Pediatric vs. Faces of Hope.

	n	**% no**	**% yes**
NSQIPP*			
Plastic surgery	406	96.6	3.4
Congenital anomaly	2,734	94.3	5.7
Nutritional/immune hx	758	86.4	13.6
FOH+	371	88	12.3

*National Surgical Quality Improvement Program Pediatric=NSQIPP: +Faces of Hope=FOH

Table 5.9 provides the average time of surgical procedure completion by surgeon in hours ranging from 0.83 to 2.03 hours. ANOVA analysis comparing the mean surgical time between providers was statistically significant; F(7,n=292)=22.68, p=.00. Post hoc analysis was undertaken to determine which procedures accounted for the variation in means. Results indicated that a statistically significant difference was observed in operating time to complete a primary repair of the lip and palate together; [F (7, n=157) =17.06, p=.000].

Table 5.9. Output: Mean Time of Surgery by Surgeon

Surgeon	n	Mean	SD
1	62	1.60	.67
2	53	2.03	.94
3	16	1.41	.58
4	78	.83	.39
5	27	.87	.48
6	33	.91	.53
7	20	1.06	.65
8	4	1.70	.28
Total	293	1.28	.77

Regarding research aim #3 procedure times showed statistically significant variability by surgeons completing a primary closure of both cleft lip and palate. The proportion of surgical complications were statistically significantly higher in 2006 (11.8%) compared to 2009. FOH surgical complication rates are similar to those put forth by NSQIPP pediatric standards.

Results to answer research Aim #4:

Determine the independent predictors (out of RA 1-2) of eligibility for receiving surgical treatment.

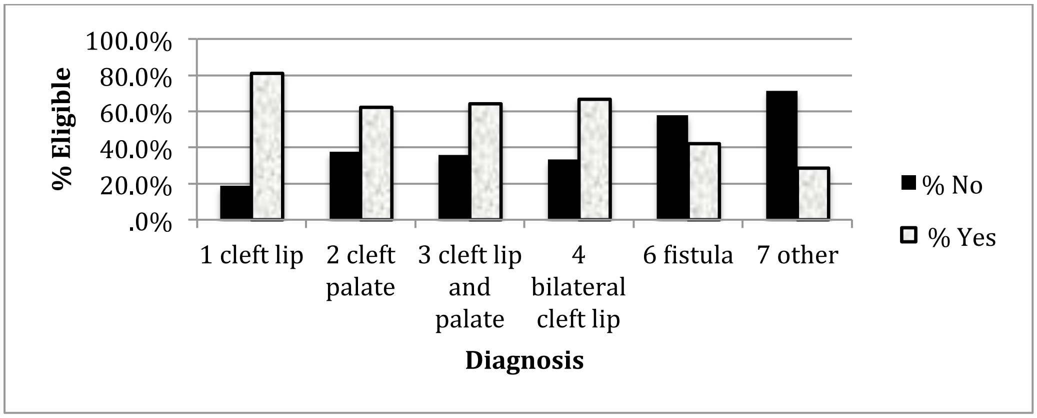


Figure 5.11. Percentage eligible by diagnosis2005-2009(n=531).

In this sample (n=488) previous surgery, illness and diagnosis are statistically significant predictors of eligibility (receiving surgery). Analysis (see tables 5.10 and 5.111) revealed that patients who had received previous surgery for cleft lip and palate (group 3) were less likely to be eligible, compared to those who had not received previous surgery; OR .17(.08,.35), p=.00. This suggests that for patients who had previous surgery for cleft lip and palate (group 3) the odds of not being eligible to receive surgical services increased 5.8 times. This estimate could be as low as 12 times less likely to be eligible, or as high as 2.8 times less likely.

Growth and development also predicted eligibility in this sample. Patients seeking surgical services that were diagnosed with growth and development problems were 20 times less likely to be eligible for surgery; OR 0.05(0.01, 0.25), p=00. This estimate could be as low as 100 times less likely and as high as 4 times. For the group with other diagnosis of illness or co-morbidity (gastro-intestinal, scabies, syndromes, cardio vascular), the odds of ineligibility were increased 2.5 times compared to patients without illness or co-morbidity; OR .397 (.174-.905), p=.028. This estimate could be as low as 5.7 times less likely or as high as 1.1 times.

For patients who had an illness or co-morbidity, their odds of being eligible to receive surgical service was decreased compared to patients without illness or co-morbidity. For those who had an upper respiratory illness the odds of not receiving surgery (ineligibility) increased ten times; OR .10(.05,.20, p=.00). This estimate could be as low as 18 times less likely, or as high as 5 times.

Pre-operative diagnosis was a statistically significant predictor of eligibility for patients with orofacial clefts. Compared to patients with non-cleft related diagnosis (in the diagnosis group other), patients with cleft lip were 11.6 times more likely to receive surgery; OR 11.6(4.5, 30.1), p=.00. Individuals with a diagnosis of cleft palate were 4 times more likely than those with non cleft diagnosis to receive surgery; OR 4.1(1.5, 11), p=.05. Patients diagnosed with cleft lip and palate were 6.3 times more likely to receive surgery than those with non cleft diagnosis; OR 6.3 (3.1, 13), p=.000, with a low estimate of 3 times more likely and a high of 13 times more likely. For those with fistulas, their odds for receiving surgery were increased 4.7 times compared to patients without cleft related diagnosis; OR 4.7(1.9-11.1), p=.001. Hemoglobin (Hgb) measured in grams per 100 ml (n=100) was not a significant predictor of eligibility (see table 5.10).

Table 5.10. Logistic Regression to determine whether Hgb is a factor in eligibility.

Independent variable		**df**	**O. R.**	**95% C.I.**		**P**
				Lower	***Upper***	
	Hgb	1	1.25	0.93	1.66	0.136
	Constant	1	.78			0.883

Table 5.11 Frequencies of Independent variables (n=488)

Independent variables		n	%
Previous surgery	None	208	14
	Cleft lip repair	129	9
	Cleft palate repair	17	1
	Cleft lip and palate repair	105	7
	Other	29	2
Diagnosis	Cleft lip	70	5
	Cleft palate	35	2
	Cleft lip and palate	251	17
	Fistula	73	5
	Other	59	4
illness	None	377	26
	Upper respiratory infection	67	5
	Growth and development	13	1
	Other	31	2
Temperature ˚C		488	100
Pulse		488	100
Wt/age ratio in weeks		488	100
Oxygen Saturation %		488	100

Eligibility (receiving surgical treatment) was predicted by previous surgery, lack of illness and diagnosis. Pre-operative vital signs and blood hemoglobin level did not predict eligibility.

Table 5.11. Logistic Regression of seven predictors of eligibility to receive surgical treatment (n=488).

Independent variables	B	df	O. R.	95% C.I.		P
				Lower	Upper	
Temp	-.226	1	0.80	.55	1.16	.236
Pulse	-.007	1	0.99	.99	1.00	.101
Wtage	-.575	1	0.56	.07	4.83	.600
Diag5grp		4				.000
Diag5grp(1)	2.451	1	11.60	4.47	30.08	.000
Diag5grp(2)	1.414	1	4.11	1.54	10.98	.005
Diag5grp(3)	1.839	1	6.29	3.02	13.10	.000
Diag5grp(4)	1.533	1	4.63	1.94	11.05	.001
Illness3grp		3				.000
Illness3grp(1)	-2.266	1	0.11	.05	.20	.000
Illness3grp(2)	-3.007	1	0.05	.01	.25	.000
Illness3grp(3)	-.925	1	0.40	.17	.91	.028
Prevsurg6grp		4				.000
Prevsurg6grp(1)	-.293	1	0.75	.40	1.38	.352
Prevsurg6grp(2)	-1.158	1	0.35	.09	1.04	.059
Prevsurg6grp(3)	-1.765	1	0.17	.08	.35	.000
Prevsurg6grp(4)	-.900	1	0.41	.15	1.09	.074
Oxygen	-.028	1	0.97	.89	1.06	.550
Constant	25.632	1	135492576241.33			.180

Results to answer research Aim # 5:

For a subset of subjects who have data available, determine the predictors (out of RA 1-3) of receiving additional procedures, in subsequent years, from Faces of Hope volunteers.

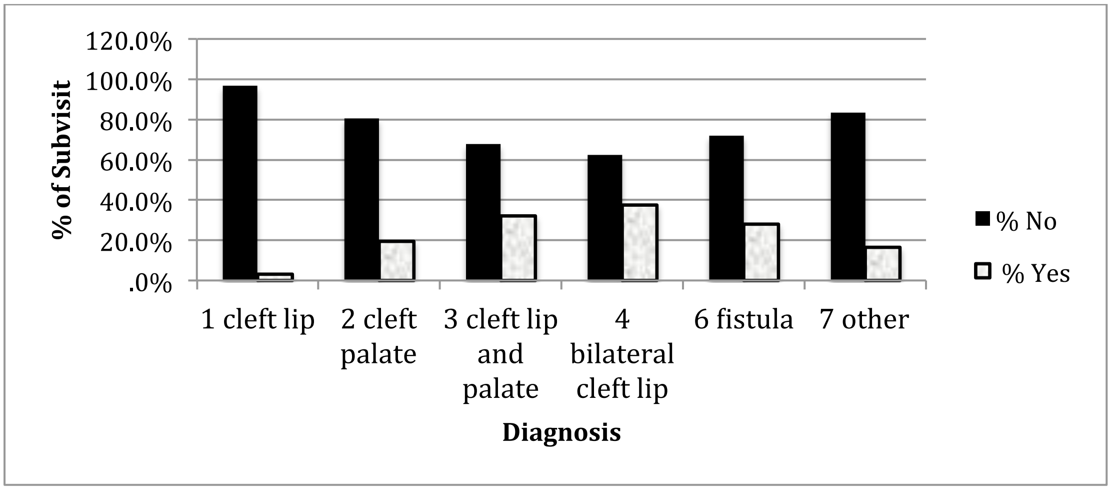


Figure 5.12. Percent of subsequent visit by diagnosis (n=426).

Table 5.12. Logistic regression of 2 predictors of subsequent visits (n=426).

Independent variables	df	O. R.	95% C.I		p
			Lower	Upper	
Diag6grp	5				.000
Cleft palate	1	9.86	1.85	52.38	.007
Cleft lip & palate	1	20.25	4.59	89.55	.000
Bilateral cleft lip	1	21.00	2.70	162.79	.004
Fistula	1	21.27	3.76	120.86	.001
Other	1	8.61	1.76	42.12	.008
Wt/age-in-month z score	1	1.47	1.14	1.90	.003
Constant	1	.022			.000

In this sample (n=426) diagnosis and wt/age-in-months z scores were statistically significant predictors of subsequent visits (patients who returned to FOH in subsequent years). Compared to patients with cleft lip, patients with cleft palate were 9.86 times more likely to return to FOH in a subsequent year. This estimate could be as low as 1.85 times more likely, or as high as 52.3 times more likely to return. Patients with both cleft lip and palate were 20.25 more likely to return for a subsequent visit when compared to those with cleft lip alone (CI 4.59, 89.55, p=00). Patients with a diagnosis of bilateral cleft lip were 21 times more likely to return in subsequent years than patients diagnosis with a unilateral cleft lip (CI 2.70, 162.79, p=.004). For patients with a post-operative fistula, their likelihood of returning was increased by 21.27 times compared to patients with cleft lip (CI 3.76, 120.86, p=.00). Patients with non cleft related diagnosis (other), have an 8.61 times increased likelihood of subsequent visits compared to patients with cleft lip (CI 1.76, 42.12, p=.008). Weight/age-in-month Z scores were a significant predictor of subsequent visits. For patients with weight/age-in-month Z scores above the 50% cut of, their likelihood of returning for a subsequent visit was increased 1.47 times (CI 1.14, 1.90, p=.00). For children who returned for subsequent visits their age-in-months and weight in kilograms were plotted against WHO international growth standards for weight-for-age (see Figures 5.14 & 5.15).

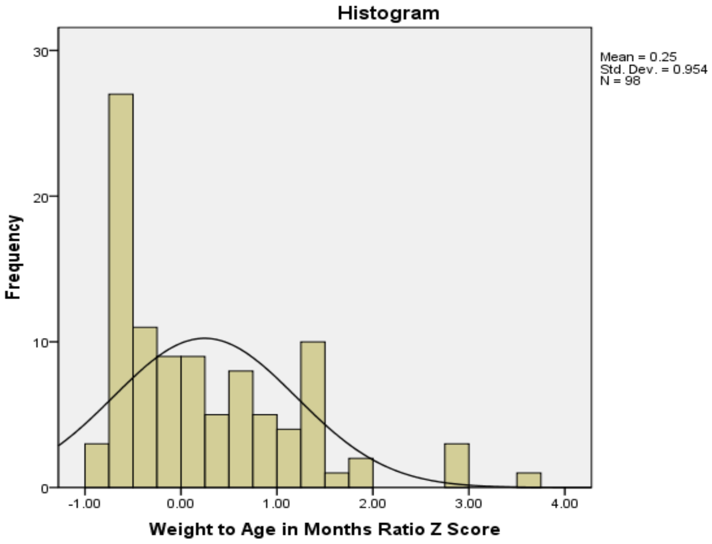


Figure 5.13 Histogram of wt/age-in-month z scores for patients that returned for a subsequent visit (n=98).

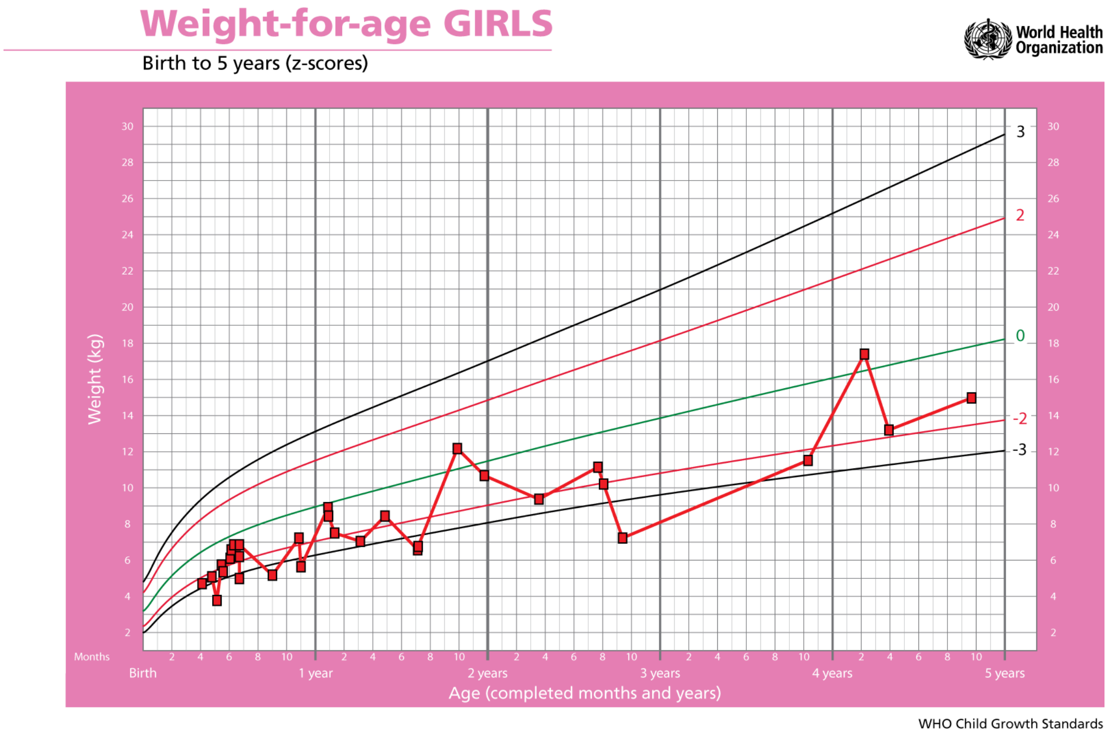


Figure 5.14. Weight for age profiles for girls under age 5 that returned for a subsequent visit.

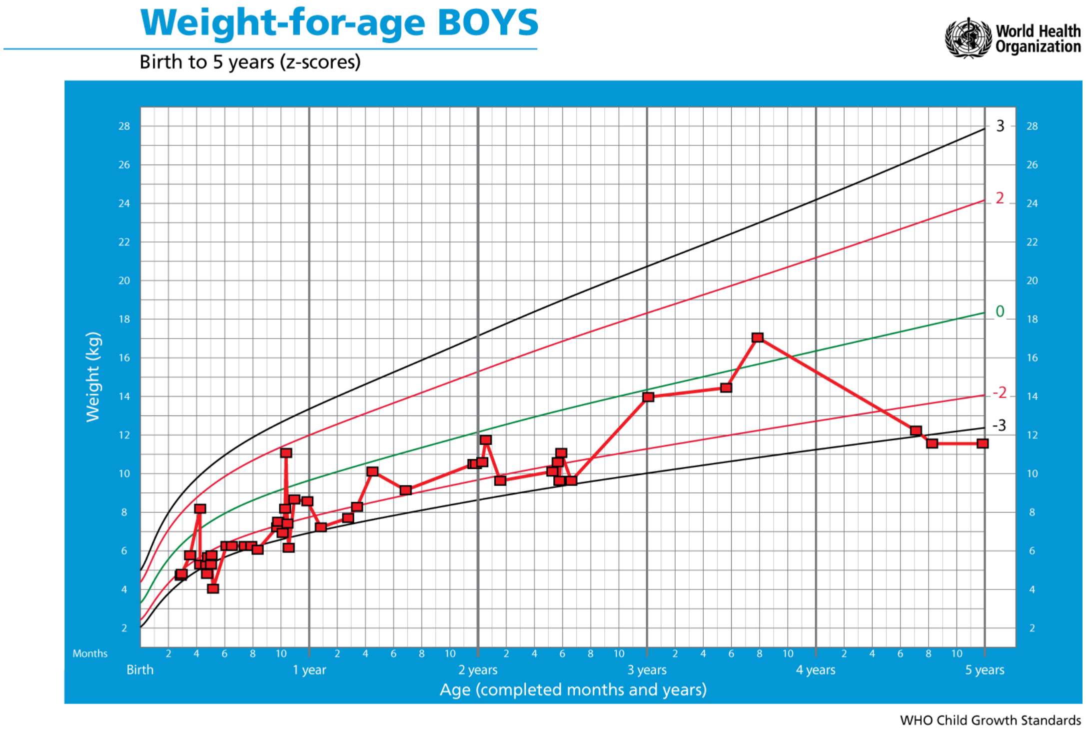


Figure 5.15. Weight for age profiles for boys under age 5 that returned for a subsequent visit.

Table 5.13. Logistic regression of future needs as a predictor of subsequent visits.

Independent variables	df	O.R.	95% C.I.		p
			Lower	Upper	
Future4grps	3				.005
Future4grps(1)	1	11.85	1.55	90.78	.017
Future4grps(2)	1	9.60	1.20	77.05	.033
Future4grps(3)	1	3.81	0.47	31.38	.214
Constant	1	.04			.002

In this sample, future needs documented by FOH physicians in the medical record were significant predictors of subsequent visits. Patients whose records stated that future surgical needs (1) were indicated had an 11.85 times higher odds of returning in subsequent years compared to patients that had no future needs documented in their record (CI1.55,90.78, p=.02). If follow up (2) was a documented need on a patient's record, patients were 9.6 times more likely to return compared to those with no future needs documented (CI 1.20,77.05, p=.03). In the sample, the charts were future surgical needs were documented, 59.7% of the patients returned for a subsequent visit. For physicians that indicated follow up was needed, 28.6% of those patients returned. Records that did not specify future needs had 13.7% of the patients return for a subsequent visit.

Summary

Comparing demographic and clinical data including the average age-in-weeks, weight-in-kilograms and wt/age ratio over the years did not demonstrate a statistically significant difference. The frequency of surgical treatment by diagnosis, when compared per year, did not reveal an association. Annual percentages of pre-operative diagnosis were not found to have a statistical association with year. The medical record data did not provide evidence for earlier treatment across the time course. The ANOVA analysis of average age in weeks, compared by diagnosis and year, was not statistically significant. Gender differences were demonstrated to exist between the frequency of a pre-operative diagnosis of both cleft lip and palate.

Surgical and anesthetic complications compared by years suggest that a statistically significant difference exists between surgical complication in 2006 and 2009, with the rates of surgical complications declining in 2009. Anesthetic complications did not have an association by year. Diagnosis, illness and wt/age- in-month Z scores were statistically significant predictors of complications.

Eligibility (receiving surgical treatment) was predicted by previous surgery, lack of illness and diagnosis. Pre-operative vital signs and blood hemoglobin level did not predict eligibility. Returning in subsequent years to seek additional services was predicted by diagnosis, wt/age-in-month Z scores and documented futures needs.

Important observations and findings based on these results will be addressed in the next chapter, with a focus on the implications for clinical practice, future research, education and policy. The limitations of the results will be discussed in relation to study findings and lessons learned.

Chapter 6: Discussion

Organizing Framework for *Faces of Hope* Evaluation

The CDC Framework (described in chapter 3) proved a useful tool in guiding the development of the organizing framework for *Faces of Hope* evaluation (Figure 3.1). The organizing framework was used to focus evaluation of FOH and structure the development of research aims. Steps three and four of the CDC framework are to: focus evaluation design (step three) and to gather credible evidence (step 4). Developing the organizing framework for *Faces of Hope* (OFE) evaluation achieved step three; evaluation focus. Gathering credible evidence was accomplished by medical record abstractions methodology in which cross sectional data from FOH clinical records were used to measure and describe relationships within the structure of the OFE. The OFE (figure 3.1) was very beneficial in situating the research aims within the concepts of access, output and outcomes.

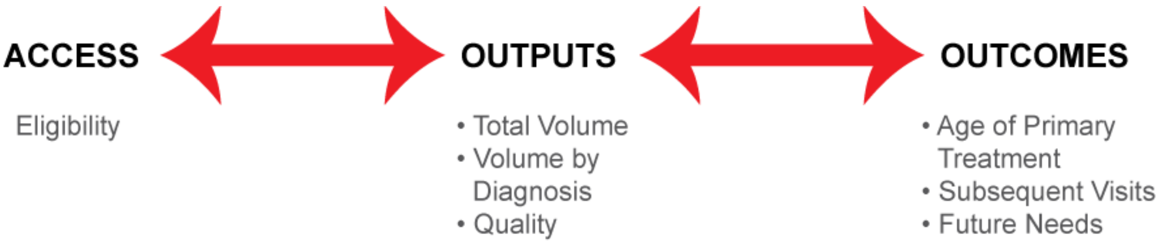


Figure 3.1 Organizing Framework for *Faces of Hope* Evaluation (OFE)

The CDC Framework recognizes that programs mature over time. Three stages of development are defined within the CDC Framework and include planning, implementation, and effects (CDC, 1999). FOH has been in operation since 2005 and is in the implementation stage of development. The goals of evaluation activities at this stage include the characterization of the actual activities (outputs and outcomes) of the program. These activities are then compared with the programs aims and missions (access).

Access

The CDC Framework describes the use of logic models to map the sequence of events that are designed to bring about change (CDC, 1999). This process was of great benefit. By employing logic modeling techniques the relationship between patients accessing services from FOH and eligibility to receive those services was diagramed (Chapter 3, Figure 3.0). Using logic modeling enabled the concept of access to be operationalized and measured as eligibility: a patient was eligible if they received surgical services and ineligible if they did not (see Figure 3.1). Evaluating the predictors of eligibility provided information to design and implement procedures to maximize it.

Eligibility

Anthropometric and clinical data from FOH records were placed in a logistic regression model to estimate predictors of eligibility. Previous surgery, illness and diagnosis were statistically significant predictors. For example, the odds of being ineligible to receive surgical services were increased 5.8 times for patients who had received previous surgery for cleft lip and palate. Conversely, the likelihood of eligibility increased depending on the pre-operative diagnosis. Compared to patients with non-cleft related diagnosis, patients with cleft lip were 11.6 times more likely to be eligible, patients with a diagnosis of cleft palate were 4 times more likely, and those with cleft lip and palate were 6.3 times more likely to receive surgery.

Sharing this data with organizations partnering with FOH to provide services, such as Partners for Surgery, whose workers screen and refer patients, is essential. Sharing lessons learned is fundamental to the CDC Framework, and likewise must underpin FOH evaluation efforts. In this example, referral efforts directed towards patients that have not received previous surgery for cleft lip and palate may increase their likelihood of being eligible. Likewise, screening efforts focused on patients with cleft related diagnosis may enhance eligibility based on these findings. Offering families information regarding their likelihood of receiving services based on predictors of eligibility will allow them to make an informed decision before they embark on a costly trip to Antigua, which may prove unnecessary.

For patients who had an upper respiratory illness (URI) their odds of ineligibility increased 10 times. For some of these patients the URI occurred after their arrival in Antigua. When told their child was ineligible because of it, families have angrily reported that their children were healthy when they arrived and only became sick while living in the family dormitories provided by FOH. At times parents identified specific children or families as the “culprit,” the one that made their child ill. This is not a trivial concern. Bringing together children and families from all regions of Guatemala and housing them in dormitory style accommodations may foster outbreaks of viral or bacterial infections, which then prohibit eligibility. Every effort must be made to implement public health strategies to minimize cross infection. It is crucial to teach families about the importance of hand washing and how to cover the mouth while coughing. Provisions to isolate sick children or parents are being considered by FOH administrators based on this data. This includes having more than one housing accommodation, or developing strategies to stagger the arrival of families to Antigua.

Growth and development also predicted eligibility. Patients seeking surgical services who were diagnosed with growth and development problems were 20 times less likely to be eligible for surgery. This represents a sad irony. Patients, whose growth and development may be impacted by having an orofacial cleft, are ineligible to receive surgery which may correct the problem that is inhibiting proper nutrition. By working with our partners in the field, children can be screened for growth and development concerns, and their families offered feeding support and education before the surgical team arrives. An example would be the provision of bottles with nipples designed to enhance intake and prevent food loss through the nose. Nutritional support can also be provided to nursing mothers. By implementing these or similar interventions, eligibility for children with growth and development concerns may be fostered.

The importance of partnerships within Guatemala is highlighted in this study. Working together to insure that assessment and referral policies are geared toward maximizing eligibility is essential. Without these relationships the ability to implement adaptive strategies to foster eligibility may be constrained.

Local partnerships were identified as key to the overall success of STSM (Eberlin, 2008; Zbar, Rai, & Dingman, 2000; Uetani, et al., 2005). Drawing from 20 years of experiences Eberlin, et al. (2008) recommended guidelines for quality assurance and bridge building to ensure continuity for surgical outreach programs. Three phases of Medical Missions for Children were identified in this study: the pre-operative phase, the peri-operative phase and post-operative care. At each phase interface with the local community was emphasized.

The philosophical model developed by Zbar, Rai, & Dingman (2000) also includes three phases: observation, integration and independence, and local partnerships are identified as crucial for each. Phase III, the final phase, is independence. This occurs when the host providers maintain the site during absence of visiting teams and offer ongoing access to clinical services. Site maintenance by local providers was likewise identified as an important factor influencing patient access to surgical services in the study by Uetani, et al. (2005). In this study financial barriers (fees charged) and the perception that the local providers lacked the training of visiting teams, were reported to be impediments to families accessing care from local providers. Access to services was increased in Ghana when financial barriers were eliminated and partnerships with STSM forged (Donkor, et al., 2007). All of the studies suggest that both access and quality are influenced by local partnerships.

For families traveling to Antigua seeking surgery, significant risks may have been undertaken and at considerable expenditure. Patients have reported being robbed while in transit. A van used to transport patients was reported stolen, and a female volunteer reported that she was assaulted returning home. When children are deemed ineligible, they are faced with the prospect of a costly and potentially dangerous return home.

Families with children that have increased odds of eligibility, those with cleft related diagnosis who are seeking primary repair, should be the focus of referral activities within the current organizational structure of FOH. Additionally, families with a decreased likelihood of receiving surgery need to be informed. If a child is ill, making the trip to Antigua may be unwarranted. Strategies to promote wellness before the journey can be offered to families. This includes exploring the possibility of isolating pre-operative candidates from family members who may be ill. Sharing information on the contributions of hand washing and covering a cough to prevent viral spread among families is also important. For those without easy access to water, waterless hand cleaning products can be provided.

Causation beliefs inform surgical experience and need to be considered to optimize access to care. Patients, caregivers and traditional providers define the socio-cultural dimensions relating to access and utilization of cleft care services. Studies in the published literature regarding causation beliefs held in developing countries concerning oral cleft etiology fall into four primary categories: contagion, God's will, personal conduct, or the actions of ancestors (Dagher & Ross, 2004; El-

Shazly et al., 2010; Reeve et al, 2004; Weatherly-White et al., 2005). Investigating the interplay between personal or cultural beliefs and presentation for surgery will deepen understanding of the factors involved in accessing FOH services. Developing the concept of access beyond mere eligibility to receive services is an area for further study that may facilitate presentation for corrective surgery.

Outputs

Volume and diagnosis

Measuring surgical volume, the total volume and that of particular diagnoses, was a useful descriptor of the services provided. Volume data taken at a single data point described output for a particular year. For private organizations such as FOH, where a large percentage of annual funds come from individual or group donations, volume data per year provides donors with information regarding the utilization of their contributions. Measuring surgical volume over successive years allowed trending. For this study, no statistically significant association between the frequencies of surgical volume and year was detected by Chi square analysis (see Chapter 5, Figure 5.4). Though not statistically significant, volume trends have clinical significance for evaluating FOH activities. For example, the percentage of fistulas repaired in 2007 was 6.7% of total case volume and in 2009 fistulas accounted for 26.7%. Similarly, lip revision accounted for 15.4% of case volume in 2005 and 34.6% in 2009. This indicates a potential maturity of FOH output over time, as volume trends are increasing for surgeries that are not primary, but revisional or correctional. If volume data suggests that a change is occurring in output, the mission of the organization needs to be re-assessed, and potentially revised to reflect the change in services provided.

The annual percentages of the specific pre-operative diagnosis for patients seeking services for the years 2005 through 2009 was also measured (Chapter 5, Figure 5.5). A significant association between a particular diagnosis and year was not suggested by Chi-square analysis. However, yearly trends in the diagnosis of patients seeking services have relevance for program evaluation. For example, patients with non-cleft related diagnosis seeking surgery decreased from 31.3% in 2005, to 10% in 2009. This suggests that our partners in the field were informed about the mission of FOH, and were incorporating the knowledge into referral practices. This had a positive effect as it appears that the number of patients with non-cleft diagnosis traveling to Antigua for FOH services declined from 2005 to 2009.

The number patients seeking services with a diagnosis of fistula increased from 2.3% in 2007, to 14% in 2009. This data supports volume trends for fistula surgeries which were identified previously. If the frequency of patients with fistulas continues to increases, planning for the impact of a change in case mix for future missions is indicated. For instance, additional surgical instrumentation and suture material specific for fistula repairs may be required.

Planning for post-operative care is also informed, as fistulas often involve the palate and recovery can be more complicated due to the proximity of the surgery site to the patient's airway. Because of concerns with airway management post-operatively, fistulas are usually repaired early in the mission. The impact of this on the teams' ability to provide primary cleft repair must be assessed vis a' vis the mission of FOH.

These results demonstrate that changes in program activities occur and evaluation efforts should include analysis of program output. Output data analyzed each year allows continuous assessment of who the stakeholders of the mission are Step 1 of the evaluation framework developed by the CDC (1999) is to engage stakeholders. Since trends in the patients who seek and receive services change over the years, processes to monitor and identify stakeholders is necessary so they can be included in the ongoing process of evaluation.

Quality:

Complications were measured and benchmarked with the indices being developed by the American College of Surgeons National Surgical Quality Improvement Program Pediatrics: Phase 1 (Raval, et al., 2011). Taken at a single point the data had value in evaluating the activities of a given year. Analysis of data over time was used to identify trends. This triggered further analysis, including the complications associated with a particular year, individual provider and per diagnoses.

There was a statistically significant difference in the frequency of surgical complications between the years 2006 and 2009.Surgical cases performed in 2006 had an 11.8% complication rate compared to 2009 when there were no documented surgical complications. All of the children with complications in 2006 (n=8) were diagnosed with both cleft lip and palate, 4 of the cases were bilateral cleft lip and palate.

For the children with surgical complications in 2006 their weight in kilograms and age-in-months were compared to WHO international growth standards (Fig.5.9 and 5.10). The growth standards were developed by the WHO Multicenter Growth Reference Study (2003). In the WHO study primary growth data was collected and used to produce a curve that represents an international standard to best describe the physical growth for children from birth to five years of age. The comparison for boys (n=4) indicated that three had scores at or below -2z , and one a score of -1z. For the girls (n=4), two had scores of -2z and two -1z. Scores of -2z place the children below the 3rd percentile and -1z below the 15th percentile. Children with weight-for-age z scores < -2 are termed undernourished, measured as being underweight for their age. This reflects both acute and chronic malnutrition (WHO, 2003). For the children in the FOH sample that had complications in 2006, 5 out of the 8 (62.5%) are considered undernourished based on WHO growth standards.

The American College of Surgeons National Surgical Quality Improvement Program Pediatric (NSQIPP) Phase 1 targeted quality improvement in children's surgical care (Raval, et al., 2011). Benchmarking FOH surgical complications rates in 2006 with NSQIPP outcomes was undertaken for quality analysis. The overall complication rate reported by NSQIPP was 3.9%. Subcategories were created to measure rates associated with specific indicators, including type of surgery and preexisting conditions. Children in the NSQIPP study that had congenital malformations had a 5.7% complication rate and those with nutritional deficiency/immunological histories, 13.6%. Logistic regression revealed that patients in the study with nutritional deficiency/immunological history had a 1.49 times higher likelihood of post-operative complications (Raval et al., 2011). The 2006 complication rate for FOH was 11.8%, which is less than that reported by NSQIPP for children with nutritional deficiency/immunological history (13.6%). The average FOH complication rate for the years 2005-2009, including surgical and anesthetic complications, was 5.08% which falls under the 5.7% rate reported by NSQIPP for children with congenital malformations. Cell size limitations precluded proportional analysis of associations between providers and complications.

Measuring the frequency of anesthetic complications for the years 2005 through 2009 did not indicate a statistically significant association between the frequency of anesthetic complications and year. For children with reported anesthetic complications that were under 5 years of age, n=31, their age-in-months and weight in kilograms were plotted against WHO international growth standards for weight-for-age (Figures 5.12 and 5.13).For children in the sample that experienced anesthetic complications, six of the girls had WHO weight-for-age z scores<-2, indicating they were undernourished, and two had z scored at -2z. For the boys, 9 had weight-for-age z scores<-2z, 3 at -2z and 6 at -1z. This suggests that 64.5% of children experiencing anesthetic complications were undernourished as measured by their weight-for-age Z scores. This supports the statement in the report of outcomes for STSM by McQueen et al. (2009), that there is an increased risk of complications for nutritionally compromised children.

American Society of Anesthesiologists (ASA) physical status classification system was developed and is used for pre-operative assessments. A person with an ASA of 1 is classified as a normal healthy individual. A person with an ASA of 2 is classified as having mild systemic disease. For the FOH sample ASA classification, when documented, ranged from 1 to 2, and was a significant predictor of anesthetic complications. Compared to patients with an ASA of 1, those with ASA of 2 were 3.3 times more likely to have an anesthetic complication. Increasing ASA physical status classification rates were associated with higher post-operative complications rates documented by the NSQIPP (Raval, et al., 2011) and confirm the usefulness of the ASA classification system in identifying patients who are at a higher risk for developing anesthetic complications.

Evaluation research regarding the anesthetic practices of volunteer providers abroad was assessed and adverse anesthesia events, e.g. laryngospasm, upper airway obstruction, bronchospasm or arrhythmia, were included as measures of quality (Fischer, et al., 2001). The study was important in profiling anesthetic complications and how the age of the patient relates to them. For example, airway complication while the children were in the operating room were the most frequent challenges and children less than 5 years old were reported have a relative risk of airway complication 3.4 times that of children greater than 5 years. Inadvertent extubation was strongly correlated with younger age.

The anesthetic complications documented in FOH records were all related to the patient's airway. The mean age of FOH patients who had an anesthetic complication (n=33) was 109.11 weeks, SD 142.31. The mean age of FOH patients overall was 317.51, SD 426.72. The lower mean age-in-weeks for patients who experienced anesthetic complications supports the results reported by Fischer, et al.(2001), in which age was a factor related to anesthetic complications.

Three predictors of peri-operative complications for FOH patients, including anesthetic and surgical complications, were measured. For patients with an upper respiratory infection, their odds of having a complication were increased 2.65 times. Those with wt/age- in- month Z scores above the .50 cut value were 1.5 times less likely to have a documented peri-operative complication.

Diagnosis was a statistically significant predictor when both surgical and anesthetic complications were combined. Patients with a diagnosis of cleft palate were 6.47 times more likely to have a peri-operative complication compared to patients with cleft lip. Patients with both cleft lip and palate were 3 times more likely to have a complication. This data is aligned with the results of STSM surgical outcomes measured by McQueen, et al. (2009). In their research anesthetic and surgical complications occurred with significantly greater frequency for patients with cleft palate or those with cleft lip and palate, 43.8% and 21.9% of the total complications, respectively. NSQIPP data indicated that in the subcategory of plastic surgery, cleft lip/palate accounted for 42.9% of the total complications.

These outcomes suggest that children who are younger, have a diagnosis of both cleft lip and palate, are nutritionally compromised, and have illness/co morbidity are more likely to have complications. Children in these groups can be considered to have a higher risk of complications compared to children without these concerns.

This is important information for STSM to consider and plan for. In addition to applying public health procedures to prevent illnesses pre-operatively, organizations must consider the nutritional status of their patients. The rule-of-tens, often applied to screen pre-operative for STSM, may be insufficient. The rule of tens states that children who are 10 lbs, 10 weeks old and have hemoglobin measured of at least 10gm/dl can be considered as surgical candidates for repair of a cleft lip. These guidelines were suggested by Eberlin, et al. (2008) to be used as specific criteria to determine operative eligibility. If children who are malnourished have an increased likelihood of experiencing complications, screening for overall nutritional status instead of the rule of tens may be warranted. STSM teams must prepare to collect data specific to each child's nutritional status. This can be accomplished by plotting weight for age against WHO standards to identify children who are undernourished and therefore at a higher risk for complications. Weight, age and measured hemoglobin may be insufficient as data points alone in pre-operative screening, and stringent guidelines need to be developed to identify children who are nutritionally compromised. Collecting height data is an important additional measure of nutritional status, and will be incorporated into the medical records of FOH based on these results.

As agencies worldwide work to establish timing guidelines for repair of orofacial clefts, the nutritional effects of the malformation should inform policy. Eberlin, et al. (2008) proposes that cleft palate patients should be around 15 to 18 months old at the time of primary repair. These age guidelines are within the timeframe reported for primary cleft palate repair in developed countries (Bearn, et al., 2001; Shaw, et al., 2001 & Yang & Lia 2010).

Research measuring the outcomes of all-in-one closure of the lip, hard and soft palate at a mean age of 3 months was conducted and the results compared to children that had later staged closures of the lip and palate (DeMey, et al., 2006). The study authors reported that early complete closure of the cleft allowed for earlier intelligibility of speech. Further research regarding this approach, with the incorporation of nutritional status as an outcome measure, would do much to inform STSM models.

Yang & Lia (2010) completed a literature review on the effect of 1 stage versus 2 stage palate repair on facial growth for patients with cleft lip and palate. A wide variation in timing of hard palate repair was reported. Better speech was associated with earlier repair, though the authors reported that little data in the published literature was discovered specific to definitive age guidelines for closure versus optimal speech outcomes.

Studies with focus on better speech outcomes with earlier repair can be expanded to include measuring any nutritional benefits associated with it. All in one repair techniques may also prove advantageous in the STSM model, where difficulties associated with travel and expense for families seeking services are often encountered. The results of the FOH evaluation suggest that nutrition and speech may both be important determinants in establishing guidelines for the age of palate repair.

Outcomes

Without treatment individuals with orofacial clefts must live with their condition, enduring both functional and social challenges (Damiano et al, 2006: Noor & Musa, 2006: Turner et al., 1997). Delayed closure of cleft lip or palate can lead to poorer outcomes, e.g., impaired speech development, nutritional deficiencies with subsequent developmental delays, and stigmatization (Bermundez et al., 2010; Eberlin, et al., 2008; Shaw, et. al, 2001; Schwarz & Khadka, 2004). Teasing is a major concern for both children with orofacial clefts and their parents, and is thought to affect their self-confidence and social relations (Noor & Musa, 2006; Turner et al., 1997). Age guidelines vary in developed countries, though primary lip repair usually occurs at approximately three months and palate repair at 15-18months (Bearn, et al., 2001; Shaw, et al., 2001).

Medical record data obtained from FOH over its 5 year history was used to analyze age profiles over the time course. The results did not provide evidence for earlier treatment (see Chapter 5). These results are in contrast to those reported by Uetani, et al. (2005) where age profiles were decreased over the study period of 10 years. For example, the primary age of cleft lip repair was reduced from 14.0 years in 1993 to 1.3 years in 2001. The age range also narrowed (Uetani, et al., 2005).Given that data collection for FOH only spanned 5 years, continuing to collect age profiling data may ultimately provide evidence for earlier treatment.

Comparing the age findings of FOH clinical profiling to the average age of treatment in developing countries demonstrated that the average age of primary treatment was higher for the children in this sample. For example, in this study the average age of primary lip repair was 39 months, and the average of primary palate repair was 45 months; compared to children in developed countries where lip repair occurs at approximately 3 months, and palate repair at 18 months (Bearn, et al., 2001; Shaw, et al., 2001). This suggests that the children in this study were living with their condition for longer periods of time. Documented challenges exist for patients with untreated clefts (Damiano et al, 2006: Noor & Musa, 2006: Turner et al., 1997), and provision of surgical services is an important step in treatment for these children.

Lack of local providers and high cost of services were reported to inhibit access to surgical care for impoverished people in developing countries (Donkor et al., 2007; Moore & Fernandes, 2006, Schwarz & Khadka, 2004; Uetani et al., 2005). Delayed treatment due to the cost, paired with necessary travel and accommodations, influenced the age of primary repair of cleft lip and /or palate (Donkor et al., 2007; Moore & Fernandes, 2006; Schwarz & Khadka, 2004; Uetani et al., 2005). Necessary time away from work for family members to access distant services, and the time required for treatment and recovery, compound expenses and delays in seeking treatment.

Schwarz & Khadka (2004) described and categorized reasons for late presentation for cleft treatment in central Nepal. Children arriving for initial treatment who were over three months of age met the study definition of late presentation. The results underscored that many barriers to accessing care exist. The most common reasons cited by the authors were a lack of knowledge of availability of repair (31%), lack of accessibility of care (29%) and lack of finances (24%). A multidirectional approach to improving access was deemed necessary, and itinerant services were posited as a viable way in which obstacles to oral cleft care could be overcome (Schwarz & Khadka, 2004).

The children seeking surgical care from FOH meet the definition of late presentation presented by Schwarz & Khadka (2004), with the average age of a child arriving with cleft lip was 39 months and for cleft palate 45 months. Incorporating the results of this study with those of Schwarz & Khadka (2004) informs further FOH evaluation efforts directed at determining the availability of information about FOH services for families who may require them, and if lack of information was a factor in late presentation. The degree to which financial concerns delayed accessing FOH services is also necessary to plan for strategies to reduce financial barriers. An example would be to consider offering services in multiple sites throughout Guatemala, to diminish travel and time away from work.

Predictors of subsequent visits

Anthropometric and clinical data from FOH records were placed in a logistic regression model to estimate predictors of patients returning to Antigua in subsequent years seeking surgical services. In the sample (n=426) diagnosis and wt/age-in-months z scores were statistically significant predictors of subsequent visits. Compared to patients with cleft lip, patients with cleft palate were 9.86 times more likely to return in a subsequent year and patients with both cleft lip and palate were 20.25 more likely to return. Patients with a diagnosis of bilateral cleft lip were 21 times more likely to return in subsequent years, and post-operative fistulas predicted a likelihood of return that was increased 21 times.

These results underscore that treatment for orofacial clefts is an ongoing process, particularly for clefts which involve both the lip and palate, and bilateral clefts. If subsequent visits are foreshadowed by diagnosis, families need to be informed of this probability. It is crucial that families understand the potential nature of treatment when they begin it, including the need to return to Antigua for future surgical correction or revision. Informed consent is as relevant in developing countries as in wealthy ones, and poses particular challenges for surgical missions. These include language barriers, literacy and cultural expectations regarding outcomes. Informed and valid consent is required for all surgical patients in the United States, and STSM cannot make exceptions based on the prerogative that providing surgical assistance to those who lack resources and access is an exemption to this practice.

Describing the relationships between additional surgical needs and pre-operative diagnosis, and integrating this information with volume data, informs materials management activities in which provisions are made available to meet anticipated future demands. For example, patients with both cleft lip and palate were found to significantly predict additional surgical needs, therefore frequency data regarding this diagnoses can be used to foresee probable case mix and case load for future missions. Program stability can also be assessed. In Ghana, researchers suggested that children who returned for additional treatment, as in the case of children with both cleft lip and palate, was a demonstration of satisfaction with initial care (Donkor. et al.,2007).

For children under the age of 5 years who returned for subsequent visits, their age-in-months and weight in kilograms were plotted against WHO international growth standards for weight-for-age (see Chapter 5, figures 5.17 and 5.18). The majority of the children in the sample, had z scores indicating they were malnourished or severely malnourished (WHO, 2003). A relationship between nutritional deficiencies and orofacial clefts is suggested, highlighting the clinical significance of providing surgical repair for optimal growth and development. Parents, in making the decision to return for subsequent treatment, may also be aware of the potential benefits of correction for proper nutrition, as they are presumably the ones with firsthand experience concerning the difficulties these children encounter when eating.

Predictors of future needs

Future needs documented by FOH physicians in the medical record were significant predictors of subsequent visits. Compared to patients without documented future needs, patients with future surgical needs were 11.85 times more likely to return, and patients needing follow up were 9.6 times more likely to return. For the group of patients needing additional surgery, 59.7% returned for a subsequent visit. For patients whose physicians indicated follow up was necessary, 28.6% of the patients returned. These percentages suggest that more effort is needed to follow up on patients with documented needs for further evaluation or care.

For teams such as FOH that fly into developing countries, perform surgical procedures and return home, lack of information regarding outcomes for patients where follow up is indicated precludes measuring functional outcomes. Determining the outcomes of surgical repair is necessary to address best practices as global healthcare models are developed. Research has demonstrated that cleft care is ongoing and requires a multidisciplinary team approach (Bearn, et al., 2001; Shaw, et al., 2001). Functional and psychosocial outcomes are supported when follow-up services are provided and necessary treatment is available. Without provisions for aftercare, STSM were characterized as aesthetic rather than functional undertakings (Zbar, Rai & Dingman, 2000).

Conclusion

The duty to assist includes the obligations of individuals, national and international entities to provide help when human flourishing and health is threatened (Ruger, 2006). Poverty, malnutrition, cultural beliefs and limited access to healthcare contribute to widening gaps between the life expectancies of people in industrialized nations and those in developing ones (Benatar & Singer, 2000).

Morbidities which are virtually unknown in wealthier nations continue to devastate the world's poor (Wall et al., 2006). Morbidities in which surgical correction may offer functional and/or corrective improvement, but access issues hinder treatment, are the focus of STSM. Providing access to plastic surgeries in Guatemala, particularly cleft lip and palate surgery, is the mission of FOH, together with building the capacity of Guatemalan health care programs.

Concerns regarding the limits of short-term surgical missions have been raised. Projects which perform highly technical cases without the ability to follow up are being questioned in the current debate. Repair of oral clefts may require additional procedures for satisfactory results with appearance and functional outcomes. Post operative speech and hearing assessments may be necessary and speech therapy and revisional surgeries required (Noor & Musa, 2006). The evaluation study of FOH activities for the years 2005-2009 was undertaken as an initial effort to address some of these concerns.

Limitations

Because of non-random sampling the findings of the study cannot be generalized. The design was not longitudinal, so program changes over time cannot be inferred. Retrospective data collection for evaluation purposes prevents immediate intervention for quality concerns. Medical record abstraction methodology precludes recovery of missing data, and it was not possible to re-verify the accuracy of outliers within the data. However, despite the study limitations important findings and lessons learned will be shared with the surgical outreach community, and incorporated into FOH practices.

Policy

The study findings suggest that trends in program outputs changed over the course of FOH activities. For example, there was a significant change in the number of surgical complications in 2006 compared to 2009. In the United States, organizations providing surgical services must meet accreditation standards in order to be awarded governmental and private payor contracts. Part of the accreditation process is the establishment of oversight committees that monitor quality outputs, such as Medical Audit and Continuous Process Improvement Committees. These governing bodies have established policies and procedures to ensure quality, which includes peer review of specific complications, with timely feedback provided. If necessary, clinical practice guidelines may be implemented, including systematic changes or the requirement for providers to be proctored or receive additional training. Developing policies to insure that STSM organizations have an infrastructure in place for timely oversight of output data should be compulsory.

Practice

The study results indicate that children with orofacial clefts who are nutritionally compromised have an increased likelihood of surgical complications. Based on these findings, stringent guidelines need to be developed to identify children who are nutritionally compromised. Appropriate pre-operative interventions, including nutritional support and education, underscore the value of partnerships with local healthcare providers and international teams.

Because cleft care is often an ongoing process, especially for children with both cleft lip and palate or bilateral cleft lip, informed consent concerning the nature of surgical intervention for the families of affected children should be a fundamental practice of STSM. Informed consent is as relevant in poor countries as in wealthy ones. Though it poses particular challenges for surgical missions including language barriers, literacy and cultural expectations, it cannot be abrogated.

Future Research

Future research designed to measure additional outcomes of FOH activities, including patient and caregiver satisfaction with surgery, is an important next step. Collecting outcome data to capture long term functional status is necessary to measure the overall success of the surgery. Photographic data of surgical results presented for independent review will also provide important information to evaluate the surgical outcomes of FOH activities.

Expanding the OFE (figure 3.1) to include and measure the concept of impact will deepen the evaluation framework. An example would be to measure the effect of repaired oral cleft on the lives of people in developing countries. The concept of impact carriers beyond particular clinical factors to include more general sociological impacts. In developing countries oral clefts can result in near total performance limitations, with consequences beyond those inherent in the deformity. An orofacial cleft, when unrepaired, can prevent schooling, socialization, social education, and employment opportunities (Reeve, et al., 2004). These issues have their own sequalae. Lack of education leads to even further restrictions. Repair of a cleft, therefore, may carry benefit far beyond the anatomic correction. Inquiries into the impacts of cleft repair for patients in underserved countries is important for overall program evaluation.

Strengths

Charitable aims can be a vehicle for promoting scientific inquiry and are germane to the goals of humanitarian missions, where engagement in the relief of the poor and kindness for the needy are integral to the projects. This was the case with this study, where research aims created to evaluate the activities of FOH, resulted in knowledge development as a manifestation of charitable acts.

There are few guidelines regarding evaluation of volunteer programs and it is appropriate to develop them to optimize benefits (Uetani et al., 2004). The literature is sparse concerning the clinical, primary and psychosocial results of surgical missions. It is significant to analyze these programs, and this study may be viewed as an important first step.

References

Aaronson, L, & Burman, M (1994). Use of health records in research: reliability and validity issues. *Research in Nursing and Health, 17,* 67-76.

Allisson, JJ, Wall t, Spettell C, Calhoun J, Fargason C JR, Kobylinski R, Farmer, R & Kiefe, C (2000). The art and science of chart review. *Joint Commission Journal of Quality Improvement, 26(3),* 115-136.

Ballard, D (2003). Indicator to improve clinical quality across an integrated health system. *International Journal for Quality in Health Care, 15*, 13-23.

Banatvla, N. & Zwi, A.B., (2000). Public health and humanitarian interventions: developing the evidence base. *Conflict and Health, 321(7253),* 101-105.

Banks, N. (1998). Designing medical record abstraction forms. *International Journal for Quality in Health Care 10(2),* 163-167.

Bearn, D., Nildinhall, J., Murphy, T., Murray, J., Shell, D., Shaw, W., Williams, A., & Sandy, J (2001). Cleft lip and palate care in the United Kingdom: the clinical standards advisory groups (CSAG) study Part 4: outcome comparisons, training and conclusions. *Cleft Palate-Craniofacial Journal 38(1),* 38-43.

Benatar, S.R., Daar, A.B., & Singer, P. A. (2005). Global health challenges: the need for an expanded discourse on bioethics. *PloS Medicine 2(7),* 142-146.

Benatar, S.R., Singer, P.A. (2000). A new look at international research ethics. *BMJ* 321, 824-826.

Bermudez, L., Carter, V., Maggee, Wm Jr., Sherman, R & Ayala, R (2010). Surgical outcomes auditing systems in humanitarian organizations. *World Journal of Surgery* 34, 403-410.

Bunyavanich, S., Walkup, R. B. (2001). US public health leaders shift toward a new paradigm of global health. *American Journal of Public Health, 91(10),* 1556-1558.

CDC (1999), Framework for Program Evaluation in Public Health. *MMWR, 48(RR11),* 1-40.

Dagher, D & Ross, E (2004). Approaches of South African traditional healers regarding treatment of cleft lip and palate. *The Cleft Palate- Craniofacial Journal 41(5)*, 461-469.

Damiano, P. C., Tyler, M.C., Romitti, P.A., Momany E.T., Canady, J. W., Karnell, M. P., Murray, J. C. (2006). Type of oral cleft and mothers' perceptions of care, health status, and outcomes for preadolescent children. *Cleft Palate-Craniofacial Journal 43(6),* 715-721.

DeMey, Al., Sweennen, G., Malevey, C., George M., Mansbach, A.L. (2006). Long-term follow up of UCLP at the Reine Fabiola Children's Hospital. *B-ENT 2 Suppl 4,* 44-50.

Donkor, P., Bankas, D., Agbenorky, P., Plange-Rhule, G. MB., & Ansah, SK (2007). Cleft lip and palate surgery in Kumasi, Ghana: 2001-2005. *Journal of Craniofacial surgery 18(6),* 1376-1379.

Dresselhaus, T., Luck, J. & Peabody, J (2002). The ethical problem of false positives: a prospective evaluation of physician reporting in the medical record. *Journal of medical ethics 28 (5),* 291-294.

Dupuis, C. C. (2004). Humanitarian missions in the third world: a polite dissent. *Plastic and Reconstructive Surgery 113(1)*, 433-435.

Eberlin, K., Zaleski, K., Snyder, H, Hamdan, U (2008). Quality assurance guidelines for surgical outreach programs: a 20 year experience. *Cleft Palate-Craniofacial Journal 45(3),* 246.255.

El-Shazly, M., Bakry, R., Tohamy, A., Ali, WM., Elbakry, S., Brown, S.E., & Weatherley-White, R.C. (2010). Attitudes toward children with cleft in rural Muslim and Hindu societies. *Annals of Plastic Surgery 64(6),* 780-783.

Fischer, QA, Nichols, D, Stewart, F, Finley, G, Magee, W, JR, Nelson, K (2001). Assessing pediatric anesthesia practices for volunteer medical services abroad. *Anesthesiology 95(6),* pp1315-1322.

Gilbert, Lowenstern, Koziol-McClaine, Barta, & Steiner (1996). Chart reviews in Emergency Medicine Research: where are the methods? *Annals of Emergency Medicine 27(3),* 305-308.

Huff, E. (1997). Comprehensive reliability assessment and comparison of quality indicators and their components. *Journal of Clinical Epidemiology 50(12),* 1395-1404.

Kaafarani, HM. & Rosen, AK. (2009). Using administrative data to identify surgical adverse events: an introduction to the Patient Safety Indicators. *American Journal of Surgery 198(5 Suppl),* S63-68.

Logan S., Boutotte, J., Wilce, M. & Etkind, S (2003). Using the CDC framework for program evaluation in public health to assess tuberculosis contact investigation programs. *International Journal of Tuberculosis and Lung Disease* 7 *(12 Suppl 3),* S375-83.

Luck, J, Peabody, J, Dresselhaus, T, Lee, M & Glassma, P (2000). How well does chart abstraction measure quality? A prospective comparison of standardized patients with the medical record. *The American Journal of Medicine 108(8)*, 642-645.

Maki, J., Qualls, M., White, B., Kleefield, S., Cron, R. (2008). Health impact assessment and short-term medical missions: a methods study to evaluate quality of care. *BMC Health Services Research 2(8),* 121-135.

McQueen, K.A., Magee, W., Crabtree, T., Romano, C., & Burkle, FM Jr (2009). Application of outcome measures in international humanitarian aid: comparing indices through retrospective analysis of corrective surgical cases. *Prehospital Disaster Medicine 24 (1),* 39-46.

Mills EJ., Robinson, J., Altaran, A., Clark, m., Singh, S., Upshur, RE., Herrmann, KJ. & Yusuf, S. (2005). Sharing evidence of humanitarian relief. *British Medical Journal 331 (17531),* 1485-6.

Moore, M. & Fernandez, A (2006). Cleft Surgery in East Timor: the first four years. *ANZ Journal of Surgery 76(8),* 683-687.

Mulliken, JB. (2004). The changing faces of children with cleft lip and palate. *The New England Journal of Medicine 351(8),* 745-747.

Nagurmey, J, Brown, D, Jan, S, Weiner, J, Wang, A, Chang, Y (2005). The accuracy and completeness of data collected by prospective and retrospective methods. *Academic Emergency Medicine 12 (9),* 884-95.

Noor, S. N., Musa, S. (2007). Assessment of patients' level of satisfaction with cleft treatment using the Cleft Evaluation Profile. *Cleft Palate-Craniofacial Journal 44(3)*, 292-303.

Pan, L., Fergusson, D., Schweitzer, I., Hebert, P (2005). Ensuring high accuracy of data abstracted from patient charts: the use of a standardized medical record as a training tool. *Journal of Clinical Epidemiology 58(9),* 918-923.

Panatela, N. & Zwi, A.B., (2000). Public health and humanitarian interventions: developing the evidence base. *Conflict and Health 321(7253)*, 101-105.

Patel, Z., Ross, E. (2003). Reflection on the cleft experience by South African adults: use of qualitative methodology. *Cleft Palate-Craniofacial Journal 40(5),* 471-480.

Peabody, J, Luck, J Glassman, P Dresselhaus, T & Lee (2000). Comparison of vignettes, standardized patient, and chart abstraction: a prospective validation study of 3 methods for measuring quality. *JAMA 282(13).* 1715-1722.

Pezzella, AT., (2006). Volunteerism and humanitarian efforts in surgery. *Current Problems in Surgery 43(12),* 848-929.

Raval, M.V., Dillom, P.W., Bruny, J.L., Ko, C.Y., Hall, B.L., Moss, R.L., Oldham, K.T., Richards, K.E., Vinour, C.D., Ziegler, M.M.; ACS NSQIP Pediatric Steering Committee (2011). American College of surgeons National Surgical Qualilty Improvement Program Pediatric; a phase 1 report. *Journal of the American College of Surgeons 21(1),* 1-11.

Reeve, M.E., Groce, N.E., Persing, J.A., & Magge, S.N. (2004). An international surgical exchange program for children with cleft lip/cleft palate in Manaus, Brazil: patient and family expectations of outcome. *Journal of Craniofacial surgery 15(1),* 170-4.

Ruger, J.P. (2006). Ethics and governance of global health inequalities. *Journal of Epidemiology and Community health, 60,* 998-1002.

Rugg, D., Carael, M., Boerma, J.T., & Novak, J. (2004). Global advances in monitoring and evaluation of HIV?AIDS: from AIDS case reporting to program improvement. *New Directions for Evaluation 103*, 33-47.

Schwarz, R. & Khadka, S. (2003). Reason for late presentation of cleft deformity in Nepal. *Clef Palate-Craniofacial Journal 41 (2),* 199-201.

Sharp, H. M., Canady, J. W., Ligot, F. A., Hague, R. A., Gutierrez, J., Gutierrez, J., (2008). Caregiver and patient reported outcomes after repair of cleft lip and/or palate In the Philippines. *Cleft Palate-Craniofacial Journal 45(2),* 163-171.

Shaw, W., Semb, B., Nelson, P., Brattstrom, V., Molsted, K., Prahl-Andersen B., & Gundlack K (2001). The Eurocleft Project 1996-2000: overview. *Journal Craniomaxillofacial Surgery 29(3),* 131-140.

Smoot EC 3rd, Johnson, m., Graham, DR., & Draper GA., (1992). Operating safety in an underdeveloped country. *Cleft Palate and Craniofacial Journal 29(5),* 445-450.

Smoot, EC., (1998). Over serving the underserved. *Plastic and Reconstructive Surgery 101(5),* 1420-1.

Tilkian, S.M., Conover, M. B. & Tilkian, A.G.(1979). Clinical implications of laboratory tests, 2nd Ed. Mosby: Missouri

Turner, S. R., Thomas, P. W., Dowell, T., Rumsey, N., Sandy, J. R. (1997). Psychological outcomes amongst cleft patients and their families. *British Journal of Plastic Surgery 50(1)*, 1-9.

Uetani, M., Jimba, M., Natsume N., Katsuki T., Xuan le T. T., Wakai, S. (2006). Effects of a long- term volunteer surgical program I a developing country: the case in Vietnam from 1993-2003. *Cleft Palate - Craniofacial Journal* 43(5), 616-619.

Wall, L.L., Arrowsmith, S.D., Lassey. A.T. & Danso, K. (2006). Humanitarian ventures or ‘fistula tourism?’: the ethical perils of pelvic surgery in the developing world. *International Urogynecolgical Journal, 17,* 559-562.

Walsh, D. S. (2004). A framework for short-term humanitarian health care projects. *International Nursing Review 51(1)*, 23-26.

Weatherly-White, R.C., Eiserman, W., Beddoe, M., & Vanderberg, R. (2005). Perception, expectations and reactions to cleft lip and palate surgery in native populations: a pilot study in rural India. *Cleft Palate/Craniofacial Journal 42(5),* 560-4.

Weiser TG., Regenbogen SE., Thompson KD., Haynes AB., Lipsitz, SR., Berry WR. & Gawande, AA., (2008). An estimation of the global volume of surgery: a modeling strategy based on available date. *Lancet 372 (9633),* 139-144.

WHO multicentre growth reference study group, WHO child growth standards: length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: methods and development, World Health Organization, Geneva (2006).

World Health Organization (2006). WHO child growth standards: methods and development: length/height-for-age, weight-for-age, weight-for-length and body mass index. *World Health Organization: Geneva.*

Wu, L & Ashton, C. (1997). Chart Review: a need for reappraisal. *Evaluation and the health professions 20(2),* 146-163.

Yang, I-Y & Lia, Y-F(2010). The effect of 1-stage vs. 2-stage palate repair on facial growth in patients with cleft lip and palate: a review. *International Journal of oral and maxillofacial surgery 39(18),* 945-950.

Yeow UK., Lee ST., Lambrecht, T. J., Barnett, J., Gorney, M., Hardjowasito, W., Lemperie, G., McComb, H., Natume, N., Strac, M., & Wilson, L., (2002). International Task Force on Volunteer Cleft Missions. *Journal of Craniofacial Surgery 13 (1),* 18-25.

Zbar, R. I., Rai, S. M., Dingman, D.L., (2000). Establishing cleft malformation surgery in developing nations: a model for the new millennium. *Plastic and Reconstructive Surgery 106(4),* 886-889.

Zhan, C., Battles, J, Chiang, Y. & Hunt, D (2007). The validity of ICD-9-CM codes in identifying post-operative deep vein thrombosis and pulmonary embolism. *Joint Commission Journal of Quality and Patient Safety 33 (6),* 326-331.

Appendix A. Histograms of clinical profiles

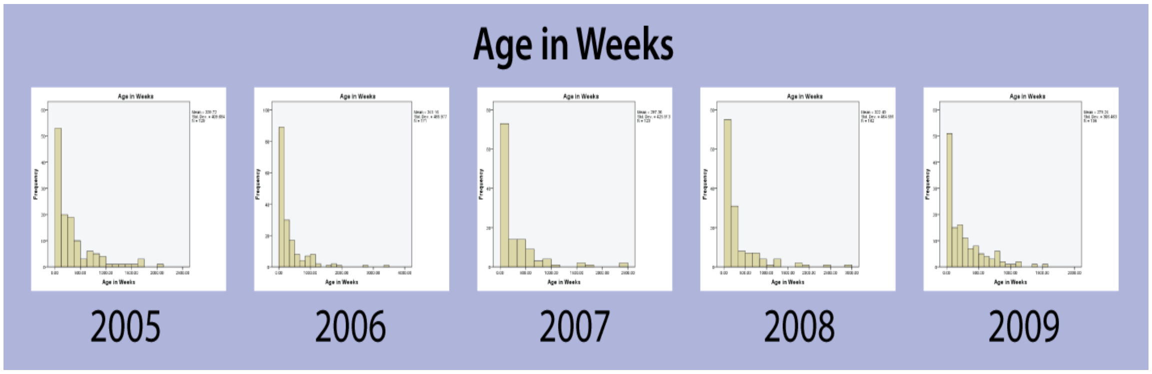


Figure 1.0. Histogram of age in weeks for years 2005-2009.

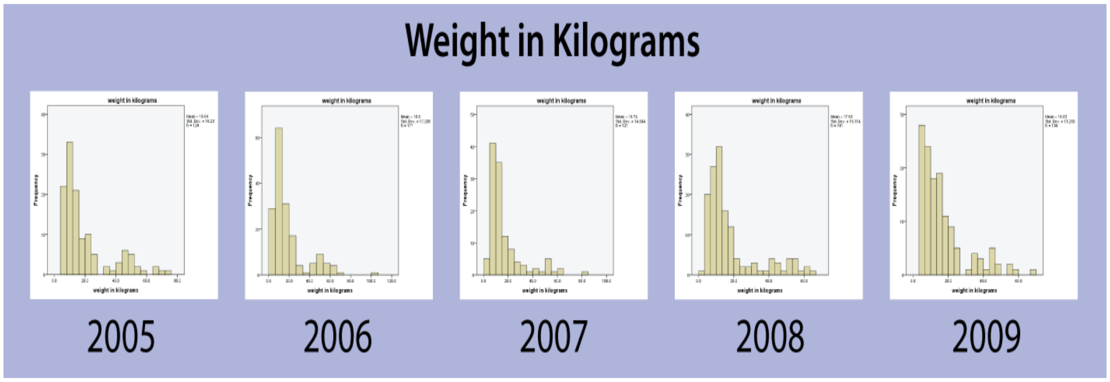


Figure 2.0. Histograms for weight in kilograms for years 2005-2009.

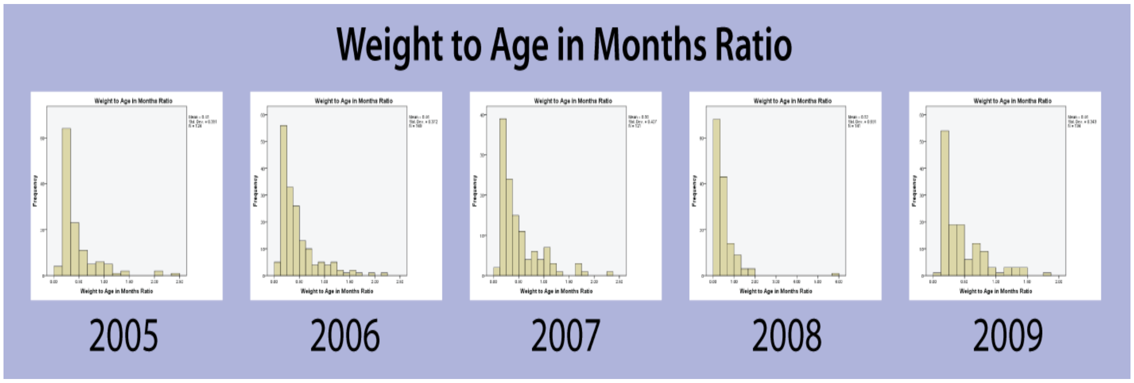


Figure 3.0. Histograms for weight to age in months ratio for the years 2005-2009.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

	12/31/2011
Author Signature	Date