Journal of Hospital Medicine®

www.journalofhospitalmedicine.com

AUGUST 2018 NO. 8 VOL. 13

An Official Publication of the Society of Hospital Medicine

shm
Society of Hospital Medicine

Play a bigger role at SHM

Your voice and participation are an integral part of SHM. Make a difference and help shape the future of hospital medicine.

Apply for a Committee

SHM is now accepting applications for 2019 committee members.
hospitalmedicine.org/committees

Join or Start a Special Interest Group (SIG)

Applications are now being considered for members interested in launching new SIGs.
hospitalmedicine.org/sigs

shm
Society of Hospital Medicine

Empowering hospitalists.
Transforming patient care.

Journal of Hospital Medicine®

Journal of Hospital Medicine®

Publisher: The Journal of Hospital Medicine (Print ISSN 1553-5592; E-ISSN 1553-5606) is published monthly for the Society of Hospital Medicine by Frontline Medical Communications, with business offices at 7 Century Drive, Suite 302, Parsippany, NJ 07054-4609, telephone 973-206-3434, fax 973-206-9378. Periodicals postage paid at Parsippany, NJ and at additional mailing offices.

Postmaster: Send address changes to Journal of Hospital Medicine, Subscription Services, P.O. Box 3000, Denville, NJ 07834-3000.

Reprints: For article reprint requests in the United States and Canada, please contact Wright's Media, toll free: (877) 652-5295; frontline@wrightsmedia.com. For those outside the US/Canada, contact Ray Thibodeau, Content Ed Net, at (267) 895-1758; ray.thibodeau@contentednet.com.

Subscriptions: Individual, USA: $406.00 p.a.; institution, USA: $845.00 p.a.; student, resident, intern, USA: $118.00; $35.00 per copy, USA. Individual, Canada/Mexico: $438.00 p.a.; institution, Canada/Mexico: $925.00 p.a.; $47.00 per copy, Canada/Mexico. Individual, all other nations: $739.00 (surface mail), $950.00 (air mail); institution, all other nations: $925.00; $59.00 per copy, all other nations. For back issues, call (833) 836-2705 to charge to your credit card. Written requests will be accepted and must be accompanied by check or money order. Send payment and request to Journal of Hospital Medicine, Subscription Services, P.O. Box 3000, Denville, NJ 07834-3000.

For subscriptions, please call (833) 836-2705, Journal of Hospital Medicine, Subscription Services, P.O. Box 3000, Denville, NJ 07834-3000 or e-mail custsvc.jhm@fulcoinc.com.

All claims for replacement of missing copies of Journal of Hospital Medicine, without charge, must be made within 3 months of date of missing issue requested. Otherwise, cost of replacement is $47.00 USA, $59.00 Canada/Mexico, $71.00 Other per copy. Produced in the USA.

WEBSITE
www.journalofhospitalmedicine.com

EDITORIAL STAFF

Managing Editor
Susan D. Hite
(240) 221-2471
susanhite@mdedge.com

ART

Creative Director
Mary Ellen Niatas

Art Director
John J. DeNapoli

PRODUCTION AND MANUFACTURING

Production Manager
Mike Wendt
(973) 206-8010
mwendt@mdedge.com

CIRCULATION

Subscription Service:
(833) 836-2705
custsvc.jhm@fulcoinc.com

PUBLISHING STAFF

Publisher
Mark Branca
(973) 290-8246
mbranca@mdedge.com

Director, Business Development
Angela Labrozzi, MBA
(973) 206-8971
Cell (917) 455-6071
alabrozzi@mdedge.com

Classified Sales
Linda Wilson
(973) 290-8243
lwilson@mdedge.com
Heather Gonroski
(973) 290-8259
hgentile@mdedge.com

Accounts Receivable
Joan Friedman
(973) 290-8211
jfriedman@mdedge.com

Journal of Hospital Medicine

shm Society of Hospital Medicine
Empowering hospitalists. Transforming patient care.

TABLE OF CONTENTS

Volume 13 | Number 8 | August 2018

ORIGINAL RESEARCH

BRIEF REPORTS

Continued >

CHOOSING WISELY®: THINGS WE DO FOR NO REASON

CLINICAL CARE CONUNDRUM

LETTERS TO THE EDITOR

How Does Your PICCOMPARE? A Pilot Randomized Controlled Trial Comparing Various PICC Materials in Pediatrics

Tricia Kleidon, RN, MNursSci[1,2]*, Amanda J. Ullman, RN, MAppSci, PhD[2,3], Li Zhang, MBBS, PhD[2], Gabor Mihala, MEng[2,4,5], Brett Chaseling, MBBS (Hons), FANZCA[1,6], Jason Schoutrop, BSc (Hons), MBBS, FANZCA[1,6], Claire M. Rickard, RN, PhD[2,7]

[1]Children's Health Queensland, Lady Cilento Children's Hospital, South Brisbane, Queensland, Australia; [2]Alliance for Vascular Access Teaching and Research Group, Menzies Health Institute Queensland, Nathan Campus, Queensland, Australia; [3]School of Nursing and Midwifery, Griffith University, Nathan Campus, Queensland, Australia; [4]Centre for Applied Health Economics, Menzies Health Institute Queensland, Nathan Campus, Queensland, Australia; [5]School of Medicine, Griffith University, Queensland, Australia; [6]University of Queensland, St. Lucia, Queensland, Australia; [7]School of Nursing and Midwifery, Griffith University, Nathan Campus, Queensland, Australia.

BACKGROUND: Despite the popularity of peripherally inserted central catheters (PICCs), recent literature highlights their potential injurious complications. Innovative PICC materials have been developed to prevent thrombosis and infection formation (Endexo®) and antireflux valves to prevent occlusion (pressure-activated safety valve®). No large randomized controlled trial has assessed these technologies. Our primary aim was to evaluate the feasibility of a large randomized controlled efficacy trial of PICC materials and design to reduce PICC complication in pediatrics.

METHODS: A randomized controlled feasibility trial was undertaken at the Lady Cilento Children's Hospital in South Brisbane, Australia, between March 2016 and November 2016. Consecutive recruitment of 150 pediatric participants were randomly assigned to receive either (1) polyurethane PICC with a clamp or (2) BioFlo® PICC (AngioDynamics Inc, Queensbury, New York). Primary outcomes were trial feasibility, including PICC failure (thrombosis, occlusion, infection, breakage, or dislodgement). Secondary outcomes were PICC complications during use.

RESULTS: Protocol feasibility was established, including staff and patient acceptability, timely recruitment, no missing primary outcome data, and 0% attrition. PICC failure was 22% (16 of 74, standard care) and 11% (8 of 72, BioFlo®) corresponding to 12.6 and 7.3 failures per 1000 hours (risk ratio 0.58; 95% confidence interval, 0.21-1.43; $P = .172$). PICC failures were primarily due to thrombosis (standard care 7% versus BioFlo® 3%) and complete occlusion (standard care 7% versus BioFlo® 1%). No blood stream infections occurred. Significantly fewer patients with BioFlo® had PICC complications during use (15% vs 34%; $P = .009$).

CONCLUSION: BioFlo® PICCs appear potentially safer for pediatrics than traditional standard care PICCs with a clamp. Further research is required to definitively identify clinical, cost-effective methods to prevent PICC failure and improve reliability. *Journal of Hospital Medicine* 2018;13:517-525. Published online first February 8, 2018.

Peripherally inserted central catheters (PICCs) have evolved since their inception in the early 1970s and are used with increasing frequency for pediatric inpatients and outpatients.[1-3] Emerging literature, including a meta-analysis of international observational studies,[4] reports PICC failure (complications necessitating premature removal) occurs in up to 30% of PICCs, most commonly due to infection, thrombosis, occlusion, and fracture.[4-7] Raffini et al.[7] report the increasing incidence of pediatric PICC-related thrombosis increases morbidity and mortality[8] and negatively impacts future vessel health and preservation.[9]

PICCs have progressed from relatively simple, silicone-based catheters with an external clamp to chemically engineered polyurethane with pressure-activated valves placed at the proximal or distal catheter hub with the intent to reduce occlusion.[10] Further modernization of PICC material occurred with the incorporation of antithrombogenic (AT) material (Endexo®). These PICCs are designed to contain a nonstick polymer, which is designed to reduce the adherence of blood components (platelets and clotting factors) and inhibit thrombus formation (and hence prevent deep vein thrombosis and occlusion, as well as inhibit microbial biofilm attachment [and subsequent infection]).[11]

In addition to new materials, other aspects of this PICC design have been the addition of a pressure-activated safety valve (PASV®) built into the proximal hub. Pressure-activated valve technology promises to prevent catheter occlusion by reducing blood reflux into the PICC; the valve opens with

***Address for correspondence:** Tricia Kleidon, Lady Cilento Children's Hospital, level 7, Department of Anaesthetics, 501 Stanley Street, South Brisbane, 4101, Qld, Australia; Telephone: +61(0)407175301; Fax: +61(0)730684419; E-mail: tricia.kleidon@health.qld.gov.au

Additional Supporting Information may be found in the online version of this article.

Received: August 1, 2017; **Revised:** September 30, 2017; **Accepted:** October 16, 2017

© 2018 Society of Hospital Medicine DOI 10.12788/jhm.2911

pressure during infusion and aspiration and remains closed with normal venous pressure, circumventing the need for clinicians to manually clamp the PICC and reducing human error and the potential for thrombosis, occlusion, and fracture development.[12] Hoffer et al.[13] reported half as many occlusions of valved PICCs (3.3%) compared with nonvalved or clamped PICCs (7.1%); although not statistically significant ($P = .10$), perhaps due to the small sample, overall complications, including occlusion and infection, were significantly lessened with the valved PICC (35% vs 79%; $P = .02$). Comparatively, Pittiruti et al.[14] conducted a trial of 2 types of valved PICCs with an open-ended, nonvalved PICC and found no reduction in PICC occlusion or catheter malfunction.

Today, PICC use is common for patients who require short-to-medium intravenous therapy. PICCs are increasingly recognized for their significant complications, including thrombosis and infection.[15] Novel PICC technology, including the incorporation of AT material such as Endexo® and PASV®, may reduce complications; however, the clinical efficacy, cost-effectiveness, and acceptability of these innovations have not been tested through randomized trials in pediatric patients. In accordance with Medical Research Council guidelines[16] for developing interventions, we pilot tested the feasibility of the BioFlo® PICC, including intervention acceptability, compliance, recruitment, and initial estimates of effect, in anticipation of a subsequent full-scale efficacy randomized controlled trial. Our secondary aim was to compare the effectiveness of the BioFlo® PICC with Endexo® and PASV® technology in reducing PICC complications and failure.

METHODS

Design

We undertook a pilot randomized controlled trial comparing the standard polyurethane PICC (with external clamp) with the BioFlo® PICC (with internal valve) in preventing catheter failure in pediatric patients. The study was prospectively registered with the Australian Clinical Trials Registry (ACTRN12615001290583), and the research protocol was published.[17]

Study Setting

The study commenced in March 2016 at the Lady Cilento Children's Hospital in South Brisbane, Australia, a tertiary-level, specialist, pediatric teaching hospital in Queensland, Australia, providing full-spectrum health services to children and young people from birth to 18 years of age. Recruitment, including data collection, was completed in November 2016.

Sample

The target sample size was 110 participants, 50 participants per group plus 10% for potential attrition, as determined by standard pilot trial sample size recommendations.[18] With ethics approval, the sample size was later increased to 150 participants in order to adequately pilot a microbiological substudy method (published separately).[17] Participants were consecutively recruited if they met the inclusion criteria: PICC insertion, age <18 years, predicted hospital stay >24 hours, single-lumen PICC, and written informed consent by an English-speaking, legal parent or guardian. Patients were excluded if they had a current (<48 hours) blood stream infection (BSI), vessel size <2 mm, could not speak English without an interpreter, required a multilumen PICC, or were previously enrolled in the study.

Interventions

Participants were randomized to receive either of the following PICCs: (1) standard care: Cook™ polyurethane, turbo-ject, power-injectable PICC (Cook Medical, Bloomington, IN) or (2) comparison: BioFlo® polyurethane with Endexo® technology (AngioDynamics Inc, Queensbury, New York).

Outcomes

The primary outcome was feasibility of a full-efficacy trial established by composite analysis of the elements of eligibility (>70% of patients will be eligible), recruitment (>70% of patients will agree to enroll), retention and attrition (<15% of participants are lost to follow-up or withdraw from the study), protocol adherence (>80% of participants receive their allocated, randomly assigned study product), missing data (<10% of data are missed during data collection), parent and healthcare staff satisfaction, and PICC failure effect size estimates to allow sample size calculations.[18,19] PICC failure was defined as the following complications associated with PICC removal: (1) catheter-associated BSI,[8,20-22] (2) local site infection,[22] (3) venous thrombosis,[23] (4) occlusion,[24,25] (5) PICC fracture, or (6) PICC dislodgement.[25,26] Parents (or caregivers) and healthcare staff were asked to rate their level of confidence with the study product and ease of PICC removal by using a 0 to 100 numeric rating scale (NRS) of increasing confidence and/or ease. These data were collected at the time of PICC removal. Operators were also asked to rate their levels of satisfaction with the insertion equipment and ease of PICC insertion immediately upon completion of the insertion procedure (both 0-100 NRS of increasing satisfaction and/or ease). Secondary outcomes included individual PICC complications (eg, occlusion) occurring at any time point during the PICC dwell (including at removal), adverse events, pain, redness at the insertion site, and overall PICC dwell.

Study Procedures

The research nurse (ReN) screened operating theater lists for patients, obtained written informed consent, and initiated the randomization. Randomization was computer generated, and web based via Griffith University (https://www151.griffith.edu.au/random) to ensure allocation concealment until study entry. Patients were randomly assigned in a 1:1 ratio with computer-generated and randomly varied block sizes of 2 and 4. Data were collected by the ReN on the day of insertion, at day 1 postinsertion, then every 2 to 3 days thereafter so that PICCs were checked at least twice per week until study completion. Participants were included in the trial until 12 weeks post-PICC insertion, study withdrawal or PICC removal (whichever came first), with an additional 48 hours follow-up for infection outcomes. Patient review was face to face during the inpatient

stay, with discharged patients' follow-up occurring via outpatient clinics, hospital-in-the-home service, or telephone.

Data collection was via Research Electronic Data Capture (http://project-redcap.org/). The ReN collected data on primary and secondary outcomes by using the predefined criteria. Demographic and clinical data were collected to assess the success of randomization, describe the participant group, and display characteristics known to increase the risk of PICC complication and thrombosis. A blinded radiologist and infectious disease specialist reviewed and diagnosed thrombosis of deep veins and catheter-associated BSI outcomes, respectively.

PICC Procedures

Extensive prestudy education for 2 months prior to trial commencement was provided to all clinicians involved with the insertion and care of PICCs, including the study products. PICCs were inserted in an operating theater environment by a qualified consultant pediatric anesthetist, a senior anesthetic registrar or fellow in an approved anesthetic training program, or pediatric vascular access nurse practitioner. Ultrasound guidance was used to assess a patient's vasculature and puncture the vessel. The operator chose the PICC size on the basis of clinical judgment of vessel size and patient needs and then inserted the allocated PICC.[27] Preferred PICC tip location was the cavoatrial junction. All PICC tip positions were confirmed with a chest x-ray before use.

Postinsertion, PICCs were managed by local interdisciplinary clinicians in accordance with local practice guidelines.[27-31] PICC care and management includes the use of 2% chlorhexidine gluconate in 70% alcohol for site antisepsis and neutral displacement needleless connectors (TUTA Pulse; Medical Australia Limited, Lidcombe, New South Wales, Australia); normal saline was used to flush after medication administration, and if the device was not in use for 6 hours or longer, heparin is instilled with securement via bordered polyurethane dressing (Tegaderm 1616; 3M, St Paul, Minnesota) and a sutureless securement device (Statlock VPPCSP; Bard, Georgia).

Statistical Analyses

Data were exported to Stata 15[32] for cleaning and analysis. Data cleaning of outlying figures and missing and implausible data was undertaken prior to analysis. Missing data were not imputed. The PICC was the unit of measurement, and all randomly assigned patients were analyzed on an intention-to-treat basis.[33] Descriptive statistics (frequencies and percentages) were used to ascertain the primary outcome of feasibility for the larger trial. Incidence rates (per 1000 catheter days) and rate ratios, including 95% confidence intervals (CIs), were calculated. The comparability of groups at baseline was described across demographic, clinical, and device characteristics. Kaplan-Meier survival curves (with log-rank tests) were used to compare PICC failure between study groups over time. Associations between baseline characteristics and failure were described by calculating hazard ratios (HRs). Univariable Cox regression was performed only due to the relatively low number of outcomes. *P* values of <.05 were considered statistically significant.

Ethics

The Children's Health Service District, Queensland (Human Research Ethics Committee/15/QRCH/164), and Griffith University (2016/077) Human Research Ethics Committees provided ethics and governance approval. Informed consent was obtained from parents or legal guardians, with children providing youth assent if they were 7 years or older, dependent upon cognitive ability.

RESULTS

Participant and PICC Characteristics

Participant and PICC characteristics are described in Table 1. The majority of participant and PICC characteristics were balanced between intervention groups. The mean patient age was 7.3 years (standard deviation 5.0; range 0-18). PICC insertion was most commonly for a respiratory diagnosis (n = 98; 65%). Most PICCs were placed in the basilica vein (n = 115; 79%), with insertion being successful on the first attempt (n = 125; 86%). There was some imbalance (>10% absolute difference between groups) in nurse practitioner and registrar insertions (standard care 35% and 23% vs BioFlo® 51% and 8%, respectively) and patients with leucocytes <1000 µl (standard care 10% vs BioFlo® 22%). Optimal PICC tip location at the cavoatrial junction was higher with BioFlo® than standard care, although this difference was <10%.

Feasibility Outcomes

As shown in Figure 1, the majority of feasibility criteria were met, with 94% of 188 screened patients being eligible to participate and 97% of eligible patients consenting to enroll. Of 150 patients randomly assigned, 4 (1 in standard care and 3 in BioFlo®) were unable to have a PICC inserted or the procedure was cancelled. Demographic data only were collected for these 4 patients. No participants were lost to follow-up, and no primary outcome data were missing. Staff satisfaction with insertion kit and ease of insertion, ease of removal of the PICC, and parental confidence in the PICC product were similar across both groups (Table 2).

PICC Failure and Complications

In total, 24 of 146 participants (16%) experienced PICC failure. There were 16 (22%) failures of standard care PICCs and 8 (11%) failures of BioFlo® PICCs. This corresponded to incident rates of 12.6 and 7.3 per 1000 catheter days (incident rate ratio 0.58; 95% CI, 0.21-1.43; *P* = .172; Table 2). Failure was most commonly from thrombosis (n = 5; 7%) or occlusion (n = 5; 7%) in the standard care group, with lower incidences in the BioFlo® group (n = 2 [3%] and n = 1 [1%], respectively). Figure 2 displays survival from PICC failure.

Considering the entire PICC dwell, of the 74 standard care patients, 49 (66%) had no complications, 9 (12%) had complications during the dwell but none at removal, 2 (3%) had no complications during the dwell but had a complication (ie, failure) at removal, and 14 (19%) had complications during the dwell and at removal. For the 72 BioFlo® patients, 61 (85%) had no PICC complications, 3 (4%) had complications during the dwell but none at removal, 4 (5.5%) had no complications during

TABLE 1. **Participant (n = 150) and PICC Characteristics (n = 144)**

		Standard Care	BioFlo®
Participant characteristics	**N**	**n = 75**	**n = 75**
Age (years)[a]	150	7.5 (4.9) (0.0-18.0)	7.1 (5.1) (0.0-17.0)
Weight (kg)[a]	133	27.6 (16.2) (5.2-78.0)	28.4 (17.8) (4.9-70.0)
Sex (male)	149	41 (55)	45 (61)
Insertion on dominant side	101	17 (34)	16 (31)
Comorbidities:	150		
none		19 (25)	12 (16)
1		47 (63)	53 (71)
2 or more		9 (12)	10 (13)
Diagnosis:	150		
respiratory		51 (68)	47 (63)
medical		6 (8)	5 (7)
oncology		4 (5)	5 (7)
surgical		5 (7)	3 (4)
gastroenterology		1 (1)	5 (7)
haematology		0 (0)	1 (1)
other		8 (11)	9 (12)
Previous deep vein thrombosis	136	4 (6)	6 (9)
Leucocytes <1000/µl	112	6 (10)	12 (22)
Assistance to mobilise	150	7 (9)	4 (5)
Confused, agitated, or drowsy	150	3 (4)	4 (5)
Placement:	146	59 (80)	56 (78)
basilica		7 (9)	11 (15)
brachial			
cephalic		6 (8)	3 (4)
axilla		2 (3)	2 (3)
Subsequent insertion	138	40 (57)	36 (53)

Continued on page 521

the dwell but had a complication (ie, failure) at removal, and 4 (5.5%) had complications during the dwell and at removal.

More than twice as many standard care patients as BioFlo® patients had a complication during the PICC dwell, and this difference was statistically significant (25 of 74, 34% vs 11 of 72, 15%; P = .009; Table 2). These results are consistent with the Kaplan-Meier curve, which shows longer complication-free survival with BioFlo® (Figure 2A and 2B). The median BioFlo® dwell was 1 day longer (13.8 vs 12.9 days), and the median time to first complication was 1 day later (4.0 BioFlo® vs 3.0 standard care; Table 2).

As per supplementary Table 1, univariate Cox regression identified PICC failure as significantly associated with tip placement in the proximal superior vena cava (SVC) compared to the SVC–right atrium junction (HR 2.61; 95% CI, 1.17-5.82; P = .024). Reduced risk of PICC failure was significantly associated with any infusion during the dwell (continuous fluid infusion, P = .007; continuous antibiotic, P = .042; or intermittent infusion, P = .046) compared to no infusion. Other variables potentially influencing the risk of failure included PICC insertion by nurse specialist compared to consultant anesthetist (HR 2.61; 95% CI, 0.85-5.44) or registrar (HR 1.97; 95% CI, 0.57-6.77). These differences were not statistically significant; however, baseline imbalance between study groups for this variable and the feasibility design preclude absolute conclusions.

DISCUSSION

This is the first pilot feasibility trial of new PICC materials and valve design incorporated in the BioFlo® PICC in the pediatric population. The trial incorporated best practice for randomized trials, including using a concurrent control group, centralized and concealed randomization, predetermined feasibility criteria, and a registered and published trial protocol.[17] As in other studies,[15,24,34] PICC failure and complication prevalence was unacceptably high for this essential device. Standard care PICCs failed twice as often as the new BioFlo® PICCs (22% vs

TABLE 1. **Participant (n = 150) and PICC Characteristics (n = 144) (continued)**

		Standard Care	BioFlo®
Participant characteristics	**N**	**n = 75**	**n = 75**
Multiple insertion attempts:	146		
1 (success at first insertion)		60 (81)	65 (90)
2		11 (15)	5 (7)
3 or more		3 (4)	2 (3)
Ultrasound guidance used	146	74 (100)	72 (100)
Catheter tip location:	146	52 (70)	57 (79)
SVC–RA junction		22 (30)	15 (21)
SVC			
Vein size (mm)[b]	103	3.7 (0.9)	3.1 (0.9)
Catheter size:	146		
3 Fr		57 (77)	58 (81)
4 Fr		17 (23)	14 (19)
Catheter-to-vein ratio (%, n = 103)[d]	103	32.3 (27.6-35.0)	36.2 (30.9-43.7)
Treatment[c]:			
intermittent bolus medication	146	63 (84)	61 (85)
cont. antibiotic infusion	146	27 (36)	31 (43)
intermittent infusion	146	23 (31)	27 (38)
cont. nonantibiotic infusion	146	11 (15)	15 (21)
Thrombolytic treatment	146	4 (5)	5 (7)
Number of blood samples[b]	146	0.57 (0.50)	0.71 (0.70)
Infection at recruitment:			
respiratory	150	31 (41)	30 (40)
wound	150	2 (3)	7 (9)
osteomyelitis	150	3 (4)	2 (3)
positive BC (48 hours prior)	150	2 (3)	2 (3)
shunt, cerebrospinal fluid	150	1 (1)	1 (1)
urinary	150	1 (1)	0 (0)
other	150	4 (5)	9 (12)

[a] Mean, standard deviation, and range shown.

[b] Mean and standard deviation.

[c] Median, interquartile range (shown as 25th and 75th percentiles) and maximum shown.

[d] Ever received, multiple treatment types possible per patient.

NOTE: Frequencies (column percentages) are shown unless otherwise noted. Percentages are calculated with the number of nonmissing observations in the denominator. Abbreviations: BC, blood culture; cont, continuous; Fr, French gauge size; N, number of nonmissing observations; PICC, peripherally inserted central catheter; RA, right atrium; SVC, superior vena cava; tx, therapy; µl, microliter.

11%), which is a clinically important difference. As researchers in a pilot study, we did not expect to detect statistically significant differences; however, we found that overall complications during the dwell occurred significantly more with the standard care than BioFlo® PICCs ($P = .009$).

BioFlo® PICC material offers a major advancement in PICC material through the incorporation of AT technologies into catheter materials, such as PICCs. Endexo® is a low molecular–weight, fluoro-oligomeric additive that self-locates to the top few nanometers of the material surface. When added to power-injectable polyurethane, the additive results in a strong but passive, nonstick, fluorinated surface in the base PICC material. This inhibits platelet adhesion, suppresses protein procoagulant conformation, and thereby reduces thrombus formation in medical devices. Additionally, Endexo® is not a catheter coating; rather, it is incorporated within the polyurethane of the PICC, thereby ensuring these AT properties are present on the internal, external, and cut surfaces of the PICC. If this technology can reduce complication during treatment and reduce failure from infection, thrombosis, occlusion, fracture, and dislodgement, it will improve patient outcomes considerably and lower health system costs. Previous studies investigating valve

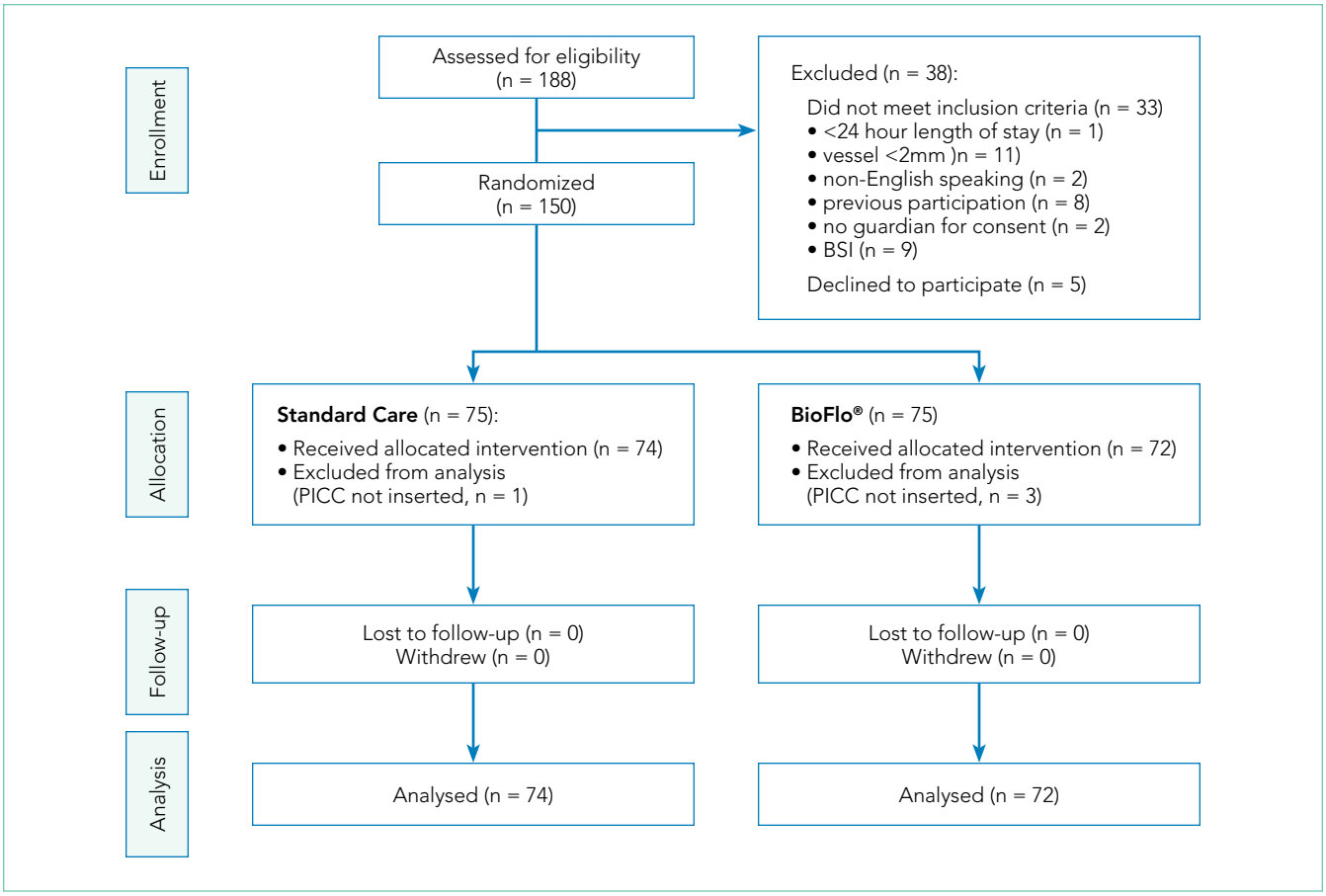

FIG 1. CONSORT Flowchart of study participants

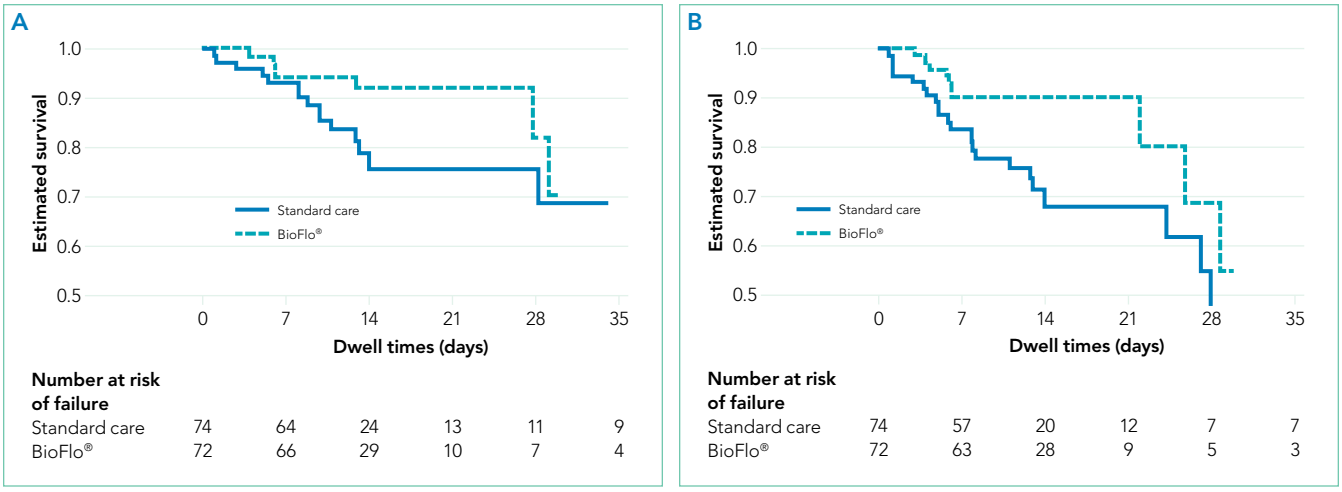

FIG 2. (A) Kaplain-Meier curves of PICC failure. (B) Kaplain-Meier curves of PICC complication

technology in PICC design to reduce occlusion have been inconclusive.[12-14,35,36] Occlusion (both partial and complete) was less frequent in our study with the BioFlo® group (n = 3; 4%) compared to the standard care group (n = 6; 8%). The results of this pilot study suggest that either the Endexo® material or PASV® technology has a positive association with occlusion reduction during PICC treatment.

Thrombosis was the primary failure type for the standard care PICCs, comprising one-third of failures. All but one patient with radiologically confirmed thrombosis required the removal of the PICC prior to completion of treatment. The decision to remove the PICC or retain and treat conservatively remained with the treating team. Raffini et al.[7] found thrombosis to increase in patients with one or more coexisting chronic medical condition.

TABLE 2. **Study Outcomes**

Outcomes	Standard Care n = 74	BioFlo® n = 72	*P* Value
Failure necessitating removal	16 (22)	8 (11)	.087[a]
Catheter-days	1268	1097	
IRR (per 1000 days, 95% CI)	Referent	0.58 (0.21-1.43)	.172[b]
Dwell time (days)[c]	12.9 (9.0-14.1; 104)	13.8 (10.0-17.3; 44)	
Complications resulting in failure[d]:			
CVAD-associated thrombosis	5 (7)	2 (3)	
occlusion, complete	5 (7)	1 (1)	
occlusion, partial	3 (4)	2 (3)	
CVAD breakage	3 (4)	0 (0)	
dislodgement, complete	0 (0)	2 (3)	
dislodgement, partial	1 (1)	0 (0)	
local infection, confirmed	0 (0)	1 (1)	
CVAD-associated BSI	0 (0)	0 (0)	
Complications during treatment[d]:			
any, per patient	25 (34)	11 (15)	.009[a]
occlusion, partial	11 (15)	5 (7)	
occlusion, complete	10 (14)	2 (3)	
thrombosis	6 (8)	2 (3)	
CVAD breakage	3 (4)	0 (0)	
dislodgement, complete	0 (0)	2 (3)	
dislodgement, partial	2 (3)	0 (0)	
local infection, confirmed	0 (0)	1 (1)	
CVAD-associated BSI	0 (0)	0 (0)	
Pain (at any time)	9 (12)	4 (6)	
Redness (at any time)	8 (11)	1 (1)	
Time to first complication (days, n = 36)[c,e]	7.8 (4.0-13; 92)	6.1 (4.2-26; 38)	
Ease of insertion (0 = worst, 100 = best)[f]	85.2 (16.0)	88.3 (14.9)	
Satisfaction with ins. kit (0 = lowest, 100 = highest)[f]	86.2 (12.2)	93.6 (7.7)	
Difficulty of removal (0 = worst, 100 = best)[f]	91.3 (5.0)	90.6 (7.1)	
Confidence in product (0 = lowest, 100 = highest)[f]	84.4 (14.2)	89.3 (7.2)	

[a] Chi-square test.

[b] Log-rank test.

[c] Median and interquartile range (25th and 75th percentiles) shown.

[d] More than 1 complication per patient possible.

[e] Maximum value shown.

[f] Mean (and standard deviation) shown.

NOTE: Frequencies (and column percentages) are shown unless otherwise noted. Percentages are calculated with the number of nonmissing observations in the denominator. Abbreviations: BSI, bloodstream infection; CI, confidence interval; CVAD, central venous access device; ins, insertion; IRR, incidence rate ratio; PICC, peripherally inserted central catheter.

Slightly more standard care than BioFlo® patients were free of such comorbidities (25% vs 16%), yet standard care patients still had the higher number of thromboses (7% vs 3%). Morgenthaler and Rodriguez[37] reported vascular access-associated thrombosis in pediatrics to be less common than in adults but higher in medically complex children. Worryingly, Menendez et al.[38] reported pediatric thrombosis to be largely asymptomatic, so the true incidence in our study is likely higher because only radiologically confirmed thromboses were recorded.

Occlusion (partial or complete) was the predominant complication across the study, being associated with one-third of all failures. When occlusion complications during the dwell (some of which were resolved with treatment), in addition to those causing failure, were considered, this number was even great-

er. Occlusion complications are prevalent and costly. Smith et al.[24] reported that occlusion was the most common reason for PICC removal and the most likely complication to delay treatment. Both the BioFlo® and standard care PICCs are pressure rated with good tensile strength; however, fracture occurred in 4% (n = 3) of standard care PICCs compared to no fractures in BioFlo® PICCs. Although the numbers are small, it may suggest a superior tensile strength of the BioFlo® material.

This study reinforces previously published results[24,38] that PICC tip position is important and can influence complications, such as occlusion and thrombosis. In addition, we found a significant association with failure when PICCs did not have a continuous infusion. These findings reinforce the need for optimal tip location at insertion and ongoing flushing and maintenance of PICCs not used for infusions.

Limitations of this study include the small sample size, which was not designed to detect statistical differences in the primary outcome between groups. Despite randomization, there were slight imbalances at baseline for inserter type and leukocyte count, although these were not significantly associated with PICC failure in the Cox regression (data not shown), and thus were unlikely to influence findings. Additionally, a difference of <10% was associated with PICC tip position, favoring the BioFlo® group. PICC tip position outside the cavoatrial junction was positively associated with failure; therefore, the effect of tip positioning on outcomes is difficult to ascertain given the small sample size and feasibility nature of the study. Further study is warranted to further explore this effect. The population sampled was pediatric medical and surgical inpatients with a vessel size >2 mm attending the operating theater suite for PICC insertion, thereby limiting the study's generalizability to adults and other populations, including neonates and those with PICCs inserted in the pediatric intensive care unit. The study could not be blinded because study products had to be visible to the clinical and research staff. However, it is unlikely that staff would intentionally sabotage PICCs to bias the study. Blinding was possible for the assessment of blood culture and ultrasound reports to diagnose infection and thrombosis. Strengths of this study included 100% protocol adherence, and no patients were lost to follow-up.

CONCLUSION

These results confirm that PICC failure is unacceptably high and suggest that the innovative BioFlo® PICC material and design holds promise to improve PICC outcomes by reducing complications and overall PICC failure. Trials of this technology are feasible, safe, and acceptable to healthcare staff and parents. Further trials are required, including in other patient populations, to definitively identify clinical, cost-effective methods to prevent PICC failure and improve reliability during treatment.

Acknowledgments

The authors thank the children and parents of Lady Cilento Children's Hospital for participating in this important research. A special thank you goes to the nurses within the Vascular Assessment and Management Service and to Karen Turner, Julieta Woosley, and Anna Dean for their efforts in data collecting and ensuring protocol adherence.

Disclosure: Griffith University has received unrestricted, investigator-initiated research or educational grants to support the research of T.K., A.J.U., and C.R.M. from product manufacturers 3M, Adhezion Inc, AngioDynamics, Bard Medical, Baxter, B. Braun Medical Inc, Becton Dickinson, CareFusion, Centurion Medical Products, Cook Medical, Entrotech, FloMedical, ICU Medical Inc, Medical Australia Limited, Medtronic, Smiths Medical, and Teleflex. Griffith University has received consultancy payments on behalf of C.R.M., A.J.U., and T.K. from manufacturers 3M, AngioDynamics, Bard Medical, B. Braun Medical Inc, Becton Dickinson, CareFusion, Mayo Healthcare Inc, ResQDevices, and Smiths Medical. AngioDynamics (the BioFlo® PICC manufacturer) provided partial funds to undertake this research via an unrestricted donation to Griffith University (but not the study authors). Queensland Health provided in-kind support to fund the remainder of the trial. The funders had no role in the study design, collection, analysis, or interpretation of the data, writing of the report, or decision to submit the article for publication.

References

1. Chopra V, Flanders SA, Saint S. The problem with peripherally inserted central catheters. *JAMA*. 2012;308(15):1527-1528.
2. Gibson C, Connolly BL, Moineddin R, Mahant S, Filipescu D, Amaral JG. Peripherally inserted central catheters: use at a tertiary care pediatric center. *J Vasc Interv Radiol*. 2013;24(9):1323-1331.
3. Ullman AJ, Cooke M, Kleidon T, Rickard CM. Road map for improvement: point prevalence audit and survey of central venous access devices in paediatric acute care. *J Paediatr Child Health*. 2017;53(2):123-130.
4. Ullman AJ, Marsh N, Mihala G, Cooke M, Rickard CM. Complications of central venous access devices: a systematic review. *Pediatrics*. 2015;136(5):e1331-e1344.
5. Barrier A, Williams DJ, Connelly M, Creech CB. Frequency of peripherally inserted central catheter complications in children. *Pediatr Infect Dis J*. 2012;31(5):519-521.
6. Jumani K, Advani S, Reich NG, Gosey L, Milstone AM. Risk factors for peripherally inserted central venous catheter complications in children. *JAMA Pediatr*. 2013;167(5):429-435.
7. Raffini L, Huang YS, Witmer C, Feudtner C. Dramatic increase in venous thromboembolism in children's hospitals in the United States from 2001 to 2007. *Pediatrics*. 2009;124(4):1001-1008.
8. Chopra V, Anand S, Krein SL, Chenoweth C, Saint S. Bloodstream infection, venous thrombosis, and peripherally inserted central catheters: reappraising the evidence. *Am J Med*. 2012;125(8):733-741.
9. Moureau NL, Trick N, Nifong T, et al. Vessel health and preservation (part 1): a new evidence-based approach to vascular access selection and management. *J Vasc Access*. 2012;13(3):351-356.
10. Poli P, Scocca A, Di Puccio F, Gallone G, Angelini L, Calabro EM. A comparative study on the mechanical behavior of polyurethane PICCs. *J Vasc Access*. 2016;17(2):175-181.
11. Interface Biologics. Surface modification technology platform. 2017. http://www.interfacebiologics.com/endexo.htm. Accessed April 5, 2017.
12. Hoffer EK, Bloch RD, Borsa JJ, Santulli P, Fontaine AB, Francoeur N. Peripherally inserted central catheters with distal versus proximal valves: prospective randomized trial. *J Vasc Interv Radiol*. 2001;12(10):1173-1177.
13. Hoffer EK, Borsa J, Santulli P, Bloch R, Fontaine AB. Prospective randomized comparison of valved versus nonvalved peripherally inserted central vein catheters. *AJR Am J Roentgenol*. 1999;173(5):1393-1398.
14. Pittiruti M, Emoli A, Porta P, Marche B, DeAngelis R, Scoppettuolo G. A prospective, randomized comparison of three different types of valved and nonvalved peripherally inserted central catheters. *J Vasc Access*. 2014;15(6):519-523.
15. Chopra V, Flanders SA, Saint S, et al. The Michigan Appropriateness Guide for Intravenous Catheters (MAGIC): Results From a Multispecialty Panel Using the RAND/UCLA Appropriateness Method. *Ann Intern Med*. 2015;163(6 Suppl):S1-S40.
16. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ*. 2008;337.
17. Kleidon TM, Ullman AJ, Zhang L, Mihala G, Rickard CM. How does your PIC-COMPARE? A pilot randomized controlled trial comparing PICC materials in pediatrics. *J Hosp Med*. 2017;(under review).
18. Hertzog MA. Considerations in determining sample size for pilot studies. *Res Nurs Health*. 2008;31(2):180-191.
19. Thabane L, Ma J, Chu R, et al. A tutorial on pilot studies: the what, why and

how. *BMC Med Res Methodol*. 2010;10:1.
20. Chopra V, O'Horo JC, Rogers MA, Maki DG, Safdar N. The risk of bloodstream infection associated with peripherally inserted central catheters compared with central venous catheters in adults: a systematic review and meta-analysis. *Infect Control Hosp Epidemiol*. 2013;34(9):908-918.
21. Kramer RD, Rogers MA, Conte M, Mann J, Saint S, Chopra V. Are antimicrobial peripherally inserted central catheters associated with reduction in central line-associated bloodstream infection? A systematic review and meta-analysis. *Am J Infect Control*. 2017;45(2):108-114.
22. Centers for Disease Control and Prevention. National Healthcare Safety Network Device Associated Module: CLABSI. 2014.
23. Lobo BL, Vaidean G, Broyles J, Reaves AB, Shorr RI. Risk of venous thromboembolism in hospitalized patients with peripherally inserted central catheters. *J Hosp Med*. 2009;4(7):417-422.
24. Smith SN, Moureau N, Vaughn VM, et al. Patterns and Predictors of Peripherally Inserted Central Catheter Occlusion: The 3P-O Study. *J Vasc Interv Radiol*. 28(5):749.e742-756.e742.
25. Chow LML, Friedman JN, MacArthur C, et al. Peripherally inserted central catheter (PICC) fracture and embolozation in the pediatric population. *Pediatrics*. 2003;142(2):141-144.
26. Chopra V, Kuhn L, Ratz D, Flanders SA, Krein SL. Vascular nursing experience, practice knowledge, and beliefs: Results from the Michigan PICC1 survey. *J Hosp Med*. 2016;11(4):269-275.
27. Frasca D, Dahyot-Fizelier C, Mimoz O. Prevention of central venous catheter-related infection in the intensive care unit. *Crit Care*. 2010;14(2):212.
28. Centre for Healthcare Related Infection Surveilance and Prevention and Tuberculosis Control. Guideline: Peripherally inserted central catheter (PICC). 2013.
29. Services Children's Health Service. Central venous catheters: nursing care and management of peripherally inserted central catheter (PICC) in paediatric patients. 2011. http://qheps.health.qld.gov.au/childrenshealth/resources/nursestand/docs/ns_03452.pdf. Accessed Februrary 1, 2016.
30. Services CsH. Central Venous Access Device Insertion and Management. 2014.
31. Central venous access device insertion and management. Queensland Government; 2014. http://qheps.health.qld.gov.au/childrenshealth/resources/proc/docs/proc_03450.pdf Accessed March 13, 2014.
32. StatCorp. Stata Statistical Software: Release 12.1 College Station. 2006.
33. Rothwell PM. Factors that can affect the external validity of randomised controlled trials. *PLoS Clin Trials*. 2006;1(1):e9.
34. Chopra V, Ratz D, Kuhn L, Lopus T, Lee A, Krein S. Peripherally inserted central catheter-related deep vein thrombosis: contemporary patterns and predictors. *J Thromb Haemost*. 2014;12(6):847-854.
35. Alport B, Burbridge B, Lim H. Bard PowerPICC Solo2 vs Cook Turbo-Ject: A Tale of Two PICCs. *Can Assoc Radiol J*. 2012;63(4):323-328.
36. Johnston AJ, Streater CT, Noorani R, Crofts JL, Del Mundo AB, Parker RA. The effect of peripherally inserted central catheter (PICC) valve technology on catheter occlusion rates—the 'ELeCTRiC' study. *J Vasc Access*. 2012;13(4):421-425.
37. Morgenthaler TI, Rodriguez V. Preventing acute care-associated venous thromboembolism in adult and pediatric patients across a large healthcare system. *J Hosp Med*. 2016;11(Suppl 2):S15-S21.
38. Menendez JJ, Verdu C, Calderon B, et al. Incidence and risk factors of superficial and deep vein thrombosis associated with peripherally inserted central catheters in children. *J Thromb Haemost*. 2016;14(11):2158-2168.

A Single, Post-ACTH Cortisol Measurement to Screen for Adrenal Insufficiency in the Hospitalized Patient

Devin C. Odom, MD[1]*, Ann M. Gronowski, PhD[2], Elizabeth Odom, MD, MPHS[3], William Clutter, MD[4], Mark Thoelke, MD, SFHM[1]

[1]Division of Hospital Medicine, School of Medicine, Washington University, St. Louis, Missouri; [2]Division of Laboratory and Genomic Medicine, School of Medicine, Washington University, St. Louis, Missouri; [3]Department of Emergency Medicine, School of Medicine, Washington University, St. Louis, Missouri; [4]Division of Endocrinology, Metabolism, and Lipid Research, School of Medicine, Washington University, St. Louis, Missouri.

BACKGROUND: Cosyntropin stimulation testing (CST) is used to screen patients for adrenal insufficiency (AI). Traditionally, CST includes baseline cortisol concentration, the administration of cosyntropin, and cortisol concentration at 30 and 60 minutes poststimulation. There is debate surrounding the utility of testing and cut-off points for concentrations at each time point.

OBJECTIVE: To determine if a single cortisol measurement at 30 or 60 minutes could replace the traditional approach.

DESIGN: We looked retrospectively at inpatients who underwent standard, high-dose CST (n = 702) and evaluated the number of patients who would screen positive for AI by using a single time point (30 or 60 minutes) compared with the traditional CST.

SETTING: A tertiary-care, academic medical center.

PATIENTS: Hospital inpatients present between January 2012 and September 2013.

RESULTS: Of tests, 84.3% were normal, which was defined as at least one cortisol concentration of 18 mcg/dL or higher at any time after stimulation. The average 60-minute concentration was higher than the average 30-minute concentration ($P < .001$). A single 60-minute concentration is 100% concordant with the full CST in the intensive care unit (ICU) subgroup and 99.6% concordant in floor patients. A single 30-minute concentration is significantly less concordant, 91.9% and 86.9%, in the ICU and floor subgroups, respectively.

CONCLUSIONS: Overall, a single 60-minute cortisol concentration to screen for AI was 99.7% concordant with the traditional CST, and the positive percent agreement was 98%. Fewer false-positive screens would occur with a single 60-minute cortisol concentration compared with a single 30-minute concentration ($P < .001$). High-dose CST screening may safely be interpreted with single 60-minute poststimulation cortisol serum concentrations. *Journal of Hospital Medicine* 2018;13:526-530. Published online first February 8, 2018. © 2018 Society of Hospital Medicine.

Testing for adrenal insufficiency (AI) is common in the hospital setting. The gold standard remains the insulin tolerance test (ITT), in which cortisol concentration is measured after the induction of hypoglycemia to <35 mg/dL.[1] Alternatively, metyrapone testing works by blocking cortisol synthesis. If pretest adrenocorticotropic hormone (ACTH) concentrations are low and ACTH concentrations do not rise after the administration of metyrapone, the patient is given a diagnosis of AI. Both assays pose some risk to patients with AI and are typically only performed as confirmatory tests. Morning random cortisol concentrations can be used to suggest AI if concentrations are <3 mcg/dL, but they often provide indeterminate results if concentrations are between 3 and 15 mcg/dL.[2] Thus, morning cortisol concentrations in isolation are not diagnostic of AI. For these reasons, most experts recommend a dynamic, high-dose cosyntropin stimulation testing (CST) with 250 mcg of intravenous cosyntropin to screen for AI. The test can be done any time of day.[3] Historically, an incremental response to cosyntropin, or "delta," was also required to indicate a normal response to stimulation.[4] However, the baseline cortisol concentration is dependent on circadian rhythm and level of stress. For this reason, a delta, whether large or small, has been abandoned as a requisite for the diagnosis of AI.[5-7] A normal CST is widely accepted to be identified by any cortisol concentration >18 mcg/dL during the test (basal or poststimulation).[8]

The seminal studies by Lindholm, Kehlet, and coauthors[9-11] validated the CST against the gold standard ITT and utilized only 0- and 30-minute cortisol concentrations. A later study in patients with pituitary disease demonstrated that 30-minute concentrations had a stronger correlation with the ITT than 60-minute concentrations (false-negative rate: 10% vs 27%).[12] However, in that study, a higher threshold was used for the 60-minute concentration than for what was obtained at 30 minutes (25.4 vs 21.8 mcg/dL, respectively). Multiple studies have shown that the 60-minute concentration is higher than the 30-minute concentration after cosyntropin stimulation.[4,5,13]

***Address for correspondence:** Devin C. Odom, MD, Campus Box 8058, 660 S. Euclid Ave., St. Louis, MO, 63109; Telephone: 314-362-1700; Fax: 314-362-7898; E-mail: dodom@wustl.edu

Received: July 28, 2017; **Revised**: October 28, 2017; **Accepted**: October 30, 2017

Subsequent, small studies of patients who were known to have AI have shown that 60-minute concentrations are as useful as 30-minute concentrations.[5,14,15] Because 30-minute cortisol concentrations are often lower than 60-minute concentrations, a single 30-minute result may lead to a falsely abnormal test.[16,17] As such, the use of a single 60-minute test may be more appropriate. Indeed, some authors have suggested that measuring only 30-minute concentrations may lead to overdiagnosis of AI by missing an appropriate response, serum cortisol >18 mcg/dL, at 60 minutes.[17-19] Peak cortisol concentrations after low-dose cosyntropin stimulation (1 mcg) are seen at 60 minutes, and low-dose stimulation has been shown to be more variable than in the high-dose test (250 mg).[19,20]

There is a lack of consensus to guide clinicians as to when cortisol concentrations should be measured after stimulation, and standard references lack uniformity. Commonly accessed medical resources – such as *UpToDate* and Jameson's *Endocrinology* – recommend basal, 30-minute, and 60-minute cortisol concentrations, while *Williams Textbook of Endocrinology* recommends basal and 30-minute concentrations, and the *Washington Manual* recommends only a single 30-minute concentration.[7,21,22] *Goldman-Cecil Medicine*[8] recommends checking a cortisol concentration between 30 and 60 minutes and recommends the same 18 mcg/dL cutoff for any test obtained in this time period. As a result of these variable recommendations, all 3 time points are often obtained. Prominent review articles continue to recommend checking all three concentrations while presenting evidence of peak cortisol response at 60 minutes poststimulation.[13]

In this study, we retrospectively examined CSTs in hospitalized, adult patients both in the intensive care unit (ICU) and hospital ward and/or floor settings to evaluate for significant differences in 30- and 60-minute cortisol concentrations and compare the concordance of screening at each time point alone with traditional CST at all 3 time points. By using these results, we discuss the utility of obtaining 3 cortisol samples.

METHODS

After receiving approval from the institutional review board, we retrospectively reviewed all standard, high-dose CSTs performed on adult inpatients at the Barnes-Jewish Hospital laboratory from January 1, 2012, to August 31, 2013. All patients received the same standard dose (250 mcg cosyntropin, a synthetic ACTH, at a concentration of 1 mcg/mL administered over 2 minutes) regardless of age or weight. We collected patient gender; age; time of baseline cortisol measurement; cortisol results at baseline, 30, and 60 minutes; and patient location (inpatient floor vs ICU status). Tests were included if results from all 3 time points (0, 30, 60 minute) were available.

Cortisol concentrations were assessed by the laboratory according to the manufacturer's instructions by using the ADVIA Centaur Cortisol assay (Siemens Healthcare Diagnostics Inc, Tarrytown, NY), a competitive chemiluminescent immunoassay. For the traditional CST, a cortisol concentration ≥18 mcg/dL at any time point during the test was used to define normal (negative). Patients with a positive (no results >18 mcg/mL) CST were defined as "screen positives" for the purposes of this analysis. Patient location data were available that allowed for an ICU vs non-ICU comparison.

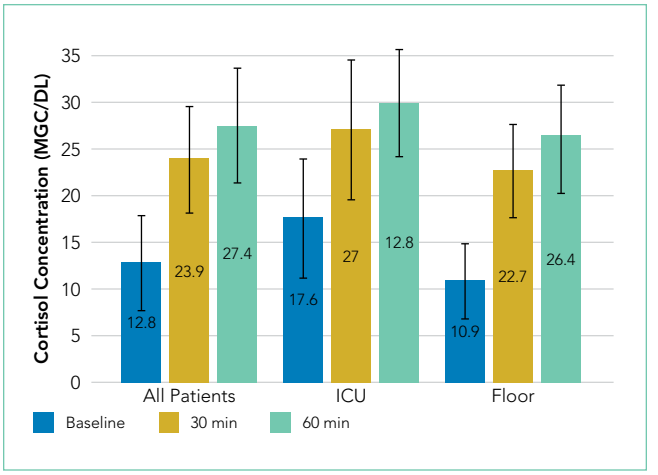

FIG. Cortisol concentration by patient location. Cortisol concentrations for 702 inpatient CSTs are shown in separate ICU and floor subgroups. Baseline (blue), 30-minute post-adrenocorticotrophic hormone (ACTH; gold), and 60-minute post-ACTH (green) cortisol concentration averages are shown for each subgroup. Differences in cortisol concentrations at each time point are significant ($P < .001$) in both subgroups in all cases.
NOTE: Abbreviations: ACTH, adrenocorticotropic hormone; CST, cosyntropin stimulation testing; ICU, intensive care unit.

Statistical analyses were performed in SAS version 9.4 (SAS Institute Inc, Cary, North Carolina). Continuous variables were compared by using a 2-tailed Student t test. Percentiles and proportions were compared by using χ^2 tests or Fisher's exact tests when appropriate. The concordance of screening at each time point compared with the traditional CST was calculated. Positive percent agreement (PPA) with the traditional CST in each subgroup (ICU and floor) and combined was also evaluated. A *P* value of .05 was used to determine significance.

RESULTS

A total of 702 complete cosyntropin tests on separate patients were included in the analysis. This included 198 ICU patients and 504 non-ICU (floor) patients. Fifty-one percent of patients were male in both the floor and ICU subgroups. The average age of ICU patients was 60.2 ± 13.2 years compared to 57.3 ± 17.3 years for patients on a general medicine floor ($P = .02$).

Cortisol concentrations obtained at 30 minutes were significantly higher than baseline cortisol concentrations (baseline: 12.8 mcg/dL; 30 minutes: 23.9 mcg/dL; $P < .001$) for all patients. The average cortisol concentrations obtained at 60 minutes (27.4 mcg/dL) were significantly higher than those at baseline and 30 minutes ($P < .001$). This trend was seen in each subgroup of patients in the ICU and on the floor (Figure). The average baseline cortisol concentration was higher for ICU patients compared to floor patients (17.6 mcg/dL vs 10.9 mcg/dL, respectively).

By using the traditional CST, there were 26 (13.1%) positive tests for AI in ICU patients and 84 (16.7%) positive tests in floor patients (Table).

The Table shows the number of patients who screened positive at each time point and compares the concordance of these

TABLE. **Concordance of 30- and 60-Minute Poststimulation Cortisol Concentrations with Traditional CST**

ICU		Traditional CST +	Traditional CST -	Total
30 minutes only	+	26	16	42
	-	0	156	156
	Total	26	172	198
		Concordance = 91.9% PPA = 62%		

ICU		Traditional CST +	Traditional CST -	Total
60 minutes only	+	26	0	26
	-	0	172	172
	Total	26	172	198
		Concordance = 100% PPA = 100%		

Floor		Traditional CST +	Traditional CST -	Total
30 minutes only	+	84	66	150
	-	0	354	354
	Total	84	420	504
		Concordance = 86.9% PPA = 56%		

Floor		Traditional CST +	Traditional CST -	Total
60 minutes only	+	84	2	86
	-	0	418	418
	Total	84	420	504
		Concordance = 99.6% PPA = 98%		

All		Traditional CST +	Traditional CST -	Total
30 minutes only	+	110	82	192
	-	0	510	510
	Total	110	592	702
		Concordance = 88.0% PPA = 57%		

All		Traditional CST +	Traditional CST -	Total
60 minutes only	+	110	2	112
	-	0	590	590
	Total	110	592	702
		Concordance = 99.7% PPA = 98%		

NOTE: A total of 702 traditional CSTs were analyzed. Because illness acuity can directly impact cortisol concentrations, results are subdivided into ICU and general floor patients. The traditional CST was considered positive for AI if the cortisol concentrations were <18 mcg/dL at both time points (30 and 60 minutes). This is considered screen positive. The traditional CST was considered negative for AI if the cortisol concentrations were >18 mcg/dL at any time point. This is considered screen negative. Concordance and Percent Positive Agreement resultsare in bold. The difference between the 30-minute and 60-minute results are significant ($P < .001$) in all groups. Abbreviations: AI, adrenal insufficiency; CST, cosyntropin stimulation test; ICU, intensive care unit; PPA, positive percent agreement.

results with the results of the overall CST in each subgroup (ICU and floor). The 60-minute concentration demonstrated higher concordance with the traditional CST than the 30-minute concentration overall (99.7% vs 88.0%, respectively), in ICU patients (100% vs 91.9%, respectively), and in floor patients (99.6% vs 86.9%, respectively). In the ICU subgroup, 60-minute concentrations were 100% concordant with the traditional CSTs. The PPA of a 60-minute–only screening compared to a traditional CST was better than a 30-minute–only screening overall (98% vs 57%, respectively), in ICU patients (100% vs 62%, respectively), and in floor patients (98% vs 56%, respectively). A 60-minute concentration was required to prevent false-positive screening in 11.7% of all screening tests, but the 30-minute concentration only prevented false-positive screening in 0.3% of screening tests. Of all 30-minute concentrations screening positive for AI alone, 42.7% were negative for AI at 60 minutes. Conversely, only 1.8% of all 60-minute concentrations screening positive for AI alone were negative for AI at 30 minutes. The likelihood of a false-positive screening test at 30 minutes was higher in floor patients (13.1%) than in ICU patients (8.1%). The difference between the false-positive screening rate of a single 30-minute cortisol concentration and a single 60-minute concentration was significant ($P < .0001$) for both floor and ICU patients. There were no instances of basal cortisol concentrations >18 mcg/dL that were subsequently <18 mcg/dL at 30 and 60 minutes after cosyntropin stimulation.

Only 13% of CSTs were started in the recommended 3-hour window from 6:00 AM to 8:59 AM. The remaining tests were begun outside this window.

DISCUSSION

Our investigation of 702 CSTs, the largest retrospective analysis to date, finds that the 60-minute cortisol concentration is significantly higher than the 30-minute concentration in a standard, high-dose CST. Sixty-minute cortisol concentrations are more concordant with traditional CST results than the 30-minute concentrations in both critically ill ICU and noncritically ill floor patients. This suggests that a single 60-minute measurement is sufficient for AI screening. The use of only 30-minute concentrations would lead to a significant increase in false-positive screening tests and significantly lower PPA (98% vs 57%). With peak cortisol concentrations occurring at

60-minutes poststimulation, measuring both 30- and 60-minute poststimulation concentrations does not appear to be of significant clinical benefit. The cost-saving from reduced phlebotomy and laboratory expenses would be significant, especially in locations with limited staff or financial resources. Our findings are similar to other recent results by Chitale et al.,[17] Mansoor et al.,[16] and Zueger et al.[18]

Zueger et al.[18] evaluated the results of high-dose CST in 73 patients and found 13.7% of patients with inadequate cortisol response (<18 mcg/dL) at 30 minutes had normal concentrations at 60 minutes (>18 mcg/dL). Their study did not identify a single case of normal cortisol concentration at 30 minutes that would have inappropriately screened positive for AI if cortisol concentrations were only checked at 60 minutes. Similarly, they suggested that the 30-minute test did not add any additional diagnostic value; however, no confirmatory testing was performed.

Higher cortisol concentrations at 60 minutes poststimulation may result in improved specificity for AI without reducing sensitivity, but it may also indicate that the cutoff value may need to be raised from 18 mcg/dL at 60 minutes to maintain an appropriate clinical sensitivity. Continued research should resolve this clinical question with gold-standard confirmatory testing. Furthermore, there is debate about an appropriate screening cortisol concentration threshold for critically ill patients. Researchers have compared concentrations of 25 mcg/dL to the traditional 18 mcg/dL to improve sensitivity for AI, but these studies do not involve comparisons to confirmatory testing and often result in reduced specificity.[23,24]

In our study, only a small fraction of testing was performed in the early-morning hours, when basal cortisol results are of value. There may be indications to perform traditional CSTs with a basal concentration, such as for suspected secondary AI, but testing must be performed in the early morning for interpretable results per current recommendations. However, poststimulation cortisol concentrations may be interpreted regardless of the time of day at which the test was initiated.[3]

Our study is limited by its scope because it is a retrospective analysis. It is also limited by a lack of gold-standard, clinical confirmatory testing or analysis of other clinical data. Our method of testing and interpretation is considered the screening standard and is often used to plan treatment for AI without confirmatory testing, as ITT is not routinely available for hospitalized patients. The validation of the traditional CST to the ITT has been performed extensively, but a randomized trial comparing a single 60-minute concentration to the ITT may be useful. The exact timing of blood draws may have introduced error in the concentration measurements, and this is critical to screening accuracy. Total serum cortisol is 10% bound to albumin,[25] and medications such as steroids or opioids and medical conditions such as obesity or liver disease can affect cortisol concentrations.[26] Albumin and free cortisol concentrations that may be used to adjust for these variables were not available.

CONCLUSION

We recommend changes to the standard CST to exclude a basal cortisol concentration unless it is indicated for the evaluation of secondary AI or obtained at the appropriate early-morning hour. A single 60-minute poststimulation cortisol concentration may be an appropriate screening test for AI and demonstrates high concordance with the traditional CST. The use of a 30-minute poststimulation concentration alone may lead to a significantly higher number of false-positive results. Alternatively, the stimulated cortisol threshold used to define a normal test may need to be higher at 60 minutes to maintain the appropriate sensitivity. Further study and comparison with confirmatory testing are needed.

Disclosure: The authors have no relevant conflicts of interest to disclose.

References

1. Ajala O, Lockett H, Twine G, Flanagan DE. Depth and duration of hypoglycaemia achieved during the insulin tolerance test. *Eur J Endocrinol*. 2012;167(1):59-65.
2. Erturk E, Jaffe CA, Barkan AL. Evaluation of the integrity of the hypothalamic-pituitary-adrenal axis by insulin hypoglycemia test. *J Clin Endocrinol Metab*. 1998;83(7):2350-2354.
3. Azziz R, Bradley E Jr, Huth J, et al. Acute adrenocorticotropin-(1-24) (ACTH) adrenal stimulation in eumenorrheic women: reproducibility and effect of ACTH dose, subject weight, and sampling time. *J Clin Endocrinol Metab*. 1990;70(5):1273-1279.
4. Wood, JB, Frankland AW, James VH, Landon J. A Rapid Test of Adrenocortical Function. *Lancet*. 1965;1(7379):243-245.
5. Speckart PF, Nicoloff JT, Bethune JE. Screening for adrenocortical insufficiency with cosyntropin (synthetic ACTH). *Arch Intern Med*. 1971;128(5):761-763.
6. Grinspoon SK, Biller BM. Clinical review 62: Laboratory assessment of adrenal insufficiency. *J Clin Endocrinol Metab*. 1994;79(4):923-931.
7. Melmed S, Polonksy K, Larsen PR, Kronenberg H. *Williams Textbook of Endocrinology*. 13th ed. Elsevier: Amsterdam, Netherlands; 2016.
8. Nieman LK. *Adrenal Cortex*, in *Goldman-Cecil Medicine*. ed. L. Goldman. 2016, Elsevier: Amsterdam, Netherlands; 2016:1514-1521.
9. Kehlet H, Blichert-Toft M, Lindholm J, Rasmussen P. Short ACTH test in assessing hypothalamic-pituitary-adrenocortical function. *Br Med J*. 1976;1(6004):249-251.
10. Lindholm J, Kehlet H. Re-evaluation of the clinical value of the 30 min ACTH test in assessing the hypothalamic-pituitary-adrenocortical function. *Clin Endocrinol (Oxf)*. 1987;26(1):53-59.
11. Lindholm J, Kehlet H, Blichert-Toft M, Dinesen B, Riishede J. Reliability of the 30-minute ACTH test in assessing hypothalamic-pituitary-adrenal function. *J Clin Endocrinol Metab*. 1978;47(2):272-274.
12. Hurel SJ, Thompson CJ, Watson MJ, Harris MM, Baylis PH, Kendall-Taylor P. The short Synacthen and insulin stress tests in the assessment of the hypothalamic-pituitary-adrenal axis. *Clin Endocrinol (Oxf)*. 1996;44(2):141-146.
13. Dorin RI, Qualls CR, Crapo LM. Diagnosis of adrenal insufficiency. *Ann Intern Med*. 2003;139(3):194-204.
14. Oelkers W, Diederich S, Bahr V. Diagnosis and therapy surveillance in Addison's disease: rapid adrenocorticotropin (ACTH) test and measurement of plasma ACTH, renin activity, and aldosterone. *J Clin Endocrinol Metab*. 1992;75(1):259-264.
15. Gonzalez-Gonzalez JG, De la Garza-Hernandez NE, Mancillas-Adame LG, Montes-Villarreal J, Villarreal-Perez JZ. A high-sensitivity test in the assessment of adrenocortical insufficiency: 10 microg vs 250 microg cosyntropin dose assessment of adrenocortical insufficiency. *J Endocrinol*. 1998;159(2):275-280.
16. Mansoor S, Islam N, Siddiqui I, Jabbar A. Sixty-minute post-Synacthen serum cortisol level: a reliable and cost-effective screening test for excluding adrenal insufficiency compared to the conventional short Synacthen test. *Singapore Med J*. 2007;48(6):519-523.
17. Chitale A, Musonda P, McGregor AM, Dhatariya KK. Determining the utility of the 60 min cortisol measurement in the short synacthen test. *Clin Endocrinol (Oxf)*. 2013;79(1):14-19.
18. Zueger T, Jordi M, Laimer M, Stettler C. Utility of 30 and 60 minute cortisol samples after high-dose synthetic ACTH-1-24 injection in the diagnosis of adrenal insufficiency. *Swiss Med Wkly*. 2014;144:w13987.
19. Cartaya J, Misra M. The low-dose ACTH stimulation test: is 30 minutes long enough? *Endocr Pract*. 2015;21(5):508-513.

20. Gonzálbez, Villabona, Ramon, et al. Establishment of reference values for standard dose short synacthen test (250 µg), low dose short synacthen test (1 µg) and insulin tolerance test for assessment of the hypothalamo–pituitary–adrenal axis in normal subjects. *Clin Endocrinol.* 2000;53(2):199-204.
21. McGill J, Clutter W, Baranski T. *The Washington Manual of Endocrinology Subspecialty Consult.* 3rd ed. *Washington Manual*, ed. Henderson K, De Fer T. Lippincott Williams and Wilkins: Philadelphia, PA; 2012:384.
22. Nieman L. Diagnosis of adrenal insufficiency in adults. In *UpToDate*, ed. Post T. Wolters Klewer: Waltham, MA; 2017.
23. Marik PE, Kiminyo K, Zaloga GP. Adrenal insufficiency in critically ill patients with human immunodeficiency virus. *Crit Care Med.* 2002;30(6):1267-1273.
24. Marik PE, Zaloga GP. Adrenal insufficiency during septic shock. *Crit Care Med.* 2003;31(1):141-145.
25. Lewis JG, Bagley CJ, Elder PA, Bachmann AW, Torpy DJ. Plasma free cortisol fraction reflects levels of functioning corticosteroid-binding globulin. *Clinica Chemica Acta.* 2005;359(1-2):189-194.
26. Torpy DJ, Ho JT. Value of Free Cortisol Measurement in Systemic Infection. *Horm Metab Res.* 2007;39(6):439-444.

Decrease in Inpatient Telemetry Utilization Through a System-Wide Electronic Health Record Change and a Multifaceted Hospitalist Intervention

Karli Edholm, MD[1]*, Polina Kukhareva, MS, MPH[2], Claire Ciarkowski, MD[1], Jason Carr, MD[3], David Gill, MD[3], Austin Rupp, MD[3], Jack Morshedzadeh, MD[4], Nathan Wanner, MD[1], Kensaku Kawamoto, MD, PhD, MHS[2]

[1]Division of General Internal Medicine, Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, Utah; [2]Department of Biomedical Informatics, University of Utah Health, Salt Lake City, Utah; [3]Department of Medicine, University of Utah School of Medicine, Salt Lake City, Utah; [4]Division of Cardiovascular Medicine, Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, Utah.

BACKGROUND: Unnecessary telemetry monitoring contributes to healthcare waste.

OBJECTIVE: To evaluate the impact of 2 interventions to reduce telemetry utilization.

DESIGN, SETTING, AND PATIENTS: A 2-group retrospective, observational pre- to postintervention study of F35,871 nonintensive care unit (ICU) patients admitted to 1 academic medical center.

INTERVENTION: On the hospitalist service, we implemented a telemetry reduction intervention including education, process change, routine feedback, and a financial incentive between January 2015 and June 2015. In July 2015, a system-wide change to the telemetry ordering process was introduced.

MEASUREMENTS: The primary outcome was telemetry utilization, measured as the percentage of daily room charges for telemetry. Secondary outcomes were mortality, escalation of care, code event rate, and appropriateness of telemetry utilization. Generalized linear models were used to evaluate changes in outcomes while adjusting for patient factors.

RESULTS: Among hospitalist service patients, telemetry utilization was reduced by 69% (95% confidence interval [CI], −72% to −64%; $P < .001$), whereas on other services the reduction was a less marked 22% (95% CI, −27% to −16%; $P < .001$). There were no significant increases in mortality, code event rates, or care escalation, and there was a trend toward improved utilization appropriateness.

CONCLUSION: Although electronic telemetry ordering changes can produce decreases in hospital-wide telemetry monitoring, a multifaceted intervention may lead to an even larger decline in utilization rates. Whether these changes are durable cannot be ascertained from our study. *Journal of Hospital Medicine* 2018;13:531-536. Published online first February 9, 2018. © 2018 Society of Hospital Medicine

Wasteful care may account for between 21% and 34% of the United States' $3.2 trillion in annual healthcare expenditures, making it a prime target for cost-saving initiatives.[1,2] Telemetry is a target for value improvement strategies because telemetry is overutilized, rarely leads to a change in management, and has associated guidelines on appropriate use.[3-10] Telemetry use has been a focus of the Joint Commission's National Patient Safety Goals since 2014, and it is also a focus of the Society of Hospital Medicine's *Choosing Wisely*® campaign.[11-13]

Previous initiatives have evaluated how changes to telemetry orders or education and feedback affect telemetry use. Few studies have compared a system-wide electronic health record (EHR) approach to a multifaceted intervention. In seeking to address this gap, we adapted published guidelines from the American Heart Association (AHA) and incorporated them into our EHR ordering process.[3] Simultaneously, we implemented a multifaceted quality improvement initiative and compared this combined program's effectiveness to that of the EHR approach alone.

***Address for correspondence:** Karli Edholm, MD, Division of General Internal Medicine, University of Utah School of Medicine, 30 N 1900 E, Room 5R218, Salt Lake City, UT 84132; Telephone: 801-581-7822; Fax: 801-585-9166; E-mail: karli.edholm@hsc.utah.edu

Additional Supporting Information may be found in the online version of this article.

Received: April 10, 2017; **Revised**: October 29, 2017; **Accepted**: November 10, 2017

 DOI 10.12788/jhm.2933

METHODS

Study Design, Setting, and Population

We performed a 2-group observational pre- to postintervention study at University of Utah Health. Hospital encounters of patients 18 years and older who had at least one inpatient acute care, non-intensive care unit (ICU) room charge and an admission date between January 1, 2014, and July 31, 2016, were included. Patient encounters with missing encounter-level covariates, such as case mix index (CMI) or attending provider identification, were excluded. The Institutional Review Board classified this project as quality improvement and did not require review and oversight.

Intervention

On July 6, 2015, our Epic (Epic Systems Corporation, Madison, Wisconsin) EHR telemetry order was modified to discourage unnecessary telemetry monitoring. The new order required providers ordering telemetry to choose a clinical indication and select

a duration for monitoring, after which the order would expire and require physician renewal or discontinuation. These were the only changes that occurred for nonhospitalist providers. The nonhospitalist group included all admitting providers who were not hospitalists. This group included neurology (6.98%); cardiology (8.13%); other medical specialties such as pulmonology, hematology, and oncology (21.30%); cardiothoracic surgery (3.72%); orthopedic surgery (14.84%); general surgery (11.11%); neurosurgery (11.07%); and other surgical specialties, including urology, transplant, vascular surgery, and plastics (16.68%).

Between January 2015 and June 2015, we implemented a multicomponent program among our hospitalist service. The hospitalist service is composed of four teams with internal medicine residents and two teams with advanced practice providers, all staffed by academic hospitalists. Our program was composed of five elements, all of which were made before the hospital-wide changes to electronic telemetry orders and maintained throughout the study period, as follows: (1) a single provider education session reviewing available evidence (eg, AHA guidelines, *Choosing Wisely*® campaign), (2) removal of the telemetry order from hospitalist admission order set on March 23, 2015, (3) inclusion of telemetry discussion in the hospitalist group's daily "Rounding Checklist,"[14] (4) monthly feedback provided as part of hospitalist group meetings, and (5) a financial incentive, awarded to the division (no individual provider payment) if performance targets were met. See supplementary Appendix ("Implementation Manual") for further details.

Data Source

We obtained data on patient age, gender, Medicare Severity-Diagnosis Related Group, Charlson comorbidity index (CCI), CMI, admitting unit, attending physician, admission and discharge dates, length of stay (LOS), 30-day readmission, bed charge (telemetry or nontelemetry), ICU stay, and inpatient mortality from the enterprise data warehouse. Telemetry days were determined through room billing charges, which are assigned based on the presence or absence of an active telemetry order at midnight. Code events came from a log kept by the hospital telephone operator, who is responsible for sending out all calls to the code team. Code event data were available starting July 19, 2014.

Measures

Our primary outcome was the percentage of hospital days that had telemetry charges for individual patients. All billed telemetry days on acute care floors were included regardless of admission status (inpatient vs observation), service, indication, or ordering provider. Secondary outcomes were inpatient mortality, escalation of care, code event rates, and appropriate telemetry utilization rates. Escalation of care was defined as transfer to an ICU after initially being admitted to an acute care floor. The code event rate was defined as the ratio of the number of code team activations to the number of patient days. Appropriate telemetry utilization rates were determined via chart review, as detailed below.

In order to evaluate changes in appropriateness of telemetry monitoring, 4 of the authors who are internal medicine physicians (K.E., C.C., J.C., D.G.) performed chart reviews of 25 randomly selected patients in each group (hospitalist and nonhospitalist) before and after the intervention who received at least 1 day of telemetry monitoring. Each reviewer was provided a key based on AHA guidelines for monitoring indications and associated maximum allowable durations.[3] Chart reviews were performed to determine the indication (if any) for monitoring, as well as the number of days that were indicated. The number of indicated days was compared to the number of telemetry days the patient received to determine the overall proportion of days that were indicated (Telemetry appropriateness per visit). Three reviewers (K.E,. A.R., C.C.) also evaluated 100 patients on the hospitalist service after the intervention who did not receive any telemetry monitoring to evaluate whether patients with indications for telemetry monitoring were not receiving it after the intervention. For patients who had a possible indication, the indication was classified as Class I (Cardiac monitoring is indicated in most, if not all, patients in this group) or Class II (Cardiac monitoring may be of benefit in some patients but is not considered essential for all patients).[3]

Adjustment Variables

To account for differences in patient characteristics between hospitalist and nonhospitalist groups, we included age, gender, CMI, and CCI in statistical models. CCI was calculated according to the algorithm specified by Quan et al.[15] using all patient diagnoses from previous visits and the index visit identified from the facility billing system.

Statistical Analysis

The period between January 1, 2014, and December 31, 2014, was considered preintervention, and August 1, 2015, to July 31, 2016, was considered postintervention. January 1, 2015, to July 31, 2015, was considered a "run-in" period because it was the interval during which the interventions on the hospitalist service were being rolled out. Data from this period were not included in the pre- or postintervention analyses but are shown in Figure 1.

We computed descriptive statistics for study outcomes and visit characteristics for hospitalist and nonhospitalist visits for pre- and postintervention periods. Descriptive statistics were expressed as n (%) for categorical patient characteristics and outcome variables. For continuous patient characteristics, we expressed the variability of individual observations as the mean ± the standard deviation. For continuous outcomes, we expressed the precision of the mean estimates using standard error. Telemetry utilization per visit was weighted by the number of total acute care days per visit. Telemetry appropriateness per visit was weighted by the number of telemetry days per visit. Patients who did not receive any telemetry monitoring were included in the analysis and noted to have 0 telemetry days. All patients had at least one acute care day. Categorical variables were compared using χ^2 tests, and continuous variables were compared using t tests. Code event rates were compared using the binomial probability mid-p exact test for person-time data.[16]

We fitted generalized linear regression models using generalized estimating equations to evaluate the relative change in outcomes of interest in the postintervention period compared with the preintervention period after adjusting for study covariates. The models included study group (hospitalist and nonhospitalist), time period (pre- and postintervention), an interaction term between study group and time period, and study covariates (age, gender, CMI, and CCI). The models were defined using a binomial distributional assumption and logit link function for mortality, escalation of care, and whether patients had at least 1 telemetry day. A gamma distributional assumption and log link function were used for LOS, telemetry acute care days per visit, and total acute care days per visit. A negative binomial distributional assumption and log link function were used for telemetry utilization and telemetry appropriateness. We used the log of the acute care days as an offset for telemetry utilization and the log of the telemetry days per visit as an offset for telemetry appropriateness. An exchangeable working correlation matrix was used to account for physician-level clustering for all outcomes. Intervention effects, representing the difference in odds for categorical variables and in amount for continuous variables, were calculated as exponentiation of the beta parameters for the covariate minus 1.

P values <.05 were considered significant. We used SAS version 9.4 statistical software (SAS Institute Inc., Cary, North Carolina) for data analysis.

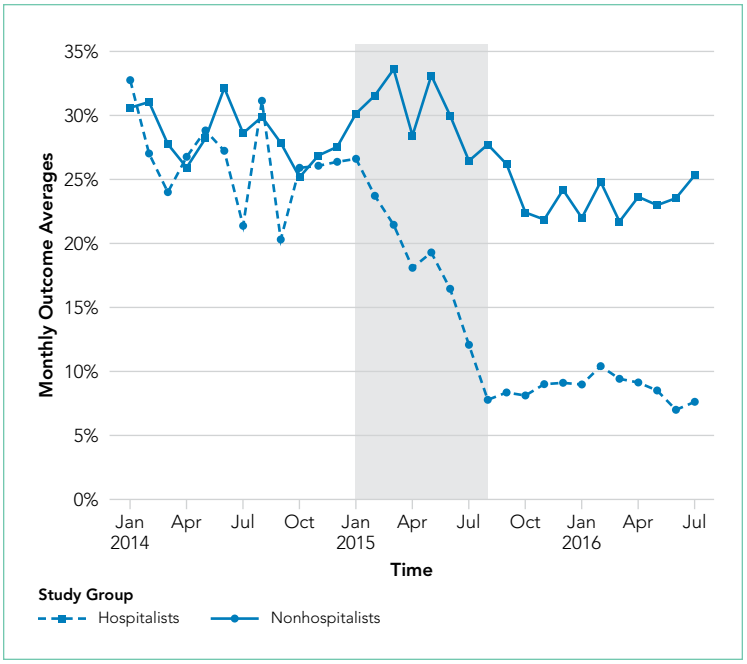

FIG. Primary outcome: telemetry utilization per patient visit. Gray area represents the "run-in period" during which the interventions were being rolled out on the hospitalist service. Removal of the telemetry order from the hospitalist admission order set occurred on March 23, 2015. System-wide change to the EHR telemetry order occurred on July 6, 2015

RESULTS

There were 46,215 visits originally included in the study. Ninety-two visits (0.2%) were excluded due to missing or invalid data. A total of 10,344 visits occurred during the "run-in" period between January 1, 2015, and July 31, 2015, leaving 35,871 patient visits during the pre- and postintervention periods. In the hospitalist group, there were 3,442 visits before the intervention and 3,700 after. There were 13,470 visits in the nonhospitalist group before the intervention and 15,259 after.

The percent of patients who had any telemetry charges decreased from 36.2% to 15.9% ($P < .001$) in the hospitalist group and from 31.8% to 28.0% in the nonhospitalist group ($P < .001$; Table 1). Rates of code events did not change over time ($P = .9$).

Estimates from adjusted and unadjusted linear models are shown in Table 2. In adjusted models, telemetry utilization in the postintervention period was reduced by 69% (95% confidence interval [CI], −72% to −64%; $P < .001$) in the hospitalist group and by 22% (95% CI, −27% to −16%; $P < .001$) in the nonhospitalist group. Compared with nonhospitalists, hospitalists had a 60% greater reduction in telemetry rates (95% CI, −65% to −54%; $P < .001$).

In the randomly selected sample of patients pre- and postintervention who received telemetry monitoring, there was an increase in telemetry appropriateness on the hospitalist service (46% to 72%, $P = .025$; Table 1). In the nonhospitalist group, appropriate telemetry utilization did not change significantly. Of the 100 randomly selected patients in the hospitalist group after the intervention who did not receive telemetry, no patient had an AHA Class I indication, and only four patients had a Class II indication.[3,17]

DISCUSSION

In this study, implementing a change in the EHR telemetry order produced reductions in telemetry days. However, when combined with a multicomponent program including education, audit and feedback, financial incentives, and changes to remove telemetry orders from admission orders sets, an even more marked improvement was seen. Neither intervention reduced LOS, increased code event rates, or increased rates of escalation of care.

Prior studies have evaluated interventions to reduce unnecessary telemetry monitoring with varying degrees of success. The most successful EHR intervention to date, from Dressler et al.,[18] achieved a 70% reduction in overall telemetry use by integrating the AHA guidelines into their EHR and incorporating nursing discontinuation guidelines to ensure that telemetry discontinuation was both safe and timely. Other studies using stewardship approaches and standardized protocols have been less successful.[19,20] One study utilizing a multidisciplinary approach but not including an EHR component showed modest improvements in telemetry.[21]

Although we are unable to differentiate the exact effect of each component of the intervention, we did note an immediate decrease in telemetry orders after removing the telemetry

TABLE 1. **Visit Characteristics and Outcomes Pre- and Postintervention by Study Group**

	Hospitalists			Nonhospitalists		
Characteristic[a]	**Preintervention**	**Postintervention**	***P* Value[b]**	**Preintervention**	**Postintervention**	***P* Value[b]**
Number of admissions	3442	3700		13,470	15,259	
Number of unique patients	2821	3060		10,514	12,055	
Patient Characteristics						
Age (yr.)	57.28 ± 19.39	56.72 ± 18.90	.22	55.85 ± 16.96	55.98 ± 17.04	.54
Female gender	1740 (50.6%)	1874 (50.6%)	.93	6287 (46.7%)	7028 (46.1%)	.30
CCI	3.72 ± 3.35	3.65 ± 3.32	.40	3.65 ± 3.61	3.82 ± 3.72	<.001
CMI	1.56 ± 1.14	1.64 ± 1.14	.001	2.39 ± 2.25	2.41 ± 2.34	.33
Initially admitted to an ICU	515 (15.0%)	635 (17.2%)	.011	1403 (10.4%)	1708 (11.2%)	.034
Outcomes						
LOS (days)	4.26 ± 0.07	4.45 ± 0.07	.054	5.34 ± 0.06	5.27 ± 0.05	.37
Required escalation of care	125 (3.6%)	162 (4.4%)	.11	964 (7.2%)	1141 (7.5%)	.30
Mortality	57 (1.7%)	46 (1.2%)	.14	195 (1.4%)	227 (1.5%)	.78
At least 1 telemetry day	1245 (36.2%)	590 (15.9%)	<.001	4280 (31.8%)	4267 (28.0%)	<.001
Telemetry acute care days per visit	0.95 ± 0.03	0.33 ± 0.02	<.001	1.18 ± 0.02	0.99 ± 0.02	<.001
Total acute care days per visit	3.56 ± 0.05	3.77 ± 0.06	.007	4.16 ± 0.04	4.16 ± 0.04	.92
Telemetry utilization per visit (%)	26.56 ± 0.64	8.66 ± 0.36	<.001	28.42 ± 0.36	23.82 ± 0.31	<.001
Telemetry appropriateness per visit (%)[c]	46.43 ± 8.45	72.00 ± 7.11	.025	69.57 ± 8.50	50.00 ± 9.33	.13

[a]Values expressed as n (%) for categorical variables (female gender, initially admitted to an ICU, required escalation of care, mortality, at least 1 telemetry day), as mean ± standard deviation for continuous patient characteristics (age, CCI, CMI), and as mean ± standard error for continuous outcomes (LOS, telemetry acute care days per visit, total acute care days per visit, telemetry utilization per visit, telemetry appropriateness per visit).

[b]*P* values are based on χ^2 tests for categorical variables and on t tests for continuous variables.

[c]Telemetry appropriateness per visit was determined from 100 chart reviews, 25 for each group.

NOTE: Abbreviations: CCI, Charlson comorbidity index; CMI, case mix index; ICU, intensive care unit; LOS, length of stay.

order from our admission order set, a trend that was magnified after the addition of broader EHR changes (Figure 1). Important additional contributors to our success seem to have been the standardization of rounds to include daily discussion of telemetry and the provision of routine feedback. We cannot discern whether other components of our program (such as the financial incentives) contributed more or less to our program, though the sum of these interventions produced an overall program that required substantial buy in and sustained focus from the hospitalist group. The importance of the hospitalist program is highlighted by the relatively large differences in improvement compared with the nonhospitalist group.

Our study has several limitations. First, the study was conducted at a single center, which may limit its generalizability. Second, the intervention was multifaceted, diminishing our ability to discern which aspects beyond the system-wide change in the telemetry order were most responsible for the observed effect among hospitalists. Third, we are unable to fully account for baseline differences in telemetry utilization between hospitalist and nonhospitalist groups. It is likely that different services utilize telemetry monitoring in different ways, and the hospitalist group may have been more aware of the existing guidelines for monitoring prior to the intervention. Furthermore, we had a limited sample size for the chart audits, which reduced the available statistical power for determining changes in the appropriateness of telemetry utilization. Additionally, because internal medicine residents rotate through various services, it is possible that the education they received on their hospitalist rotation as part of our intervention had a spillover effect in the nonhospitalist group. However, any effect should have decreased the difference between the groups. Lastly, although our postintervention time period was one year, we do not have data beyond that to monitor for sustainability of the results.

CONCLUSION

In this single-site study, combining EHR orders prompting physicians to choose a clinical indication and duration for monitoring with a broader program – including upstream changes in ordering as well as education, audit, and feedback – produced reductions in telemetry usage. Whether this reduction improves the appropriateness of telemetry utilization or reduces other effects of telemetry (eg, alert fatigue, calls for benign arrhythmias) cannot be discerned from our study. However, our results support the idea that multipronged approaches to telemetry use are most likely to produce improvements.

Acknowledgments

The authors thank Dr. Frank Thomas for his assistance with process engineering and Mr. Andrew Wood for his routine provision of data. The statistical analysis

TABLE 2. **Impact of Intervention on Outcomes by Study Group**

		Hospitalists		Nonhospitalists		Difference Between Hospitalists and Nonhospitalists	
Outcome Variables		**Intervention Effect[a]**	***P* Value[b]**	**Intervention Effect[a]**	***P* Value[b]**	**Intervention Effect[a]**	***P* Value[b]**
LOS (days)	Unadjusted	2% (–3% to 8%)	.43	–2% (–6% to 2%)	.27	5% (–2% to 12%)	.20
	Adjusted	2% (–3% to 7%)	.54	–2% (–5% to 1%)	.14	4% (–2% to 10%)	.20
Required escalation of care	Unadjusted	16% (–12% to 55%)	.30	5% (–6% to 16%)	.41	11% (–18% to 51%)	.49
	Adjusted	13% (–4% to 34%)	.14	2% (–6% to 11%)	.60	11% (–8% to 34%)	.27
Mortality	Unadjusted	–32% (–49% to –8%)	.013	–5% (–19% to 11%)	.51	–28% (–49% to 1%)	.061
	Adjusted	–24% (–42% to –0%)	.050	–8% (–22% to 9%)	.34	–17% (–40% to 13%)	.24
At least 1 telemetry day	Unadjusted	–65% (–70% to –60%)	<.001	–22% (–28% to –16%)	<.001	–56% (–62% to –48%)	<.001
	Adjusted	–65% (–69% to –59%)	<.001	–23% (–29% to –17%)	<.001	–54% (–60% to –46%)	<.001
Telemetry acute care days per visit	Unadjusted	–21% (–31% to –10%)	<.001	–10% (–16% to –4%)	.002	–12% (–24% to 2%)	.091
	Adjusted	–21% (–31% to –11%)	<.001	–10% (–15% to –5%)	<.001	–12% (–23% to 0%)	.059
Total acute care days per visit	Unadjusted	4% (–1% to 10%)	.093	–1% (–5% to 2%)	.41	6% (–0% to 12%)	.063
	Adjusted	3% (–2% to 8%)	.21	–2% (–5% to 1%)	.14	5% (–0% to 11%)	.067
Telemetry utilization per visit	Unadjusted	–67% (–71% to –63%)	<.001	–21% (–26% to –16%)	<.001	–58% (–63% to –53%)	<.001
	Adjusted	–69% (–72% to –64%)	<.001	–22% (–27% to –16%)	<.001	–60% (–65% to –54%)	<.001
Telemetry appropriateness per visit[c]	Unadjusted	62% (20% to 117%)	.001	53% (–15% to 176%)	.16	5% (–45% to 104%)	.87
	Adjusted	37% (–9% to 105%)	.13	56% (–10% to 171%)	.11	–12% (–56% to 75%)	.71

[a]Intervention effect represents relative change in odds for categorical variables (required escalation of care, mortality, at least 1 telemetry day) and in amount for continuous variables (LOS, telemetry acute care days per visit, total acute care day per visit, telemetry utilization per visit) and was calculated as exponentiation of the beta parameter for the variable minus 1. Shown in parentheses are 95% CIs. Minus (–) sign represents decrease in odds or quantity.

[b]*P* values are based on generalized linear models.

[c]Telemetry appropriateness per visit was determined from 100 chart reviews, 25 for each group.

NOTE: Telemetry utilization decreased in the postimplementation period as compared with the preimplementation period by 69% in the hospitalist group and by 22% in the nonhospitalist group. The reduction in telemetry utilization in the hospitalist group was 60% greater than in the non-hospitalist group. Abbreviations: CI, confidence interval; LOS, length of stay.

was supported by the University of Utah Study Design and Biostatistics Center, with funding in part from the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant 5UL1TR001067-05 (formerly 8UL1TR000105 and UL1RR025764).

Disclosure: The authors have no conflicts of interest to report.

References

1. National Health Expenditure Fact Sheet. 2015; https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NHE-Fact-Sheet.html. Accessed June 27, 2017.
2. Berwick DM, Hackbarth AD. Eliminating waste in US health care. *JAMA*. 2012;307(14):1513-1516.
3. Drew BJ, Califf RM, Funk M, et al. Practice standards for electrocardiographic monitoring in hospital settings: an American Heart Association scientific statement from the Councils on Cardiovascular Nursing, Clinical Cardiology, and Cardiovascular Disease in the Young: endorsed by the International Society of Computerized Electrocardiology and the American Association of Critical-Care Nurses. *Circulation*. 2004;110(17):2721-2746.
4. Sandau KE, Funk M, Auerbach A, et al. Update to Practice Standards for Electrocardiographic Monitoring in Hospital Settings: A Scientific Statement From the American Heart Association. *Circulation*. 2017;136(19):e273-e344.
5. Mohammad R, Shah S, Donath E, et al. Non-critical care telemetry and in-hospital cardiac arrest outcomes. *J Electrocardiol*. 2015;48(3):426-429.
6. Dhillon SK, Rachko M, Hanon S, Schweitzer P, Bergmann SR. Telemetry monitoring guidelines for efficient and safe delivery of cardiac rhythm monitoring to noncritical hospital inpatients. *Crit Pathw Cardiol*. 2009;8(3):125-126.
7. Estrada CA, Rosman HS, Prasad NK, et al. Evaluation of guidelines for the use of telemetry in the non-intensive-care setting. *J Gen Intern Med*. 2000;15(1):51-55.
8. Estrada CA, Prasad NK, Rosman HS, Young MJ. Outcomes of patients hospitalized to a telemetry unit. *Am J Cardiol*. 1994;74(4):357-362.
9. Atzema C, Schull MJ, Borgundvaag B, Slaughter GR, Lee CK. ALARMED: adverse events in low-risk patients with chest pain receiving continuous electrocardiographic monitoring in the emergency department. A pilot study. *Am J Emerg Med*. 2006;24(1):62-67.
10. Schull MJ, Redelmeier DA. Continuous electrocardiographic monitoring and cardiac arrest outcomes in 8,932 telemetry ward patients. *Acad Emerg Med*. 2000;7(6):647-652.
11. The Joint Commission 2017 National Patient Safety Goals https://www.jointcommission.org/hap_2017_npsgs/. Accessed on February 15, 2017.
12. Joint Commission on Accreditation of Healthcare Organizations. The Joint Commission announces 2014 National Patient Safety Goal. *Jt Comm Perspect*. 2013;33(7):1, 3-4.
13. Bulger J, Nickel W, Messler J, et al. Choosing wisely in adult hospital medicine: five opportunities for improved healthcare value. *J Hosp Med*. 2013;8(9):486-492.
14. Yarbrough PM, Kukhareva PV, Horton D, Edholm K, Kawamoto K. Multifaceted intervention including education, rounding checklist implementation, cost feedback, and financial incentives reduces inpatient laboratory costs. *J Hosp Med*. 2016;11(5):348-354.
15. Quan H, Li B, Couris CM, et al. Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. *Am J Epidemiol*. 2011;173(6):676-682.
16. Greenland S, Rothman KJ. Introduction to categorical statistics In: Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology*. Vol 3. Philadelphia, PA: Lippincott Williams & Wilkins; 2008: 238-257.
17. Henriques-Forsythe MN, Ivonye CC, Jamched U, Kamuguisha LK, Olejeme KA, Onwuanyi AE. Is telemetry overused? Is it as helpful as thought? *Cleve Clin J Med*. 2009;76(6):368-372.
18. Dressler R, Dryer MM, Coletti C, Mahoney D, Doorey AJ. Altering overuse of cardiac telemetry in non-intensive care unit settings by hardwiring the use of American Heart Association guidelines. *JAMA Intern Med*. 2014;174(11):1852-1854.
19. Boggan JC, Navar-Boggan AM, Patel V, Schulteis RD, Simel DL. Reductions

in telemetry order duration do not reduce telemetry utilization. *J Hosp Med*. 2014;9(12):795-796.
20. Cantillon DJ, Loy M, Burkle A, et al. Association Between Off-site Central Monitoring Using Standardized Cardiac Telemetry and Clinical Outcomes Among Non-Critically Ill Patients. *JAMA*. 2016;316(5):519-524.
21. Svec D, Ahuja N, Evans KH, et al. Hospitalist intervention for appropriate use of telemetry reduces length of stay and cost. *J Hosp Med*. 2015;10(9): 627-632.

Effect of Hospital Readmission Reduction on Patients at Low, Medium, and High Risk of Readmission in the Medicare Population

Saul Blecker, MD, MHS[1,2,3], Jeph Herrin, PhD[4,5], Ji Young Kwon, MPH[6], Jacqueline N. Grady, MS[6], Simon Jones, PhD[1,3], Leora I. Horwitz, MD, MHS[1,2,3]

[1]Division of Healthcare Delivery Science, Department of Population Health, NYU School of Medicine, New York, New York; [2]Division of General Internal Medicine and Clinical Innovation, Department of Medicine, NYU School of Medicine, New York, New York; [3]Center for Healthcare Innovation and Delivery Science, NYU Langone Medical Center, New York, New York; [4]Section of Cardiovascular Medicine, Yale School of Medicine, New Haven, Connecticut; [5]Health Research & Educational Trust, Chicago, Illinois; [6]Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, New Haven, Connecticut.

BACKGROUND: Hospitalization and readmission rates have decreased in recent years, with the possible consequence that hospitals are increasingly filled with high-risk patients.

OBJECTIVE: We studied whether readmission reduction has affected the risk profile of hospitalized patients and whether readmission reduction was similarly realized among hospitalizations with low, medium, and high risk of readmissions.

DESIGN: Retrospective study of hospitalizations between January 2009 and June 2015.

PATIENTS: Hospitalized fee-for-service Medicare beneficiaries, categorized into 1 of 5 specialty cohorts used for the publicly reported hospital-wide readmission measure.

MEASUREMENTS: Each hospitalization was assigned a predicted risk of 30-day, unplanned readmission using a risk-adjusted model similar to publicly reported measures. Trends in monthly mean predicted risk for each cohort and trends in monthly observed to expected readmission for hospitalizations in the lowest 20%, middle 60%, and highest 20% of risk of readmission were assessed using time series models.

RESULTS: Of 47,288,961 hospitalizations, 16.2% (n = 7,642,161) were followed by an unplanned readmission within 30 days. We found that predicted risk of readmission increased by 0.24% (P = .03) and 0.13% (P = .004) per year for hospitalizations in the surgery/gynecology and neurology cohorts, respectively. We found no significant increase in predicted risk for hospitalizations in the medicine (0.12%, P = .12), cardiovascular (0.32%, P = .07), or cardiorespiratory (0.03%, P = .55) cohorts. In each cohort, observed to expected readmission rates steadily declined, and at similar rates for patients at low, medium, and high risk of readmission.

CONCLUSIONS: Hospitals have been effective at reducing readmissions across a range of patient risk strata and clinical conditions. The risk of readmission for hospitalized patients has increased for 2 of 5 clinical cohorts. *Journal of Hospital Medicine* 2018;13:537-543. Published online first February 12, 2018. © 2018 Society of Hospital Medicine

Given the high cost of readmissions to the healthcare system, there has been a substantial push to reduce readmissions by policymakers.[1] Among these is the Hospital Readmissions Reduction Program (HRRP), in which hospitals with higher than expected readmission rates receive reduced payments from Medicare.[2] Recent evidence has suggested the success of such policy changes, with multiple reports demonstrating a decrease in 30-day readmission rates in the Medicare population starting in 2010.[3-8]

Initiatives to reduce readmissions can also have an effect on total number of admissions.[9,10] Indeed, along with the recent reduction in readmission, there has been a reduction in all admissions among Medicare beneficiaries.[11,12] Some studies have found that as admissions have decreased, the burden of comorbidity has increased among hospitalized patients,[3,11] suggesting that hospitals may be increasingly filled with patients at high risk of readmission. However, whether readmission risk among hospitalized patients has changed remains unknown, and understanding changes in risk profile could help inform which patients to target with future interventions to reduce readmissions.

Hospital efforts to reduce readmissions may have differential effects on types of patients by risk. For instance, low-intensity, system-wide interventions such as standardized discharge instructions or medicine reconciliation may have a stronger effect on patients at relatively low risk of readmission who may have a

***Address for correspondence:** Saul Blecker, MD, MHS, NYU School of Medicine, 227 E. 30th St., Room 734, New York, NY 10016; Telephone: 646-501-2513; Fax: 646-501-2706; E-mail: Saul.blecker@nyumc.org

Additional Supporting Information may be found in the online version of this article.

Received: July 25, 2017; **Revised**: November 10, 2017; **Accepted**: November 22, 2017

few important drivers of readmission that are easily overcome. Alternatively, the impact of intensive care transitions management might be greatest for high-risk patients, who have the most need for postdischarge medications, follow-up, and self-care.

The purpose of this study was therefore twofold: (1) to observe changes in average monthly risk of readmission among hospitalized Medicare patients and (2) to examine changes in readmission rates for Medicare patients at various risk of readmission. We hypothesized that readmission risk in the Medicare population would increase in recent years, as overall number of admissions and readmissions have fallen.[7,11] Additionally, we hypothesized that standardized readmission rates would decline less in highest risk patients as compared with the lowest risk patients because transitional care interventions may not be able to mitigate the large burden of comorbidity and social issues present in many high-risk patients.[13,14]

METHODS

We performed a retrospective cohort study of hospitalizations to United States nonfederal short-term acute care facilities by Medicare beneficiaries between January 2009 and June 2015. The design involved four steps. First, we estimated a predictive model for unplanned readmissions within 30 days of discharge. Second, we assigned each hospitalization a predicted risk of readmission based on the model. Third, we studied trends in mean predicted risk of readmission during the study period. Fourth, we examined trends in observed to expected (O/E) readmission for hospitalizations in the lowest, middle, and highest categories of predicted risk of readmission to determine whether reductions in readmissions were more substantial in certain risk groups than in others.

Data were obtained from the Centers for Medicare and Medicaid Services (CMS) Inpatient Standard Analytic File and the Medicare Enrollment Data Base. We included hospitalizations of fee-for-service Medicare beneficiaries age ≥65 with continuous enrollment in Part A Medicare fee-for-service for at least one year prior and 30 days after the hospitalization.[15] Hospitalizations with a discharge disposition of death, transfer to another acute hospital, and left against medical advice (AMA) were excluded. We also excluded patients with enrollment in hospice care prior to hospitalization. We excluded hospitalizations in June 2012 because of an irregularity in data availability for that month.

Hospitalizations were categorized into five specialty cohorts according to service line. The five cohorts were those used for the CMS hospital-wide readmission measure and included surgery/gynecology, medicine, cardiovascular, cardiorespiratory, and neurology.[15] Among the three clinical conditions tracked as part of HRRP, heart failure and pneumonia were a subset of the cardiorespiratory cohort, while acute myocardial infarction was a subset of the cardiovascular cohort. Our use of cohorts was threefold: first, the average risk of readmission differs substantially across these cohorts, so pooling them produces heterogeneous risk strata; second, risk variables perform differently in different cohorts, so one single model may not be as accurate for calculating risk; and, third, the use of disease cohorts makes our results comparable to the CMS model and similar to other readmission studies in Medicare.[7,8,15]

For development of the risk model, the outcome was 30-day unplanned hospital readmission. Planned readmissions were excluded; these were defined by the CMS algorithm as readmissions in which a typically planned procedure occurred in a hospitalization with a nonacute principal diagnosis.[16] Independent variables included age and comorbidities in the final hospital-wide readmission models for each of the five specialty cohorts.[15] In order to produce the best possible individual risk prediction for each patient, we added additional independent variables that CMS avoids for hospital quality measurement purposes but that contribute to risk of readmission: sex, race, dual eligibility status, number of prior AMA discharges, intensive care unit stay during current hospitalization, coronary care unit stay during current hospitalization, and hospitalization in the prior 30, 90, and 180 days. We also included an indicator variable for hospitalizations with more than 9 discharge diagnosis codes on or after January 2011, the time at which Medicare allowed an increase of the number of International Classification of Diseases, 9th Revision-Clinical Modification diagnosis billing codes from 9 to 25.[17] This indicator adjusts for the increased availability of comorbidity codes, which might otherwise inflate the predicted risk relative to hospitalizations prior to that date.

Based on the risk models, each hospitalization was assigned a predicted risk of readmission. For each specialty cohort, we pooled all hospitalizations across all study years and divided them into risk quintiles. We categorized hospitalizations as high risk if in the highest quintile, medium risk if in the middle three quintiles, and low risk if in the lowest quintile of predicted risk for all study hospitalizations in a given specialty cohort.

For our time trend analyses, we studied two outcomes: monthly mean predicted risk and monthly ratio of observed readmissions to expected readmissions for patients in the lowest, middle, and highest categories of predicted risk of readmission. We studied monthly predicted risk to determine whether the average readmission risk of patients was changing over time as admission and readmission rates were declining. We studied the ratio of O/E readmissions to determine whether the decline in overall readmissions was more substantial in particular risk strata; we used the ratio of O/E readmissions, which measures number of readmissions divided by number of readmissions predicted by the model, rather than crude observed readmissions, as O/E readmissions account for any changes in risk profiles over time within each risk stratum. Independent variables in our trend analyses were year – entered as a continuous variable – and indicators for postintroduction of the Affordable Care Act (ACA, March 2010) and for postintroduction of HRRP (October 2012); these time indicators were included because of prior studies demonstrating that the introduction of ACA was associated with a decrease from baseline in readmission rates, which leveled off after introduction of HRRP.[7] We also included an indicator for calendar quarter to account for seasonal effects.

Statistical Analysis

We developed generalized estimating equation models to predict 30-day unplanned readmission for each of the five specialty cohorts. The five models were fit using all patients in each cohort for the included time period and were adjusted for clustering by hospital. We assessed discrimination by calculating area under the receiver operating characteristic curve (AUC) for the five models; the AUCs measured the models' ability to distinguish patients who were readmitted versus those who were not.[18] We also calculated AUCs for each year to examine model performance over time.

Using these models, we calculated predicted risk for each hospitalization and averaged these to obtain mean predicted risk for each specialty cohort for each month. To test for trends in mean risk, we estimated 5 time series models, one for each cohort, with the dependent variable of monthly mean predicted risk. For each cohort, we first estimated a series of 12 empty autoregressive models, each with a different autoregressive term (1, 2...12). For each model, we calculated χ^2 for the test that the autocorrelation was 0; based on a comparison of chi-squared values, we specified an autocorrelation of 1 month for all models. Accordingly, a one-month lag was used to estimate one final model for each cohort. Independent variables included year and indicators for post-ACA and post-HRRP; these variables captured the effect of trends over time and the introduction of these policy changes, respectively.[19]

To determine whether changes in risk over time were a result of changes in particular risk groups, we categorized hospitalizations into risk strata based on quintiles of predicted risk for each specialty cohort for the entire study period. For each individual year, we calculated the proportion of hospitalizations in the highest, middle, and lowest readmission risk strata for each cohort.

We calculated the monthly ratio of O/E readmission for hospitalizations in the lowest 20%, middle 60%, and highest 20% of readmission risk by month; O/E reflects the excess or deficit observed events relative to the number predicted by the model. Using this monthly O/E as the dependent variable, we developed autoregressive time series models as above, again with a one-month lag, for each of these 3 risk strata in each cohort. As before, independent variables were year as a continuous variable, indicator variables for post-ACA and post-HRRP, and a categorical variable for calendar quarter.

All analyses were done in SAS version 9.3 (SAS Institute Inc., Cary, North Carolina) and Stata version 14.2 (StataCorp LLC, College Station, Texas).

RESULTS

We included 47,288,961 hospitalizations in the study, of which 11,231,242 (23.8%) were in the surgery/gynecology cohort, 19,548,711 (41.3%) were in the medicine cohort, 5,433,125 (11.5%) were in the cardiovascular cohort, 8,179,691 (17.3%) were in the cardiorespiratory cohort, and 2,896,192 (6.1%) were in the neurology cohort. The readmission rate was 16.2% (n = 7,642,161) overall, with the highest rates observed in the cardiorespiratory (20.5%) and medicine (17.6%) cohorts and the lowest rates observed in the surgery/gynecology (11.8%) and neurology (13.8%) cohorts.

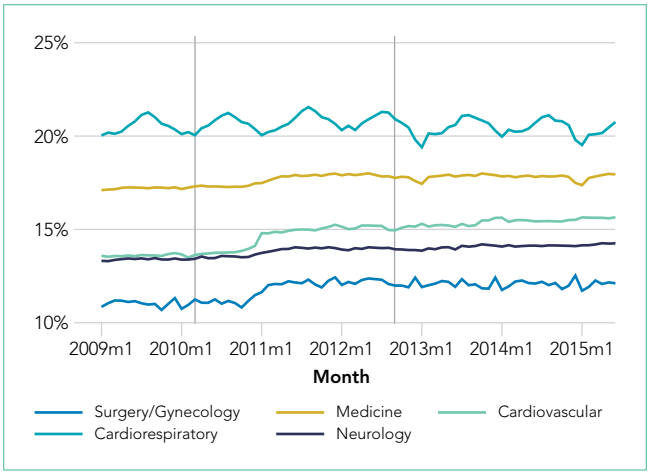

FIG 1. Trends in monthly mean predicted risk of readmission, by specialty cohort. Vertical lines represent introduction of the Affordable Care Act in March 2010 and introduction of the Hospital Readmissions Reduction Program in October 2012.

The final predictive models for each cohort ranged in number of parameters from 56 for the cardiorespiratory cohort to 264 for the surgery/gynecology cohort. The models had AUCs of 0.70, 0.65, 0.67, 0.65, and 0.63 for the surgery/gynecology, medicine, cardiovascular, cardiorespiratory, and neurology cohorts, respectively; AUCs remained fairly stable over time for all disease cohorts (Appendix Table 1).

We observed an increase in the mean predicted readmission risk for hospitalizations in the surgery/gynecology and cardiovascular hospitalizations in early 2011 (Figure 1), a period between the introduction of ACA in March 2010 and the introduction of HRRP in October 2012. In time series models, the surgery/gynecology, cardiovascular, and neurology cohorts had increased predictive risks of readmission of 0.24%, 0.32%, and 0.13% per year, respectively, although this difference did not reach statistical significance for the cardiovascular cohort (Table 1). We found no association between introduction of ACA or HRRP and predicted risk for these cohorts (Table 1). There were no trends or differences in predicted readmission risk for hospitalizations in the medicine cohort. We observed a seasonal variation in predicted readmission risk for the cardiorespiratory cohort but no notable change in predicted risk over time (Figure 1); in the time series model, there was a slight decrease in risk following introduction of HRRP (Table 1).

After categorizing hospitalizations by predicted readmission risk, trends in the percent of hospitalizations in low, middle, and high risk strata differed by cohort. In the surgery/gynecology cohort, the proportion of hospitalizations in the lowest risk stratum increased only slightly, from 20.1% in 2009 to 21.1% of all surgery/gynecology hospitalizations in 2015 (Appendix Table 2). The proportion of surgery/gynecology hospitalizations in the high risk stratum (top quintile of risk) increased from 16.1% to 21.6% between 2009 and 2011 and remained

TABLE 1. **Beta Coefficients from the Times Series Models Examining the Association Between Time and Predicted Risk of Readmission Following Hospitalization**

	Surgery/Gynecology		Medicine		Cardiovascular		Cardiopulmonary		Neurology	
Model Variable	Beta Coefficient (SE)	*P* Value	Beta Coefficient (SE)	*P* Value	Beta Coefficient (SE)	*P* Value	Beta Coefficient (SE)	*P* Value	Beta Coefficient (SE)	*P* Value
Time (years)	0.236 (0.109)	.03	0.118 (0.076)	.12	0.317 (0.171)	.065	0.027 (0.044)	.547	0.131 (0.045)	.004
post-ACA	–0.108 (0.237)	.649	0.006 (0.202)	.976	0.103 (0.224)	.646	0.116 (0.169)	.491	0.103 (0.100)	.302
post-HRRP	–0.125 (0.361)	.729	–0.008 (0.194)	.967	0.090 (0.447)	.841	–0.319 (0.153)	.038	–0.018 (0.122)	.883
Intercept	10.963 (0.322)	<.001	17.196 (0.284)	<.001	13.639 (0.608)	<.001	20.142 (0.168)	<.001	13.326 (0.128)	<.001

NOTE: Years represents change with each year; post-ACA represents change with introduction of the ACA in March 2010; post-HRRP represents change with introduction of the HRRP in October 2012. Adjusted for calendar quarter. Abbreviations: ACA, Affordable Care Act; HRRP, Hospital Readmissions Reduction Program; SE, standard error.

at 21.8% in 2015, and the proportion of surgery/gynecology hospitalizations in the middle risk stratum (middle three quintiles of risk) decreased from 63.7% in 2009 to 59.4% in 2011 to 57.1% in 2015. Low-risk hospitalizations in the medicine cohort decreased from 21.7% in 2009 to 19.0% in 2015, while high-risk hospitalizations increased from 18.2% to 20.7% during the period. Hospitalizations in the lowest stratum of risk steadily declined in both the cardiovascular and neurology cohorts, from 24.9% to 14.8% and 22.6% to 17.3% of hospitalizations during the period, respectively; this was accompanied by an increase in the proportion of high-risk hospitalizations in each of these cohorts from 16.0% to 23.4% and 17.8% to 21.6%, respectively. The proportion of hospitalizations in each of the 3 risk strata remained relatively stable in the cardiorespiratory cohort (Appendix Table 2).

In each of the five cohorts, O/E readmissions steadily declined from 2009 to 2015 for hospitalizations with the lowest, middle, and highest predicted readmission risk (Figure 2). Each risk stratum had similar rates of decline during the study period for all cohorts (Table 2). Among surgery/gynecology hospitalizations, the monthly O/E readmission declined by 0.030 per year from an initial ratio of 0.936 for the lowest risk hospitalizations, by 0.037 per year for the middle risk hospitalizations, and by 0.036 per year for the highest risk hospitalizations (Table 2). Similarly, for hospitalizations in the lowest versus highest risk of readmission, annual decreases in O/E readmission rates were 0.018 versus 0.015, 0.034 versus 0.033, 0.020 versus 0.015, and 0.038 versus 0.029 for the medicine, cardiovascular, cardiorespiratory, and neurology cohorts, respectively. For all cohorts and in all risk strata, we found no significant change in O/E readmission risk with introduction of ACA or HRRP (Table 2).

DISCUSSION

In this six-year, national study of Medicare hospitalizations, we found that readmission risk increased over time for surgical and neurological patients but did not increase in medicine or cardiorespiratory hospitalizations, even though those cohorts are known to have had substantial decreases in admissions and readmissions over the same time period.[7,8] Moreover, we found that O/E readmissions decreased similarly for all hospitalized Medicare patients, whether of low, moderate, or high risk of readmission. These findings suggest that hospital efforts have resulted in improved outcomes across the risk spectrum.

A number of mechanisms may account for the across-the-board improvements in readmission reduction. Many hospitals have instituted system-wide interventions, including patient education, medicine reconciliation, and early postdischarge follow-up,[20] which may have reduced readmissions across all patient risk strata. Alternatively, hospitals may have implemented interventions that disproportionally benefited low-risk patients while simultaneously utilizing interventions that only benefited high-risk patients. For instance, increasing threshold for admission[7] may have the greatest effect on low-risk patients who could be most easily managed at home, while many intensive transitional care interventions have been developed to target only high-risk patients.[21,22]

With the introduction of HRRP, there have been a number of concerns about the readmission measure used to penalize hospitals for high readmission rates. One major concern has been that the readmission metric may be flawed in its ability to capture continued improvement related to readmission.[23] Some have suggested that with better population health management, admissions will decrease, patient risk of the remaining patients will increase, and hospitals will be increasingly filled with patients who have high likelihood of readmission. This potential for increased risk with HRRP was suggested by a recent study that found that comorbidities increased in hospitalized Medicare beneficiaries between 2010 and 2013.[11] Our results were mixed in supporting this potential phenomenon because we examined global risk of readmission and found that some of the cohorts had increased risk over time while others did not. Others have expressed concern that readmission measure does not account for socioeconomic status, which has been associated with readmission rates.[24-27] Although we did not directly examine socioeconomic status in our study, we found that hospitals have been able to reduce readmission across all levels of risk, which includes markers of socioeconomic status, including race and Medicaid eligibility status.

Although we hypothesized that readmission risk would increase as number of hospitalizations decreased over time, we

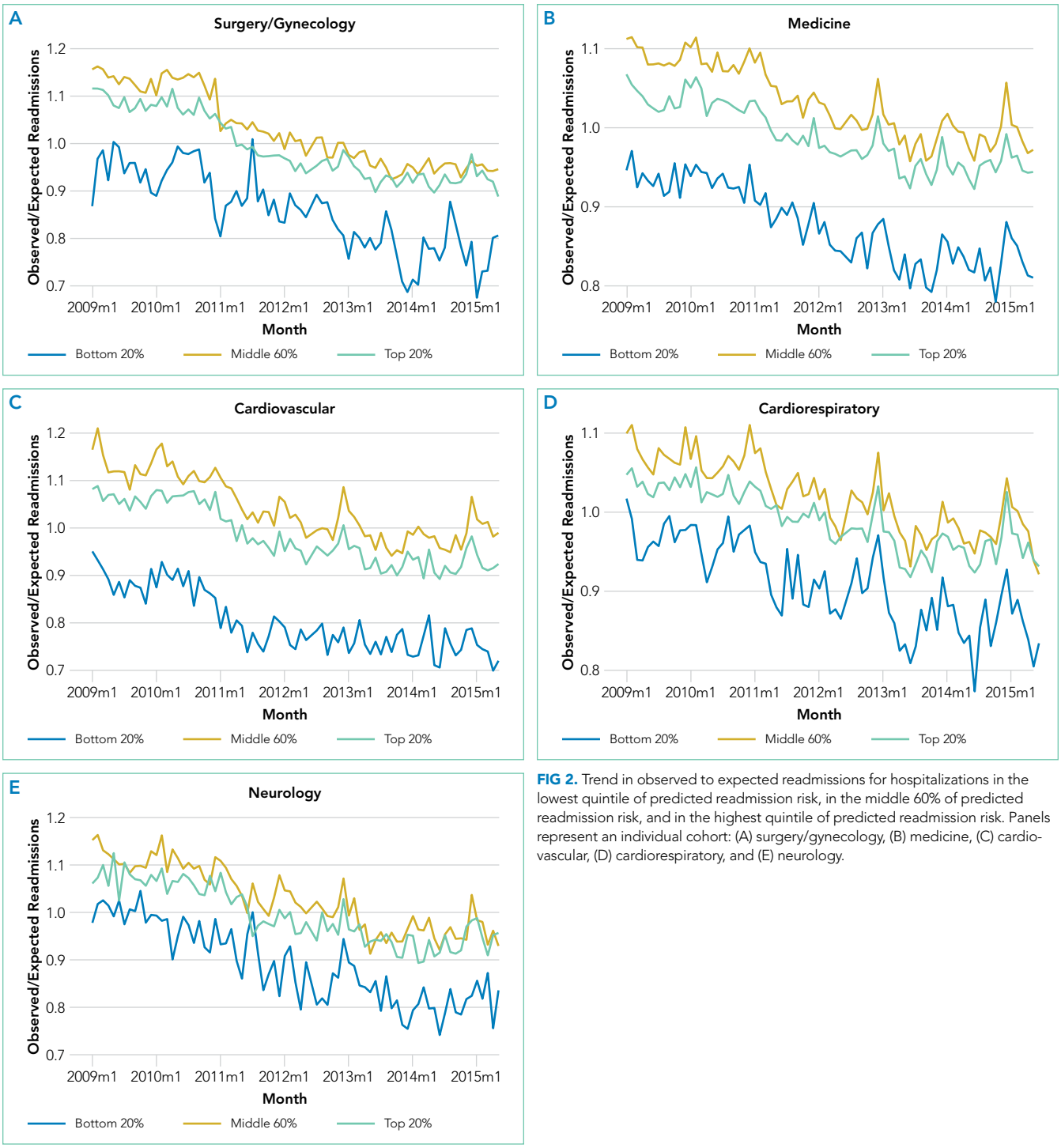

FIG 2. Trend in observed to expected readmissions for hospitalizations in the lowest quintile of predicted readmission risk, in the middle 60% of predicted readmission risk, and in the highest quintile of predicted readmission risk. Panels represent an individual cohort: (A) surgery/gynecology, (B) medicine, (C) cardiovascular, (D) cardiorespiratory, and (E) neurology.

found no increase in readmission risk among the cohorts with HRRP diagnoses that had the largest decrease in readmission rates.[7,8] Conversely, readmission risk did increase – with a concurrent increase in the proportion of high-risk hospitalizations – in the surgery/gynecology and neurology cohorts that were not subject to HRRP penalties. Nonetheless, rehospitalizations were reduced for all risk categories in these two cohorts. Notably, surgery/gynecology and neurology had the lowest readmission rates overall. These findings suggest that initiatives to prevent initial hospitalizations, such as increasing the threshold for postoperative admission, may have had a greater effect on low- versus high-risk patients in low-risk hospitalizations. However, once a patient is hospitalized, multidisciplinary strategies appear to be effective at reducing readmissions for all risk classes in these cohorts.

For the three cohorts in which we observed an increase in readmission risk among hospitalized patients, the risk appeared to increase in early 2011. This time was about 10 months after

TABLE 2. **Beta Coefficients for Time Series Models Examining Observed to Expected Readmission for Hospitalizations in the Lowest 20%, Middle 60%, and Highest 20% of Predicted Risk for Readmission**

	Lowest Risk Quintile		Middle 60% of Risk		Highest Risk Quintile	
Cohorts	**Coefficient (SE)**	***P* Value**	**Coefficient (SE)**	***P* Value**	**Coefficient (SE)**	***P* Value**
Surgery/Gynecology Cohort						
Time (years)	−0.030 (0.009)	.001	−0.037 (0.010)	<.001	−0.036 (0.009)	<.001
Post-ACA	0.005 (0.022)	.835	−0.021 (0.025)	.41	0.039 (0.020)	.05
Post-HRRP	−0.033 (0.029)	.261	0.006 (0.030)	.848	−0.016 (0.027)	.559
Intercept	0.936 (0.012)	<.001	1.147 (0.025)	<.001	1.083 (0.027)	<.001
Medicine Cohort						
Time (years)	−0.018 (0.005)	<.001	−0.022 (0.005)	<.001	−0.015 (0.006)	.014
Post-ACA	−0.008 (0.015)	.567	−0.001 (0.016)	.946	−0.020 (0.019)	.276
Post-HRRP	−0.017 (0.014)	.221	−0.001 (0.017)	.971	−0.000 (0.018)	.982
Intercept	0.949 (0.012)	<.001	1.103 (0.015)	<.001	1.051 (0.011)	<.001
Cardiovascular Cohort						
Time (years)	−0.034 (0.008)	<.001	−0.029 (0.007)	<.001	−0.033 (0.008)	<.001
Post-ACA	−0.021 (0.021)	.311	−0.005 (0.024)	.85	0.000 (0.022)	.994
Post-HRRP	0.032 (0.028)	.246	−0.004 (0.022)	.857	0.014 (0.024)	.543
Intercept	0.908 (0.016)	<.001	1.146 (0.018)	<.001	1.073 (0.019)	<.001
Cardiopulmonary Cohort						
Time (years)	−0.020 (0.006)	.001	−0.019 (0.006)	.002	−0.014 (0.004)	.001
Post-ACA	−0.010 (0.019)	.59	−0.002 (0.020)	.925	−0.004 (0.014)	.798
Post-HRRP	−0.003 (0.017)	.85	−0.007 (0.017)	.686	−0.012 (0.015)	.424
Intercept	0.985 (0.012)	<.001	1.080 (0.013)	<.001	1.038 (0.012)	<.001
Neurology Cohort						
Time (years)	−0.038 (0.008)	<.001	−0.032 (0.007)	<.001	−0.029 (0.007)	<.001
Post-ACA	−0.023 (0.026)	.379	0.002 (0.017)	.89	−0.004 (0.018)	.849
Post-HRRP	0.019 (0.025)	.442	−0.006 (0.022)	.791	0.004 (0.023)	.866
Intercept	1.036 (0.019)	<.001	1.134 (0.016)	<.001	1.088 (0.012)	<.001

NOTE: Years represents change with each year; post-ACA represents change with introduction of the ACA in March 2010; post-HRRP represents change with introduction of the HRRP in October 2012. Adjusted for calendar quarter. Abbreviations: ACA, Affordable Care Act; HRRP Hospital Readmissions Reduction Program; SE, standard error.

passage of ACA, the timing of which was previously associated with a drop in readmission rates,[7,8] but well before HRRP went into effect in October 2012. The increase in readmission risk coincided with an increase in the number of diagnostic codes that could be included on a hospital claim to Medicare.[17] This increase in allowable codes allowed us to capture more diagnoses for some patients, potentially resulting in an increase in apparent predicted risk of readmissions. While we adjusted for this in our predictive models, we may not have fully accounted for differences in risk related to coding change. As a result, some of the observed differences in risk in our study may be attributable to coding differences. More broadly, studies demonstrating the success of HRRP have typically examined risk-adjusted rates of readmission.[3,7] It is possible that a small portion of the observed reduction in risk-adjusted readmission rates may be related to the increase in predicted risk of readmission observed in our study. Future assessment of trends in readmission during this period should consider accounting for change in the number of allowed billing codes.

Other limitations should be considered in the interpretation of this study. First, like many predictive models for readmission,[14] ours had imperfect discrimination, which could affect our results. Second, our study was based on older Medicare patients, so findings may not be applicable to younger patients. Third, while we accounted for surrogates for socioeconomic status, including dual eligibility and race, our models lacked other socioeconomic and community factors that can influence readmission.[24-26] Nonetheless, 1 study suggested that easily measured socioeconomic factors may not have a strong influence on the readmission metric used by Medicare.[28] Fourth, while our study included over 47 million hospitalizations, our time trend analyses used calendar month as the primary independent variable. As our study included 77 months, we may not have had sufficient power to detect small changes in risk over time.

Medicare readmissions have declined steadily in recent years, presumably at least partly in response to policy changes including HRRP. We found that hospitals have been effective at reducing readmissions across a range of patient risk strata and clinical conditions. As a result, the overall risk of readmission for hospitalized patients has remained constant for some but not all conditions. Whether institutions can continue to reduce readmission rates for most types of patients remains to be seen.

Acknowledgments

This study was supported by the Agency for Healthcare Research and Quality (AHRQ) grant R01HS022882. Dr. Blecker was supported by the AHRQ grant K08HS23683. The authors would like to thank Shawn Hoke and Jane Padikkala for administrative support.

Disclosure: This study was supported by the Agency for Healthcare Research and Quality (AHRQ) grants R01HS022882 and K08HS23683. The authors have no conflicts to report.

References

1. Jha AK. Seeking Rational Approaches to Fixing Hospital Readmissions. *JAMA*. 2015;314(16):1681-1682.
2. Centers for Medicare & Medicaid Services. Readmissions Reduction Program. https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program.html. Accessed on January 17, 2017.
3. Suter LG, Li SX, Grady JN, et al. National patterns of risk-standardized mortality and readmission after hospitalization for acute myocardial infarction, heart failure, and pneumonia: update on publicly reported outcomes measures based on the 2013 release. *J Gen Intern Med*. 2014;29(10):1333-1340.
4. Gerhardt G, Yemane A, Hickman P, Oelschlaeger A, Rollins E, Brennan N. Medicare readmission rates showed meaningful decline in 2012. *Medicare Medicaid Res Rev*. 2013;3(2):pii:mmrr.003.02.b01.
5. Centers for Medicare and Medicaid Services. New Data Shows Affordable Care Act Reforms Are Leading to Lower Hospital Readmission Rates for Medicare Beneficiaries. http://blog.cms.gov/2013/12/06/new-data-shows-affordable-care-act-reforms-are-leading-to-lower-hospital-readmission-rates-for-medicare-beneficiaries/. Accessed on January 17, 2017.
6. Krumholz HM, Normand SL, Wang Y. Trends in hospitalizations and outcomes for acute cardiovascular disease and stroke, 1999-2011. *Circulation*. 2014;130(12):966-975.
7. Zuckerman RB, Sheingold SH, Orav EJ, Ruhter J, Epstein AM. Readmissions, Observation, and the Hospital Readmissions Reduction Program. *N Engl J Med*. 2016;374(16):1543-1551.
8. Desai NR, Ross JS, Kwon JY, et al. Association Between Hospital Penalty Status Under the Hospital Readmission Reduction Program and Readmission Rates for Target and Nontarget Conditions. *JAMA*. 2016;316(24):2647-2656.
9. Brock J, Mitchell J, Irby K, et al. Association between quality improvement for care transitions in communities and rehospitalizations among Medicare beneficiaries. *JAMA*. 2013;309(4):381-391.
10. Jencks S. Protecting Hospitals That Improve Population Health. http://medicaring.org/2014/12/16/protecting-hospitals/. Accessed on January 5, 2017.
11. Dharmarajan K, Qin L, Lin Z, et al. Declining Admission Rates And Thirty-Day Readmission Rates Positively Associated Even Though Patients Grew Sicker Over Time. *Health Aff (Millwood)*. 2016;35(7):1294-1302.
12. Krumholz HM, Nuti SV, Downing NS, Normand SL, Wang Y. Mortality, Hospitalizations, and Expenditures for the Medicare Population Aged 65 Years or Older, 1999-2013. *JAMA*. 2015;314(4):355-365.
13. Amarasingham R, Moore BJ, Tabak YP, et al. An automated model to identify heart failure patients at risk for 30-day readmission or death using electronic medical record data. *Med Care*. 2010;48(11):981-988.
14. Kansagara D, Englander H, Salanitro A, et al. Risk prediction models for hospital readmission: a systematic review. *JAMA*. 2011;306(15):1688-1698.
15. Horwitz LI, Partovian C, Lin Z, et al. Development and use of an administrative claims measure for profiling hospital-wide performance on 30-day unplanned readmission. *Ann Intern Med*. 2014;161(10 Suppl):S66-S75.
16. 2016 Condition-Specific Measures Updates and Specifications Report Hospital-Level 30-Day Risk-Standardized Readmission Measures. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/AMI-HF-PN-COPD-and-Stroke-Readmission-Updates.zip. Accessed on January 19, 2017.
17. Centers for Medicare & Medicaid Services. Pub 100-04 Medicare Claims Processing, Transmittal 2028. https://www.cms.gov/Regulations-and-Guidance/Guidance/Transmittals/downloads/R2028CP.pdf. Accessed on November 28, 2016.
18. Martens FK, Tonk EC, Kers JG, Janssens AC. Small improvement in the area under the receiver operating characteristic curve indicated small changes in predicted risks. *J Clin Epidemiol*. 2016;79:159-164.
19. Blecker S, Goldfeld K, Park H, et al. Impact of an Intervention to Improve Weekend Hospital Care at an Academic Medical Center: An Observational Study. *J Gen Intern Med*. 2015;30(11):1657-1664.
20. Hansen LO, Young RS, Hinami K, Leung A, Williams MV. Interventions to reduce 30-day rehospitalization: a systematic review. *Ann Intern Med*. 2011;155(8):520-528.
21. Cavanaugh JJ, Jones CD, Embree G, et al. Implementation Science Workshop: primary care-based multidisciplinary readmission prevention program. *J Gen Intern Med*. 2014;29(5):798-804.
22. Jenq GY, Doyle MM, Belton BM, Herrin J, Horwitz LI. Quasi-Experimental Evaluation of the Effectiveness of a Large-Scale Readmission Reduction Program. *JAMA Intern Med*. 2016;176(5):681-690.
23. Lynn J, Jencks S. A Dangerous Malfunction in the Measure of Readmission Reduction. http://medicaring.org/2014/08/26/malfunctioning-metrics/. Accessed on January 17, 2017.
24. Calvillo-King L, Arnold D, Eubank KJ, et al. Impact of social factors on risk of readmission or mortality in pneumonia and heart failure: systematic review. *J Gen Intern Med*. 2013;28(2):269-282.
25. Barnett ML, Hsu J, McWilliams JM. Patient Characteristics and Differences in Hospital Readmission Rates. *JAMA Intern Med*. 2015;175(11):1803-1812.
26. Singh S, Lin YL, Kuo YF, Nattinger AB, Goodwin JS. Variation in the risk of readmission among hospitals: the relative contribution of patient, hospital and inpatient provider characteristics. *J Gen Intern Med*. 2014;29(4):572-578.
27. American Hospital Association. American Hospital Association (AHA) Detailed Comments on the Inpatient Prospective Payment System (PPS) Proposed Rule for Fiscal Year (FY) 2016. http://www.aha.org/advocacy-issues/letter/2015/150616-cl-cms1632-p-ipps.pdf. Accessed on January 10, 2017.
28. Bernheim SM, Parzynski CS, Horwitz L, et al. Accounting For Patients' Socioeconomic Status Does Not Change Hospital Readmission Rates. *Health Aff (Millwood)*. 2016;35(8):1461-1470.

The Design and Evaluation of the Comprehensive Hospitalist Assessment and Mentorship with Portfolios (CHAMP) Ultrasound Program

Benji K. Mathews, MD FACP SFHM[1,2]*, Kreegan Reierson, MD[1], Khuong Vuong, MD[1], Ankit Mehta, MBBS, FHM, FACP[1], Paula Miller, MPH[3], Seth Koenig, MD, FCCP[4,5], Mangala Narasimhan, DO, FCCP[5,6]

[1]Department of Hospital Medicine, HealthPartners, Bloomington, Minnesota; [2]University of Minnesota Medical School, Minneapolis, Minnesota; [3]Critical Care Research Center, Regions Hospital, Saint Paul, Minnesota; [4]Long Island Jewish Medical Center of Northwell Health, New Hyde Park, New York; [5]Department of Medicine, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, Hofstra University, Hempstead, New York; [6]Critical Care Medicine, Northwell Health System, New Hyde Park, New York.

BACKGROUND: Literature supports the use of point-of-care ultrasound performed by the treating hospitalist in the diagnosis of common diseases. There is no consensus on the training paradigm or the evaluation of skill retention for hospitalists.

OBJECTIVE: To evaluate the effectiveness of a comprehensive bedside ultrasound training program with postcourse competency assessments for hospitalists.

DESIGN: A retrospective report of a training program with 53 hospitalists. The program consisted of online modules, a 3-day in-person course, portfolios, 1-day refresher training, monthly scanning, and assessments. Hospitalists were rated by using similar pre- and postcourse competency assessments and self-rating parameters during the 3-day and refresher courses.

SETTING: A large tertiary-care center.

RESULTS: Skills increased after the 3-day course from a median preassessment score of 15% correct (interquartile range [IQR] 10%-25%) to a median postassessment score of 90% (IQR 80%-95%; $P < .0001$). At the time of the refresher course, the median precourse skills score had decreased to 65% correct (IQR 35%-90%), which improved to 100% postcourse (IQR 85%-100%; $P < .0001$). Skills scores decreased significantly less between the post 3-day course assessment and pre 1-day refresher course for hospitalists who completed portfolios (mean decrease 13.6% correct; $P < .0001$) and/or monthly scanning sessions (mean decrease 7.3% correct; $P < .0001$) compared with hospitalists who did not complete these items.

CONCLUSIONS: A comprehensive longitudinal ultrasound training program including competency assessments improved ultrasound acquisition skills with hospitalists. Skill retention remained high in those who completed portfolios and/or monthly scanning sessions along with a 1-day in-person refresher course. *Journal of Hospital Medicine* 2018;13:544-550. Published online first February 27, 2018. © 2018 Society of Hospital Medicine

Point-of-care ultrasound (POCUS) is a valuable tool to assist in the diagnosis and treatment of many common diseases.[1-11] Its use has increased in clinical settings over the years, primarily because of more portable, economical, high-quality devices and training availability.[12] POCUS improves procedural success and guides the diagnostic management of hospitalized patients.[2,9-12] Literature details the training of medical students,[13,14] residents,[15-21] and providers in emergency medicine[22] and critical care,[23,24] as well as focused cardiac training with hospitalists.[25-27] However, no literature exists describing a comprehensive longitudinal training program for hospitalists or skills retention.

***Address for correspondence:** Benji K. Mathews, MD, FACP, SFHM, CLHM Regions Hospital, 640 Jackson Street, Mail Stop 11109E, Saint Paul, MN 55101; Telephone: 651-254-9555; Fax: 651-254-9673; E-mail: benji.k.mathews@healthpartners.com

Additional Supporting Information may be found in the online version of this article.

Received: April 3, 2017; **Revised:** November 15, 2017; **Accepted:** November 23, 2017

 DOI 10.12788/jhm.2938

This document details the hospital medicine department's ultrasound training program from Regions Hospital, part of HealthPartners in Saint Paul, Minnesota, a large tertiary care medical center. We describe the development and effectiveness of the Comprehensive Hospitalist Assessment and Mentorship with Portfolios (CHAMP) Ultrasound Program. This approach is intended to support the development of POCUS training programs at other organizations.

The aim of the program was to build a comprehensive bedside ultrasound training paradigm for hospitalists. The primary objective of the study was to assess the program's effect on skills over time. Secondary objectives were confidence ratings in the use of ultrasound and with various patient care realms (volume management, quality of physical exam, and ability to narrow the differential diagnosis). We hypothesized there would be higher retention of ultrasound skills in those who completed portfolios and/or monthly scanning sessions as well as increased confidence through all secondary outcome measures (see below).

MATERIALS AND METHODS

This was a retrospective descriptive report of hospitalists who entered the CHAMP Ultrasound Program. Study participants

were providers from the 454-bed Regions Hospital in Saint Paul, Minnesota. The study was deemed exempt by the HealthPartners Institutional Review Board. Three discrete 3-day courses and two 1-day in-person courses held at the Regions Hospital Simulation Center (Saint Paul, Minnesota) were studied.

Program Description

In 2014, a working group was developed in the hospital medicine department to support the hospital-wide POCUS committee with a charter to provide standardized training for providers to complete credentialing.[28] The goal of the hospital medicine ultrasound program was to establish the use of ultrasound by credentialed hospitalists into well-defined applications integrated into the practice of hospital medicine. Two providers were selected to lead the efforts and completed additional training through the American College of Chest Physicians (CHEST) Certificate of Completion Program.[29] An overall director was designated with the responsibilities delineated in supplementary Appendix 1. This director provided leadership on group practice, protocols, and equipment, creating the organizational framework for success with the training program. The hospital medicine training program had a 3-day in-person component built off the CHEST Critical Care Ultrasonography Program.[24] The curriculum was adapted from the American College of Chest Physicians/Société de Réanimation de Langue Française Statement on Competence in Critical Care Ultrasonography.[30] See Table 1 for the components of the training program.

All components of the training program are required to receive the certificate of completion with the exception of the refresher training. Learner feedback after each 3-day course and refresher training was incorporated into subsequent iterations of the training program. During initial phases, additional hands-on faculty were recruited from emergency medicine and critical care who had extensive experience with bedside ultrasound. Subsequently, faculty consisted of former course participants. All faculty followed a standard set of ultrasound and educational principles to guide the hands-on training of participants (supplementary Appendix 2).

Online Modules

As a prerequisite to the 3-day introductory course, hospitalists were required to complete modules for precourse knowledge involving a set of focused-topic online reading and videos with quizzes (supplementary Appendix 3).

3-Day In-Person Course with Assessments

The 3-day course provided 6 hours of didactics, 8 hours of image interpretation, and 9 hours of hands-on instruction (supplementary Appendix 4). Hospitalists first attended a large group didactic, followed by divided groups in image interpretation and hands-on scanning.[24]

Didactics were provided in a room with a 2-screen set up. Providers used 1 screen to present primary content and the other for simultaneously scanning a human model.

Image interpretation sessions were interactive smaller group learning forums in which participants reviewed high-yield images related to the care of hospital medicine patients and received feedback. Approximately 45 videos with normal and abnormal findings were reviewed during each session.

The hands-on scanning component was accomplished with human models and a faculty-to-participant ratio between 1:2 and 1:3. Human models for this course were paid community models. A variety of ultrasound machine platforms were provided for participants. Learning objectives were clearly delineated prior to each scanning session to ensure the coverage of required content.

TABLE 1. **The Components of the Overall Training Program**

Component	
1.	Online modules
2.	3-day in-person course with pre and post written and skills assessment
3.	Portfolio development
4.	1-day refresher training with pre and post skills assessment (optional)
5.	Monthly scanning sessions (optional)
6.	Final knowledge and skills assessments

Portfolios

Portfolio development was a key aspect in overall POCUS competency for each participant. The hospital medicine department's required portfolio files are presented in the Figure, with standards coinciding with the quality assurance grading rubric as developed by the POCUS committee at Regions Hospital and described by Mathews and Zwank.[28] Images taken with real patients were submitted without patient identifiers to a shared online portal. Faculty provided regular cycling feedback by entering the status of submission (accepted or declined) and specific comments on images and interpretations. Learners worked off of the feedback, practiced their skills, and resubmitted files. An image was considered acceptable if it met criteria of depth, axis, and gain and showed the required organ. Participants could use the same patient for different views but could not use the same patient for multiple images of the same view.

Refresher Training: 1-Day In-Person Course with Assessments and Monthly Scanning Sessions (Optional)

Only hospitalists who completed the 3-day course were eligible to take the 1-day in-person refresher course (supplementary Appendix 5). The first half of the course incorporated scanning with live human models, while the second half of the course had scanning with hospitalized patients focusing on pathology (pleural effusion, hydronephrosis, reduced left ventricular function, etc.). The course was offered at 3, 6, and 12 months after the initial 3-day course.

Monthly scanning sessions occurred for 2 hours every third Friday and were also available prior to the 1-day refresher. The first 90 minutes had a hands-on scanning component with hos-

pitalized patients with faculty supervision (1:2 ratio). The last 30 minutes had an image interpretation component.

Assessments

Knowledge and skills assessment were adapted from the CHEST model (supplementary Appendix 6).[24] Before and after the 3-day and 1-day in-person courses, the same hands-on skills assessment with a checklist was provided (supplementary Appendix 7). Before and after the 3-day course, a written knowledge assessment with case-based image interpretation was provided (supplementary Appendix 6). A final knowledge and skills assessment was given at either of the in-person courses to those who completed the required components of the training. Passing scores for the final knowledge assessment were established at 85% items correct by an expert panel by using the Angoff method.[31] This same standard was applied to the final skills examination. Participants who do not pass the final assessments are provided opportunities for further training and allowed to reattempt the assessments. In this regard, there is a standard training outcome but variances in length of training time for each participant. Pre- and postcourse skills assessments used the same faculty, checklist, and ultrasound device. Raters received an orientation the day prior to each in-person course, reviewing common learner pitfalls, reviewing the checklist, and discussing specific examples.

Measurement

Participant demographic and clinical information was collected at the initial 3-day course for all participants, including age, gender, specialty, years of experience, and number and type of ultrasound procedures personally conducted or supervised in the past year. For skills assessment, a 20-item dichotomous checklist was developed and scored as done correctly or not done/done incorrectly. This same assessment was provided both before and after each of the 3-day and 1-day courses. A 20-question image-based knowledge assessment was also developed and administered both before and after the 3-day course only. The same 20-item checklist was used for the final skills examination. However, a new more detailed 50-question examination was written for the final examination after the portfolio of images was complete. Self-reported measures were confidence in the use of ultrasound, volume management, quality of physical exam, and ability to narrow the differential diagnosis. Confidence in ultrasound use, confidence in volume management, and quality of physical exam were assessed by using a questionnaire both before and after the 3-day course and 1-day course. Participants rated confidence and quality on a 5-point scale, 1 being least confident and 5 being most confident.

Statistical Analysis

Demographics of the included hospitalist population and pre and post 3-day assessments, including knowledge score, skills score, confidence in ultrasound use, confidence in volume management, and quality of physical exam, were summarized. Values for all assessment variables are presented as percentages. Confidence scores were reported as a percentage of the Likert scale (eg, 4/5 was reported as 80%). Skills and written examinations were expressed as percentages of items correct. Data were reported as median and interquartile range or means and standard deviation based on variable distributions. Differences between pre- and postvalues for 3-day course variables were assessed by using 2-sample paired Wilcoxon signed rank tests with a 95% confidence level.

For the subset of hospitalists who also completed the 1-day course, pre and post 1-day course assessments, including skills score, confidence in ultrasound use, confidence in volume management, and quality of physical exam, were summarized. Differences between pre- and postvalues for 1-day assessment variables were assessed by using 2-sample paired Wilcoxon signed rank tests with a 95% confidence level.

For hospitalists who completed both the 3-day and 1-day courses, the change in course assessments, including skills score, confidence in ultrasound use, confidence in volume management, and quality of physical exam, was assessed by summarizing the change from post 3-day metrics to pre 1-day metrics (Table 2). The differences between these 2 assessments were evaluated by using 2-sample paired Wilcoxon signed rank tests with a 95% confidence level. Changes in skills score from post 3-day assessment to pre 1-day assessment were also compared for hospitalists completing any of the portfolio and those completing none, and for hospitalists attending any monthly scanning sessions and those who did not attend any,

Cardiac Study (20 studies with the following images per study)
Total: 100 images
1. Parastemal long axis view
2. Parastemal short axis view
3. Apical four-chamber view
4. Subcostal long axis view
5. Inferior vena cava longitudinal view

Lung/Pleural Study (5 studies with the following images per study)
Total: 20 images
1. Pleural effusion (any size)
2. Sliding lung with A-lines
3. Consolidation
4. B-lines

Abdominal Study (5 studies with the following images per study)
Total: 20 images
1. Left kidney longitudinal view with splenorenal space
2. Right kidney longitudinal view with hepatorenal recess
3. Abdominal aorta longitudinal view
4. Bladder transverse view

Vascular Diagnostic DVT Study (3 studies with the following images per study—including right and left legs)
Total: 24 images
1. Right common femoral vein with compression
2. Left common femoral vein with compression
3. Right common femoral vein at saphenous intake with compression
4. Left common femoral vein at saphenous intake with compression
5. Right superficial femoral vein with compression
6. Left superficial femoral vein with compression
7. Right popliteal vein with compression
8. Left popliteal vein with compression

FIG. CHAMP Ultrasound Program Portfolio Requirements

TABLE 2. **Difference in Assessment Scores Before and After the 3-Day and 1-Day Courses and Difference in Mean Skills Score Between the Post 3-day and Pre 1-day Assessment by Skills Retention Action**

Assessment	Median Score Pre (IQR)	Median Score Post (IQR)	*P* value
Pre to post 3-day course change			
Skills (%)	15.0 (15.0)	90.0 (15.0)	<.0001
Knowledge (%)	40.0 (20.0)	90.0 (15.0)	<.0001
Confidence in US use (%)	20.0 (0.0)	60.0 (40.0)	<.0001
Confidence in volume management (%)	60.0 (40.0)	80.0 (0.0)	<.0001
Quality of physical exam (%)	60.0 (20.0)	80.0 (0.0)	<.0001
Pre to post 1-day course change			
Skills (%)	65.0 (55.0)	100.0 (15.0)	<.0001
Confidence in US use (%)	40.0 (20.0)	80.0 (20.0)	<.0001
Confidence in volume management (%)	40.0 (20.0)	80.0 (20.0)	<.0001
Quality of physical exam (%)	40.0 (20.0)	80.0 (20.0)	<.0001
	Median Post 3-day (IQR)	**Median Pre 1-day (IQR)**	***P* value**
Post 3-day to pre 1-day course change			
Skills (%)	90.0 (15.0)	65.0 (55.0)	<.0001
Confidence in US use (%)	60.0 (40.0)	40.0 (20.0)	.0058
Confidence in volume management (%)	80.0 (20.0)	40.0 (20.0)	<.0001
Quality of physical exam (%)	80.0 (20.0)	40.0 (20.0)	<.0001
	Mean Skills Post 3-day (SD)	**Mean Skills Pre 1-day (SD)**	***P* value**
Skills retention action			
Portfolio completed			
Any	92.0 (6.5)	78.4 (14.0)	
None	82.5 (6.0)	32.5 (10.4)	<.0001
Monthly scanning sessions			
Any	94.5 (6.3)	87.2 (7.9)	
None	85.0 (6.2)	46.5 (19.2)	<.0001

NOTE: All values are displayed as percentages. Abbreviations: IQR, interquartile range; SD, standard deviation; US, ultrasound.

by using analysis of variance and Scheffe tests.

Multiple linear regression was performed with the change in skills assessment score from postcompletion of the 3-day course to precompletion of the 1-day course as the dependent variable. Hospitalists were split into two age groups (30-39 and 40-49) for the purpose of this analysis. The percent of monthly scanning sessions attended, age category, timing of 1-day course, and percent portfolio were assessed as possible predictors of the skills score by using simple linear regression with a $P = .05$ cutoff. A final model was chosen based on predictors significant in simple linear regression and included the percent of the portfolio completed and attendance of monthly scanning sessions.

RESULTS

Demographics

Of the 56 3-day course participants, 53 had complete data (Table 3). Three participants with incomplete data completed most of the course but left prior to postcourse assessments and were excluded from the analysis. Twenty-three hospitalists also completed the 1-day in-person course. Seven hospitalists completed the 1-day course 3 months after the initial course, 8 completed it at 6 months, and 8 completed it at 12 months. Completed portfolios required 164 approved video images. Fifteen of the 23 hospitalists at the 1-day course have started and are working towards completion of the online portfolio, while 9 of the 23 participated in the monthly scanning sessions.

3-Day In-Person Course

For the 53 hospitalists who completed skills-based assessments, performance increased significantly after the 3-day course. Knowledge scores also increased significantly from preassessment to postassessment. Self-reported confidence ratings for ultrasound use, confidence in volume management, and quality of physical exam all increased significantly from preassessment to postassessment (Table 2).

Refresher Training: 1-Day In-Person Course

Because the refresher training was encouraged but not required, only 25 of 53 hospitalists, 23 with complete data, completed the 1-day course. For the 23 hospitalists who completed skills-based assessments before and after the 1-day course, mean skills scores increased significantly (Table 2). Self-reported confidence ratings for ultrasound use, confidence in volume management, and quality of physical exam all increased significantly from preassessment to postassessment (Table 2).

Monthly Scanning Sessions and Portfolio Development

The skills retention from initial course to refresher course by portfolio completion and monthly scanning sessions is shown in Table 2. Multiple regression analysis showed that for every 10% increase in the percent of monthly sessions attended, the mean change in skills score was 3.7% ($P = .017$), and for every 10% increase in the percent of portfolio completed, the mean change in skills score was 2.5% ($P = .04$), showing that both monthly scanning session attendance and portfolio completion are significantly predictive of skills retention over time.

Final Assessments

Four providers met mastery at initial attempt. No providers to date have needed remediation. Many others are going through different stages of the process and are expected to attain mastery in a short period of time.

DISCUSSION

This is the first description of a successful longitudinal training program with assessments in POCUS for hospital medicine providers that shows an increase in skill retention with the use of a follow-up course and bedside scanning.

The CHAMP Ultrasound Program was developed to provide hospital medicine clinicians with a specialty focused in-house training pathway in POCUS and to assist in sustained skills acquisition by providing opportunities for regular feedback and practice. Practice with regular expert feedback is a critical aspect to develop and maintain skills in POCUS.[32,33] Arntfield[34] described the utility of remote supervision with feedback for ultrasound training in critical care, which demonstrated varying learning curves in the submission of portfolio images.[35,36] The CHAMP Ultrasound training program provided expert oversight, longitudinal supervision, and feedback for course participants. The educational method of mastery learning was employed by setting minimum standards and allowing learners to practice until they met that standard.[37-39]

This unique program is made possible by the availability of expert-level faculty. Assessment scores improved with an initial 3-day course; however, they also decayed over time, most prominently with hospitalists that did not continue with POCUS scanning after their initial course. Ironically, those who performed more ultrasounds in the year prior to beginning the 3-day course had lower confidence ratings, likely explained by their awareness of their limitations and opportunities for improvement. The incorporation of refresher training to supplement the core 3-day course and portfolio development are key additions that differentiate this training program. These additions and the demonstration of successful training make this a durable pathway for other hospitalist programs. There are many workshops and short courses for medical students, residents, and practicing providers in POCUS.[40-43] However, without an opportunity for longitudinal supervision and feedback, there is a noted decrease in the skills for participants. The refresher training with its 2 components (1-day in-person course and monthly scanning sessions) provides evidence of the value of mentored training.

TABLE 3. **Demographic Characteristics of Hospitalists Completing the 3-Day Training Course**

Characteristic	N (%)
Age (years)	
20-29	3 (5.9)
30-39	24 (47.1)
40-49	15 (29.4)
50-59	6 (11.8)
60+	3 (5.88)
Gender	
Female	21 (41.2)
Male	30 (58.8)
Years of practice	
0-5	19 (37.3)
6-10	15 (29.4)
11-15	9 (17.7)
16-20	4 (7.8)
20+	4 (7.8)
Have you supervised trainees in the past year?	
Yes	38 (74.5)
No	13 (25.5)
Number of ultrasound procedures done or supervised in the past year	
0	19 (37.3)
1-5	19 (37.3)
6-10	10 (19.6)
11-20	2 (3.9)
20+	1 (2.0)

In the initial program development, refresher training was encouraged but optional. We intentionally tracked those that completed refresher training compared with those that did not. Based on the results showing significant skills retention among those attending some form of refresher training, the program is planned to change to make this a requirement. We recommend refresher training within 12 months of the initial introductory course. There were several hospitalists that were unable to accommodate taking a full-day refresher course and, therefore, monthly scanning sessions were provided as an alternative.

The main limitation of the study is that it was completed in a single hospital system with available training mentors in POCUS. This gave us the ability to perform longitudinal training but may make this less reproducible in other hospital systems. Another limitation is that our course participants did not complete the pre- and postknowledge assessments for the refresher training components of the program, though they did for the initial 3-day course. Our pre- and postassessments have not been externally shown to produce valid data, though they are based on the already validated CHEST ultrasound data.[44]

Finally, our CHAMP Ultrasound Program required a significant time commitment by both faculty and learners. A relatively small percentage of hospitalists have completed the final assessments. The reasons are multifactorial, including program rigor, desire by certain hospitalists to know the basics but not pursue more expertise, and the challenges of developing a skillset that takes dedicated practice over time. We

have aimed to address these barriers by providing additional hands-on scanning opportunities, giving timely feedback with portfolios, and obtaining more ultrasound machines. We expect more hospitalists to complete the final assessments in the coming year as evidenced by portfolio submissions to the shared online portal and many choosing to attend either the monthly scanning sessions and/or the 1-day course. We recognize that other institutions may need to adapt our program to suit their local environment.

CONCLUSION

A comprehensive longitudinal ultrasound training program including competency assessments significantly improved ultrasound acquisition skills with hospitalists. Those attending monthly scanning sessions and participating in the portfolio completion as well as a refresher course significantly retained and augmented their skills.

Acknowledgments

The authors would like to acknowledge Kelly Logue, Jason Robertson, MD, Jerome Siy, MD, Shauna Baer, and Jack Dressen for their support in the development and implementation of the POCUS program in hospital medicine.

Disclosure: The authors do not have any relevant financial disclosures to report.

References

1. Spevack R, Al Shukairi M, Jayaraman D, Dankoff J, Rudski L, Lipes J. Serial lung and IVC ultrasound in the assessment of congestive heart failure. *Crit Ultrasound J*. 2017;9:7-13.
2. Soni NJ, Franco R, Velez M, et al. Ultrasound in the diagnosis and management of pleural effusions. *J Hosp Med*. 2015 Dec;10(12):811-816.
3. Boyd JH, Sirounis D, Maizel J, Slama M. Echocardiography as a guide for fluid management. *Crit Care*. 2016;20(1):274-280.
4. Mantuani D, Frazee BW, Fahimi J, Nagdev A. Point-of-care multi-organ ultrasound improves diagnostic accuracy in adults presenting to the emergency department with acute dyspnea. *West J Emerg Med*. 2016;17(1):46-53.
5. Glockner E, Christ M, Geier F, et al. Accuracy of Point-of-Care B-Line Lung Ultrasound in Comparison to NT-ProBNP for Screening Acute Heart Failure. *Ultrasound Int Open*. 2016;2(3):E90-E92.
6. Bhagra A, Tierney DM, Sekiguchi H, Soni NH. Point-of-Care Ultrasonography for Primary Care Physicians and General Internists. *Mayo Clin Proc*. 2016 Dec;91(12):1811-1827.
7. Crisp JG, Lovato LM, Jang TB. Compression ultrasonography of the lower extremity with portable vascular ultrasonography can accurately detect deep venous thrombosis in the emergency department. *Ann Emerg Med*. 2010;56(6):601-610.
8. Squire BT, Fox JC, Anderson C. ABSCESS: Applied bedside sonography for convenient. Evaluation of superficial soft tissue infections. *Acad Emerg Med*. 2005;12(7):601-606.
9. Narasimhan M, Koenig SJ, Mayo PH. A Whole-Body Approach to Point of Care Ultrasound. *Chest*. 2016;150(4):772-776.
10. Copetti R, Soldati G, Copetti P. Chest sonography: a useful tool to differentiate acute cardiogenic pulmonary edema from acute respiratory distress syndrome. *Cardiovasc Ultrasound*. 2008;6:16-25.
11. Soni NJ, Arntfield R, Kory P. *Point of Care Ultrasound*. Philadelphia: Elsevier Saunders; 2015.
12. Moore CL, Copel JA. Point-of-Care Ultrasonography. *N Engl J Med*. 2011;364(8):749-757.
13. Rempell JS, Saldana F, DiSalvo D, et al. Pilot Point-of-Care Ultrasound Curriculum at Harvard Medical School: Early Experience. *West J Emerg Med*. 2016;17(6):734-740. doi:10.5811/westjem.2016.8.31387.
14. Heiberg J, Hansen LS, Wemmelund K, et al. Point-of-Care Clinical Ultrasound for Medical Students. *Ultrasound Int Open*. 2015;1(2):E58-E66. doi:10.1055/s-0035-1565173.
15. Razi R, Estrada JR, Doll J, Spencer KT. Bedside hand-carried ultrasound by internal medicine residents versus traditional clinical assessment for the identification of systolic dysfunction in patients admitted with decompensated heart failure. *J Am Soc Echocardiogr*. 2011;24(12):1319-1324.
16. Alexander JH, Peterson ED, Chen AY, Harding TM, Adams DB, Kisslo JA Jr. Feasibility of point-of-care echocardiography by internal medicine house staff. *Am Heart J*. 2004;147(3):476-481.
17. Hellmann DB, Whiting-O'Keefe Q, Shapiro EP, Martin LD, Martire C, Ziegelstein RC. The rate at which residents learn to use hand-held echocardiography at the bedside. *Am J Med*. 2005;118(9):1010-1018.
18. Kimura BJ, Amundson SA, Phan JN, Agan DL, Shaw DJ. Observations during development of an internal medicine residency training program in cardiovascular limited ultrasound examination. *J Hosp Med*. 2012;7(7):537-542.
19. Akhtar S, Theodoro D, Gaspari R, et al. Resident training in emergency ultrasound: consensus recommendations from the 2008 Council of Emergency Medicine Residency Directors Conference. *Acad Emerg Med*. 2009;16(s2):S32-S36.
20. Jacoby J, Cesta M, Axelband J, Melanson S, Heller M, Reed J. Can emergency medicine residents detect acute deep venous thrombosis with a limited, two-site ultrasound examination? *J Emerg Med*. 2007;32(2):197-200.
21. Jang T, Docherty M, Aubin C, Polites G. Resident-performed compression ultrasonography for the detection of proximal deep vein thrombosis: fast and accurate. *Acad Emerg Med*. 2004;11(3):319-322.
22. Mandavia D, Aragona J, Chan L, et al. Ultrasound training for emergency physicians—a prospective study. *Acad Emerg Med*. 2000;7(9):1008-1014.
23. Koenig SJ, Narasimhan M, Mayo PH. Thoracic ultrasonography for the pulmonary specialist. *Chest*. 2011;140(5):1332-1341. doi: 10.1378/chest.11-0348.
24. Greenstein YY, Littauer R, Narasimhan M, Mayo PH, Koenig SJ. Effectiveness of a Critical Care Ultrasonography Course. *Chest*. 2017;151(1):34-40. doi:10.1016/j.chest.2016.08.1465.
25. Martin LD, Howell EE, Ziegelstein RC, Martire C, Shapiro EP, Hellmann DB. Hospitalist performance of cardiac hand-carried ultrasound after focused training. *Am J Med*. 2007;120(11):1000-1004.
26. Martin LD, Howell EE, Ziegelstein RC, et al. Hand-carried ultrasound performed by hospitalists: does it improve the cardiac physical examination? *Am J Med*. 2009;122(1):35-41.
27. Lucas BP, Candotti C, Margeta B, et al. Diagnostic accuracy of hospitalist-performed hand-carried ultrasound echocardiography after a brief training program. *J Hosp Med*. 2009;4(6):340-349.
28. Mathews BK, Zwank M. Hospital Medicine Point of Care Ultrasound Credentialing: An Example Protocol. *J Hosp Med*. 2017;12(9):767-772.
29. Critical Care Ultrasonography Certificate of Completion Program. American College of Chest Physicians. http://www.chestnet.org/Education/Advanced-Clinical-Training/Certificate-of-Completion-Program/Critical-Care-Ultrasonography. Accessed March 30, 2017
30. Mayo PH, Beaulieu Y, Doelken P, et al. American College of Chest Physicians/Société de Réanimation de Langue Française statement on competence in critical care ultrasonography. *Chest*. 2009;135(4):1050-1060.
31. Donlon TF, Angoff WH. The scholastic aptitude test. The College Board Admissions Testing Program; 1971:15-47.
32. Ericsson KA, Lehmann AC. Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annu Rev Psychol*. 1996;47:273-305.
33. Ericcson KA, Krampe RT, Tesch-Romer C. The role of deliberate practice in the acquisition of expert performance. *Psychol Rev*. 1993;100(3):363-406.
34. Arntfield RT. The utility of remote supervision with feedback as a method to deliver high-volume critical care ultrasound training. *J Crit Care*. 2015;30(2):441.e1-e6.
35. Ma OJ, Gaddis G, Norvell JG, Subramanian S. How fast is the focused assessment with sonography for trauma examination learning curve? *Emerg Med Australas*. 2008;20(1):32-37.
36. Gaspari RJ, Dickman E, Blehar D. Learning curve of bedside ultrasound of the gallbladder. *J Emerg Med*. 2009;37(1):51-66. doi:10.1016/j.jemermed.2007.10.070.
37. Barsuk JH, McGaghie WC, Cohen ER, Balachandran JS, Wane DB. Use of simulation-based mastery learning to improve quality of central venous catheter placement in a medical intensive care unit. *J Hosp Med*. 2009:4(7):397-403.
38. McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB. A critical review of simulation-based mastery learning with translational outcomes. *Med Educ*. 2014:48(4):375-385.
39. Guskey TR. The essential elements of mastery learning. *J Classroom Interac*. 1987;22:19-22.
40. Ultrasound Institute. Introduction to Primary Care Ultrasound. University of South Carolina School of Medicine. http://ultrasoundinstitute.med.sc.edu/Ulcme.asp. Accessed October 24, 2017.
41. Society of Critical Care Medicine. Live Critical Care Ultrasound: Adult. http://

www.sccm.org/Education-Center/Ultrasound/Pages/Fundamentals.aspx. Accessed October 24, 2017.
42. Castlefest Ultrasound Event. Castlefest 2018. http://castlefest2018.com/. Accessed October 24, 2017.
43. Office of Continuing Medical Education. Point of Care Ultrasound Workshop. UT Health San Antonio Joe R. & Teresa Lozano Long School of Medicine. http://cme.uthscsa.edu/ultrasound.asp. Accessed October 24, 2017.
44. Patrawalla P, Eisen LA, Shiloh A, et al. Development and Validation of an Assessment Tool for Competency in Critical Care Ultrasound. *J Grad Med Educ*. 2015;7(4):567-573.

Focusing on Inattention: The Diagnostic Accuracy of Brief Measures of Inattention for Detecting Delirium

Annachiara Marra, MD, PhD[1,2], James C. Jackson, PsyD[1,3,4,5], E. Wesley Ely, MD, MPH[1,3,4,5,6], Amy J. Graves, MS[7], John F. Schnelle, PhD[4,5,6,8], Robert S. Dittus, MD, MPH[3,5,8], Amanda Wilson, MD[9], Jin H. Han, MD, MSc[3,6,10]*

[1]Division of Allergy, Pulmonary, and Critical Care Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee; [2]Department of Neurosciences, Reproductive and Odontostomatological Sciences, University of Naples, Federico II, Naples, Italy; [3]Center for Health Services Research, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee; [4]Research Service, Department of Veterans Affairs Medical Center, Tennessee Valley Healthcare System, Nashville, Tennessee; [5]Geriatric Research, Education and Clinical Center (GRECC) Service, Department of Veterans Affairs Medical Center, Tennessee Valley Healthcare System, Nashville, Tennessee; [6]Center for Quality Aging, Vanderbilt University Medical Center, Nashville, Tennessee; [7]Department of Urology, Vanderbilt University Medical Center, Nashville, Tennessee; [8]Division of General Internal Medicine and Public Health, Vanderbilt University Medical Center, Nashville, Tennessee; [9]Department of Psychiatry, Vanderbilt University Medical Center, Nashville, Tennessee; [10]Department of Emergency Medicine, Vanderbilt University Medical Center, Nashville, Tennessee.

BACKGROUND: Delirium is frequently missed in most clinical settings. Brief delirium assessments are needed.

OBJECTIVE: To determine the diagnostic accuracy of reciting the months of year backwards (MOTYB) from December to July (MOTYB-6) and December to January (MOTYB-12) for delirium as diagnosed by a psychiatrist and to explore the diagnostic accuracies of the following other brief attention tasks: (1) spell the word "LUNCH" backwards, (2) recite the days of the week backwards, (3) 10-letter vigilance "A" task, and (4) 5 picture recognition task.

DESIGN: Preplanned secondary analysis of a prospective observational study.

SETTING: Emergency department located within an academic, tertiary care hospital.

PARTICIPANTS: 234 acutely ill patients who were ≥65 years old.

MEASUREMENTS: The inattention tasks were administered by a physician. The reference standard for delirium was a comprehensive psychiatrist assessment using Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision criteria. Sensitivities and specificities were calculated.

RESULTS: Making any error on the MOTYB-6 task had a sensitivity of 80.0% (95% confidence interval [CI], 60.9%-91.1%) and specificity of 57.1% (95% CI, 50.4%-63.7%). Making any error on the MOTYB-12 task had a sensitivity of 84.0% (95% CI, 65.4%-93.6%) and specificity of 51.9% (95% CI, 45.2%-58.5%). The best combination of sensitivity and specificity was reciting the days of the week backwards task; if the patient made any error, this was 84.0% (95% CI, 65.4%-93.6%) sensitive and 81.9% (95% CI, 76.1%-86.5%) specific.

CONCLUSION: MOTYB-6 and MOTYB-12 had very good sensitivities but had modest specificities for delirium, limiting their use as a standalone assessment. Reciting the days of the week backwards appeared to have the best combination of sensitivity and specificity for delirium. *Journal of Hospital Medicine* 2018;13:551-557. Published online first March 26, 2018. © 2018 Society of Hospital Medicine

Delirium is an acute neurocognitive disorder[1] that affects up to 25% of older emergency department (ED) and hospitalized patients.[2-4] The relationship between delirium and adverse outcomes is well documented.[5-7] Delirium is a strong predictor of increased length of mechanical ventilation, longer intensive care unit and hospital stays, increased risk of falls, long-term cognitive impairment, and mortality.[8-13] Delirium is frequently missed by healthcare professionals[2,14-16] and goes undetected in up to three out of four patients by bedside nurses and medical practitioners in many hospital settings.[14,17-22] A significant barrier to recognizing delirium is the absence of brief delirium assessments.

In an effort to improve delirium recognition in the acute care setting, there has been a concerted effort to develop and validate brief delirium assessments. To address this unmet need, 4 'A's Test (4AT), the Brief Confusion Assessment Method (bCAM), and the 3-minute diagnostic assessment for CAM-defined delirium (3D-CAM) are 1- to 3-minute delirium assessments that were validated in acutely ill older patients.[23] However, 1 to 3 minutes may still be too long in busy clinical environments, and briefer (<30 seconds) delirium assessments may be needed.

One potential more-rapid method to screen for delirium is to specifically test for the presence of inattention, which is a

***Address for correspondence:** Jin H. Han, MD, MSc, Suite 312, 1313 21st Ave S, Nashville, TN 37232; Telephone: 615-322-0253; Fax: 615-936-1316 ; E-mail: jin.h.han@vanderbilt.edu

Received: August 4, 2017; **Revised:** November 17, 2017; **Accepted:** December 5, 2017

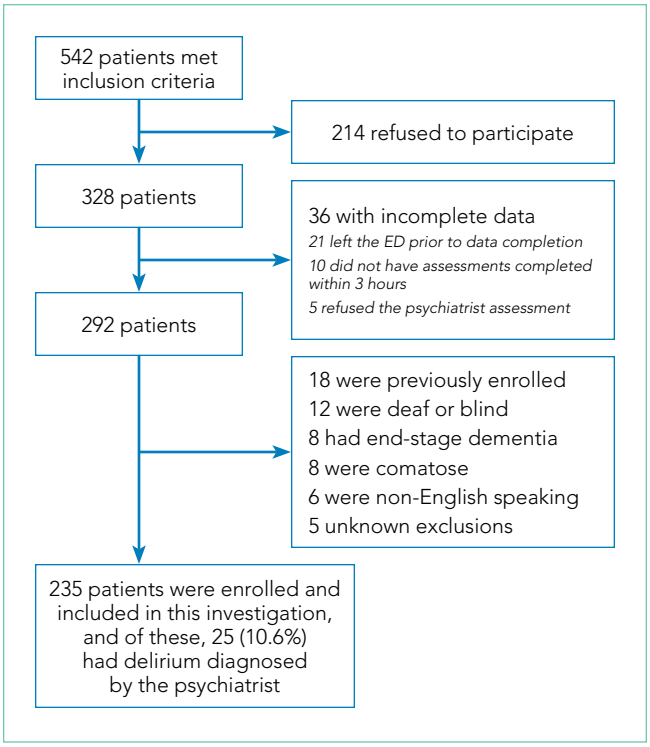

FIG. Enrollment and flow diagram.

cardinal feature of delirium.[24,25] Inattention can be ascertained by having the patient recite the months backwards, recite the days of the week backwards, or spell a word backwards.[26] Recent studies have evaluated the diagnostic accuracy of reciting the months of the year backwards for delirium. O'Regan et al.[27] evaluated the diagnostic accuracy of the month of the year backwards from December to July (MOTYB-6) and observed that this task was 84% sensitive and 90% specific for delirium in older patients. However, they performed the reference standard delirium assessments in patients who had a positive MOTYB-6, which can overestimate sensitivity and underestimate specificity (verification bias).[28] Fick et al.[29] examined the diagnostic accuracy of 20 individual elements of the 3D-CAM and observed that reciting the months of the year backwards from December to January (MOTYB-12) was 83% sensitive and 69% specific for delirium. However, this was an exploratory study that was designed to identify an element of the 3D-CAM that had the best diagnostic accuracy.

To address these limitations, we sought to evaluate the diagnostic performance of the MOTYB-6 and MOTYB-12 for delirium as diagnosed by a reference standard. We also explored other brief tests of inattention such as spelling a word ("LUNCH") backwards, reciting the days of the week backwards, 10-letter vigilance "A" task, and five-picture recognition task.

METHODS

Study Design and Setting

This was a preplanned secondary analysis of a prospective observational study that validated three delirium assessments.[30,31] This study was conducted at a tertiary care, academic ED. The local institutional review board (IRB) reviewed and approved this study. Informed consent from the patient or an authorized surrogate was obtained whenever possible. Because this was an observational study and posed minimal risk to the patient, the IRB granted a waiver of consent for patients who were both unable to provide consent and were without an authorized surrogate available in the ED or by phone.

Selection of Participants

We enrolled a convenience sample of patients between June 2010 and February 2012 Monday through Friday from 8 AM to 4 PM. This enrollment window was based upon the psychiatrist's availability. Because of the extensiveness of the psychiatric evaluations, we limited enrollment to one patient per day. Patients who were 65 years or older, not in a hallway bed, and in the ED for less than 12 hours at the time of enrollment were included. We used a 12-hour cutoff so that patients who presented in the evening and early morning hours could be included. Patients were excluded if they were previously enrolled, non-English speaking, deaf or blind, comatose, suffered from end-stage dementia, or were unable to complete all the study assessments. The rationale for excluding patients with end-stage dementia was that diagnosing delirium in this patient population is challenging.

Research assistants approached patients who met inclusion criteria and determined if any exclusion criteria were present. If none of the exclusion criteria were present, then the research assistant reviewed the informed consent document with the patient or authorized surrogate if the patient was not capable of providing consent. If a patient was not capable of providing consent and no authorized surrogate was available, then the patient was enrolled (under the waiver of consent) as long as the patient assented to be a part of the study. Once the patient was enrolled, the research assistant contacted the physician rater and reference standard psychiatrists to approach the patient.

Measures of Inattention

An emergency physician (J.H.H.) who had no formal training in the mental status assessment of elders administered a cognitive battery to the patient, including tests of inattention. The following inattention tasks were administered:

- Spell the word "LUNCH" backwards.[30] Patients were initially allowed to spell the word "LUNCH" forwards. Patients who were unable to perform the task were assigned 5 errors.
- Recite the months of the year backwards from December to July.[23,26,27,30,32] Patients who were unable to perform the task were assigned 6 errors.
- Recite the days of the week backwards.[23,26,33] Patients who were unable to perform the task were assigned 7 errors.
- Ten-letter vigilance "A" task.[34] The patient was given a series of 10 letters ("S-A-V-E-A-H-A-A-R-T") every 3 seconds and was asked to squeeze the rater's hand every time the patient heard the letter "A." Patients who were unable to perform the task were assigned 10 errors.

- Five-picture recognition task.[34] Patients were shown 5 objects on picture cards. Afterwards, patients were shown 10 pictures with the previously shown objects intermingled. The patient had to identify which objects were seen previously in the first 5 pictures. Patients who were unable to perform the task were assigned 10 errors.
- Recite the months of the year backwards from December to January.[29] Patients who were unable to perform the task were assigned 12 errors.

Reference Standard for Delirium

A comprehensive consultation-liaison psychiatrist assessment was the reference standard for delirium; the diagnosis of delirium was based on Diagnostic and Statistical Manual of Mental Disorders, *Fourth Edition, Text Revision (DSM-IV-TR)* criteria.[35] Three psychiatrists who each had an average of 11 years of clinical experience and regularly diagnosed delirium as part of their daily clinical practice were available to perform these assessments. To arrive at the diagnosis of delirium, they interviewed those who best understood the patient's mental status (eg, the patient's family members or caregivers, *physician, and nurses). They also* reviewed the patient's medical record and radiology and laboratory test results. They performed bedside cognitive testing that included, but was not limited to, the Mini-Mental State Examination, Clock Drawing Test, Luria hand sequencing task, and tests for verbal fluency. A focused neurological examination was also performed (ie, screening for paraphasic errors, tremors, tone, asterixis, frontal release signs, etc.), and they also evaluated the patient for affective lability, hallucinations, and level of alertness. If the presence of delirium was still questionable, then confrontational naming, proverb interpretation or similarities, and assessments for apraxias were performed at the discretion of the psychiatrist. The psychiatrists were blinded to the physician's assessments, and the assessments were conducted within three hours of each other.

Additional Variables Collected

Using medical record review, comorbidity burden, severity of illness, and premorbid cognition were ascertained. The Charlson Comorbidity Index, a weighted index that takes into account the number and seriousness of 19 preexisting comorbid conditions, was used to quantify comorbidity burden; higher scores indicate higher comorbid burden.[36,37] The Acute Physiology Score of the Acute Physiology and Chronic Health Evaluation II was used to quantify severity of illness.[38] This score is based upon the initial values of 12 routine physiologic measurements such as vital sign and laboratory abnormalities; higher scores represent higher severities of illness.[38] The medical record was reviewed to ascertain the presence of premorbid cognitive impairment; any documentation of dementia in the patient's clinical problem list or physician history and physical examination from the outpatient or inpatient settings was considered positive. The medical record review was performed by a research assistant and was double-checked for accuracy by one of the investigators (JHH).

Data Analyses

Measures of central tendency and dispersion for continuous variables were reported as medians and interquartile ranges. Categorical variables were reported as proportions. Receiver operating characteristic curves were constructed for each inattention task. Area under the receiver operating characteristic curves (AUC) was reported to provide a global measure of diagnostic accuracy. Sensitivities, specificities, positive likelihood ratios (PLRs), and negative likelihood ratios (NLRs) with their 95% CIs were calculated using the psychiatrist's assessment as the reference standard.[39] Cut-points with PLRs greater than 10 (strongly increased the likelihood of delirium) or NLRs less than 0.1 (strongly decreased the likelihood of delirium) were preferentially reported whenever possible.

All statistical analyses were performed with open source R statistical software version 3.0.1 (http://www.r-project.org/), SAS 9.4 (SAS Institute, Cary, North Carolina), and Microsoft Excel 2010 (Microsoft Inc., Redmond, Washington).

RESULTS

A total of 542 patients were screened; 214 patients refused to participate, and 93 were excluded, leaving 235 patients. The patient characteristics can be seen in Table 1. Compared with all patients (N = 15,359) who presented to the ED during the study period, enrolled patients were similar in age but more likely to be male, have cardiovascular chief complaints, and be admitted to the hospital. Of those enrolled, 25 (10.6%) were delirious. Delirious patients were older, more likely to be nonwhite, have a past history of dementia, have a graduate school degree, and have a chief complaint of altered mental status.

Making any error on the MOTYB-6 task had a sensitivity of 80.0% (95% CI, 60.9%-91.1%), specificity of 57.1% (95% CI, 50.4%-63.7%), PLR of 1.87 (95% CI, 1.45-2.40) and NLR of 0.35 (95% CI, 0.16-0.77) for delirium as diagnosed by a psychiatrist. Making any error on the MOTYB-12 task had a sensitivity of 84.0% (95% CI, 65.4%-93.6%), specificity of 51.9% (95% CI, 45.2%-58.5%), PLR of 1.75 (95% CI, 1.40-2.18), and NLR of 0.31 (95% CI, 0.12-0.76) for delirium. The AUCs for the MOTYB-6 and MOTYB-12 tasks were 0.77 and 0.78, respectively, indicating very good diagnostic performance.

The diagnostic performances of the other inattention tasks and additional cutoff values for the MOTYB-6 and MOTYB-12 tasks can be seen in Table 2. Increasing the MOTYB-6 cut-off to two or more errors and MOTYB-12 cut-off to three or more errors increased the specificity to 70.0% and 70.5%, respectively, without decreasing their sensitivity. The best combination of sensitivity and specificity was reciting the days of the week backwards task; if the patient made any error, this was 84.0% (95% CI, 65.4%-93.6%) sensitive and 81.9% (95% CI, 76.1%-86.5%) specific for delirium. The inattention tasks that strongly increased the likelihood of delirium (PLR > 10) were the vigilance "A" and picture recognition tasks. If the patient made two or more errors on the vigilance task or three or more errors on the picture recognition task, then the likelihood of delirium strongly increased, as evidenced by a PLR of 16.80 (95% CI, 8.01-35.23) and 23.10 (95% CI, 7.95-67.12), respectively. No other inattention tasks

TABLE 1. **Demographic and Clinical Characteristics of Patients**

Patient Characteristics	Enrolled Patients (n = 235)	All Potentially Eligible Patients (N = 15,359)	Nondelirious N = 210	Delirious N = 25
Median age (IQR)	74 (69, 79)	74 (69, 81)	74 (69, 79)	84 (67, 86)
Female gender	107 (45.5%)	8,198 (53.4%)	96 (45.7%)	11 (44.0%)
Nonwhite race	31 (13.9%)	-	24 (11.4%)	7 (28.0%)
Residence		-		
Home	215 (91.5%)		192 (91.4%)	23 (92.0%)
Assisted living	15 (6.4%)		14 (6.7%)	1 (4.0%)
Rehabilitation/postacute	2 (0.9%)		2 (1.0%)	0 (0.0%)
Nursing home	3 (1.3%)		2 (1.0%)	1 (4.0%)
Education		-		
Elementary or below	6 (2.6%)		5 (2.4%)	1 (4.0%)
Middle school	25 (10.6%)		21 (10.0%)	4 (16.0%)
High school	95 (40.4%)		86 (41.0%)	9 (36.0%)
College	71 (30.2%)		65 (31.0%)	6 (24.0%)
Graduate school	38 (16.2%)		33 (15.7%)	5 (20.0%)
Dementia in medical record	17 (7.2%)	-	10 (4.8%)	7 (28.0%)
Median Charlson (IQR)	2 (1, 4)	-	2 (1, 4)	3 (2, 6)
Median APS (IQR)	3 (1, 5)	-	3 (1, 5)	4 (1, 5)
ED chief complaint				
Abdominal pain	8 (3.4%)	854 (5.6%)	7 (3.3%)	1 (4.0%)
Altered mental status	14 (6.0%)	617 (4.0%)	2 (1.0%)	12 (48.0%)
Chest pain	43 (18.3%)	1575 (10.3%)	39 (18.6%)	4 (16.0%)
General weakness	21 (8.9%)	1,101 (7.2%)	18 (8.6%)	3 (12.0%)
Shortness of breath	31 (13.2%)	1,377 (9.0%)	29 (13.8%)	2 (8.0%)
Syncope	14 (6.0%)	422 (2.3%)	14 (6.7%)	0 (0.0%)
Admitted to the hospital	168 (71.5%)	9491 (61.8%)		

NOTE: The APS is part of the APACHE II. Abbreviations: APACHE II, Acute Physiology and Chronic Health Evaluation II; APS, Acute Physiology Score; ED, emergency department; IQR, interquartile range.

were able to achieve a PLR of greater than 10, regardless of what cutoff was used. No inattention task was able to achieve a NLR of less than 0.10, which would have strongly decreased the likelihood of delirium. The best NLRs were if the patient made no errors spelling the word "LUNCH" backwards (NLR, 0.16; 95% CI, 0.04-0.60), no errors on the vigilance "A" task (NLR, 0.18; 95% CI, 0.07-0.43), and no errors on the days of the week backwards task (NLR, 0.20; 95% CI, 0.08-0.48).

DISCUSSION

Delirium is frequently missed by healthcare providers because it is not routinely screened for in the acute care setting. To help address this deficiency of care, we evaluated several brief measures of inattention that take less than 30 seconds to complete. We observed that any errors made on the MOTYB-6 and MOTYB-12 tasks had very good sensitivities (80% and 84%) but were limited by their modest specificities (approximately 50%) for delirium. As a result, these assessments have limited clinical utility as standalone delirium screens. We also explored other commonly used brief measures of inattention and at a variety of error cutoffs. Reciting the days of the week backwards appeared to best balance sensitivity and specificity. None of the inattention measures could convincingly rule out delirium (NLR < 0.10), but the vigilance "A" and picture recognition tasks may have clinical utility in ruling in delirium (PLR > 10). Overall, all the inattention tasks, including MOTYB-6 and MOTYB-12, had very good diagnostic performances based upon their AUC. However, achieving a high sensitivity often had to be sacrificed for specificity or, alternatively, achieving a high specificity had to be sacrificed for sensitivity.

Inattention has been shown to be the cardinal feature for delirium,[40] and its assessment using cognitive testing has been recommended to help identify the presence of delirium according to an expert consensus panel.[26] The diagnostic performance of the MOTYB-12 observed in our study is similar to a study by Fick et al., who reported that MOTYB-12 had very good sensitivity (83%) but had modest specificity (69%) with a cutoff of 1 or more errors. Hendry et al. observed that the MOTYB-12 was 91% sensitive and 50% specific using a cutoff of 4 or more errors. With regard to the MOTYB-6, our reported specificity was different from what was observed by O'Regan et al.[27] Using 1 or more errors as a cutoff, they observed a much higher specificity for delir-

TABLE 2. **Diagnostic Accuracy of Inattention Measures**

Inattention Test	AUC	Sensitivity	Specificity (95% CI)	PLR (95% CI)	NLR (95% CI)
MOTYB-6					
1 or more errors	0.77	80.0% (60.9%-91.1%)	57.1% (50.4%-63.7%)	1.87 (1.45-2.40)	0.35 (0.16-0.77)
2 or more errors		80.0% (60.9%-91.1%)	70.0% (63.5%-75.8%)	2.67 (2.01-3.55)	0.29 (0.13-0.63)
3 or more errors		60.0% (40.7%-76.6%)	80.0% (74.1%-84.9%)	3.00 (1.97-4.56)	0.50 (0.31-0.81)
4 or more errors		56.0% (37.1%-73.3%)	86.2% (80.9%-90.2%)	4.06 (2.50-6.58)	0.51 (0.33-0.88)
5 or more errors		52.0% (33.5%-70.0%)	89.6% (83.6%-92.2%)	4.55 (2.67-7.75)	0.54 (0.36-0.82)
6 errors or UTD		44.0% (26.7%-62.9%)	91.9% (87.4%-94.8%)	4.48 (2.46-8.17)	0.62 (0.44-0.88)
MOTYB-12					
1 or more errors	0.78	84.0% (65.4%-93.6%)	51.9% (45.2%-58.5%)	1.75 (1.40-2.18)	0.31 (0.12-0.76)
2 or more errors		84.0% (65.4%-93.6%)	65.2% (58.6%-71.4%)	2.42 (1.88-3.11)	0.25 (0.10-0.61)
3 or more errors		84.0% (65.4%-93.6%)	70.5% (64.0%-76.2%)	2.85 (2.17-3.73)	0.23 (0.09-0.56)
4 or more errors		76.0% (56.6%-88.5%)	73.8% (67.5%-79.3%)	2.90 (2.11-3.98)	0.33 (0.16-0.66)
12 errors or UTD		48.0% (30.0%-66.5%)	90.0% (85.2%-93.4%)	4.80 (2.70-8.53)	0.58 (0.40-0.84)
Spelling "LUNCH" backwards					
1 or more errors	0.81	92.0% (75.0%-97.8%)	50.5% (43.8%-57.2%)	1.86 (1.55-2.22)	0.16 (0.04-0.60)
2 or more errors		88.0% (70.0%-95.8%)	64.8% (58.1%-70.9%)	1.57 (1.32-1.87)	0.27 (0.09-0.80)
3 or more errors		80.0% (60.9%-91.1%)	73.3% (67.0%-78.9%)	3.00 (2.23-4.04)	0.27 (0.12-0.60)
4 or more errors		56.0% (37.1%-73.3%)	81.9% (76.1%-86.5%)	3.39 (2.15-5.33)	0.53 (0.34-0.82)
5 errors or UTD		56.0% (37.1%-73.3%)	87.1% (81.9%-91.0%)	4.56 (2.78-7.49)	0.50 (0.32-0.78)
Days of the week backwards					
1 or more errors	0.85	84.0% (65.4%-93.6%)	81.9% (76.1%-86.5%)	4.64 (3.32-6.49)	0.20 (0.08-0.48)
2 or more errors		60.0% (40.7%-76.6%)	90.5% (85.8%-93.8%)	6.30 (3.72-10.66)	0.44 (0.27-0.72)
3 or more errors		56.0% (37.1%-73.3%)	91.4% (86.9%-94.5%)	6.53 (3.72-11.46)	0.48 (0.31-0.75)
4 or more errors		56.0% (37.1%-73.3%)	91.9% (87.4%-94.9%)	6.92 (3.90-12.27)	0.48 (0.31-0.75)
5 or more errors		56.0% (37.1%-73.3%)	92.4% (88.0%-95.3%)	7.35 (4.09-13.20)	0.48 (0.31-0.74)
6 or more errors		56.0% (37.1%-73.3%)	92.9% (88.6%-95.6%)	7.84 (4.31-14.27)	0.47 (0.30-0.74)
7 errors or UTD		44.0% (26.7%-62.9%)	95.2% (91.5%-97.4%)	9.24 (4.37-19.55)	0.59 (0.41-0.83)
10-letter vigilance "A" task					
1 or more errors	0.84	84.0% (65.4%-93.6%)	63.8% (57.1%-70.0%)	9.80 (6.10-15.74)	0.18 (0.07-0.43)
2 or more errors		64.0% (44.5%-79.8%)	91.4% (85.9%-94.5%)	16.80 (8.01-35.23)	0.37 (0.22-0.63)
3 or more errors		60.0% (40.7%-76.7%)	96.2% (92.7%-98.1%)	126.00 (17.37-913.82)	0.40 (0.25-0.65)
Picture recognition task					
1 or more errors	0.81	64.0% (44.5%-79.8%)	80.0% (74.1%-84.9%)	1.40 (1.02-1.92)	0.66 (0.39-1.13)
2 or more errors		60.0% (40.7%-76.7%)	93.8% (89.7%-96.4%)	9.69 (5.23-17.95)	0.43 (0.26-0.69)
3 or more errors		44.0% (27.7%-62.9%)	98.1% (95.2%-99.3%)	23.10 (7.95-67.12)	0.57 (0.40-0.81)

NOTE: Sensitivities, specificities, PLRs, and NLRs of several brief measures of inattention with their AUC. Abbreviations: AUC, area under the receiver operating characteristic curve; CI, confidence interval; MOTYB-6, months of the year backwards December to July; MOTYB-12, months of the year backwards December to January; NLR, negative likelihood ratio; PLR, positive likelihood ratio; UTD, unable to do.

ium than we did (90% vs 57%). Discordant observations regarding the diagnostic accuracy for other inattention tasks also exist. We observed that making any error on the days of the week backwards task was 84% sensitive and 82% specific for delirium, whereas Fick et al. observed a sensitivity and specificity of 50% and 94%, respectively. For the vigilance "A" task, we observed that making two or more errors over a series of 10 letters was 64.0% sensitive and 91.4% specific for delirium, whereas Pompei et al.[41] observed that making two or more errors over a series of 60 letters was 51% sensitive and 77% specific for delirium.

The abovementioned discordant findings may be driven by spectrum bias, wherein the sensitivities and specificities for each inattention task may differ in different subgroups. As a result, differences in the age distribution, proportion of college graduates, history of dementia, and susceptibility to delirium can influence overall sensitivity and specificity. Objective measures of delirium, including the inattention screens studied, are particularly prone to spectrum bias.[31,34] However, the strength of this approach is that the assessment of inattention becomes less reliant upon clinical judgment and allows it to be used by raters from a wide range of clinical backgrounds. On the other hand, a subjective interpretation of these inattention tasks may allow the rater to capture the subtleties of inattention (ie, decreased speed of performance in a highly intelligent and well-educated patient without dementia). The disadvantage of this approach, however, is that it is more dependent on clinical judgment and may have decreased diagnostic accuracy in those with less clinical experience or with limited training.[14,42,43] These factors must be carefully considered when determining which delirium assessment to use.

Additional research is required to determine the clinical utility of these brief inattention assessments. These findings need to be further validated in larger studies, and the optimal cutoff of each task for different subgroup of patients (eg, demented vs nondemented) needs to be further clarified. It is not completely clear whether these inattention tests can serve as standalone assessments. Depending on the cutoff used, some of these assessments may have unacceptable false negative or false positive rates that may lead to increased adverse patient outcomes or increased resource utilization, respectively. Additional components or assessments may be needed to improve the diagnostic accuracy of these assessments. In addition to understanding these inattention assessments' diagnostic accuracies, their ability to predict adverse outcomes also needs to be investigated. While a previous study observed that making any error on the MOTYB-12 task was associated with increased physical restraint use and prolonged hospital length of stay,[44] these assessments' ability to prognosticate long-term outcomes such as mortality or long-term cognition or function need to be studied. Lastly, studies should also evaluate how easily implementable these assessments are and whether improved delirium recognition leads to improved patient outcomes.

This study has several notable limitations. Though planned a priori, this was a secondary analysis of a larger investigation designed to validate three delirium assessments. Our sample size was also relatively small, causing our 95% CIs to overlap in most cases and limiting the statistical power to truly determine whether one measure is better than the other. We also asked the patient to recite the months backwards from December to July as well as recite the months backwards from December to January. It is possible that the patient may have performed better at going from December to January because of learning effect. Our reference standard for delirium was based upon DSM-IV-TR criteria. The new DSM-V criteria may be more restrictive and may slightly change the sensitivities and specificities of the inattention tasks. We enrolled a convenience sample and enrolled patients who were more likely to be male, have cardiovascular chief complaints, and be admitted to the hospital; as a result, selection bias may have been introduced. Lastly, this study was conducted in a single center and enrolled patients who were 65 years and older. Our findings may not be generalizable to other settings and in those who are less than 65 years of age.

CONCLUSIONS

The MOTYB-6 and MOTYB-12 tasks had very good sensitivities but modest specificities (approximately 50%) using any error made as a cutoff; increasing cutoff to 2 errors and 3 errors, respectively, improved their specificities (approximately 70%) with minimal impact to their sensitivities. Reciting the days of the week backwards, spelling the word "LUNCH" backwards, and the 10-letter vigilance "A" task appeared to perform the best in ruling out delirium but only moderately decreased the likelihood of delirium. The 10-letter Vigilance "A" and picture recognition task appeared to perform the best in ruling in delirium. Days of the week backwards appeared to have the best combination of sensitivity and specificity.

Disclosure: The authors report no financial conflicts of interest.

Funding: This study was funded by the Emergency Medicine Foundation Career Development Award, National Institutes of Health K23AG032355, and National Center for Research Resources, Grant UL1 RR024975-01.

References

1. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders: DSM-5.* Washington, DC: American Psychiatric Association; 2013.
2. Hustey FM, Meldon SW, Smith MD, Lex CK. The effect of mental status screening on the care of elderly emergency department patients. *Ann Emerg Med.* 2003;41(5):678-684.
3. Inouye SK, Rushing JT, Foreman MD, Palmer RM, Pompei P. Does delirium contribute to poor hospital outcomes? A three-site epidemiologic study. *J Gen Intern Med.* 1998;13(4):234-242.
4. Pitkala KH, Laurila JV, Strandberg TE, Tilvis RS. Prognostic significance of delirium in frail older people. *Dement Geriatr Cogn Disord.* 2005;19(2-3):158-163.
5. Han JH, Shintani A, Eden S, et al. Delirium in the emergency department: an independent predictor of death within 6 months. *Ann Emerg Med.* 2010;56(3):244-252.
6. Gross AL, Jones RN, Habtemariam DA, et al. Delirium and long-term cognitive trajectory among persons with dementia. *Arch Intern Med.* 2012;172(17):1324-1331.
7. Davis DH, Muniz Terrera G, Keage H, et al. Delirium is a strong risk factor for dementia in the oldest-old: a population-based cohort study. *Brain.* 2012;135(Pt 9):2809-2816.
8. Ely EW, Baker AM, Dunagan DP, et al. Effect on the duration of mechanical ventilation of identifying patients capable of breathing spontaneously. *N Engl J Med.* 1996;335(25):1864-1869.
9. Ely EW, Shintani A, Truman B, et al. Delirium as a predictor of mortality in mechanically ventilated patients in the intensive care unit. *JAMA.* 2004;291(14):1753-1762.
10. Lin SM, Liu CY, Wang CH, et al. The impact of delirium on the survival of mechanically ventilated patients. *Crit Care Med.* 2004;32(11):2254-2259.
11. Salluh JI, Wang H, Schneider EB, et al. Outcome of delirium in critically ill patients: systematic review and meta-analysis. *BMJ.* 2015;350:h2538.
12. Klein Klouwenberg PM, Zaal IJ, Spitoni C, et al. The attributable mortality of delirium in critically ill patients: prospective cohort study. *BMJ.* 2014;349:g6652.
13. Han JH, Vasilevskis EE, Chandrasekhar R, et al. Delirium in the Emergency Department and Its Extension into Hospitalization (DELINEATE) Study: Effect on 6-month Function and Cognition. *J Am Geriatr Soc.* 2017;65(6):1333-1338.
14. Inouye SK, Foreman MD, Mion LC, Katz KH, Cooney LM Jr. Nurses' recognition of delirium and its symptoms: comparison of nurse and researcher ratings. *Arch Intern Med.* 2001;161(20):2467-2473.
15. Han JH, Zimmerman EE, Cutler N, et al. Delirium in older emergency department patients: recognition, risk factors, and psychomotor subtypes. *Acad Emerg Med.* 2009;16(3):193-200.
16. Elie M, Cole MG, Primeau FJ, Bellavance F. Delirium risk factors in elderly hospitalized patients. *J Gen Intern Med.* 1998;13(3):204-212.
17. Spronk PE, Riekerk B, Hofhuis J, Rommes JH. Occurrence of delirium is severely underestimated in the ICU during daily care. *Intensive Care Med.* 2009;35(7):1276-1280.
18. van Eijk MM, van Marum RJ, Klijn IA, de Wit N, Kesecioglu J, Slooter AJ. Comparison of delirium assessment tools in a mixed intensive care unit. *Crit Care Med.* 2009;37(6):1881-1885.
19. Devlin JW, Fong JJ, Schumaker G, O'Connor H, Ruthazer R, Garpestad E. Use of a validated delirium assessment tool improves the ability of physicians to identify delirium in medical intensive care unit patients. *Crit Care Med.* 2007;35(12):2721-2724.
20. Han JH, Eden S, Shintani A, et al. Delirium in Older Emergency Department Patients Is an Independent Predictor of Hospital Length of Stay. *Acad Emerg Med.* 2011;18(5):451-457.
21. Pun BT, Gordon SM, Peterson JF, et al. Large-scale implementation of sedation and delirium monitoring in the intensive care unit: a report from two medical centers. *Crit Care Med.* 2005;33(6):1199-1205.

22. Grossmann FF, Hasemann W, Graber A, Bingisser R, Kressig RW, Nickel CH. Screening, detection and management of delirium in the emergency department - a pilot study on the feasibility of a new algorithm for use in older emergency department patients: the modified Confusion Assessment Method for the Emergency Department (mCAM-ED). *Scand J Trauma Resusc Emerg Med*. 2014;22:19.
23. Marcantonio ER, Ngo LH, O'Connor M, et al. 3D-CAM: Derivation and Validation of a 3-Minute Diagnostic Interview for CAM-Defined Delirium: A Cross-sectional Diagnostic Test Study. *Ann Intern Med*. 2014;161(8):554-561.
24. Blazer DG, van Nieuwenhuizen AO. Evidence for the diagnostic criteria of delirium: an update. *Curr Opin Psychiatry*. 2012;25(3):239-243.
25. Meagher DJ, Maclullich AM, Laurila JV. Defining delirium for the International Classification of Diseases, 11th Revision. *J Psychosom Res*. 2008;65(3):207-214.
26. Huang LW, Inouye SK, Jones RN, et al. Identifying indicators of important diagnostic features of delirium. *J Am Geriatr Soc*. 2012;60(6):1044-1050.
27. O'Regan NA, Ryan DJ, Boland E, et al. Attention! A good bedside test for delirium? *J Neurol Neurosurg Psychiatry*. 2014;85(10):1122-1131.
28. Richardson ML, Petscavage JM. Verification bias: an under-recognized source of error in assessing the efficacy of MRI of the meniscii. *Acad Radiol*. 2011;18(11):1376-1381.
29. Fick DM, Inouye SK, Guess J, et al. Preliminary development of an ultrabrief two-item bedside test for delirium. *J Hosp Med*. 2015;10(10):645-650.
30. Han JH, Wilson A, Vasilevskis EE, et al. Diagnosing delirium in older emergency department patients: validity and reliability of the delirium triage screen and the brief confusion assessment method. *Ann Emerg Med*. 2013;62(5):457-465.
31. Han JH, Wilson A, Graves AJ, et al. Validation of the Confusion Assessment Method for the Intensive Care Unit in older emergency department patients. *Acad Emerg Med*. 2014;21(2):180-187.
32. Inouye SK. Delirium in hospitalized older patients. *Clin Geriatr Med*. 1998;14(4):745-764.
33. Hamrick I, Hafiz R, Cummings DM. Use of days of the week in a modified mini-mental state exam (M-MMSE) for detecting geriatric cognitive impairment. *J Am Board Fam Med*. 2013;26(4):429-435.
34. Ely EW, Inouye SK, Bernard GR, et al. Delirium in mechanically ventilated patients: validity and reliability of the confusion assessment method for the intensive care unit (CAM-ICU). *JAMA*. 2001;286(21):2703-2710.
35. American Psychiatric Association. Task Force on DSM-IV. *Diagnostic and statistical manual of mental disorders: DSM-IV-TR*. 4th ed. Washington, DC: American Psychiatric Association; 2000.
36. Murray SB, Bates DW, Ngo L, Ufberg JW, Shapiro NI. Charlson Index is associated with one-year mortality in emergency department patients with suspected infection. *Acad Emerg Med*. 2006;13(5):530-536.
37. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40(5):373-383.
38. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med*. 1985;13(10):818-829.
39. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J Clin Epidemiol*. 1991;44(8):763-770.
40. Blazer DG, Wu LT. The epidemiology of substance use and disorders among middle aged and elderly community adults: national survey on drug use and health. *Am J Geriatr Psychiatry*. 2009;17(3):237-245.
41. Pompei P, Foreman M, Cassel CK, Alessi C, Cox D. Detecting delirium among hospitalized older patients. *Arch Intern Med*. 1995;155(3):301-307.
42. Lemiengre J, Nelis T, Joosten E, et al. Detection of delirium by bedside nurses using the confusion assessment method. *J Am Geriatr Soc*. 2006;54(4):685-689.
43. Ryan K, Leonard M, Guerin S, Donnelly S, Conroy M, Meagher D. Validation of the confusion assessment method in the palliative care setting. *Palliat Med*. 2009;23(1):40-45.
44. Yevchak AM, Doherty K, Archambault EG, Kelly B, Fonda JR, Rudolph JL. The association between an ultrabrief cognitive screening in older adults and hospital outcomes. *J Hosp Med*. 2015;10(10):651-657.

Evaluating the Benefits of Hospital Room Artwork for Patients Receiving Cancer Treatment: A Randomized Controlled Trial

Daniel R. George, PhD, MSc[1]*, Claire de Boer, MS[2], Joel Hammer, BA[1], Margaret Hopkins, MEd, MA[1], Tonya King, PhD[3], Michael J. Green, MD, MS[1,4]

[1]Department of Humanities, Pennsylvania State University College of Medicine, Hershey, Pennsylvania; [2]Milton S. Hershey Medical Center, Pennsylvania State University College of Medicine, Hershey, Pennsylvania; [3]Public Health Sciences, Pennsylvania State University College of Medicine, Hershey, Pennsylvania; [4]Department of Medicine, Pennsylvania State University College of Medicine, Hershey, Pennsylvania.

We examined whether placing a painting in the line of vision of a hospitalized patient improves patient outcomes and satisfaction and whether having patients choose their paintings offers greater benefit. From 2014 to 2016, we enrolled 186 inpatients with cancer diagnoses from Pennsylvania State University Cancer Institute and randomly assigned them to three groups: those who chose paintings displayed in rooms, those whose paintings were randomly selected, and those with no paintings. We assessed anxiety, mood, depression, quality of life, perceptions of hospital environment, sense of control and/or influence, self-reported pain, and length of stay and compared patients with paintings versus those without paintings, as well as those with an artwork choice versus those with no choice. There were no differences in psychological and/or clinical outcomes across the groups, but patients in the three groups with paintings reported significantly improved perceptions of the hospital environment. Integrating artwork into inpatient rooms may represent one means of improving perceptions of the institution. *Journal of Hospital Medicine* 2018;13:558-561. Published online first February 5, 2018.

With hospital reimbursement increasingly being linked to patient satisfaction,[1] about half of United States hospitals have embraced arts programs as a means of humanizing clinical environments and improving the patient experience.[2,3] There is emerging evidence that integrating such programs into clinical settings is associated with less pain, stress, and anxiety[4-10] as well as improved mood,[11] greater levels of interaction,[12] and feeling less institutionalized.[13] However, it has been observed that existing studies have been undertaken with variable methodological rigor,[14] and few randomized controlled trials (RCTs) have linked specific design features or interventions directly to healthcare outcomes. We designed a RCT to test the hypotheses that (1) placing a painting by a local artist in the line of vision of hospitalized patients would improve psychological and clinical outcomes and patient satisfaction and (2) letting patients choose their own painting would offer even greater benefit in these areas.

METHODS

From 2014 to 2016, our research team recruited inpatients who were being treated in the Pennsylvania State University Hershey Cancer Institute in Hershey, Pennsylvania. Patients were eligible if they were English speaking, over the age of 19, not cognitively impaired, and had been admitted for cancer-related treatments that required at least a 3-day inpatient stay. During recruitment, patients were told that the study was on patient care and room décor, and thus those who were not being given artwork did not know about the artwork option. By using a permuted block design with mixed block size, we randomly assigned consenting patients to one of the following three groups: (1) those who chose the painting displayed in their rooms, (2) those whose painting was randomly selected, and (3) those with no painting in their rooms, only white boards in their line of vision (see Figure 1). All paintings were created by artists in central Pennsylvania and reproduced as high-quality digital prints for the study, costing approximately $90 apiece. Members of the research team visited patients in the designated rooms three times during their stay – within 24 hours of being admitted, within 24 to 48 hours of the first visit, and within 24 to 48 hours of the second visit – with each visit lasting from 5 to 10 minutes. Patients who were given the opportunity to select art for their rooms were shown a catalogue of approximately 20 available paintings from which to choose a desired print; as with the group whose paintings were randomly selected for them, patients who made a choice had a print immediately hung in their room by members of the research team for the entirety of their inpatient stay.

Outcomes and Measures

The primary outcomes were psychological and included the following: anxiety, mood, depression, and sense of control and/or influence. These were measured using the validated

***Address for correspondence:** Daniel R. George, PhD, MSc, Milton S. Hershey Medical Center, 500 University Drive, Hershey, PA 17033; Telephone: 717-531-8778; Fax: 717-531-3894; E-mail: dgeorge1@pennstatehealth.psu.edu

Additional Supporting Information may be found in the online version of this article.

Received: July 6, 2017; **Revised**: October 9, 2017; **Accepted**: October 19, 2017

2018 Society of Hospital Medicine DOI 10.12788/jhm.2915

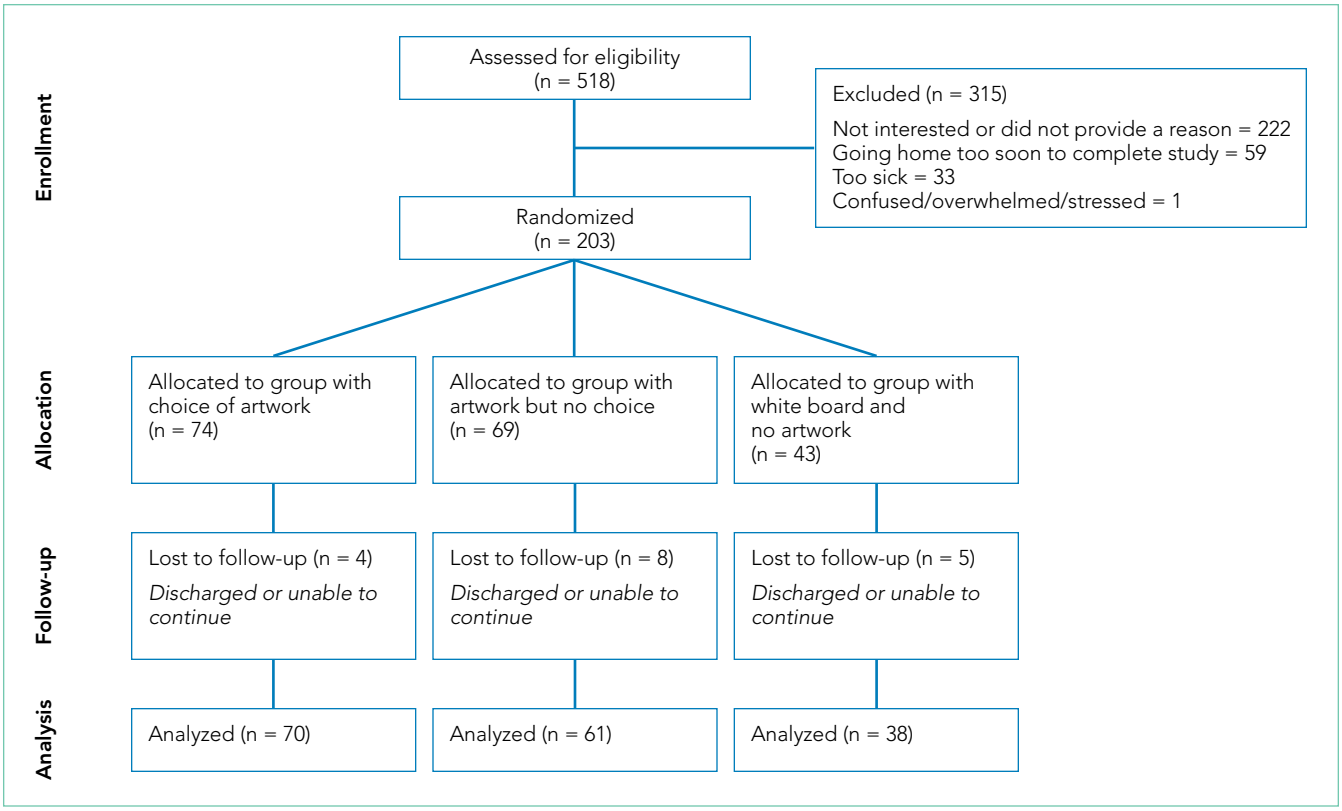

FIG 1. Study Flow Chart

State-Trait Anxiety Inventory (STAI)[15] an emotional thermometer instrument (ETI)[16], and a self-designed instrument measuring one's sense of control and influence over the environment. Secondary outcomes were clinical, encompassing pain, quality of life (QOL), length of stay (LOS), and related to perceptions of the hospital environment. These were assessed using data extracted from the electronic medical record (EMR) as well as the Room Assessment (RA) survey, a validated instrument used in prior clinical studies to assess inpatient settings.[17] The RA survey uses the Semantic Differential scale, a rating scale designed to measure emotional associations by using paired attributes.[18] In our scale, we listed 17 paired and polar opposite attributes, with one descriptor reflecting a more positive impression than the other. Anxiety, emotional state, and control and/or influence were assessed at baseline and prior to discharge; emotional state was assessed every 1 to 2 days; and perceptions of the room and overall patient experience were measured once, prior to discharge, using the RA survey.

Data Analysis

A sample of 180 participants were chosen, with a 2:1 ratio of art group to no-art control group to provide at least 80% power to detect a difference in anxiety score of 4 units, for the comparisons of interest among the groups. The calculations assumed a 2-sided test with $\alpha = 0.05$.

Comparisons were made between (1) those with paintings versus those without and (2) those with a choice of paintings versus those with no choice. For the primary psychological outcome, the average anxiety score at discharge was compared between groups of interest by using analysis of covariance, with adjustment for baseline score. Items measuring mood, depression, control, and influence that were collected more frequently were compared between groups by using repeated measures analysis of covariance, with adjustment for corresponding score at baseline. For clinical outcomes, median LOS was compared between groups by using the Wilcoxon rank sum test due to the skewed distribution of data, and QOL and pain were compared between groups by using repeated measures analysis of covariance. The model for patient-reported pain included covariates for pain medication received and level of pain tolerance. Outcomes measuring perceptions of hospital environment were collected at a single time point and compared between groups by using the 2-sample t-test. Results were reported in terms of means and 95% confidence intervals or medians and quartiles. Significance was defined by $P < .05$. All facets of this study were approved by the Pennsylvania State University College of Medicine Institutional Review Board.

RESULTS

We approached 518 patients to participate in the study, and 203 elected to enroll. Seventeen patients withdrew from the study because they had been discharged from the hospital or were unable to continue. Of the 186 participants who completed the study, 74 chose the painting displayed in their rooms, 69 had paintings randomly selected for them, and 43 had no paintings in their rooms, only white boards in their line

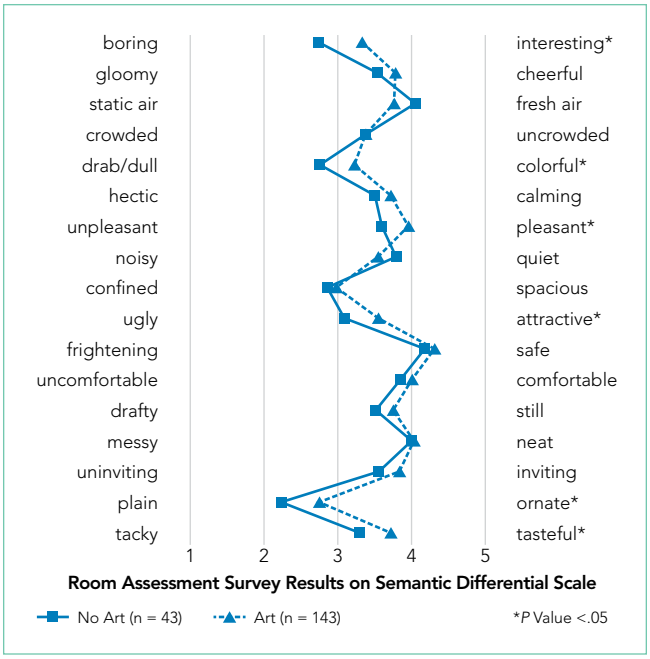

FIG 2. Those with artwork indicated their rooms were more interesting 3.3 (95% CI 3.2-3.5) vs 2.7 (95%CI 2.4-3.1), $P = .002$), colorful 3.2 (95% CI 3.0-3.5) vs. 2.8 (95% CI 2.4-3.1), $P = .026$), pleasant 4.0 (95% CI 3.8-4.1) vs 3.6 (95% CI 3.3-3.9), $P = .044$), attractive 3.6 (95% CI 3.4-3.7) vs 3.1 (95% CI 2.8-3.4; $P = .005$), ornate 2.8 (95% CI 2.6-2.9) vs 2.2 (95% CI 1.8-2.6; $P = .007$), and tasteful 3.7 (95% CI 3.6-3.9) vs 3.3 (95% CI 3.0-3.6; $P = .016$).

of vision. The average age of participants in the trial was 56 years, 49% were male, and 89% were Caucasian. There were no significant differences between participants and decliners in terms of race ($P = .13$) and mean age ($P = .08$). However, they did differ by gender, with 49% of participants being male compared with 68% of decliners ($P < .001$). There were no significant differences among the three study groups with respect to these demographic characteristics. No harms were observed for any patients; however, several patients in the group whose artwork was randomly selected expressed distaste for the image and/or color scheme of their painting.

Psychological Outcomes: Anxiety (STAI), Mood and Depression (ETI), and Sense of Control and/or Influence (Self-Designed Instrument)

There were no differences in anxiety for the primary comparison of artwork versus no artwork or the secondary comparison of choice versus no choice. Likewise, there were no differences in mood, depression, or sense of control and/or influence across the three groups.

Clinical Outcomes: Self-Reported Pain, LOS, and QOL (from EMR)

There were no differences in self-reported pain, LOS, or QOL across the three groups. With regard to LOS, the median (quartile 1 [Q1], quartile 3 [Q3]) stay was 6 days for the choice group (4.0, 12.0), 6 days for the no-choice group (5.0, 9.5), and 9.5 days for the group with no artwork (5.0, 20.0; see supplementary Table).

Perceptions of Hospital Environment (RA Survey)

As shown in Figure 2, participants who had art in their rooms generally had more positive impressions of the hospital environment than those who did not. For 6 of the 17 paired attributes, participants with artwork were significantly more likely to choose the positive attribute – specifically, such patients indicated their rooms were more interesting, colorful, pleasant, attractive, ornate, and tasteful. With regard to the other attributes, though not reaching levels of significance, the overall pattern clearly reflected a more positive impression of rooms with art than without it.

DISCUSSION

While having paintings in cancer inpatient rooms did not affect the psychological or clinical outcomes we assessed, patients who had paintings in their rooms had more positive impressions of the hospital environment. Given that healthcare administrators are under strong pressures to control costs while increasing care quality and patient satisfaction to maximize reimbursement, integrating local artwork into inpatient rooms may represent a simple and relatively inexpensive way (approximately $90 per room) to humanize clinical environments, systematically improve perceptions of the institution, and perhaps contribute to increased patient satisfaction scores. While more work must be done to establish a positive link between access to artwork and improved standardized patient satisfaction outcomes, our results suggest that there may be potential benefit in giving patients an opportunity to engage artwork as a therapeutic resource during the physical, emotional, and spiritual challenges that arise during inpatient treatment.

These findings also have implications for inpatients with illnesses other than cancer. Though we did not explicitly study noncancer patients, we know that nearly 40 million Americans are admitted annually to institutional care (ie, acute hospitalizations, rehabilitation hospitals, and skilled nursing facilities) and often find themselves in environments that can be stark and medicalized. We would anticipate that providing art in these patients' rooms would likewise improve perceptions of the institutions where they receive their inpatient medical care.

This study had several limitations that could affect the generalizability of our findings. First, it was difficult to enroll patients, with greater than 50% of persons approached declining to participate. Second, nonparticipants were more likely to be male, and this clearly provides a biased sample. Third, we have incomplete data for some patients who were unavailable or changed rooms during the study. Fourth, while each patient room had standardized features (eg, windows, televisions, etc.), there were logistical challenges with placing paintings in the exact same location (ie, in the patient's direct line of vision) in every hospital room because the shape, size, and idiosyncratic decorating of hospital rooms varied, so we were not able to fully control for all room décor features. Fifth, the study was conducted at a single site and only among patients with cancer; other populations could respond very differently. It is possible that other confounding factors (such as prior hospital ex-

perience, patient predilection for artwork, and usage of digital devices during hospitalization) could have affected outcomes, but these were not measured in this study.

In conclusion, as patient satisfaction continues to influence hospital reimbursement, identifying novel and effective approaches to improving patient perceptions can play a meaningful role in patient care. Future research should focus on different inpatient populations and venues; new strategies to effectively evaluate relevant clinical outcomes; comparisons with other nonpharmacological, arts-based interventions in inpatient settings (eg, music, creation of artwork, etc.); and assessment of aggregate scores on standardized patient satisfaction instruments (eg, Press Ganey, Hospital Consumer Assessment of Healthcare Providers and Systems). There may also be an additive benefit in providing "coaching" to healthcare providers on how to engage with patients regarding the artwork they have chosen. Such approaches might also examine the value of giving patients control over multiple opportunities to influence the aesthetics in their room versus a single opportunity during the course of their stay.

Acknowledgments

The authors would like to acknowledge the contributions of Lorna Davis, Lori Snyder, and Renee Stewart to this work.

Disclosure: This work was supported by funding from the National Endowment for the Arts (grant 14-3800-7008). ClinicalTrials.gov Identifier for Penn State Milton S. Hershey Medical Center Protocol Record STUDY00000378: NCT02357160. The authors report no conflicts of interest.

References

1. Mehta SJ. Patient Satisfaction Reporting and Its Implications for Patient Care. *AMA J Ethics*. 2015;17(7):616-621.
2. Hathorn KN. *A Guide to Evidence-based Art*. The Center for Health and Design; 2008. https://www.healthdesign.org/sites/default/files/Hathorn_Nanda_Mar08.pdf . Accessed November 5, 2017.
3. Sonke J, Rollins J, Brandman R, Graham-Pole J. The state of the arts in healthcare in the United States. *Arts & Health*. 2009;1(2):107-135.
4. Ulrich RS, Zimring C, Zhu X, et al. A Review of the Research Literature on Evidence-Based Healthcare Design. *HERD*. 2008;1(3):61-125.
5. Beukeboom CJ, Langeveld D, Tanja-Dijkstra K. Stress-reducing effects of real and artificial nature in a hospital waiting room. *J Altern Complement Med*. 2012;18(4):329-333.
6. Miller AC, Hickman LC, Lemaster GK. A Distraction Technique for Control of Burn Pain. *J Burn Care Rehabil*. 1992;13(5):576-580.
7. Diette GB, Lechtzin N, Haponik E, Devrotes A, Rubin HR. Distraction Therapy with Nature Sights and Sounds Reduces Pain During Flexible Bronchoscopy: A Complementary Approach to Routine Analgesic. *Chest*. 2003;123(3):941-948.
8. Tse MM, Ng JK, Chung JW, Wong TK. The effect of visual stimuli on pain threshold and tolerance. *J Clin Nursing*. 2002;11(4):462-469. 9. Vincent E, Battisto D, Grimes L. The Effects of Nature Images on Pain in a Simulated Hospital Patient Room. *HERD*. 2010;3(3):56-69.
10. Staricoff RL. Arts in health: the value of evaluation. *J R Soc Promot Health*. 2006;126(3):116-120.
11. Karnik M, Printz B, Finkel J. A Hospital's Contemporary Art Collection: Effects on Patient Mood, Stress, Comfort, and Expectations. *HERD*. 2014;7(3):60-77.
12. Suter E, Baylin D. Choosing art as a complement to healing. *Appl Nurs Res*. 2007;20(1):32-38.
13. Harris PB, McBride G, Ross C, Curtis L. A Place to Heal: Environmental Sources of Satisfaction among Hospital Patients. *J Appl Soc Psychol*. 2002;32(6):1276-1299.
14. Moss H, Donnellan C, O'Neill D. A review of qualitative methodologies used to explore patient perceptions of arts and healthcare. *Med Humanit*. 2012;38(2):106-109.
15. Corsini RJ, Ozaki BD. *Encyclopedia of psychology*. Vol. 1. New York: Wiley; 1994. State-Trait Anxiety Inventory.
16. Beck KR, Tan SM, Lum SS, Lim LE, Krishna LK. Validation of the emotion thermometers and hospital anxiety and depression scales in Singapore: Screening cancer patients for distress, anxiety and depression. *Asia Pac J Clin Oncol*. 2016;12(2):e241-e249.
17. Lohr VI, Pearson-Mims CH. Physical discomfort may be reduced in the presence of interior plants. *HortTechnology*. 2000;10(1):53-58.
18. Semantic Differential. http://psc.dss.ucdavis.edu/sommerb/sommerdemo/scaling/semdiff.htm. Accessed November 5, 2017.

The Use of Individual Provider Performance Reports by US Hospitals

Joshua A. Rolnick, MD, JD[1,2,3,4]*, Kira L. Ryskina, MD, MS[1,3]

[1]*Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania;* [2]*National Clinician Scholars Program, University of Pennsyvania, Philadelphia, Pennsylvania;* [3]*Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, Pennsylvania;* [4]*Corporal Michael J. Crescenz Veterans Affairs Medical Center, Philadelphia, Pennsylvania.*

Medicare reimbursement for hospitals is increasingly tied to performance. The use of individual provider performance reports offers the potential to improve clinical outcomes through social comparison, and isolated cases of clinical dashboard uses at specific institutions have been previously reported. However, little is known about overall trends in how hospitals use the electronic health record to track and provide feedback on provider performance. We used data from 2013 to 2015 from the American Hospital Association (AHA) Annual Survey Information Technology Supplement, which asked hospitals if they have used electronic data to create performance profiles. We linked these data to AHA Annual Survey responses for all general adult and pediatric hospitals. Multivariable logistic regression was used to model the odds of use as a function of hospital characteristics. In 2015, 65.8% of the 2,334 respondents used performance profiles, whereas 59.3% of the 2,077 respondents used them in 2013. Report use was associated with nonprofit status (odds ratio [OR], 2.77; 95% confidence interval [CI], 1.94-3.95) compared to for-profit, large hospital size (OR, 2.37; 95% CI, 1.56-3.60) compared to small size, highest quartile of bed-adjusted expenditures compared to bottom quartile (OR, 2.09; 95% CI, 1.55-2.82; $P < .01$), and participation in a health maintenance organization (OR, 1.50; 95% CI, 1.17-1.90; $P < .01$) or bundled payment program (OR, 1.61; 95% CI, 1.18-2.19; $P < .01$). While a majority of hospitals now use such profiles, more than a third do not. The hospitals that do not use performance profiles may be less well positioned to adapt to value-based payment reforms. *Journal of Hospital Medicine* 2018;13:562-565. Published online first February 7, 2018. © 2018 Society of Hospital Medicine

Reimbursement for hospitals and physicians is increasingly tied to performance.[1] Bundled payments, for example, require hospitals to share risk for patient outcomes. Medicare bundled payments are becoming mandatory for some surgical and medical conditions, including joint replacement, acute myocardial infarction, and coronary artery bypass graft surgery.[2] Value-based payment is anticipated to become the norm as Medicare and private payers strive to control costs and improve outcomes. Although value-based reimbursement for hospitals targets hospital-level costs and outcomes, we know that variations at the level of individual providers explain a considerable proportion of variation in utilization and outcomes.[3] However, physicians often lack awareness of their own practice patterns and relative costs, and successful participation in new payment models may require an investment by hospitals in the infrastructure needed to measure and provide feedback on performance to individual providers to affect their behavior.[4,5]

Electronic health record (EHR)-based reports or "dashboards" have been proposed as one potential tool to provide individualized feedback on provider performance.[6] Individual provider performance profiles (IPPs) offer the potential to provide peer comparisons that may adjust individual behavior by correcting misperceptions about norms.[7] Behavioral economic theory suggests that individual performance data, if combined with information on peer behavior and normative goals, may be effective in changing behavior.[8] Several studies have reported the effects of specific efforts to use IPPs, showing that such reports can improve care in certain clinical areas. For example, individual provider dashboards have been associated with better outcomes for hospitalized patients, such as increased compliance with recommendations for prophylaxis of venous thromboembolism, although evidence in other areas of practice is mixed.[9,10] A randomized controlled trial of peer comparison feedback reduced inappropriate antibiotic prescribing for upper respiratory infections by 11% among internists.[11]

Despite the promise of individualized feedback to optimize behavior, however, little has been reported on trends in the use of IPPs on a population level. It is unknown whether their use is common or rare, or what hospital characteristics are associated with adoption. Such information would help guide future efforts to promote IPP use and understand its effect on practice. We used data from a nationally representative survey of United States hospitals to examine the use of individual provider-level performance profiles.

***Address for correspondence:** Joshua A. Rolnick, MD, JD, University of Pennsylvania, National Clinician Scholars Program, Blockley Hall, 13th Floor, 423 Guardian Drive, Philadelphia, PA 19104-6021; Telephone: 617-538-5191; Fax: 610-642-4380; E-mail: rolnick@pennmedicine.upenn.edu

Additional Supporting Information may be found in the online version of this article.

Received: July 14, 2017; **Revised:** October 14, 2017; **Accepted:** October 24, 2017

 DOI 10.12788/jhm.2922

METHODS

We used data from the American Hospital Association (AHA) Annual Survey Information Technology (IT) Supplement, which asked respondents to indicate whether they have used electronic clinical data from the EHR or other electronic system in their hospital to create IPPs. The AHA survey is sent annually to all US operating hospitals. Survey results are supplemented by data from the AHA registration database, US Census Bureau, hospital accrediting bodies, and other organizations. The AHA IT supplement is also sent yearly to each hospital's chief executive officer, who assigns it to the most knowledgeable person in the institution to complete.[12]

We linked data on IPP use to AHA Annual Survey responses on hospital characteristics for all general adult and pediatric hospitals. Multivariable logistic regression was used to model the odds of individual provider performance profile use as a function of hospital characteristics, including ownership (nonprofit, for profit, or government), geographic region, teaching versus nonteaching status, rural versus urban location, size, expenditures per bed, proportion of patient days covered by Medicaid, and risk-sharing models of reimbursement (participation in a health maintenance organization or bundled payments program). Variables were chosen a priori to account for important characteristics of US hospitals (eg, size, teaching status, and geographic location). These were combined with variables representing risk-sharing arrangements based on the hypothesis that hospitals whose payments are at greater risk would be more likely to invest in tracking provider performance. We eliminated any variable with an item nonresponse rate greater than 15%, which resulted in elimination of two variables representing hospital revenue from capitated payments and any risk-sharing arrangement, respectively. All other variables had item nonresponse rates of 0%, except for 4.7% item nonresponse for the bundled payments variable.

We also measured the trend in individual provider performance report use between 2013 and 2015 by estimating the linear probability between IPP use and year. A *P* value less than .05 was considered statistically significant.

Because past work has demonstrated nonresponse bias in the AHA Survey and IT Supplement, we performed additional analyses using nonresponsive weights based on hospital characteristics. Weighting methodology was based on prior work with the AHA and AHA IT surveys.[13,14] Weighting exploits the fact that a number of hospital characteristics are derived from sources outside the survey and thus are available for both respondents and nonrespondents. We created nonresponse weights based on a logistic regression model of survey response as a function of hospital characteristics (ownership, size, teaching status, systems membership, critical access hospital, and geographic region). Our findings were similar for weighted and nonweighted models and nonweighted estimates are presented throughout.

The University of Pennsylvania Institutional Review Board exempted this study from review. Analyses were performed using Stata statistical software, version 14.0 (StataCorp, College Station, Texas).

FIG. Percent of United States hospitals with individual provider profiles

RESULTS

In 2015, 2,334 general hospitals completed all questions of interest in both surveys. Among respondents, 65.8% used individual provider performance reports. Individual provider performance use increased by 3.3% each year from 2013 to 2015 ($P = .006$; Figure).

The table shows the association between hospital characteristics and the odds of individual provider performance report use. Report use was associated with nonprofit status (odds ratio [OR], 2.77; 95% confidence interval [CI], 1.94-3.95; $P < .01$) compared to for-profit, large hospital size (OR, 2.37; 95% CI, 1.56-3.60; $P < .01$) compared to small size, highest (OR, 2.09; 95% CI, 1.55-2.82; $P < .01$) and second highest (OR, 1.43; 95% CI, 1.08-1.89; $P = .01$) quartiles of bed-adjusted expenditures compared to the bottom quartile, and West geographic region compared to Northeast (OR, 2.07; 95% CI, 1.45-2.95; $P < .01$). Individual provider performance use was also independently associated with participation in a health maintenance organization (OR, 1.50; 95% CI, 1.17-1.90; $P < .01$) or bundled payment program (OR, 1.61; 95% CI, 1.18-2.19; $P < .01$), controlling for other covariates. Adjustment for nonresponse bias did not change any coefficients by more than 10% (supplementary Table).

DISCUSSION

We found that a large and increasing proportion of US hospitals reported using electronic data to measure individual provider performance. Hospitals that reported IPP use tended to be larger and have higher expenditures than hospitals that did not use IPPs. Adjusting for other hospital characteristics, participation in a bundled payment program was associated with greater odds of using IPPs. To our knowledge, our study is the first population-level analysis of IPP use by US hospitals.

The Medicare Access and Children Health Insurance Program Reauthorization Act is accelerating the shift from quantity based toward value-based reimbursement. The proficient adoption of IT by healthcare providers has been cited as an important factor in adapting to new payment models.[15] Phy-

TABLE. **Hospital Characteristics Associated with Use of IPPs**[a]

Characteristics	All Hospitals (N = 2334)	Use IPP Reports (N = 1567)	Do Not Use IPP Reports (N = 767)	OR (95% CI)	*P* value
Type of hospital					
For-profit private	160 (6.9)	77 (4.9)	83 (10.8)	Reference	
Nonprofit private	1609 (68.9)	1192 (76.1)	417 (54.4)	2.77 (1.94-3.95)	<.001
Government (federal or nonfederal)	565 (24.2)	298 (19.0)	267 (34.8)	1.43 (0.98-2.09)	.060
Geographic Region					
Northeast	317 (13.6)	212 (13.5)	105 (13.7)	Reference	
Midwest	886 (38.0)	577 (36.8)	309 (40.3)	1.74 (1.30-2.36)	<.001
South	724 (31.0)	482 (30.8)	242 (31.6)	1.93 (1.41-2.64)	<.001
West	407 (17.4)	296 (18.9)	111 (14.5)	2.07 (1.45-2.95)	<.001
Teaching hospital	626 (26.8)	500 (31.9)	126 (16.4)	1.06 (0.80-1.41)	.673
Size					
Small (<99 beds)	1023 (43.8)	571 (36.4)	452 (58.9)	Reference	
Medium (100-399)	991 (42.5)	728 (46.5)	263 (34.3)	1.77 (1.39-2.26)	<.001
Large (≥400 beds)	320 (13.7)	268 (17.1)	52 (6.8)	2.37 (1.56-3.60)	<.001
Expenditures per 100 beds					
First quartile (mean $4,209)	533 (22.8)	277 (17.7)	256 (33.4)	Reference	
Second quartile (mean $8,003)	575 (24.6)	377 (24.1)	198 (25.8)	1.28 (0.98-1.67)	.067
Third quartile (mean $11,255)	608 (26.0)	431 (27.5)	177 (23.1)	1.43 (1.08-1.89)	.012
Fourth quartile (mean $20,091)	618 (26.5)	482 (30.8)	136 (17.7)	2.09 (1.55-2.82)	<.001
Rural location[b]	496 (21.3)	245 (15.6)	251 (32.7)	0.79 (0.61-1.02)	.069
Medicaid patient days[c]	20.7 (16.5)	21.1 (15.8)	19.8 (18.1)	1.05 (0.58-1.90)	.880
Participates in a health maintenance organization	525 (22.5)	416 (26.5)	109 (14.2)	1.50 (1.18-2.19)	.001
Participates in a bundled payment program	408 (17.5)	340 (21.7)	68 (8.9)	1.61 (1.18-2.19)	.002

[a]*P* values are derived from a multivariate logistic regression model of IPP use as a function of hospital characteristics. *P* values are based on the Wald test of the null hypothesis that the OR for each predictor is equal to 1. The Hosmer-Lemeshow goodness-of-fit test showed no statistical evidence of lack of fit (*P* = .34). The variance inflation factor was less than 10 for all covariates (mean VIF 2.71), suggesting a lack of statistical evidence of multicollinearity.

[b]Any location that is not part of a micropolitan or metropolitan core-based statistical area, from the US Census.

[c]Medicaid patient days as a percentage of total inpatient days, expressed as mean (standard deviation). All other values correspond to number (percentage).

NOTE: Abbreviations: CI, confidence interval; IPP, individual provider performance profile; OR, odds ratio; VIF, variance inflation factor.

sicians, and in particular hospitalists, who practice in an inpatient environment, may not directly access financial incentives aimed to adapt performance for value-based reimbursement. They may also have difficulty assessing their performance relative to peers and longitudinally over time. Individualized EHR-based provider-level performance reports offer one option for hospitals to measure performance and provide comparative feedback at the individual physician level. Our findings show that, in fact, a majority of US hospitals have made investments in the infrastructure necessary to create such profiles.

Nevertheless, a third of the hospitals surveyed have not adopted individualized provider performance profiles. If meeting efficiency and outcomes goals for value-based payments necessitates changes to individual provider behavior, those hospitals may be less well positioned to benefit from value-based payment models that incentivize hospitals for efficiency and outcomes. Furthermore, while we observe widespread adoption of individual performance profiles, it is unclear whether those were used to provide feedback to providers, and if so, how the feedback provided may influence its effect on behavior. Behavioral economics theory suggests, for example, that publicly reporting performance compared to peers provides stronger incentives for behavior change than "blinded" personalized reports.[16]

Our study has important limitations. We cannot exclude the possibility that unmeasured variables help explain individual provider performance adoption. These omitted variables may confound the association between hospital characteristics and individual provider performance adoption observed in this study. We were also unable to establish causality between bundled payments and individual provider performance profile use. For instance, hospitals may elect to make investments in IT infrastructure to enable individual provider performance profile adoption in anticipation of bundled payment reforms. Alternatively, the availability of IPPs may have led hospitals to enter bundled payments reimbursement arrangements. In addition, we are unable to describe how individual provider performance use affects physician practice or healthcare delivery more broadly. Finally, we are also unable to account for other sources of performance data. For exam-

ple, some physician may receive data from their physician practice groups.

Our study suggests several avenues for future research. First, more work is needed to understand why certain types of hospitals are more likely to use IPPs. Our findings indicate that IPP use may be partly a function of hospital size and resources. However, other factors not measured here may play an important role as well, such as institutional culture. Institutions with a focus on informatics and strong IT leadership may be more likely to use their EHR to monitor performance. Second, further research should explore in greater depth how profiles are used. Future research should evaluate, for example, how hospitals are using behavioral economic principles, such as peer comparison, to motivate behavior change, and if such techniques have successfully influenced practice and patient outcomes. Ultimately, multicentered, randomized evaluations of IPP use may be necessary to understand their risks and evaluate their effect on patient outcomes. This work is necessary to inform policy and practice as hospitals transition from fee-for-service to value-based reimbursement.

In sum, we observed increasing adoption of individualized electronic provider performance profiles by US hospitals from 2013 to 2015. Hospitals that did not use IPPs were more likely to be small, for profit, and less likely to participate in bundled payment programs. Those hospitals may be less well positioned to track provider performance and implement incentives for provider behavior changes needed to meet targets for value-based reimbursement.

Disclosure: Dr. Rolnick is a consultant to Tuple Health, Inc. and was a part-time employee of Acumen, LLC outside the submitted work. Dr. Ryskina has nothing to disclose.

References

1. Hussey PS, Liu JL, White C. The Medicare Access And CHIP Reauthorization Act: effects On Medicare payment policy and spending. *Health Aff*. 2017;36(4):697-705.
2. Navathe AS, Song Z, Emanuel EJ. The next generation of episode-based payments. *JAMA*. 2017;317(23):2371-2372.
3. Tsugawa Y, Jha AK, Newhouse JP, Zaslavsky AM, Jena AB. Variation in physician spending and association with patient outcomes. *JAMA Intern Med*. 2017;177(5):675-682.
4. Saint S, Wiese J, Amory JK, et al. Are physicians aware of which of their patients have indwelling urinary catheters? *Am J Med*. 2000;109(6):476-480.
5. Saturno PJ, Palmer RH, Gascón JJ. Physician attitudes, self-estimated performance and actual compliance with locally peer-defined quality evaluation criteria. *Int J Qual Health Care J Int Soc Qual Health Care*. 1999;11(6):487-496.
6. Mehrotra A, Sorbero MES, Damberg CL. Using the lessons of behavioral economics to design more effective pay-for-performance programs. *Am J Manag Care*. 2010;16(7):497-503.
7. Emanuel EJ, Ubel PA, Kessler JB, et al. Using behavioral economics to design physician Incentives that deliver high-value care. *Ann Intern Med*. 2016;164(2):114-119.
8. Liao JM, Fleisher LA, Navathe AS. Increasing the value of social comparisons of physician performance using norms. *JAMA*. 2016;316(11):1151-1152.
9. Michtalik HJ, Carolan HT, Haut ER, et al. Use of provider-level dashboards and pay-for-performance in venous thromboembolism prophylaxis. *J Hosp Med*. 2015;10(3):172-178.
10. Kurtzman G, Dine J, Epstein A, et al. Internal medicine resident engagement with a laboratory utilization dashboard: mixed methods study. *J Hosp Med*. 12(9):743-746.
11. Linder JA, Schnipper JL, Tsurikova R, et al. Electronic health record feedback to improve antibiotic prescribing for acute respiratory infections. *Am J Manag Care*. 2010;16 (12 Suppl HIT):e311-e319.
12. Jha AK, DesRoches CM, Campbell EG, et al. Use of electronic health records in U.S. hospitals. *N Engl J Med*. 2009;360(16):1628-1638.
13. Walker DM, Mora AM, Scheck McAlearney A. Accountable care organization hospitals differ in health IT capabilities. *Am J Manag Care*. 2016;22(12):802-807.
14. Adler-Milstein J, DesRoches CM, Furukawa MF, et al. More than half of US hospitals have at least a basic EHR, but stage 2 criteria remain challenging for most. *Health Aff* 2014;33(9):1664-1671.
15. Porter ME. A strategy for health care reform—toward a value-based system. *N Engl J Med*. 2009;361(2):109-112.
16. Navathe AS, Emanuel EJ. Physician peer comparisons as a nonfinancial strategy to improve the value of care. *JAMA*. 2016;316(17):1759-1760.

Collaborations with Pediatric Hospitalists: National Surveys of Pediatric Surgeons and Orthopedic Surgeons

Rebecca E. Rosenberg, MD, MPH, FAAP[1]*, Joshua M. Abzug, MD, FAAOS[2], David I. Rappaport, MD[3], Mark V. Mazziotti, MD, FACS[4], M. Wade Shrader, MD, FAAOS[5], David Zipes, MD[6], Benedict Nwomeh, MD, FACS[7], Lisa McLeod, MD, MSCE[8], and the AAP Section on Hospital Medicine Surgical Care Subcommittee

[1]NYU School of Medicine, New York, New York; [2]University of Maryland Medical Center, Baltimore, Maryland; [3]Nemours Alfred I. Dupont Hospital for Children, Wilmington, Delaware; [4]Texas Children's Hospital, Houston, Texas; [5]University of Mississippi, Oxford, Mississippi; [6]Peyton Manning Children's Hospital, Indianapolis, Indiana; [7]Nationwide Children's Hospital, Columbus, Ohio; [8]Children's Hospital Colorado, Aurora, Colorado.

To understand characteristics of pediatric hospitalist (PH) involvement in the care of children admitted to surgical services and explore surgeons' perspectives of PH effectiveness, we conducted a cross-sectional, web-based survey of pediatric surgical (PS) and pediatric orthopedic subspecialists (OS) from professional organizations. We used basic analyses to compare responses between the two surgical groups. The initial response rate was 48% (291/606) for PS and 59% (415/706) for OS. Among 185 PS and 212 OS unique programs, PH were routinely engaged (69% and 75%) in the care of surgical patients, particularly in patients with medical complexity (64% PS vs 81% OS; $P = .003$). PS and OS perceived positive PH impact on care coordination and comorbidity management but little on pain management or length of stay. OS were more likely than PS to view PH involvement positively (64% vs 42%; $P < .001$). Further research on care models, especially for children with medical complexity, is needed. *Journal of Hospital Medicine* 2018;13:566-569. Published online first February 6, 2018. © 2018 Society of Hospital Medicine

Pediatric expertise is critical in caring for children during the perioperative and postoperative periods.[1,2] Some postoperative care models involve pediatric hospitalists (PH) as collaborators for global care (comanagement),[3] as consultants for specific issues, or not at all.

Single-site studies in specific pediatric surgical populations[4-7]and medically fragile adults[8] suggest improved outcomes for patients and systems by using hospitalist-surgeon collaboration. However, including PH in the care of surgical patients may also disrupt systems. No studies have broadly examined the clinical relationships between surgeons and PH.

The aims of this cross-sectional survey of United States pediatric surgeons (PS) and pediatric orthopedic surgeons (OS) were to understand (1) the prevalence and characteristics of surgical care models in pediatrics, specifically those involving PH, and (2) surgeons' perceptions of PH in caring for surgical patients.

METHODS

The target US surgeon population was the estimated 850 active PS and at least 600 pediatric OS.[9] Most US PS (n = 606) are affiliated with the American Academy of Pediatrics (AAP) Section on Surgery (SoSu), representing at least 200 programs. Nearly all pediatric OS belong to the Pediatric Orthopedic Society of North America (POSNA) (n = 706), representing 340 programs; a subset (n = 130) also belong to the AAP SoSu.

Survey Development and Distribution

Survey questions were developed to elicit surgeons' descriptions of their program structure and their perceptions of PH involvement. For programs with PH involvement, program variables included primary assignment of clinical responsibilities by service line (surgery, hospitalist, shared) and use of a written service agreement, which defines each service's roles and responsibilities.

The web-based survey, created by using Survey Monkey (San Mateo, California), was pilot tested for usability and clarity among eight surgeons and one PH. The survey had logic points around involvement of hospitalists and multiple hospital affiliations (supplemental Appendix A). The survey request with a web-based link was e-mailed three times to surgical and orthopedic distribution outlets, endorsed by organizational leadership. Respondents' hospital ZIP codes were used as a proxy for program. If there was more than one complete survey response per ZIP code, one response with complete data was randomly selected to ensure a unique entry per program.

Classification of Care Models

Each surgical program was classified into one of the following three categories based on reported care of primary surgical patients: (1) comanagement, described as PH writing orders and/

***Address for correspondence:** Rebecca Rosenberg, MD, MPH, 550 First Ave, New York, NY 10016; Telephone: 212-263-0959; Fax: 212-263-0557; E-mail: Rebecca.rosenberg@nyumc.org

Additional Supporting Information may be found in the online version of this article.

Received: August 1, 2017; **Revised**: October 11, 2017; **Accepted**: October 24, 2017

 DOI 10.12788/jhm.2921

TABLE. **Survey Responses Grouped by Hospitals and by Individual Surgeons**

HOSPITAL PROGRAMS	Pediatric General Surgery Programs	Pediatric Orthopedic Programs	*P* value, χ^2
Total	185	212	
Primary site	145	171	
Secondary site	40	41	
Number of pediatric beds[a]			<.001
<60	14 (10%)	50 (26%)	
61-100	19 (13%)	31 (16%)	
101-200	43 (30%)	39 (20%)	
>200	65 (45%)	73 (38%)	
Hospital type			
Freestanding	93 (50%)	117 (55%)	.340
CH within general	54 (29%)	50 (24%)	
General tertiary/no CH	24 (13%)	22 (10%)	
Community/other	14 (8%)	23 (11%)	
Primary team includes			
Surgical residents	181 (98%)	180 (85%)	<.001
Surgical advanced providers	168 (91%)	136 (64%)	<.001
Type of PH involvement[a]			<.001
No involvement	49 (28%)	43 (21%)	
PH involvement	127 (69%)	158 (75%)	
Follows/no orders (consult)	54 (31%)	20 (10%)	
Follows/writes orders (comanage)	73 (42%)	138 (69%)	
SURGEONS			
Unique respondents	252	340	ns
Posttraining >10 years	158 (66%)	259 (69%)	
>75% practice children <18 years old	240 (92%)	228 (95%)	

[a]Primary hospitals.
NOTE: Data may not add up to 100% because of incomplete or inconsistent answers.
Abbreviations: CH, Children's Hospital; ns, not significant; PH, pediatric hospitalist.

or functioning as the primary service; (2) consultation, described as PH providing clinical recommendations only; and (3) no PH involvement, described as "rarely" or "never" involving PH.

Clinical Responsibility Score

To estimate the degree of hospitalist involvement, we devised and calculated a composite score of service responsibilities for each program. This score involved the following seven clinical domains: management of fluids or nutrition, pain, comorbidities, antibiotics, medication dosing, wound care, and discharge planning. Scores were summed for each domain: 0 for surgical team primary responsibility, 1 for shared surgical and hospitalist responsibility, and 2 for hospitalist primary responsibility. Composite scores could range from 0 to 14; lower scores represented a stronger tendency for surgeon management, and higher scores represented a stronger tendency toward PH management.

Data Analysis

For data analysis, simple exploratory tests with χ^2 analysis and Student *t* tests were performed by using Stata 14.2 (StataCorp LLC, College Station, Texas) to compare differences by surgical specialty programs and individuals by role assignment and perceptions of PH involvement.

The NYU School of Medicine Institutional Review Board approved this study.

RESULTS

Respondents and Programs

Of the estimated 606 PS in the AAP SoSu, 291 (49%) US-based surgeons (PS) responded with 251 (41%) sufficiently completed surveys (Table). The initial and completed survey response rate for pediatric OS through the POSNA listserv was 58% and 48% (340/706), respectively. These respondents represented 185 unique PS programs and 212/340 (62%) unique OS programs in the US (supplemental Appendix B).

Among the unique 185 PS programs and 212 OS programs represented, PH were often engaged in the care of primary surgical patients (Table).

Roles of PH in Collaborative Programs

Among programs that reported any hospitalist involvement (PS, n = 100; OS, n = 157), few (≤15%) programs involved hospitalists

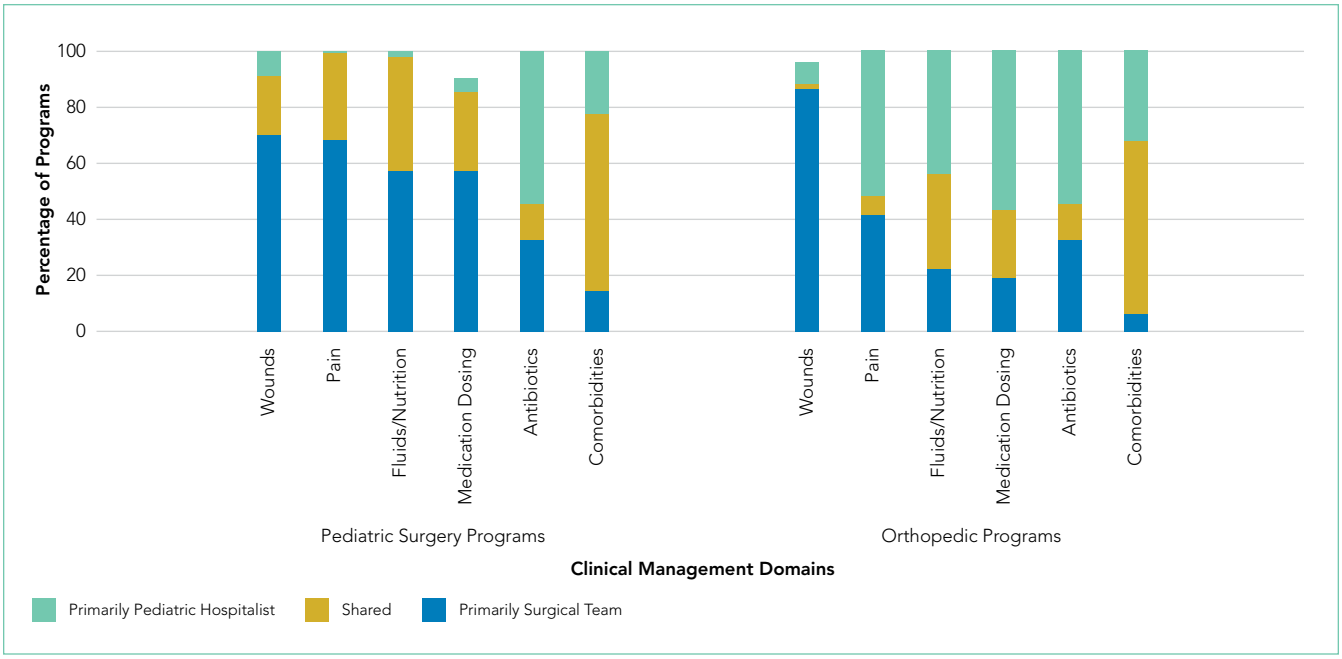

FIG 1. Distribution of clinical domain management among pediatric surgery (n=100) and orthopedic surgery programs (n=157) by clinical team responsibility: primarily surgical, hospitalist, or shared.

with all patients. Pediatric OS programs were significantly more likely than pediatric surgical programs to involve PH for healthy patients with any high-risk surgery (27% vs 9%; $P = .001$). Most PS (64%) and OS (83%) reported involving PH for all medically complex patients, regardless of surgery risk ($P = .003$).

In programs involving PH, few PS (11%) or OS programs (16%) reported using a written service agreement.

Care of Surgical Patients in PH-involved programs

Both PS and OS programs with hospitalist involvement reported that surgical teams were either primarily responsible for, or shared with the hospitalist, most aspects of patient care, including medication dosing, nutrition, and fluids (Figure). PH management of antibiotic and nonsurgical comorbidities was higher for OS programs than PS programs.

Composite clinical responsibility scores ranged from 0 to 8, with a median score of 2.3 (interquartile range [IQR] 0-3) for consultation programs and 5 (IQR 1-7) for comanagement programs. Composite scores were higher for OS (7.4; SD 3.4) versus PS (3.3; SD 3.4) programs ($P < .001$; 95% CI, 3.3-5.5; supplemental Appendix C).

Surgeons' Perspectives on Hospitalist Involvement

Surgeons in programs without PH involvement viewed PH overall impact less positively than those with PH (27% vs 58%). Among all surgeons surveyed, few perceived positive (agree/strongly agree) PH impact on pain management (<15%) or decreasing LOS (≤15%; supplemental Appendix D).

Most surgeons (n = 355) believed that PH financial support should come from separate billing (patient fee; 48%) or hospital budget (36%). Only 17% endorsed PH receiving part of the surgical global fee, with no significant difference by surgical specialty or current PH involvement status.

DISCUSSION

This study is the first comprehensive assessment of surgeons' perspectives on the involvement and effectiveness of PH in the postoperative care of children undergoing inpatient general or orthopedic surgeries. The high prevalence (>70%) of PH involvement among responding surgical programs suggests that PH comanagement of hospitalized patients merits attention from providers, systems, educators, and payors.

Collaboration and Roles are Correlated with Surgical Specialty and Setting

Forty percent of inpatient pediatric surgeries occur outside of children's hospitals.[10] We found that PH involvement was higher at smaller and general hospitals where PH may provide pediatric expertise when insufficient pediatric resources, like pain teams, exist.[7] Alternately, some quaternary centers have dedicated surgical hospitalists. The extensive involvement of PH in the bulk of certain clinical care domains, especially care coordination, in OS and in many PS programs (Figure) suggests that PH are well integrated into many programs and provide essential clinical care.

In many large freestanding children's hospitals, though, surgical teams may have sufficient depth and breadth to manage most aspects of care. There may be an exception for care coordination of medically complex patients. Care coordination is a patient- and family-centered care best practice,[11] encompasses integrating and aligning medical care among clinical services, and is focused on shared decision making and communication. High-quality care coordination processes are of great value to patients and families, especially in medically complex children,[11] and are associated with improved transitions from hospital to home.[12] Well-planned transitions likely decrease

these special populations' postoperative readmission risk, complications, and prolonged length of stay.[13] Reimbursement for these services could integrate these contributions needed for safe and patient-centered pediatric inpatient surgical care.

Perceptions of PH Impact

The variation in perception of PH by surgical specialty, with higher prevalence as well as higher regard for PH among OS, is intriguing. This disparity may reflect current training and clinical expectations of each surgical specialty, with larger emphasis on medical management for surgical compared with orthopedic curricula (www.acgme.org).

While PS and OS respondents perceived that PH involvement did not influence length of stay, pain management, and resource use, single-site studies suggest otherwise.[4,8,14] Objective data on the impact of PH involvement on patient and systems outcomes may help elucidate whether this is a perceived or actual lack of impact. Future metrics might include pain scores, patient centered care measures on communication and coordination, patient complaints and/or lawsuits, resource utilization and/or cost, readmission, and medical errors.

This study has several limitations. There is likely a (self) selection bias by surgeons with either strongly positive or negative views of PH involvement. Future studies may target a random sampling of programs rather than a cross-sectional survey of individual providers. Relatively few respondents represented community hospitals, possibly because these facilities are staffed by general OS and general surgeons[10] who were not included in this sample.

CONCLUSION

Given the high prevalence of PH involvement in caring for surgical pediatric patients in varied settings, the field of pediatric hospital medicine should support increased PH training and standardized practice around perioperative management, particularly for medically complex patients with increased care coordination needs. Surgical comanagement, including interdisciplinary communication skills, deserves inclusion as a PH core competency and as an entrustable professional activity for pediatric hospital medicine and pediatric graduate medical education programs,[15] especially orthopedic surgeries.

Further research on effective and evidence-based pediatric postoperative care and collaboration models will help PH and surgeons to most effectively and respectfully partner to improve care.

Acknowledgments

The authors thank the members of the AAP Section on Hospital Medicine Surgical Care Subcommittee, AAP SOHM leadership, and Ms. Alexandra Case.

Disclosure: The authors have no conflicts of interest relevant to this manuscript to report.

Funding: This study was supported in part by the Agency for Health Care Research and Quality (LM, R00HS022198).

References

1. Task Force for Children's Surgical Care. Optimal resources for children's surgical care in the United States. *J Am Coll Surg*. 2014;218(3):479-487, 487.e1-4.
2. Section on Hospital Medicine, American Academy of Pediatrics. Guiding principles for pediatric hospital medicine programs. *Pediatrics*. 2013;132(4):782-786.
3. Freiburg C, James T, Ashikaga T, Moalem J, Cherr G. Strategies to accommodate resident work-hour restrictions: Impact on surgical education. *J Surg Educ*. 2011;68(5):387-392.
4. Pressel DM, Rappaport DI, Watson N. Nurses' assessment of pediatric physicians: Are hospitalists different? *J Healthc Manag*. 2008;53(1):14-24; discussion 24-25.
5. Simon TD, Eilert R, Dickinson LM, Kempe A, Benefield E, Berman S. Pediatric hospitalist comanagement of spinal fusion surgery patients. *J Hosp Med*. 2007;2(1):23-30.
6. Rosenberg RE, Ardalan K, Wong W, et al. Postoperative spinal fusion care in pediatric patients: Co-management decreases length of stay. *Bull Hosp Jt Dis (2013)*. 2014;72(3):197-203.
7. Dua K, McAvoy WC, Klaus SA, Rappaport DI, Rosenberg RE, Abzug JM. Hospitalist co-management of pediatric orthopaedic surgical patients at a community hospital. *Md Med*. 2016;17(1):34-36.
8. Rohatgi N, Loftus P, Grujic O, Cullen M, Hopkins J, Ahuja N. Surgical comanagement by hospitalists improves patient outcomes: A propensity score analysis. *Ann Surg*. 2016;264(2):275-282.
9. Poley S, Ricketts T, Belsky D, Gaul K. Pediatric surgeons: Subspecialists increase faster than generalists. *Bull Amer Coll Surg*. 2010;95(10):36-39.
10. Somme S, Bronsert M, Morrato E, Ziegler M. Frequency and variety of inpatient pediatric surgical procedures in the United States. *Pediatrics*. 2013;132(6):e1466-e1472.
11. Frampton SB, Guastello S, Hoy L, Naylor M, Sheridan S, Johnston-Fleece M, eds. *Harnessing Evidence and Experience to Change Culture: A Guiding Framework for Patient and Family Engaged Care*. Washington, DC: National Academies of Medicine; 2017.
12. Auger KA, Kenyon CC, Feudtner C, Davis MM. Pediatric hospital discharge interventions to reduce subsequent utilization: A systematic review. *J Hosp Med*. 2014;9(4):251-260.
13. Simon TD, Berry J, Feudtner C, et al. Children with complex chronic conditions in inpatient hospital settings in the united states. *Pediatrics*. 2010;126(4):647-655.
14. Rappaport DI, Adelizzi-Delany J, Rogers KJ, et al. Outcomes and costs associated with hospitalist comanagement of medically complex children undergoing spinal fusion surgery. *Hosp Pediatr*. 2013;3(3):233-241.
15. Jerardi K, Meier K, Shaughnessy E. Management of postoperative pediatric patients. *MedEdPORTAL*. 2015;11:10241. doi:10.15766/mep_2374-8265.10241.

Trends in Inpatient Admission Comorbidity and Electronic Health Data: Implications for Resident Workload Intensity

Amanda V. Clark, MD[1], Charles M. LoPresti, MD[2,3], Todd I. Smith, MD, FHM[2,3]*

[1]*Department of Medicine, University of Mississippi Medical Center, Jackson, Mississippi;* [2]*Medical Service, Louis Stokes Cleveland VA Medical Center, Cleveland, Ohio;* [3]*Department of Medicine, Case Western Reserve University School of Medicine, Cleveland, Ohio.*

In the era of duty-hour regulations, there is increasing concern regarding resident workload compression. We conducted a retrospective, observational assessment of all internal medicine resident admissions to a Veterans Affairs hospital over a 15-year period to evaluate several admission components that impact resident workload and workload intensity, including electronic health record (EHR) data burden and patient comorbidity. A total of 67,346 admissions were included in the analysis. Mean patient comorbidity, as measured by the Charlson Comorbidity Index, increased throughout the study period. EHR data burden, measured by numbers of notes, medications, and discharge summaries available per patient at the time of admission, also increased over the study period. These findings suggest that EHR data burden and comorbidity have increased over time, which impacts resident workload in the era of duty hour restrictions. *Journal of Hospital Medicine* 2018;13:570-572. Published online first March 26, 2018. © 2018 Society of Hospital Medicine

Since the Accreditation Council for Graduate Medical Education (ACGME) posed new duty hour regulations in 2003 and again in 2011, there have been concerns that the substantial compression of resident workload may have resulted in a negative learning environment.[1-3] Residents are now expected to complete more work in a reduced amount of time and with less flexibility.[4] In addition to time constraints, the actual work of a resident today may differ from that of a resident in the past, especially in the area of clinical documentation.[5] Restricting resident work hours without examining the workload may result in increased work intensity and counter the potential benefits of working fewer hours.[6] Measuring workload, as well as electronic health record (EHR)–related stress, may also help combat burnout in internal medicine.[7] There are many components that influence resident workload, including patient census, patient comorbidities and acuity, EHR data and other available documentation, and ancillary tasks and procedures.[7] We define resident workload intensity as the responsibilities required to provide patient care within a specified time. There is a paucity of objective data regarding the workload intensity of residents, which are essential to graduate medical education reform and optimization. Patient census, ancillary responsibilities, number of procedures, and conference length and frequency are some of the variables that can be adjusted by each residency program. As a first step to objective measurement of resident workload intensity, we endeavored to evaluate the less easily residency program–controlled workload components of patient comorbidity and EHR data the time of patient admission.

METHODS

We conducted an observational, retrospective assessment of all admissions to the Louis Stokes Cleveland VA Medical Center (LSCVAMC) internal medicine service from January 1, 2000 to December 31, 2015. The inclusion criteria were admission to non-ICU internal medicine services and an admission note written by a resident physician. Otherwise, there were no exclusions. Data were accessed using VA Informatics and Computing Infrastructure. This study was approved by the LSCVAMC institutional review board.

We evaluated multiple patient characteristics for each admission that were accessible in the EHR at the time of hospital admission including patient comorbidities, medication count, and number of notes and discharge summaries. The Charlson Comorbidity Index (CCI) Deyo version was used to score all patients based on the EHR's active problem list at the time of admission.[8,9] The CCI is a validated score created by categorizing comorbidities using *International Classification of Diseases, Ninth and Tenth Revisions*.[8] Higher CCI scores predict increased mortality and resource usage. For each admission, we also counted the number of active medications, the number of prior discharge summaries, and the total number of notes available in the EHR at the time of patient admission. Patient admissions were grouped by calendar year, the mean numbers of active medications, prior discharge summaries, and total available notes per patient during each year were calculated (Table). Data comparisons were completed between 2003 and 2011 as well as between 2011 and 2015; median data are also

***Address for correspondence:** Todd I. Smith, MD, FHM,Louis Stokes Cleveland Department of Veterans Affairs Medical Center, 10701 East Blvd 111(W), Cleveland, OH 44106; ; E-mail: Todd.Smith@va.gov

Received: July 31, 2017; **Revised**: December 14, 2017; **Accepted**: January 5, 2018

TABLE. **Data Trends Among Internal Medicine Admissions**

	Trainee Admissions	CCI	Medical Problems	Notes	Medications	Discharge Summaries
2000	2,729	0.83	2	43	13.15	0.80
2001	3,621	1.26	4	88	12.82	1.48
2002	3,431	1.37	5	128	11.28	1.91
2003	4,304	1.60 (1)	6 (5)	193 (125)	8.37 (7)	2.29 (1)
2004	3,851	1.81	8	273	8.93	2.68
2005	3,807	2.25	10	363	9.17	3.14
2006	3,963	2.32	12	454	10.26	3.50
2007	4,309	2.61	13	536	14.24	3.89
2008	4,535	2.70	15	632	16.66	4.22
2009	4,620	2.97	15	742	17.18	4.56
2010	4,932	3.11	17	823	17.15	4.55
2011	4,977	3.05 (2)	16 (15)	841 (559)	16.90 (16)	4.42 (2)
2012	4,576	3.12	17	949	16.51	4.58
2013	4,674	3.23	18	1,056	16.56	4.77
2014	4,452	3.45	19	1,185	16.17	5.04
2015	4,565	3.77 (3)	20 (19)	1,289 (819)	16.49 (16)	5.48 (3)

Trainee admissions is the total number of admissions for the calendar year. All other data presented are mean measurements per patient at time of admission. Medians are listed in parentheses for 2003, 2011, and 2015.

provided for these years (Table). These years were chosen based on the years of the duty hour changes as well as comparing a not brand new, but still immature EHR (2003), a mature EHR (2011), and the most recent available data (2015).

RESULTS

A total of 67,346 admissions were included in the analysis. All parameters increased from 2000 to 2015. Mean CCI increased from 1.60 in 2003 (95% CI, 1.54-1.65) to 3.05 in 2011 (95% CI, 2.97-3.13) and to 3.77 in 2015 (95% CI, 3.67-3.87). Mean number of comorbidities increased from 6.21 in 2003 (95% CI, 6.05-6.36) to 16.09 in 2011 (95% CI, 15.84-16.34) and to 19.89 in 2015 (95% CI, 19.57-20.21). Mean number of notes increased from 193 in 2003 (95% CI, 186-199) to 841 in 2011 (95% CI, 815-868) and to 1289 in 2015 (95% CI, 1,243-1,335). Mean number of medications increased from 8.37 in 2003 (95% CI, 8.15-8.59) to 16.89 in 2011 (95% CI 16.60-17.20) and decreased to 16.49 in 2015 (95% CI, 16.18-16.80). Mean number of discharge summaries available at admission increased from 2.29 in 2003 (95% CI, 2.19-2.38) to 4.42 in 2011 (95% CI, 4.27-4.58) and to 5.48 in 2015 (95% CI, 5.27-5.69).

DISCUSSION

This retrospective, observational study shows that patient comorbidity and EHR data burden have increased over time, both of which impact resident workload at the time of admission. These findings, combined with the duty hour regulations, suggest that resident workload intensity at the time of admission may be increasing over time.

Patient comorbidity has likely increased due to a combination of factors. Elective admissions have decreased, and demographics have changed consistent with an aging population. Trainee admissions patterns also have changed over time, with less-acute admissions often admitted to nonacademic providers. Additionally, there are more stringent requirements for inpatient admissions, resulting in higher acuity and comorbidity.

As EHRs have matured and documentation requirements have expanded, the amount of electronic data has grown per patient, substantially increasing the time required to review a patient's medical record.[5,10] In our evaluation, all EHR metrics increased between 2003 and 2011. The only metric that did not increase between 2011 and 2015 was the mean number of medications. The number of notes per patient has shown a dramatic increase. Even in an EHR that has reached maturity (in use more than 10 years), the number of notes per patient still increased by greater than 50% between 2011 and 2015. The VA EHR has been in use for more than 15 years, making it an ideal resource to study data trends. As many EHRs are in their infancy in comparison, these data may serve as a predictor of how other EHRs will mature. While all notes are not reviewed at every admission, this illustrates how increasing data burden combined with poor usability can be time consuming and promote inefficient patient care.[11] Moreover, many argue that poor EHR usability also affects cogni-

tive workflow and clinical decision making, a task that is of utmost value to patient quality and safety as well as resident education.[12]

Common program requirements for internal medicine as set forth by the ACGME state that residency programs should give adequate attention to scheduling, work intensity, and work compression to optimize resident well-being and prevent burnout.[13] Resident workload intensity is multifaceted and encompasses many elements, including patient census and acuity, EHR data assessment, components of patient complexity such as comorbidity and psychosocial situation, and time.[13] The work intensity increases with increase in the overall patient census, complexity, acuity, or data burden. Similarly, work intensity increases with time restrictions for patient care (in the form of duty hours). In addition, work intensity is affected by the time allotted for nonclinical responsibilities, such as morning reports and conferences, as these decrease the amount of time a resident can spend providing patient care.

Many programs have responded to the duty-hour restrictions by decreasing patient caps.[14] Our data suggest that decreasing patient census alone may not adequately mitigate the workload intensity of residents. There are other alternatives to prevent the increasing workload intensity that may have already been employed by some institutions. One such method is that programs can take into account patient complexity or acuity when allocating patients to teaching teams.[14] Another method is to adjust the time spent on ancillary tasks such as obtaining outside hospital records, transporting patients, and scheduling follow-up appointments. Foregoing routine conferences such as morning reports or noon conferences would decrease work intensity, although obviously at the expense of resident education. Geographic rounding can encourage more efficient use of clinical time. One of the most difficult, but potentially impactful strategies would be to streamline EHRs to simplify and speed documentation, refocus regulations, and support and build based on the view of clinicians.[15]

The main limitations of this study include its retrospective design, single-center site, and focus on the internal medicine admissions to a VA hospital. Therefore, these findings may not be generalizable to other patient populations and training programs. Another potential limitation may be that changes in documentation practices have led to "upcoding" of patient comorbidy within the EHR. In addition, in this study, we looked only at the data available at the time of admission. To get a more complete picture of true workload intensity, understanding the day-to-day metrics of inpatient care would be crucial.

CONCLUSION

Our study demonstrates that components of resident workload (patient comorbidity and EHR data burden), specifically at the time of admission, have increased over time. These findings, combined with the duty-hour regulations, suggest resident workload intensity at the time of admission has increased over time. This can have significant implications regarding graduate medical education, patient safety, and burnout. To optimize resident workload, innovation will be required in the areas of workflow, informatics, and curriculum. Future studies to assess the workload and intensity of the course of the entire patient hospitalization are needed.

Acknowledgments

The authors thank Paul E. Drawz, MD, MHS, MS (University of Minnesota) for contributions in designing and reviewing the study.

Ethical approval: The study was approved by the Institutional Review Board at the LSCVAMC. The contents do not represent the views of the US Department of Veterans Affairs or the US government. This material is the result of work supported with resources and the use of facilities of the LSCVAMC.

Disclosures: The authors declare that they have no conflicts of interest to disclose.

References

1. Bolster L, Rourke L. The Effect of Restricting Residents' Duty Hours on Patient Safety, Resident Well-Being, and Resident Education: An Updated Systematic Review. *J Grad Med Educ.* 2015;7(3):349-363.
2. Fletcher KE, Underwood W, Davis SQ, Mangrulkar RS, McMahon LF, Saint S. Effects of work hour reduction on residents' lives: a systematic review. *JAMA.* 2005; 294(9):1088-1100.
3. Amin A, Choe J, Collichio F, et al. Resident Duty Hours: An Alliance for Academic Internal Medicine Position Paper. http://www.im.org/d/do/6967. Published February 2016. Accessed November 30, 2017.
4. Goitein L, Ludmerer KM. Resident workload-let's treat the disease, not just the symptom. *JAMA Intern Med.* 2013;173(8):655-656.
5. Oxentenko AS, West CP, Popkave C, Weinberger SE, Kolars JC. Time spent on clinical documentation: a survey of internal medicine residents and program directors. *Arch Intern Med.* 2010;170(4):377-380.
6. Fletcher KE, Reed DA, Arora VM. Doing the dirty work: measuring and optimizing resident workload. *J Gen Intern Med.* 2011;26(1):8-9.
7. Linzer M, Levine R, Meltzer D, Poplau S, Warde C, West CP. 10 bold steps to prevent burnout in general internal medicine. *J Gen Intern Med.* 2014;29(1):18-20.
8. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis.* 1987;40(5):373-383.
9. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol.* 1992;45(6):613-619.
10. Kuhn T, Basch P, Barr M, Yackel T, et al; Physicians MICotACo. Clinical documentation in the 21st century: executive summary of a policy position paper from the American College of Physicians. *Ann Intern Med.* 2015;162(4): 301-303.
11. Friedberg MW, Chen PG, Van Busum KR, et al. Factors Affecting Physician Professional Satisfaction and Their Implications for Patient Care, Health Systems, and Health Policy. *Rand Health Q.* 2014;3(4):1.
12. Smith SW, Koppel R. Healthcare information technology's relativity problems: a typology of how patients' physical reality, clinicians' mental models, and healthcare information technology differ. *J Am Med Inform Assoc.* 2014; 21(1):117-131.
13. ACGME Program Requirements for Graduate Medical Education in Internal Medicine. http://www.acgme.org/Portals/0/PFAssets/ProgramRequirements/140_internal_medicine_2017-07-01.pdf. Revised July 1, 2017. Accessed July 22, 2017.
14. Thanarajasingam U, McDonald FS, Halvorsen AJ, et al. Service census caps and unit-based admissions: resident workload, conference attendance, duty hour compliance, and patient safety. Mayo Clin Proc. 2012;87(4):320-327.
15. Payne TH, Corley S, Cullen TA, et al. Report of the AMIA EHR-2020 Task Force on the status and future direction of EHRs. *J Am Med Inform Assoc.* 2015;22(5):1102-1110.

Things We Do For No Reason: Neutropenic Diet

Heather R. Wolfe, MD[1,4], Navid Sadeghi, MD[2,4], Deepak Agrawal, MD[1,3,4], David H. Johnson, MD[1], Arjun Gupta, MD[1,4]*

[1]Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, Texas; [2]Division of Hematology and Oncology, University of Texas Southwestern Medical Center, Dallas, Texas; [3]Division of Gastroenterology, University of Texas Southwestern Medical Center, Dallas, Texas; [4]Parkland Health and Hospital System, Dallas, Texas.

The "Things We Do for No Reason" series reviews practices which have become common parts of hospital care but which may provide little value to our patients. Practices reviewed in the TWDFR series do not represent "black and white" conclusions or clinical practice standards, but are meant as a starting place for research and active discussions among hospitalists and patients. We invite you to be part of that discussion.

For several decades, providers have routinely restricted the diets of neutropenic cancer patients by eliminating foods that might harbor pathogenic microbes to reduce infection rates. These diets, known as neutropenic or low-bacteria diets, are prescribed across the country with little uniformity in the extent or content of prescription. These diets are difficult to follow and force patients to omit fresh fruits and vegetables and limit dairy and meat products from their diet. These dietary omissions compromise nutritional intake in patients who are already at high risk of malnutrition. Randomized trials have shown that these restrictive diets are not superior in preventing infections than more liberalized diets. Evidence shows that adherence to the Safe Food-Handling guidelines issued by the Food and Drug Administration, a mandate for all hospital kitchens, provides adequate protection against food-borne infection, precluding the need for the neutropenic diet. Thus, routine use of the neutropenic diet should be abandoned.

CLINICAL SCENARIO

A 67-year-old man with acute myeloid leukemia who has recently completed a cycle of consolidation chemotherapy presents to the emergency room with fatigue and bruising. He is found to have pancytopenia due to chemotherapy. His absolute neutrophil count (ANC) is 380/mm^3, and he has no symptoms or signs of infection. He is admitted for transfusion support and asks for a dinner tray. The provider reflexively prescribes a neutropenic diet.

***Address for correspondence:** Arjun Gupta, MD, Chief Resident for Quality, Safety and Value, Department of Internal Medicine, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390-8852; Telephone: 214-648-9651; Fax: 214-648-9100; E-mail: guptaarjun90@gmail.com.

Published online first May 30, 2018.

Received: November 4, 2017; **Revised**: March 4, 2018; **Accepted**: March 15, 2018

BACKGROUND

Although aggressive chemotherapy regimens have significantly improved survival rates in patients with cancer, these intensive regimens put patients at risk for a number of complications, including severe, prolonged neutropenia. Patients with neutropenia, particularly those with ANC< 500/mm^3, are at a significantly increased risk for infection. Common sites of infection include the blood stream, skin, lungs, urinary tract, and, particularly, the gastrointestinal tract.[1] Oncologists and dieticians first designed neutropenic diets, or low-bacteria diets, to limit the introduction of pathogenic microbes to the gastrointestinal system. Neutropenic diets typically limit the intake of fresh fruits, fresh vegetables, raw or undercooked meats and fish, and soft cheese made from unpasteurized milk. Despite the widespread recommendation of the neutropenic diet, no standardized guidelines exist, and the utilization of the diet and its contents vary widely among and within institutions.[2]

The neutropenic diet is a national phenomenon. A survey of 156 United States members of the Association of Community Cancer Centers revealed that 120 (78%) of the members had placed patients with neutropenia on restricted diets.[2] The triggers for prescription (neutropenia, or starting chemotherapy), ANC threshold for prescription, and duration of prescription (throughout chemotherapy or just when neutropenic) were not uniform. A majority of centers restricted fresh fruits, fresh vegetables, and raw eggs, while some locations also restricted tap water, herbs and spices, and alcoholic beverages.[2] Similarly, a study of practices in 29 countries across 6 continents found that 88% of centers have some version of a neutropenic diet guideline with significant heterogeneity in their prescription and content. For example, dried fruits were unrestricted in 23% of centers but were forbidden in 43%.[3]

WHY YOU MIGHT THINK THE NEUTROPENIC DIET IS HELPFUL IN PREVENTING INFECTION

The rationale behind the neutropenic diet is to limit the bacterial load delivered to the gut. Studies have shown that organisms such as *Enterobacter, Pseudomonas, and Klebsiella* have been isolated from food, particularly fruits and vegetables.[4,5] The ingestion of contaminated food products may serve as a source of pathogenic bacteria, which may cause potentially life-threatening infections. Mucositis, a common complication among cancer patients receiving therapy, predisposes patients to infection by disrupting the mucosal barrier, allowing bacteria to translocate from the gut to the bloodstream. Given that neutropenia and mucositis often occur simultaneously, these

TABLE. **Summary of Notable Neutropenic Diet Studies**

First Author	Population	Design	Number of Patients	Percentage of Patients Neutropenic	Average duration of Neutropenia	Duration of Study
DeMille et al.[14]	Adults undergoing chemotherapy outpatient	Prospective Observational	28	Not reported	Not reported	12 weeks
Garder et al.[7]	Adults with acute myelogenous leukemia or high risk myelodysplastic syndrome undergoing induction chemotherapy	Randomized controlled trial	153	Not reported	21 days in raw food arm, 20 days in cooked food arm	25 days
Lassiter et al.[18]	Adults undergoing myeloablative hematopoietic stem-cell transplants	Randomized controlled trial	46	100%	Not reported	>4 weeks
Moody et al. (2006)[19]	Children undergoing chemotherapy	Randomized controlled trial	19	89.5%	5.9 days in neutropenic diet arm, 9.2 days in food safety arm	Single chemotherapy cycle
Tramsen et al.[9]	Children with acute myelogenous leukemia undergoing induction chemotherapy	Prospective Observational	339	Not reported	Not reported	Duration of intensive treatment of acute myelogenous leukemia
Moody et al. (2017)[8]	Children undergoing myelosuppressive chemotherapy	Randomized controlled trial	150	74% in neutropenic diet arm, 71% in food safety arm	10.5 days in neutropenic diet arm, 9.6 days in food safety arm	24.5 days
Trifilio et al.[15]	Adult hematopoietic stem-cell transplants	Retrospective	726	Not reported	Not reported	Duration of hospitalization
Van Tiel et al.[11]	Adults with acute leukemia undergoing induction chemotherapy	Randomized controlled trial	20	Not reported	Not reported	Not reported

patients are at an increased risk of infections.[6] Cooking destroys bacteria if present, rendering cooked foods safe. Thus, the avoidance of fresh fruits and vegetables and other foods considered to have high bacterial loads should theoretically decrease the risk of infections in these patients.

WHY THE NEUTROPENIC DIET IS NOT HELPFUL IN PREVENTING INFECTION

Researchers have investigated the ability of the neutropenic diet to reduce infection in adult and pediatric neutropenic patients. A study involving 153 patients receiving chemotherapy for acute myeloid leukemia or myelodysplastic syndrome randomized 78 patients to a diet that restricted raw fruits and vegetables and 75 patients to a diet that included those foods.[8] The groups had similar rates of major infection (29% in the cooked group versus 35% in the raw group, $P = .60$) with no difference in mortality.[7] In a randomized, multi-institutional trial of 150 pediatric oncology patients, 77 patients received a neutropenic diet plus a diet based on the food safety guidelines approved by the Food and Drug Administration (FDA), while 73 children received a diet based on FDA-approved food safety guidelines.[8] Infection rates between the groups were not significantly different (35% vs 33% respectively, $P = .78$).

Intensive conditioning regimens place hematopoietic stem-cell transplant (HSCT) recipients at an even greater risk of infectious complications than other patients and may increase gastrointestinal toxicity and prolong neutropenia. A study from a single academic US center included 726 HSCT recipients, 363 of whom received a neutropenic diet and 363 of whom received a general diet. Significantly fewer infections were observed in the general diet group than in the neutropenic diet group. Notably, this study was a retrospective trial, and approximately 75% of participants were autologous HSCT recipients, who traditionally have low risks of infection. A survey and analysis of nonpharmacologic anti-infective measures in 339 children with leukemia enrolled in the multicenter Acute Myeloid Leukemia Berlin-Frankfurt-Munster 2004 trial also did not show that the neutropenic diet has protective effects on infection rates.[9] A meta-analysis that compiled data from the studies mentioned above found the hazard ratio for any infection (major or minor) and fever was actually higher in the neutropenic diet arm (relative risk 1.18, 95% confidence interval: 1.05-1.34, $P = .007$) relative to that in the unrestricted arm.[10]

The inefficacy of the neutropenic diet may be attributed to the fact that many of the organisms found on fresh fruits and vegetables are part of the normal flora in the gastrointestinal tract. A Dutch prospective randomized pilot study of 20 adult patients with acute myeloid leukemia undergoing chemotherapy compared the gut flora in patients on a low-bacteria diet versus that in patients on a normal hospital diet. Gut colonization by potential pathogens or infection rates were not significantly different between the 2 groups.[11]

In addition to mucositis, the common gastrointestinal complications of chemotherapy include nausea, vomiting, diarrhea, food aversions, and changes in smells and taste, which limit oral intake.[12] Unnecessary dietary restrictions can place patients at further risk of inadequate intake and malnutrition.[13] In the outpatient setting, compliance with the neutropenic diet is also problematic. In one study of 28 patients educated about the neutropenic diet, only 16 (57%) were compliant with the diet as revealed through telephone-based assessments at 6 and 12 weeks, and infection rates were not different be-

tween compliant versus noncompliant patients.[14] Patients and family members reported that following the neutropenic diet requires considerably more effort than following a less restrictive diet.[8] Maintaining nutrition in this patient population is already challenging, and the restriction of a wide variety of food items (fresh fruits, vegetables, dairy, certain meats, eggs) can cause malnutrition, low patient satisfaction, and poor quality of life.[13,14]

WHEN MIGHT THE NEUTROPENIC DIET BE HELPFUL?

Evidence shows no benefit of the neutropenic diet in any particular clinical scenario or patient population. However, despite the dearth of evidence to support neutropenic diets, the overall data regarding neutropenic diets are sparse. Randomized control trials to date have been limited by their small size with possible confounding by the type of malignancy and cancer therapy; use of prophylactic antibiotics, growth factors, and air-filtered rooms; variation in contents and adherence to the prescribed diet; and inpatient versus outpatient status. The study that included HSCT recipients was a retrospective trial, and a majority of patients were autologous HSCT recipients.[15] Although no study has specifically investigated the neutropenic diet in preventing infection in patients with noncancer-related neutropenia, no reason exists to suspect that it is helpful. The FDA advises safe food-handling practices for other immunocompromised patients, such as transplant recipients and patients with human immunodeficiency virus/acquired immunodeficiency syndrome, and the same principles can likely be applied to patients with noncancer-related neutropenia.

WHAT WE SHOULD DO INSTEAD

Although the neutropenic diet has not been proven beneficial, the prevention of food-borne infection in this population remains important. FDA-published guidelines, which promote safe food handling to prevent food contamination in patients with cancer, should be followed in inpatient and outpatient settings.[16] These guidelines allow for fresh fruits and vegetables as long as they have been adequately washed. Cleaning (eg, cleaning the lids of canned foods before opening, hand washing), separating raw meats from other foods, cooking to the right temperature (eg, cooking eggs until the yolk and white are firm), and chilling/refrigerating food appropriately are strongly emphasized. These guidelines are also recommended by the American Dietetic Association. Despite additional flexibility, patients following the FDA diet guidelines do not have increased risk of infection.[8] At our hospitals, the neutropenic diet can no longer be ordered. Neutropenic patients are free to consume all items on the general hospital menu, including eggs, meat, soft cheeses, nuts, and washed raw fruits and vegetables. The National Comprehensive Cancer Network guidelines for the prevention and treatment of cancer-related infections do not specifically address diet.[17] We call upon them to note the lack of benefit and potential harm of the neutropenic diet in the guidelines. Such an action may persuade more institutions to abandon this practice.

RECOMMENDATIONS

- Neutropenic diets, or low-bacteria diets, should not be prescribed to neutropenic patients.
- Properly handled and adequately washed fresh fruits and vegetables can safely be consumed by patients with neutropenia.
- Patients and hospitals should follow FDA-published safe food-handling guidelines to prevent food contamination.

CONCLUSIONS

A general diet can be safely ordered for our patient in the presented clinical scenario. Available data from individual studies and pooled data provide no evidence that neutropenic diets prevent infectious complications in patients with neutropenia.

Hospital kitchens must adhere to the food-handling guidelines issued by the FDA, and following these guidelines should provide adequate protection against food-borne infection, even in patients who are immunocompromised. Instead of restricting food groups, the FDA guidelines focus on safe food-handling practices. Less dietary restrictions provide patient's additional opportunities for balanced nutrition and for food choices based on personal preferences or cultural practices.

Do you think this is a low-value practice? Is this truly a "Thing We Do for No Reason?" Share what you do in your practice and join in the conversation online by retweeting it on Twitter (#TWDFNR) and liking it on Facebook. We invite you to propose ideas for other "Things We Do for No Reason" topics by emailing TWDFNR@hospitalmedicine.org.

Disclosures: There are no financial or other disclosures for any author.

References

1. Freifeld AG, Bow EJ, Sepkowitz KA, et al. Clinical practice guideline for the use of antimicrobial agents in neutropenic patients with cancer: 2010 update by the infectious diseases society of America. *Clin Infect Dis*. 2011;52(4):e56-e93. doi: 10.1093/cid/ciq147.
2. Smith LH, Besser SG. Dietary restrictions for patients with neutropenia: a survey of institutional practices. *Oncol Nurs Forum*. 2000;27(3):515-520.
3. Mank AP, Davies M, research subgroup of the European Group for B, Marrow Transplantation Nurses Group. Examining low bacterial dietary practice: a survey on low bacterial food. *Eur J Oncol Nurs*. 2008;12(4):342-348. doi: 10.1016/j.ejon.2008.03.005.
4. Casewell M, Phillips I. Food as a source of Klebsiella species for colonization and infection of intensive care patients. *J Clin Pathol*. 1978;31(9):845-849. doi: http://dx.doi.org/10.1136/jcp.31.9.845.
5. Wright C, Kominoa SD, Yee RB. Enterobacteriaceae and Pseudomonas aeruginosa recovered from vegetable salads. *Appl Environ Microbiol*. 1976;31(3):453-454.
6. Blijlevens N, Donnelly J, De Pauw B. Mucosal barrier injury: biology, pathology, clinical counterparts and consequences of intensive treatment for haematological malignancy: an overview. *Bone Marrow Transplant*. 2000;25(12):1269-1278. doi: 10.1038/sj.bmt.1702447.
7. Gardner A, Mattiuzzi G, Faderl S, et al. Randomized comparison of cooked and noncooked diets in patients undergoing remission induction therapy for acute myeloid leukemia. *J Clin Oncol*. 2008;26(35):5684-5688. doi: 10.1200/JCO.2008.16.4681.
8. Moody KM, Baker RA, Santizo RO, et al. A randomized trial of the effectiveness of the neutropenic diet versus food safety guidelines on infection rate in pediatric oncology patients. *Pediatr Blood Cancer*. 2017;65(1). doi: 10.1002/pbc.26711.
9. Tramsen L, Salzmann-Manrique E, Bochennek K, et al. Lack of effectiveness of neutropenic diet and social restrictions as anti-infective measures in chil-

dren with acute myeloid leukemia: an analysis of the AML-BFM 2004 trial. *J Clin Oncol.* 2016;34(23):2776-2783. doi: 10.1200/JCO.2016.66.7881.
10. Sonbol MB, Firwana B, Diab M, Zarzour A, Witzig TE. The effect of a neutropenic diet on infection and mortality rates in cancer patients: a meta-analysis. *Nutr Cancer.* 2015;67(8):1230-1238. doi: 10.1080/01635581.2015.1082109.
11. van Tiel F, Harbers MM, Terporten PHW, et al. Normal hospital and low-bacterial diet in patients with cytopenia after intensive chemotherapy for hematological malignancy: a study of safety. *Ann Oncol.* 2007;18(6):1080-1084. doi: 10.1093/annonc/mdm082.
12. Murtaza B, Hichami A, Khan AS, Ghiringhelli F, Khan NA. Alteration in taste perception in cancer: causes and strategies of treatment. *Front Physiol.* 2017;8:134. doi: 10.3389/fphys.2017.00134.
13. Argiles JM. Cancer-associated malnutrition. *Eur J Oncol Nurs.* 2005;9(2):S39-S50. doi: 10.1016/j.ejon.2005.09.006.
14. DeMille D, Deming P, Lupinacci P, et al. The effect of the neutropenic diet in the outpatient setting: a pilot study. *Oncol Nurs Forum.* 2006;33(2):337-343. doi: 10.1188/ONF.06.337-343.
15. Trifilio S, Helenowski I, Giel M, et al. Questioning the role of a neutropenic diet following hematopoetic stem cell transplantation. *Biol Blood Marrow Transplant.* 2012;18(9):1385-1390. doi: 10.1016/j.bbmt.2012.02.015.
16. Safe Food Handling: What You Need to Know. https://www.fda.gov/Food/FoodborneIllnessContaminants/BuyStoreServeSafeFood/ucm255180.htm. Accessed October 29, 2017.
17. Baden LR, Swaminathan S, Angarone M, et al. Prevention and treatment of cancer-related infections, Version 2.2016, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw.* 2016;14(7):882-913.
18. Lassiter M, Schneider SM. A pilot study comparing the neutropenic diet to a non-neutropenic diet in the allogeneic hematopoietic stem cell transplantation population. *Clin J Oncol Nurs.* 2015;19(3):273-278. doi: 10.1188/15.CJON.19-03AP.
19. Moody K, Finlay J, Mancuso C, Charlson M. Feasibility and safety of a pilot randomized trial of infection rate: neutropenic diet versus standard food safety guidelines. *J Pediatr Hematol Oncol.* 2006;28(3):126-133. doi: 10.1097/01.mph.0000210412.33630.fb.

Diagnosing the Treatment

The approach to clinical conundrums by an expert clinician is revealed through the presentation of an actual patient's case in an approach typical of a morning report. Similar to patient care, sequential pieces of information are provided to the clinician, who is unfamiliar with the case. The focus is on the thought processes of both the clinical team caring for the patient and the discussant.

This icon represents the patient's case. Each paragraph that follows represents the discussant's thoughts.

Sarah A. McGuffin, MD, MS[1,2,*], Robert L. Trowbridge, MD[3,4], Ann M. O'Hare, MD[5,6], and Andrew P. J. Olson, MD[7,8,9]

[1]Division of Allergy and Infectious Diseases, University of Washington, Seattle, Washington; [2]Department of Internal Medicine, University of Washington, Seattle, Washington; [3]Department of Internal Medicine, Maine Medical Center, Portland, Maine; [4]Department of Medicine, Tufts University School of Medicine, Boston, Massachusetts; [5]Division of Nephrology, VA Puget Sound Health Care System, Seattle, Washington; [6]Division of Nephrology, University of Washington, Seattle, Washington; [7]Department of Medicine, University of Minnesota, Minneapolis, Minnesota; [8]Department of Pediatrics, University of Minnesota, Minneapolis, Minnesota; [9]University of Minnesota Medical School, Minneapolis, Minnesota.

A 70-year-old man presented to the emergency department with 5 days of decreased appetite, frequent urination, tremors, and memory difficulties. He also reported 9 months of malaise, generalized weakness, and weight loss. There was no history of fever, chills, nausea, diarrhea, constipation, pain, or focal neurologic complaints.

This patient exemplifies a common clinical challenge: an older adult with several possibly unrelated concerns. In many patients, a new presentation is usually either a different manifestation of a known condition (eg, a complication of an established malignancy) or the emergence of something they are at risk for based on health behavior or other characteristics (eg, lung cancer in a smoker). The diagnostic process in older adults can be complicated because many have, or are at risk for, multiple chronic conditions.

After reviewing the timeline of symptoms, the presence of nine months of symptoms suggests a chronic and progressive underlying process, perhaps with subsequent superimposition of an acute problem. Although it is not certain whether chronic and acute symptoms are caused by the same process, this assumption is reasonable. The superimposition of acute symptoms on a chronic process may represent progression of the underlying condition or an acute complication of the underlying disease. However, the patient's chronic symptoms of malaise, weakness, and weight loss are nonspecific.

Although malignancy is a consideration given the age of the patient and time course of symptoms, attributing the symptoms to a specific pattern of disease or building a cogent differential diagnosis is difficult until additional information is obtained. One strategy is to try to localize the findings to one or more organ systems; for example, given that tremors and memory difficulties localize to the central nervous system, neurodegenerative disorders, such as "Parkinson plus" syndromes, and cerebellar disease are possible. However, this tactic still leaves a relatively broad set of symptoms without an immediate and clear unifying cause.

The patient's medical history included hyperlipidemia, peripheral neuropathy, prostate cancer, and papillary bladder cancer. The patient was admitted to the hospital four months earlier for severe sepsis presumed secondary to a urinary tract infection, although bacterial cultures were sterile. His social history was notable for a 50 pack-year smoking history. Outpatient medications included alfuzosin, gabapentin, simvastatin, hydrocodone, and cholecalciferol. He used a Bright Light Therapy lamp for one hour per week and occasionally used calcium carbonate for indigestion. The patient's sister had a history of throat cancer.

On examination, the patient was detected with blood pressure of 104/56 mm Hg, pulse of 85 beats per minute, temperature of 98.2 °F, oxygen saturation of 97% on ambient air, and body mass index of 18 kg/m². The patient appeared frail with mildly decreased strength in the upper and lower extremities bilaterally. The remainder of the physical examination was normal. Reflexes were symmetric, no tremors or rigidity was noted, sensation was intact to light touch, and the response to the Romberg maneuver was normal.

Past medical history is the cornerstone of the diagnostic process. The history of two different malignancies is the most striking element in this case. Papillary bladder cancer is usually a local process, but additional information is needed regarding its stage and previous treatment, including whether or not the

***Address for correspondence:** Sarah A. McGuffin, MD, MS, Division of Allergy and Infectious Diseases, University of Washington, 959 NE Pacific St Health Sciences Building, BB-545 Seattle, WA 98195; Telephone: 206-685-7288; Fax: 206-616-3892; E-mail: mcguffin@uw.edu

Published online first June 27, 2018.

Received: October 25, 2017; **Revised**: March 6, 2018; **Accepted**: March 13, 2018

patient received Bacille Calmette Guerin (BCG) vaccine, which can rarely be associated with infectious and inflammatory complications. Metastatic prostate cancer could certainly account for his symptomatology, and bladder outlet obstruction could explain the history of urinary frequency and probable urosepsis. His medication list suggested no obvious causes to explain his presentation, except that cholecalciferol and calcium carbonate, which when taken in excess, can cause hypercalcemia. This finding is of particular importance given that many of the patient's symptoms, including polyuria, malaise, weakness, tremor, memory difficulties, anorexia, acute kidney injury and (indirectly) hypotension and weight loss, are also seen in patients with hypercalcemia. The relatively normal result of the neurologic examination decreases the probability of a primary neurologic disorder and increases the likelihood that his neurologic symptoms are due to a global systemic process. The relative hypotension and weight loss similarly support the possibility that the patient is experiencing a chronic and progressive process.

The differential diagnosis remains broad. An underlying malignancy would explain the chronic progressive course, and superimposed hypercalcemia would explain the acute symptoms of polyuria, tremor, and memory changes. Endocrinopathies including hyperthyroidism or adrenal insufficiency are other possibilities. A chronic progressive infection, such as tuberculosis, is possible, although no epidemiologic factors that increase his risk for this disease are present.

The patient had serum calcium of 14.5 mg/dL, ionized calcium of 3.46 mEq/L, albumin of 3.6 g/dL, BUN of 62 mg/dL, and creatinine of 3.9 mg/dL (all values were normal three months prior). His electrolytes and liver function were otherwise normal. Moreover, he had hemoglobin level of 10.5 mg/dL, white blood cell count of 4.8 × 10^9cells/L, and platelet count of 203 × 10^9 cells/L.

Until this point, only nonspecific findings were identified, leading to a broad differential diagnosis with little specificity. However, laboratory examinations confirm the suspected diagnosis of hypercalcemia, provide an opportunity to explain the patient's symptoms, and offer a "lens" to narrow the differential diagnosis and guide the diagnostic evaluation. Hypercalcemia is most commonly secondary to primary hyperparathyroidism or malignancy. Primary hyperparathyroidism is unlikely in this patient given the relatively acute onset of symptoms. The degree of hypercalcemia is also atypical for primary hyperparathyroidism because it rarely exceeds 13 mg/dL, although the use of concurrent vitamin D and calcium supplementation could explain the high calcium level. Malignancy seems more likely given the degree of hypercalcemia in the setting of weight loss, tobacco use, and history of malignancy. Malignancy may cause hypercalcemia through multiple disparate mechanisms, including development of osteolytic bone metastases, elaboration of parathyroid hormone-related Peptide (PTHrP), increased production of 1,25-dihydroxyvitamin D, or, very rarely, ectopic production of parathyroid hormone (PTH). However, none of these mechanisms are particularly common in bladder or prostate cancer, which are the known malignancies in the patient. Other less likely and less common causes of hypercalcemia are also possible given the clinical clues, including vitamin D toxicity and milk alkali syndrome (vitamin D and calcium carbonate supplementation), multiple endocrine neoplasia (a sister with throat cancer), and granulomatous disease (weight loss). At this point, further laboratory evaluations would be helpful, specifically determination of PTH and PTHrP levels and serum and urine protein electrophoresis.

With respect to the patient's past medical history, his Gleason 3 + 3 prostate cancer was diagnosed 12 years prior to admission and treated with external beam radiation therapy and brachytherapy. His bladder cancer was diagnosed 3 years before admission and treated with tumor resection followed by 2 rounds of intravesical BCG (iBCG), 1 round of mitomycin C, and 2 additional rounds of iBCG over the course of treatment spanning 2 years and 6 months. The treatment was complicated by urethral strictures requiring dilation, ureteral outlet obstruction requiring left ureteral stent placement, and multiple urinary tract infections.

The patient's last round of iBCG was delivered 6 months prior to his current presentation. The patient's hospital admission 4 months earlier for severe sepsis was presumed secondary to a urologic source considering that significant pyuria was noted on urinalysis and he was treated with meropenem, although bacterial cultures of blood and urine were sterile. From the time of discharge until his current presentation, he experienced progressive weakness and an approximately 50 lb weight loss.

The prior cancers and associated treatments of the patient may be involved in his current presentation. The simplest explanation would be metastatic disease with resultant hypercalcemia, which is atypical of either prostate or bladder cancer. The history of genitourinary surgery could predispose the patient to a chronic infection of the urinary tract with indolent organisms, such as a fungus, especially given the prior sepsis without clear etiology. However, the history would not explain the presence of hypercalcemia. Tuberculosis must thus be considered given the weight loss, hypercalcemia, and "sterile pyuria" of the patient. A more intriguing possibility is whether or not the patient's constellation of signs and symptoms might be a late effect of iBCG. Intravesical BCG for treatment of localized bladder cancer is occasionally associated with complications. BCG is a modified live form of Mycobacterium bovis which invokes an intense inflammatory reaction when instilled into the bladder. These complications include disseminated infection and local complications, such as genitourinary infections. BCG infection might also explain the severe sepsis of unclear etiology that the patient had experienced 4 months earlier. Most interestingly, hypercalcemia has been described in the setting of BCG infection. Diagnosis of disseminated BCG is best made via culture or polymerase chain reaction testing for *M. bovis* at potential sites of involvement, including the blood. Neverthe-

less, a common presentation of a common disorder is still most likely. If his current presentation is distilled down to a chronic presentation of weakness, weight loss, and hypercalcemia in the setting of known malignancy, then the underlying malignancy seems to offer the most unifying explanation. Given that neither of his known cancers are commonly associated with hypercalcemia, the possibility that he has developed a third malignancy must also be considered.

In the hospital, the patient received intravenous normal saline, furosemide, and pamidronate. Evaluation for hypercalcemia revealed appropriately suppressed PTH (8 mg/dL), and normal levels of PTHrP (<.74 pmol/L), prostate specific antigen (<.01 ng/mL), and morning cortisol (16.7 mcg/dL). Serum and urine electrophoresis did not show evidence for monoclonal gammopathy, and the 25-hydroxy vitamin D level (39.5 ng/mL) was within the normal limits (normal range 20.1-50.0 ng/mL). The patient had elevated levels of 1,25-dihydroxy vitamin D (122 ng/mL, normal range 19.9–79.3 pg/mL), lactate dehydrogenase (196 units/L, normal 50–150 units/L), and angiotensin-converting enzyme (153 units/L, normal 14-82 units/L).

The suppressed PTH level makes primary hyperparathyroidism unlikely, the low PTHrP level decreases the probability of a paraneoplastic process, and the normal protein electrophoresis makes multiple myeloma unlikely. The presence of a significantly elevated 1,25-dihydroxy vitamin D level with a normal 25-hydroxy vitamin D level indicates extrarenal conversion of 25-hydroxy vitamin D by 1-hydroxylase as the etiology of hypercalcemia. Increased activity of 1-hydroxylase is the most consistent with granulomatous diseases, including sarcoidosis, and, with the exception of lymphoma, would not be expected in hypercalcemia malignancy. This mechanism is also associated with tuberculosis, disseminated fungal infections, such as coccidioidomycosis and histoplasmosis, and as a late effect of BCG treatment, regardless of whether disseminated infection or granulomatous immune response. Elevated lactate dehydrogenase and angiotensin-converting enzyme levels may also be noted in many of these disorders.

Lymphoma would appear to be the most likely diagnosis as it accounts for most of the clinical findings observed in the patient and is a fairly common disorder. Sarcoidosis is also reasonably common and would explain the laboratory abnormalities but is not usually associated with weight loss and frailty. Disseminated infections, such as tuberculosis, histoplasmosis, and coccidioidomycosis, are all possible, but the patient lacks key risk factors for these infections. A complication of iBCG is the most intriguing possibility and could account for many of the patient's clinical findings, including the septic episode, which is an event not clearly accounted for by the other diagnostic possibilities. However, disseminated BCG and hypersensitivity reactions to BCG leading to hypercalcemia are rare. When asked to choose between the most interesting possibility and the most common possibility, the most common will usually be the best (and safest) bet. Nonetheless, the effects of prior BCG treatment, including disseminated infection or diffuse immune-mediated granulomatous disease, would be near the top of the differential diagnosis in this case.

The bone survey was normal, the renal ultrasound examination showed nodular wall thickening of the bladder with areas of calcification, and the CT scan of the chest, abdomen, and pelvis showed an area of calcification in the superior portion of the bladder but no evidence of lymphadenopathy or masses to suggest lymphoma. Aerobic and anaerobic blood and urine cultures were sterile. The patient was discharged 12 days after admission with plans for further outpatient diagnostic evaluation. At this time, his serum calcium had stabilized at 10.5 mg/dL with pamidronate, diuretics, and aggressive oral hydration.

Outpatient bone marrow biopsy revealed a normocellular marrow with multiple small epithelioid granulomas consisting of histiocytes and Touton-type giant cells. Outpatient cystoscopy with barbotage was notable for recurrent urethral stricture that required dilation but did not reveal any new lesions or tumors. At 42 days after discharge, acid-fast culture and stain from blood cultures obtained in the hospital on day 10 grew acid-fast bacilli of the *Mycobacterium* tuberculosis complex (Figure). In broth culture, the bacilli were noted to form macroscopic cords.[1,2] Given the concern for disseminated *M. bovis,* the patient was started on antituberculosis therapy with isoniazid, pyridoxine, rifampin, and ethambutol along with a short course of steroids for presumed granuloma-associated hypercalcemia. The PCR results confirmed that the organism was *M. bovis.* The patient responded well to this course of treatment. His hypercalcemia resolved rapidly, and he regained weight, strength, and energy over the ensuing months.

DISCUSSION

Hypercalcemia is a common finding in both hospital and ambulatory settings. The classic symptoms associated with hypercalcemia are aptly summarized with the mnemonic "bones, stones, abdominal groans, and psychiatric overtones" (to represent the associated skeletal involvement, renal disease, gastrointestinal symptoms, and effects on the nervous system). However, the severity and type of symptoms vary depending on the degree of hypercalcemia, acuity of onset, and underlying etiology. The vast majority (90%) of hypercalcemia cases are due to primary hyperparathyroidism and malignancy.[3] Measuring the PTH level is a key step in the diagnostic evaluation process. An isolated elevation of PTH confirms the presence of primary or possibly tertiary hyperparathyroidism. Low PTH concentrations (<20 pg/mL) occur in the settings of PTHrP or vitamin-D-mediated hypercalcemia such as hypervitaminosis D, malignancy, or granulomatous disease.

Elevated PTHrP occurs most commonly in squamous cell, renal, bladder, and ovarian carcinomas.[3,4] Elevated levels of 25-hydroxy vitamin D can occur with excessive consumption of vitamin D-containing products and some herbal supplements. In this case, neither PTHrP nor 25-hydroxy vitamin D level was

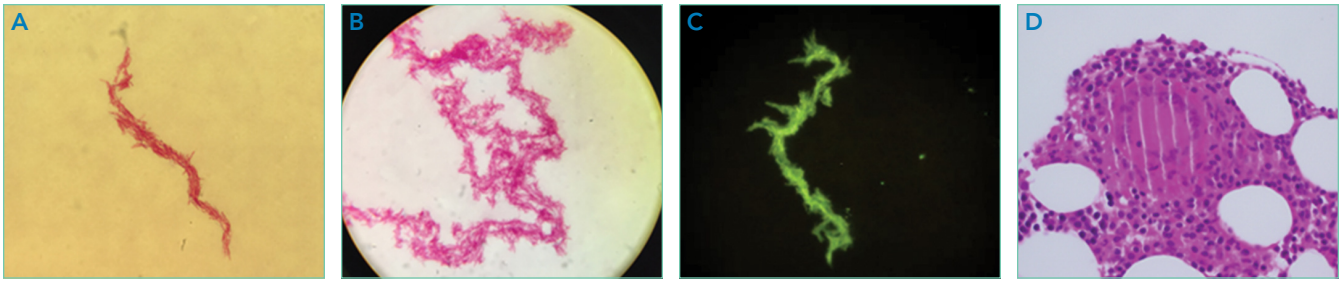

FIG. Blood cultures showing acid-fast bacteria on Kinyoun stain at low (A) and higher (B) power. The mycobacteria bacilli in the patient's cultures assembled in an end-to-end serpentine arrangement known as "cording." Cord formation is characteristic of mycobacteria and due to a specific cell wall glycolipid, namely, trehalose 6, 6'-dimycolate or cord factor, which contributes to mycobacterial virulence, granuloma formation, and humoral and cellular immune responses. The tendency to aggregate in cords actually disappeared in the attenuated form of *Mycobacterium bovis* cultured by Calmette and Guerin, suggesting that the patient had been exposed to and thus subsequently infected with a more virulent form of BCG. (C) Auramine stain of blood cultures growing *Mycobacterium bovis*, showing similar serpentine cord formation. (D) Bone marrow biopsy with multiple granulomas consisting of histocytes and Touton-type giant cells consistent with BCG infection.

elevated, leading to an exhaustive search for other causes. Although iBCG treatment is a rare cause of hypercalcemia, 2 previous reports indicated the presence of hypercalcemia secondary to granuloma formation in treated patients.[5,6]

The finding of an elevated 1,25-dihydroxy vitamin D level was unexpected. As the discussant mentioned, this finding is associated with lymphoma and with granulomatous disorders that were not initially strong diagnostic considerations in the patient. A variety of granulomatous diseases can cause hypercalcemia. Sarcoidosis and tuberculosis are the most common, but berylliosis, fungal infections, Crohn's disease, silicone exposure, and granulomatosis with polyangiitis may also be associated with hypercalcemia.[7] The mechanism for hypercalcemia in these situations is increased intestinal calcium absorption mediated by inappropriately increased, PTH-independent, extrarenal calcitriol (1,25-dihydroxy vitamin D) production. Activated monocytes upregulate 25(OH)D-alpha-hydroxylase, converting 25-hydroxy vitamin D to 1,25-dihydroxy vitamin D. Concurrently, the elevated levels of gamma-interferon render macrophages resistant to the normal regulatory feedback mechanisms, thereby promoting the production and inhibiting the degradation of 1,25-dihydroxy vitamin D.[8]

The tuberculosis vaccine BCG is an attenuated form of *M. bovis* and was originally developed by Albert Calmette and Camille Guérin at the Pasteur Institute in Paris in the early 20th century. In addition to its use as a vaccine against tuberculosis, BCG can protect against other mycobacterial infections, help treat atopic conditions via stimulation of the Th1 cellular immune response, and has been used as an antineoplastic agent. To date, BCG remains the most effective agent available for intravesical treatment of superficial bladder cancer.[9,10] Although iBCG therapy is considered relatively safe and well-tolerated, rare complications do occur. Localized symptoms (bladder irritation, hematuria) and/or flu-like symptoms are common immediately after instillation and thought to be related to the cellular immune response and inflammatory cascade triggered by mycobacterial antigens.[11] Other adverse effects, such as infectious and noninfectious complications, may occur months to years after treatment with BCG, and the associated symptoms can be quite nonspecific. Infectious complications include mycobacterial prostatitis, orchiepididymitis, balantitis, pneumonia, hepatitis, nephritis, septic arthritis, osteomyelitis, infected orthopedic and vascular prostheses, endocarditis, and bacteremia. Traumatic catheterization is the most common risk factor for infection with BCG.[11-13] Noninfectious complications include reactive arthritis, hypersensitivity pneumonitis, hemophagocytic lymphohistiocytosis (HLH), and sterile granulomatous infiltration of solid organs.

The protean and nonspecific nature of the adverse effects of iBCG treatment and the fact that complications can present weeks to years after instillation can make diagnosis quite challenging.[14] Even if clinical suspicion is high, it may be difficult to definitively identify BCG as the underlying etiology because acid fast staining, culture, and even PCR can lead to falsely negative results.[14,15] For this reason, biopsy and tissue culture are recommended to demonstrate granuloma formation and identify the presence of *M. bovis*.

Although no prospective studies have been conducted to assess the optimal therapy for BCG infection, opinion-based recommendations include cessation of BCG treatment, initiation of at least 3 tuberculostatic agents, and treatment for 3-12 months depending on the severity of the complications.[11,14] *M. bovis* is susceptible to isoniazid, rifampin, and ethambutol as well as to fluoroquinolones, clarithromycin, aminoglycosides, and doxycycline; however, this organism is highly resistant to pyrazinamide due to single-point mutation.[11,16] Interestingly, imipenem is used to treat other nontuberculous mycobacterial diseases, such as those caused by *M. abscessus*, thereby raising the possibility that the patient's exposure to meropenem during treatment for his prior sepsis may have partially treated an acute infection due to *M. bovis*.

Although treatment with steroids is a standard approach for management of hypercalcemia in other granulomatous disorders and leads to rapid reduction in circulating levels of 1,25-dihydroxy vitamin D and serum calcium., specific evidence has not been established to support its efficacy and effectiveness in treating hypercalcemia and other complications due to *M. bovis*.[17] Nevertheless, some experts recommend the use of steroids in conjunction with a multidrug tuberculostatic regimen in cases of septicemia and multiorgan failure due to *M. bovis*.[12,14,18-20]

In summary, this case illustrates the importance of making room in differential diagnosis to include iatrogenic complications. That is, when faced with an unclear diagnosis, the provider should consider common and uncommon immediate and delayed side effects of prior therapies.

TEACHING POINTS:

- Complications of intravesical BCG treatment include manifestations of granulomatous diseases, such as hypercalcemia.
- When generating a differential diagnosis, medical providers should not only consider the possibility of a new disease process or the progression of a known comorbidity but also the potential of an adverse effect related to prior treatments.
- Medical providers should be wary of accepting previously made diagnoses, particularly when key pieces of objective data are lacking.

Disclosures: The authors have no financial or other conflicts of interest that might bias this work.

References

1. Geisel RE, Sakamoto K, Russell DG, Rhoades ER. In vivo activity of released cell wall lipids of Mycobacterium bovis bacillus Calmette-Guérin is due principally to trehalose mycolates. *J Immunol*. 2005;174(8):5007-5015. https://doi.org/10.4049/jimmunol.174.8.5007.
2. Ryll R, Kumazawa Y, Yano I. Immunological properties of trehalose dimycolate (cord factor) and other mycolic acid-containing glycolipids--a review. *Microbiol Immunol*. 2001;45(12):801-811. https://doi.org/10.1111/j.1348-0421.2001.tb01319.x.
3. Carroll MF, Schade DS. A practical approach to hypercalcemia. *Am Fam Physician*. 2003;67(9):1959-1966.
4. Goldner W. Cancer-related hypercalcemia. *J Oncol Pract*. 2016;12(5):426-432. https://doi.org/10.1200/JOP.2016.011155.
5. Nayar N, Briscoe K. Systemic Bacillus Calmette-Guerin sepsis manifesting as hypercalcaemia and thrombocytopenia as a complication of intravesical Bacillus Calmette-Guerin therapy. *Intern Med J*. 2015;45(10):1091-1092. https://doi.org/10.1111/imj.12876.
6. Schattner A, Gilad A, Cohen J. Systemic granulomatosis and hypercalcaemia following intravesical bacillus Calmette–Guerin immunotherapy. *J Intern Med*. 2002;251(3):272-277. https://doi.org/10.1046/j.1365-2796.2002.00957.x.
7. Tebben PJ, Singh RJ, Kumar R. Vitamin D-mediated hypercalcemia: mechanisms, diagnosis, and treatment. *Endocr Rev*. 2016;37(5):521-547. https://doi.org/10.1210/er.2016-1070.
8. Nielsen CT, Andersen ÅB. Hypercalcemia and renal failure in a case of disseminated Mycobacterium marinum infection. *Eur J Intern Med*. 2016;20(2):e29-e31. https://doi.org/10.1016/j.ejim.2008.08.015.
9. Sylvester RJ. Bacillus Calmette-Guérin treatment of non-muscle invasive bladder cancer. *Int J Urol*. 2011;18(2):113-120. https://doi.org/10.1111/j.1442-2042.2010.02678.x.
10. Clark PE, Spiess P, Agarwal N, Al. E. NCCN Guidelines ® Insights Bladder Cancer, Version 2.2016 Featured Updates to the NCCN Guidelines. *J Natl Compr Canc Netw*. 2016;14(10):1213-1224. https://doi.org/10.6004/jnccn.2016.0131.
11. Decaestecker K, Oosterlinck W. Managing the adverse events of intravesical bacillus Calmette–Guérin therapy. *Res Reports Urol*. 2015;7:157-163. https://doi.org/10.2147/RRU.S63448.
12. Gandhi NM, Morales A, Lamm DL. Bacillus Calmette-Guerin immunotherapy for genitourinary cancer. *BJU Int*. 2013;112(3):288-297. https://doi.org/10.1111/j.1464-410X.2012.11754.x.
13. Brausi M, Oddens J, Sylvester R, et al. Side effects of bacillus calmette-guerin (BCG) in the treatment of intermediate- and high-risk Ta, T1 papillary carcinoma of the bladder: Results of the EORTC genito-urinary cancers group randomised phase 3 study comparing one-third dose with full dose and 1 year with 3 years of maintenance BCG. *Eur Urol*. 2014;65(1):69-76. https://doi.org/10.1016/j.eururo.2013.07.021.
14. Gonzalez OY, Musher DM, Brar I, et al. Spectrum of bacille Calmette-Guérin (BCG) infection after intravesical BCG immunotherapy. *Clin Infect Dis*. 2003;36(2):140-148. https://doi.org/10.1086/344908.
15. Pérez-Jacoiste Asín MA, Fernández-Ruiz M, López-Medrano F, et al. Bacillus Calmette-Guérin (BCG) infection following intravesical BCG administration as adjunctive therapy for bladder cancer. *Medicine (Baltimore)*. 2014;93(17):236-254. https://doi.org/10.1097/MD.0000000000000119.
16. Durek C, Rüsch-Gerdes S, Jocham D, Böhle A. Sensitivity of BCG to modern antibiotics. *Eur Urol*. 2000;37(Suppl 1):21-25. https://doi.org/10.1159/000052378.
17. Sharma OP. Hypercalcemia in granulomatous disorders: a clinical review. *Curr Opin Pulm Med*. 2000;6(5):442-447. https://doi.org/10.1097/00063198-200009000-00010.
18. LeMense GP, Strange C. Granulomatous pneumonitis following intravesical BCG: what therapy is needed? *Chest*. 1994;106(5):1624-1626. https://doi.org/10.1378/chest.106.5.1624.
19. Nadasy KA, Patel RS, Emmett M, et al. Four cases of disseminated Mycobacterium bovis infection following intravesical BCG instillation for treatment of bladder carcinoma. *South Med J*. 2008;101(1):91-95. https://doi.org/10.1097/SMJ.0b013e31815d4047.
20. Macleod LC, Ngo TC, Gonzalgo ML. Complications of intravesical bacillus calmette-guérin. *Can Urol Assoc J*. 2014;8(7-8):E540-E544. https://doi.org/10.5489/cuaj.1411.

Azithromycin: Short Course with Long Duration

Jonne J. Sikkens, MD, MSc*, Michiel. A. van Agtmael, MD, PhD

Department of Internal Medicine, VU University Medical Center, Amsterdam, The Netherlands.

Royer and colleagues[1] have performed a meta-analysis comparing shorter versus longer courses of antibiotics for treating infections in hospitalized patients. They conclude that shorter courses are safe. However, the authors do not address a flaw in the analysis; they included studies in which treatment with azithromycin was considered a short antibiotic course relative to treatment with another antibiotic. Azithromycin is a macrolide antibiotic that has a relatively long terminal serum half-life, which has been reported to be 35-96 hours.[2-4] Moreover, the half-life of azithromycin in lung tissue can be as long as 132 hours,[4] which is important because tissue concentrations are thought to be more indicative of the clinical efficacy of macrolides.[5] In 4 of 19 studies in the meta-analysis,[1] azithromycin was used as a short course for the treatment of pneumonia and compared with longer courses of antibiotics with a much shorter half-life. This implies that in these studies, the duration of the effective antibiotic tissue concentration in the short arms was probably not shorter than in the comparator arms. It could even be longer due to azithromycin's favorable pharmacokinetics. In our view, these studies have unfairly contributed to the clinical efficacy of short courses, thereby threatening the validity of the overall conclusions. We think that effective antibiotic blood/tissue levels determine the clinical outcome, not just shorter or longer antibiotic courses.

***Address for correspondence:** Jonne J. Sikkens, MD, MSc, Department of Internal Medicine, VU University Medical Center, PO Box 7057, 1007 MB, Amsterdam, The Netherlands; Telephone: +31(20)-444-4444; Fax: +31(20)-444-4645; E-mail: j.sikkens@vumc.nl

Received: February 9, 2018; **Accepted:** March 11, 2018

 DOI 10.12788/jhm.2990

Disclosures: The authors declare that they have no conflicts of interest to report.

References

1. Royer S, DeMerle KM, Dickson RP, Prescott HC. Shorter versus longer courses of antibiotics for infection in hospitalized patients: a systematic review and meta-analysis. *J Hosp Med*. 2018:13(5):336-342. doi: 10.12788/jhm.2905.
2. Lode H. The pharmacokinetics of azithromycin and their clinical significance. *Eur J Clin Microbiol Infect Dis*. 1991;10(10):807-812.
3. Singlas E. Clinical pharmacokinetics of azithromycin. *Pathol Biol*. 1995;43(6):505-511.
4. Di Paolo A, Barbara C, Chella A, Angeletti CA, Del Tacca M. Pharmacokinetics of azithromycin in lung tissue, bronchial washing, and plasma in patients given multiple oral doses of 500 and 1000 mg daily. *Pharmacol Res*. 2002;46(6):545-550. doi: 10.1016/S1043-6618(02)00238-4.
5. Amsden GW. Advanced-generation macrolides: tissue-directed antibiotics. *Int J Antimicrob Agents*. 2001;18(1):S11-S15.

Reply to Azithromycin: Short Course with Long Duration

Stephanie Royer, MD[1,2,3]*, Hallie C. Prescott MD[1,4]

[1]Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan; [2]Division of Hospital Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio; [3]Department of Internal Medicine, University of Cincinnati, Cincinnati, Ohio; [4]VA Center for Clinical Management Research, VA Ann Arbor Healthcare System, Ann Arbor, Michigan.

We appreciate the interest in our review of antibiotic duration in hospitalized patients. Drs. Sikkens and van Agtmael comment that drug pharmacokinetics can alter true treatment duration.[1,2] Specifically, azithromycin has a long half-life in tissues.[3] We did not consider pharmacokinetics in our prespecified protocol for study inclusion, nor require that studies compare the same drug between treatment groups. This is consistent with a systematic review of antibiotic duration in community-acquired pneumonia, which included 3 of the 4 studies comparing short-course azithromycin to a longer course of another antibiotic.[4] Similarly, in a recent pilot study of antibiotic duration in bloodstream infections, only treatment duration was prespecified.[5] We agree that the differing pharmacokinetics between drugs is a limitation to our findings.

To assess whether the inclusion of studies using short-course azithromycin biased our conclusions, we performed an additional meta-analysis for clinical efficacy excluding the 4 studies that compared azithromycin with another drug. This subgroup included 9 trials comprising 1270 patients. The overall risk difference was 0.3% (95% CI −2.7%, 3.3%), consistent with the primary findings that short-course antibiotic treatment is non-inferior to long-course antibiotic treatment. None of these 4 studies examined mortality; thus, the meta-analyses for short-term and long-term mortality are unaffected.

***Address for correspondence:** Stephanie Royer, MD, 3333 Burnet Avenue, MLC 3024, Cincinnati, OH, 45229; Telephone: (513) 636-5148; Fax: (513) 803-9245; E-mail: Stephanie.Royer@cchmc.org

Received: March 6, 2018; **Accepted**: March 11, 2018

Disclosures: Dr. Royer holds stock in Pfizer. The authors have no other potential financial conflicts of interest to report.

Funding: This work was supported by K08 GM115859 [HCP]. This manuscript does not necessarily represent the position or policy of the US government or the Department of Veterans Affairs.

References

1. Sikkens JJ, van Agtmael MA. Azithromycin: short course with long duration. *J Hosp Med*. 2018;13(7):582.
2. Royer S, DeMerle KM, Dickson RP, Prescott HC. Shorter versus longer courses of antibiotics for infection in hospitalized patients: a systematic review and meta-analysis. *J Hosp Med*. 2018;13(5):336-342. doi: 10.12788/jhm.2905.
3. Di Paolo A, Barbara C, Chella A, Angeletti CA, Del Tacca M. Pharmacokinetics of azithromycin in lung tissue, bronchial washing, and plasma in patients given multiple oral doses of 500 and 1000 mg daily. *Pharmacol Res*. 2002;46(6):545-550. doi: 10.1016/S1043661802002384.
4. Li JZ, Winston LG, Moore DH, Bent S. Efficacy of short-course antibiotic regimens for community-acquired pneumonia: a meta-analysis. *Am J Med*. 2007;120(9):783-790.
5. Daneman N, Rishu AH, Pinto R, et al. 7 versus 14 days of antibiotic treatment for critically ill patients with bloodstream infection: a pilot randomized clinical trial. *Trials*. 2018;19(1):111.

Be inspired by your career in academic hospital medicine.

Don't miss AHA, the only hands-on, small session faculty-to-attendee ratio and interactive program designed specifically for academic hospital medicine professionals.

Register Today

academichospitalist.org.

shm Society of Hospital Medicine | ACLGIM Association of Chiefs & Leaders of General Internal Medicine | SGIM Society of General Internal Medicine