Effective Leadership of Surgical Teams: A Mixed Methods Study of Surgeon Behaviors and Functions

Juliana L. Stone, MS,* Emma-Louise Aveling, PhD, Molly Frean, BA, Morgan C. Shields, MS, Cameron Wright, MD, Francesca Gino, PhD, Thoralf M. Sundt, MD, and Sara J. Singer, PhD

Department of Health Research and Policy, Harvard T. H. Chan School of Public Health, Boston, Massachusetts; Cambridge Centre for Health Services Research, University of Cambridge, Cambridge, United Kingdom; Health Care Management Department, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania; Institute for Behavioral Health, The Heller School for Social Policy and Management, Brandeis University, Waltham, Massachusetts; Department of Surgery, Massachusetts General Hospital, Boston, Massachusetts; and Negotiation, Organizations & Markets, Harvard Business School, Boston, Massachusetts

Background. **The importance of effective team leadership for achieving surgical excellence is widely accepted, but we understand less about the behaviors that achieve this goal. We studied cardiac surgical teams to identify leadership behaviors that best support surgical teamwork.**

Methods. **We observed, surveyed, and interviewed cardiac surgical teams, including 7 surgeons and 116 team members, from September 2013 to April 2015. We documented 1,926 surgeon/team member interactions during 22 cases, coded them by behavior type and valence (ie, positive/negative/neutral), and characterized them by leadership function (conductor, elucidator, delegator, engagement facilitator, tone setter, being human, and safe space maker) to create a novel framework of surgical leadership derived from direct observation. We surveyed nonsurgeon team members about their perceptions of individual surgeon's leadership effectiveness on a 7-point Likert scale and correlated survey measures with individual surgeon profiles created by calculating percentage of behavior types, leader functions, and valence.**

Results. **Surgeon leadership was rated by nonsurgeons from 4.2 to 6.2 (mean, 5.4). Among the 33 types of behaviors observed, most interactions constituted elucidating (24%) and tone setting (20%). Overall, 66% of interactions (range, 43%–84%) were positive and 11% (range, 1%–45%) were negative. The percentage of positive and negative behaviors correlated strongly ($r = 0.85$ for positive and $r = 0.75$ for negative, $p < 0.05$) with nonsurgeon evaluations of leadership. Facilitating engagement related most positively ($r = 0.80$; $p = 0.03$), and negative forms of elucidating, ie, criticism, related most negatively ($r = -0.81$; $p = 0.03$).**

Conclusions. **We identified 7 surgeon leadership functions and related behaviors that impact perceptions of leadership. These observations suggest actionable opportunities to improve team leadership behavior.**

(Ann Thorac Surg 2017;■:■–■)

Effective teamwork is essential to safe surgical care [1]. Nontechnical aspects of team performance, such as communication failures, contribute to surgical errors and adverse outcomes, especially in cardiac operations [2] and may be avoidable through improved interpersonal interactions [3]. Although research suggests that leadership impacts team performance [4], little is known about which leadership behaviors benefit surgical teamwork and which do not.

Surgeons are de facto team leaders, yet surgical training focuses on technical skills. Leadership behaviors are "picked up" by observing role models without evidence to support or refute their effectiveness in promoting team performance. An objective understanding of the impact of specific behaviors is therefore critical to optimizing surgical leadership.

We undertook an observational study of how surgeons actually lead in the operating room and created a tool for assessing surgeons' leadership. Using data from surgical observations and interviews with team members, we characterized behaviors as positive, neutral, and negative and compared these with measures of surgeons' leadership

Accepted for publication Jan 4, 2017.

*Juliana L. Stone is the recipient of the 2016 Southern Thoracic Surgical Association Carolyn Reed President's Award.

Address correspondence to Dr Singer, Department of Health Research and Policy, Harvard T. H. Chan School of Public Health, 677 Huntington Ave, Boston, MA 02115; email: ssinger@hsph.harvard.edu.

The Supplemental Materials can be viewed in the online version of this article [http://dx.doi.org/10.1016/j.athoracsur.2017.01.021] on http://www.annalsthoracicsurgery.org.

0003-4975/$36.00
http://dx.doi.org/10.1016/j.athoracsur.2017.01.021

as perceived by surgical team members. In doing so, we developed a novel and empirically based framework of leadership functions and behaviors that can be used to enhance surgeons' leadership of operating room teams (see List of Supplemental Material and Supplemental Fig 1).

Material and Methods

Research Setting/Design

We applied mixed methods to study cardiac surgery teams in an academic medical center performing more than 1,000 cardiac surgical procedures annually with outcomes meeting national benchmarks. Data were collected during two 4-month periods between September 2013 and April 2015. Each data collection period comprised: (1) a staff survey on team dynamics and surgeon leadership, (2) observations of surgeons' interactions with team members during surgical procedures, and (3) semistructured interviews with team members to gain insights on contextual influences underlying observed interactions. Data from the 2 collection periods were combined after confirmation of little substantive change over time and subjected to cross-sectional analysis. The institutional review boards of participating centers approved this study. Supplemental Material-A presents detail regarding study methods and results.

Sample

The study population included surgeons, scrub technicians/nurses, circulating nurses, physician assistants, perfusionists, anesthesiologists, and trainees (eg, surgical fellows, anesthesia residents). This included 7 surgeons and 116 nonsurgeons across the 2 data collection periods. Three nonsurgeons declined to participate and were excluded from the research.

In each data collection period, we surveyed all active surgeons and nonsurgeons in the sample, including 7 surgeons and 82 nonsurgeons in the first period and 5 surgeons and 105 nonsurgeons in the second period. We observed cases involving all 7 surgeons in the first data collection period and 4 of the 5 active surgeons in the second period (1 surgeon requested observations be discontinued). We conducted interviews with 34 surgical team members, including all surgeons and 1 to 3 team members from each discipline.

Data and Data Collection

The survey (Supplemental Material-B) used 13 constructs of 1 to 3 items drawn from previously validated scales to measure surgical staff member perceptions and attitudes about themselves, their teams, and team dynamics. Nonsurgeons were also asked to evaluate the general performance of each surgeon as a team leader. Surveys were administered electronically and used a 7-point Likert scale.

We used an observation tool (Supplemental Material-C) to collect data about interactions between surgeons and nonsurgeons during individual surgical procedures. In addition to closed-ended items about case characteristics (date/time/location, type and difficulty, checklist use, level of surgeon participation), the tool largely comprised structured space to allow investigators to record all verbal and nonverbal interactions. Each surgeon was first observed by a team of 2 to 4 investigators to calibrate use of the tool, enhance its reliability, and acclimate surgical team members to our presence. After calibration, 1 research assistant (RA) observed each case. We pilot tested the tool in 23 cases (2 to 4 cases per surgeon) outside of formal data collection in the first period and again in 13 cases (2 to 5 cases per surgeon) outside of formal data collection in the second period. In total, the analytic sample included 22 cases (14 in the first period and 8 in the second period) comprising 110 observation hours. Average case duration was 5 hours, ranging from 1 to 9 hours.

Semistructured interviews asked participants to describe operating room team dynamics at their best and worst, frequency of and factors influencing such conditions, opportunities for improvement, perceived level of shared understanding among team members, and contextual influences underlying surgeon/team member interactions (Supplemental Material-D). In the second data collection period, we asked participants to comment on preliminary findings from the first period, resulting in modifications, as needed, of our initial interpretations. Interviews were conducted by 1 or 2 investigators, were confidential, lasted 15 to 60 minutes, and were digitally recorded and transcribed. Participation in all data collection was voluntary and without incentives.

Analysis

Survey data from both collection periods were combined into a single analytic data set. For individuals who completed the survey twice, their responses were averaged and the mean taken as their score for each item. We calculated composite scores for each survey construct and generated distributions and descriptive statistics for all measures. To evaluate surgeon performance as perceived by surgical staff, we averaged responses provided by all nonsurgeons for the survey question on performance of the surgeon as a team leader.

During observation pilot testing, we performed qualitative coding to generate an initial set of behavior codes and definitions. After initial coding, we compared our empirically derived codes with previously published taxonomies for surgeon or surgical team member behaviors (Supplemental Material-E) [5–10]. Given little consensus among preexisting taxonomies and minimal overlap with our codes, we made only minor word choice changes. We then classified coded interactions for the 22 cases in our analytic sample into 33 behavior types. Each RA independently coded 5 transcripts to establish interrater reliability and coding consistency (kappa = 0.8; $p < 0.0001$) so that all remaining transcripts could be reviewed and coded by 1 RA.

We assigned a valence to each behavior type (positive, neutral, or negative) based on investigator assessment of the contribution of the behavior to more or less productive team dynamics. A neutral valence indicated ambiguity or that the behavior was contextually contingent. The 33 behavior types were then grouped into 7 distinct leadership