



The experience of academic physician coaches in a professional development coaching program

Gazi Rashid¹ · Alyssa Pradarelli² · Jacqueline Chu³ · David Chang² · Benjamin Nelson⁴ · John Mullen² · Kerri Palamara³

Received: 20 May 2022 / Revised: 14 September 2022 / Accepted: 9 October 2022 / Published online: 4 November 2022
© The Author(s), under exclusive licence to Association for Surgical Education 2022

Abstract

Purpose The Professional Development Coaching Program (PDCP) is a physician coaching program founded on the principles of positive psychology that has been shown to improve burnout and well-being in residents. The experience of the physician faculty coaches is not well understood. We studied the impact of a longitudinal coaching intervention on the experience of coaches.

Methods From 2017 to 2019, faculty from Surgery, Pediatrics, and Medicine at an academic medical center participated as coaches in the PDCP. Coaches underwent training in positive psychology and coaching skills and coached trainees for one or two years. Surveys were performed at baseline, after 1 year (EOY-1), and after 2 years of coaching (EOY-2). Outcomes include burnout and professional fulfillment (Stanford Professional Fulfillment Index), well-being (PERMA), and experience with coaching skills. Surveys from each timepoint were paired and analyzed through bivariate analyses and multivariate linear regression.

Results Of the 136 coaches who participated, 44% submitted sufficient data for paired analysis. There was no change from pre- to post- in primary outcomes on bivariate analysis. On linear regression, burnout and PERMA scores declined from baseline but did not change from EOY-1 to EOY-2. Coaches reported increased coping skills on EOY surveys compared to baseline. Confidence with coping skills was associated with low burnout, high professional fulfillment, and high PERMA.

Conclusions These findings create a profile of physician coaches and demonstrate both the benefits and changes in burnout and well-being that they experience. Larger studies with comparison groups can further explore the effects of coaching on the coach.

Keywords Coach · Coaching · Physician · Burnout · Well-being · Coping

Introduction

Coaching is an evidence-based approach to supporting physician well-being that is increasing in prevalence for trainee populations. In contrast to the more familiar “mentor,”

“advisor” or “sponsor”, coaches can serve to help trainees design and achieve their personal and professional goals while also supporting their well-being and addressing burnout. There are also technical and clinical skills coaches that are prevalent in the surgical literature; their approach differs in style, process, and intended outcome. Trainee coaching programs have been developed and implemented, including the Professional Development Coaching Program (PDCP) [1]. The PDCP was founded to guide residents and fellows to focus on their strengths and support their resilience through the development of skills in coping, positive psychology, and reflection. Evaluations of these programs have focused on the impact on the coachees and have been shown to reduce burnout, improve well-being, and increase coping skills [1, 2].

✉ Gazi Rashid
grashid@mgh.harvard.edu; gazorashid@gmail.com

¹ Department of Surgery, University of Massachusetts Chan Medical School, Worcester, MA, USA

² Department of Surgery, Massachusetts General Hospital, Boston, MA, USA

³ Department of Medicine, Massachusetts General Hospital, Boston, MA, USA

⁴ Department of Pediatrics, Mass General Hospital for Children, Boston, MA, USA

Despite the evidence of benefits for coachees, the effect of coaching on the experience of physician coaches is less clear. Business literature has described higher workplace satisfaction and reduced work-related mental strain in coaches compared to the general population [3]. In contrast, a study of 106 business coaches found that the average coach experienced nearly six negative effects from coaching, the most common of which were being negatively affected by topics discussed with coachees and insecurities about fulfilling their role as a coach. These effects were also correlated with increased emotional exhaustion, a major component of burnout [4].

Physicians—in particular, those in academic faculty positions—are known to be at high risk for burnout, psychological stress, and professional attrition [5]. The research on coaching in medicine has focused on coachees and only recently have the experiences of coaches been explored. A 2020 qualitative study by Brooks et al. of medical student coaches identified both professional benefits (helping struggling students, providing holistic guidance) and challenges (unclear expectations of the role, balancing between coaching and supervising) [6]. A mixed methods study from 2021 by Elster et al. found that a higher proportion of medical student coaches was burned out compared to physician faculty with administrative roles. The coaches in this study also felt challenged by the high emotional output required to support students [7]. Neither of these studies investigated the change from before-to-after coaching. To our knowledge, no work has investigated whether coaches gain skills or benefits like those endorsed by the coachees. Of note, the PDCP engages volunteer faculty who undergo training in positive psychology and coaching to serve as novice coaches. These are not professional certified coaches, as have been studied in the business literature; nor are they academic learning coaches that are utilized in medical schools to support students.

Given that the coaches in the PDCP are academic physician volunteers serving as novice coaches and not certified physician coaches, the objective of this study was to investigate the impact of a longitudinal coaching intervention on the experience of physician faculty who underwent training to become novice coaches. The study focused on burnout, well-being, and professional fulfillment before and after coaching. Given that non-medical coaches suffer emotional exhaustion from their work and that physician coaches have higher rates of burnout than non-coach faculty in other studies, we hypothesized that burnout and well-being might worsen after coaching. We hypothesized professional fulfillment might be higher despite that decline, given the professional benefits of coaching mentioned in the prior studies. We also sought to investigate whether the coaches benefited from the coaching program in terms of the development and use of skills with others.

Methods

Settings and participants

This was a prospective longitudinal study in a single large academic medical center spanning two year-long iterations of the coaching program from 2017 to 2019. Coaches were unpaid volunteers from faculty in the departments of medicine, pediatrics, and various surgical subspecialties including general surgery, obstetrics and gynecology, urology, oral maxillofacial surgery, vascular surgery, pediatric surgery, and thoracic surgery. Coachees were residents and fellows from the departments listed above who, as part of a randomized control trial (manuscript currently in progress), were assigned to the intervention coaching arm vs. a comparison non-coaching arm. In the intervention arm, they were assigned a coach, but their participation in coaching was voluntary.

Program description

The goals of the PDCP are to support coachees by reducing their burnout, improving their resilience, and maximizing their strengths and potential. The PDCP was founded on the principles of positive psychology to provide trainees a safe environment to reflect on their experiences and build the skills to optimize their strengths and cope with the stressors and challenges. The PDCP was developed at the senior author's home institution in 2011 through collaboration with the Institute of Coaching, and first implemented with internal medicine interns in 2012 [8]. It has since been expanded to several institutions and organizations with several publications describing the model, approach, and impact [9–11]. The research protocol and all research materials and methods were approved by this institution's Institutional Review Board (Protocol #2017P00056).

Intervention

In the spring of 2017, the first cohort of coaches was recruited via emails from program directors. They then received two hours of interactive training in the principles and practices of coaching and positive psychology led by a subject-matter expert in professional development coaching. Training focused on core coaching techniques through exercises in reflective listening, asking future-oriented questions, facilitating goal-setting conversations, and maximizing strengths and positives. After practicing these skills, coaches were led through the same exercises they would be using with their coachees. Through a

dyadic approach, coaches were able to practice and receive coaching.

Coaches also received curricular guides for their meetings with the coachees. At the beginning of the 2017 academic year, coaches were paired with at least one coachee from outside of their subspecialty. This was done to ensure a safe space for the coachees, and to assist the coaches in not defaulting to traditional mentoring conversations. Pairs were expected to meet quarterly for approximately 60 min. using the curricular guides. These sessions focused on the coachees' development of skills for processing feedback, reflecting on their experiences, and coping with stressors. The contents of these meetings were kept confidential.

Coaches were asked to complete baseline surveys after completing training but before the start of coaching, and end-of-year surveys at the end of each academic year they participated as coaches. Coaches participated for either one academic year (2017–2018 or 2018–2019) or two academic years (2017–2019). This allowed for survey data at baseline, after one year of coaching, and—for some coaches—after two years of coaching. Survey measures were the same at all time points, except that end-of-year surveys asked about length and quality of coaching meetings. Surveys were administered through and stored in REDCap. This study was reviewed and considered exempt by this institution's Institutional Review Board with exemption #45 CFR 46.101(b)(1) research conducted in established and commonly accepted educational settings, involving normal educational practices.

Outcomes

There were 3 main categories of survey metrics, in addition to demographics and department. Program measures included characteristics of the coaching relationship (number of coachees; length and quality of meetings) and perceptions of the coaching program. Primary outcomes included burnout, well-being, and professional fulfillment. The Stanford Professional Fulfillment Index (PFI) scale measures an overall burnout score (5-point Likert scale [0–4], 10 items, total range 0–40) through subscales for Workplace Exhaustion (WE) and Interpersonal Disengagement (IPD) (5-point Likert scale [0–4], 6 items for WE and 4 items for IPD) [13]. The PFI also separately measures Professional Fulfillment (5-Point Likert scale [0–4], 6 items for Professional Fulfillment, total range 0–24). Well-being was measured using the PERMA (Positive Emotions, Engagement, Meaning, Relationships, Accomplishments) (5-point Likert scale [1–5], 15 items, total range 15–75) [12]. Secondary outcomes included program-related skills and benefits including specific coping

skills, experiences with reflection and receiving feedback, and use of coaching skills in other relationships.

Statistical analyses

Coaches who had submitted both a baseline survey and at least one end-of-year survey were included. This allowed for paired pre-post analysis. To maximize sample size, surveys were included even if not all questions were answered. Baseline, end-of-year-1 (EOY-1) and end-of-year-2 (EOY-2) surveys were paired for each coach.

Demographic data were summarized with descriptive statistics. The scores from burnout (and its WE and IPD subscales), and Professional Fulfillment were calculated as averages on a 0–10 point scale and PERMA was summarized into a continuous measure. All primary outcomes, secondary outcomes, and characteristics and perceptions of the coaching program were dichotomized into categorical measures. The dichotomization was done based on existing thresholds published in the literature (averages of burnout, WE, IPD: high $> = 3.325$, low < 3.325 ; sum of Professional Fulfillment: high $> = 8$; low < 8) [13][13]. If thresholds were not available, the variables were dichotomized into the top 2 quartiles and bottom 2 quartiles based on the median value at baseline. This latter approach was also used to convert all other survey metrics into categorical variables.

Initially, bivariate analyses were performed for all 3 categories of metrics to compare change from baseline to EOY-1, baseline to EOY-2, and EOY-1 to EOY-2. The coach population at these 3 time points was not balanced, and therefore, repeated measures analysis was not performed. For the primary outcomes, the continuous measures were compared between each of the three time points (baseline to EOY-1, baseline to EOY-2, and EOY-1 to EOY-2) with paired *t*-test. The categorical measures were compared through paired analyses using McNemar's test.

Then, adjusted analyses were performed through multivariate linear regression. The dependent variable was the continuous measure of the primary outcomes (burnout, well-being, and professional fulfillment). Covariates included the value of that primary outcome at baseline, time spent coaching, gender, race, and department. Lastly, bivariate analyses were performed to understand if there were elements of the coaching relationship or skills emphasized in coaching that had an association with the primary outcomes. Chi-squared tests were performed with unpaired categorical measures of both the primary outcomes vs. coaching characteristics and secondary outcomes. A two-sided *p* value < 0.05 was considered statistically significant for all tests. All statistical analyses were performed on Stata 17.

Table 1 Demographic characteristics of the participants and survey respondents

Demographics	
Total participants	136
% with completed surveys (<i>n</i>)	44.1 (60)
Characteristics	% of responses (<i>n</i>)
Gender	
Male	37.5 (21)
Female	62.5 (35)
Race	
White	85.6 (46)
Asian	9.1 (5)
Multi-racial	1.9 (1)
Other	5.4 (3)
African American, Pacific Islander, Hawaiian, American Indian	–
Hispanic or Latin origin	
Yes	5.4 (3)
No	94.6 (53)
Department	
Pediatrics	38.3 (23)
Surgical specialties	25 (15)
Medicine	36.7 (22)
Length of time spent coaching	
Coached for 1 year	48.3 (29)
Coached for 2 years	51.7 (31)

Results

Demographics and coaching program experience

From 2017 to 2019, 136 total coaches participated in the PDCP, and 44% (*n* = 60) of coaches submitted a baseline and at least one end-of-year survey. Table 1 describes the demographics of the respondents, 63% (*n* = 35) of whom identified as female, 85% (*n* = 46) of whom identified as white, and 25% (*n* = 15) of whom were from a surgical specialty. Forty-nine percent (*n* = 29) coached for 1 year and 52% (*n* = 31) coached for 2 years. Table 2 describes characteristics and perceptions of the coaching program. In the first year of coaching, 68% (*n* = 34) of coaches had more than 1 coachee, while in the second year, 53% (*n* = 16) of coaches had more than 1 coachee. There was no significant difference in these measures between the first vs. second years of coaching. Of coaches who started in 2017–2018, 95% (104/109) coached again in 2018–2019.

Bivariate analyses of primary outcomes vs. time spent coaching

The bivariate analyses of our primary outcomes vs. time spent coaching are reported in Table 3. There was a non-significant increase in the proportion of coaches with a low burnout score (68% at baseline [*n* = 38] to 69% at EOY-1 [*n* = 36, *p* = 0.69 vs. baseline] to 84% at EOY-2 [*n* = 26, *p* = 0.45 vs. baseline]). The proportion of coaches with low WE also increased, but not significantly (54% at baseline [*n* = 31] to 61% at EOY-1 [*n* = 33, *p* = 0.61 vs. baseline] to 74% at EOY-2 [*n* = 23, *p* = 0.39 vs baseline]). There was no

Table 2 Characteristics of the coaching relationship and perceptions of the coaching program during the first and second years of coaching

Coaching program experience	First year of coaching	Second year of coaching	<i>p</i> value
Number of coachees per coach			0.202
% with 1 coachee (<i>n</i>)	32 (16)	46.7 (14)	
% with > 1 coachees (<i>n</i>)	68 (34)	53.3 (16)	
Length of quarterly meetings			0.191
% 0–60 min long (<i>n</i>)	84 (42)	89.7 (26)	
% > 60 min long (<i>n</i>)	16 (8)	10.3 (3)	
Quality of communication in meetings			0.349
% endorsing poor/fair quality (<i>n</i>)	36.7 (18)	39.3 (11)	
% endorsing good/excellent quality (<i>n</i>)	63.3 (31)	60.7 (17)	
Do you think that the coaching program is useful to address trainees' biggest challenges?			1.00
% no/not sure (<i>n</i>)	1.9 (1)	3.1 (1)	
% somewhat yes or definitely yes (<i>n</i>)	98.1 (51)	96.9 (31)	
Would you advise other training programs to implement a coaching program?			1.00
% definitely would not/probably would not (<i>n</i>)	1.9 (1)	0 (0)	
% probably would/definitely would (<i>n</i>)	98.1 (51)	100 (32)	

Table 3 Comparison of means (with standard deviation) and proportions of burnout, WE, IPD, Professional Fulfillment, and PERMA at baseline to end of first year to end of second year

Outcomes	Baseline	End of 1st year	End of 2nd year	Baseline vs. EOY-1 <i>p</i> value	Baseline vs. EOY-2	EOY-1 vs. EOY-2
PFI: burnout (0–10 scale)						
Mean (STD)	2.74 (1.64)	2.5 (1.24)	2.48 (1.32)	0.643	0.699	0.958
% high, > = 3.325 (<i>n</i>)	32.1 (18)	30.8 (16)	16.1 (5)	1.000	0.688	0.453
% low, < 3.325 (<i>n</i>)	67.9 (38)	69.2 (36)	83.9 (26)			
WE: workplace exhaustion subscale (0–10 scale)						
Mean (STD)	3.48 (2.13)	3.05 (1.57)	2.83 (1.57)	0.255	0.264	0.844
% high, > = 3.325 (<i>n</i>)	46.6 (27)	38.9 (21)	25.8 (8)	0.607	0.388	0.688
% low, < 3.325 (<i>n</i>)	53.5 (31)	61.1 (33)	74.2 (23)			
IPD: interpersonal disengagement subscale (0–10 scale)						
Mean (STD)	2.19 (1.58)	2.14 (1.38)	2.27 (1.43)	0.667	0.569	0.78
% high, > = 3.325 (<i>n</i>)	24.6 (14)	19.1 (10)	28.1 (9)	0.774	1.000	0.453
% low, < 3.325 (<i>n</i>)	75.4 (43)	80.8 (42)	71.9 (23)			
Professional Fulfillment (0–10 scale)						
Mean (STD)	6.51 (1.91)	6.96 (1.86)	6.81 (1.49)	0.162	0.43	0.337
% high, > = 8 (<i>n</i>)	45.6 (26)	50 (27)	41.9 (13)	1.000	1.000	0.688
% low, < = 8 (<i>n</i>)	54.4 (31)	50 (27)	58.1 (18)			
PERMA: well-being (15–75 scale)						
Mean (STD)	57.3 (10.3)	56.9 (12.9)	54.7 (15.6)	0.852	0.605	0.231
% In top 2 quartiles (<i>n</i>)	45.8 (27)	52.8 (28)	40.6 (13)	0.754	0.508	0.219
% in bottom 2 quartiles (<i>n</i>)	54.2 (32)	47.2 (25)	59.4 (19)			

significant change in the percentage of coaches scoring low on IPD (75% at baseline [$n=43$] to 81% at EOY-1 [$n=42$, $p=0.77$ vs. baseline] to 72% at EOY-2 [$n=23$, $p=1.0$]). The proportion of coaches with high Professional Fulfillment scores did not significantly change (46% at baseline [$n=26$] to 50% at EOY-1 [$n=27$, $p=1.0$] to 42% at EOY-2 [$n=13$, $p=1.0$]). Lastly, the proportion of coaches who scored in the top 2 quartiles for PERMA did not significantly change across timepoints (46% at baseline [$n=27$] to 52.8% EOY-1 [$n=28$, $p=0.75$ vs. baseline] to 41% at EOY-2 ($n=13$, $p=0.51$ vs. baseline)).

Benefits and skills from coaching

Table 4 demonstrates potential benefits and skills gained from coaching. A significantly higher proportion of coaches felt confident about staying emotionally balanced after coaching compared to baseline (55% at baseline [$n=32$] to 71% at EOY-1 [$n=37$, $p=0.04$ vs. baseline] to 66% [$n=21$, $p=0.03$ vs. baseline]). There was no significant change in the proportion of coaches who reported “Good/Excellent” experiences with skills emphasized in the PDCP, including practicing reflection (58% [$n=35$] at baseline vs. 68% at EOY-1 [$n=36$, $p=0.18$ vs. baseline] and 71% at EOY-2 [$n=24$, $p=1.0$]). Finally, compared to baseline, a

significantly higher proportion of coaches reported using the skills gained through the PDCP in other non-coaching relationships at EOY-1 and EOY-2 (i.e., faculty colleagues: 25% at baseline [$n=15$] vs. 69% at EOY-1 [$n=36$, $p<0.001$ vs. baseline] and 84.4% at EOY-2 [$n=27$, $p<0.001$ vs. baseline]).

Adjusted analyses of primary outcomes

We conducted multivariate linear regression to analyze the relationship between our primary outcomes and time spent coaching after adjusting for demographics, department, and the value of that outcome at baseline. These outcomes are listed in Tables 6, 7, 8, 9, 10 in the appendix.

Burnout was significantly lower if coaches had low burnout at baseline ($\beta = -1.30$, $p<0.001$) and significantly higher at EOY-1 and EOY-2 compared to baseline ($\beta = 0.0780$, $p=0.039$ and $\beta = 0.871$, $p=0.039$, respectively). There was no significant change at EOY-2 compared to EOY-1 ($\beta = 0.091$, $p=0.781$). On subscale analyses, WE was significantly lower for coaches with low WE at baseline ($\beta = -1.35$, $p=0.001$). IPD was significantly lower for coaches who had low IPD at baseline ($\beta = -1.29$, $p=0.001$). IPD was significantly higher at EOY-1 and EOY-2 compared to baseline ($\beta = 1.01$, $p=0.017$ and $\beta = 1.13$, $p=0.011$).

Table 4 Comparison of proportions of coaches endorsing program-related skills and benefits at baseline to end of first year to end of second year (* indicates $p < 0.05$)

Skills and benefits	Baseline % response chosen (n)	End of 1st year	End of 2nd year	Baseline vs. EOY-1 p value	Baseline vs. EOY-2	EOY-1 vs. EOY-2
Being able to choose the best coping responses for hard situations						
I can't do this or I can do this a little bit	39.0 (23)	20.8 (11)	25 (8)	0.119	0.070	0.688
I can do this pretty well/extremely well	61 (36)	79.2 (42)	75 (24)			
Coming up with emotionally balanced thoughts even during negative times						
I can't do this or I can do this a little bit	44.8 (26)	28.9 (15)	34.4 (11)	0.039*	0.031*	0.625
I can do this pretty well/extremely well	55.2 (32)	71.2 (37)	65.6 (21)			
Asking people in my life for support or assistance whenever I need it						
I can't do this or I can do this a little bit	49.2 (29)	50.9 (27)	43.8 (14)	1.000	0.688	0.726
I can do this pretty well/extremely well	50.9 (30)	49.1 (26)	56.3 (18)			
Rate your experiences this past year with reflecting on your own performance						
Poor/Fair	41.7 (25)	32.1 (17)	29.4 (10)	0.180	0.508	1.000
Good/Excellent	58.3 (35)	67.9 (36)	70.6 (24)			
Rate your experiences this past year with receiving feedback about your skills/performance as a physician						
Poor/Fair	65 (39)	55.6 (30)	52.9 (18)	0.146	0.070	1.000
Good/Excellent	35 (21)	44.4 (24)	47.1 (16)			
Since training, have you used the skills you have learned in PDCP with.....colleagues in medicine?						
No, I have not	74.6 (44)	30.8 (16)	15.6 (5)	<0.001*	<0.001*	0.25
Yes, I somewhat/definitely have	25.4 (15)	69.2 (36)	84.4 (27)			
...colleagues in nursing?						
No, I have not	86.4 (51)	69.2 (36)	50 (16)	0.012*	<0.001*	0.219
Yes, I somewhat/definitely have	13.6 (8)	30.8 (16)	50 (16)			
...family & friends?						
No, I have not	64.4 (38)	26.9 (14)	15.6 (5)	<0.001	<0.001*	0.125
Yes, I somewhat/definitely have	35.6 (21)	73.1 (38)	84.4 (27)			
...mentors or advisors?						
No, I have not	84.8 (50)	50 (26)	28.1 (9)	<0.001*	<0.001*	0.109
Yes, I somewhat/definitely have	15.2 (9)	50 (26)	71.9 (23)			
...with patients?						
No, I have not	68.4 (39)	39.2 (20)	25 (8)	<0.001*	0.004*	0.625
Yes, I somewhat/definitely have	31.6 (18)	60.8 (31)	75 (24)			

respectively). There was no significant change in IPD at EOY-2 compared to EOY-1 ($\beta = 0.127$, $p = 0.711$).

PERMA was significantly higher for coaches who scored highly on PERMA at baseline ($\beta = 10$, $p < 0.001$) and significantly lower at EOY-1 and EOY-2 compared to baseline ($\beta = -4.47$, $p = 0.011$ and $\beta = -7.01$, $p < 0.001$ respectively). There was no significant change in PERMA at EOY-2 compared to EOY-1 ($\beta = -2.54$, $p = 0.146$). Professional Fulfillment was significantly higher for coaches who had high Professional Fulfillment at baseline ($\beta = 2.05$,

$p < 0.001$) and for non-surgical coaches compared to surgical coaches ($\beta = 0.361$, $p = 0.0430$).

Bivariate analyses of primary outcomes vs. coaching program characteristics and skills

The categorical measures of primary outcomes were also compared with categorical measures of coaching program characteristics and skills through bivariate analyses. The results are reported in Table 5. Of coaches who scored in the bottom 2 quartiles of PERMA, 76% had more than 1 coachee

Table 5 Comparison of proportions of coaches scoring high vs. low on primary outcomes with characteristics of the coaching relationship and program-related skills (*indicates $p < 0.05$)

	PFI burnout		WE subscale		IPD subscale		Professional Fulfillment		PERMA	
	High	Low	High	Low	High	Low	Low	High	Bottom 2 quartiles	Top 2 quartiles
Number of coachees per coach										
% with 1 coachee	33.3	39.3	34.5	40	42.1	35.6	32.6	44.4	24.4	51.3
% with > 1 coachees	66.7	60.7	65.5	60	57.9	64.4	67.4	55.6	75.6	48.7
<i>p</i> value	0.631		0.626		0.609		0.278		0.013	
Length of quarterly meetings										
% 0–60 min long	80	89.3	85.7	86	83.3	88.1	88.1	83.3	92.5	79.5
% > 60 min long	20	10.7	14.3	14	16.7	11.9	11.9	16.7	7.5	20.5
<i>p</i> value	0.292		0.972		0.596		0.547		0.095	
Quality of communication in meetings										
% endorsing poor/fair quality	42.9	37.0	41.4	36.2	38.9	37.9	45.2	26.5	45	29.7
% Endorsing Good/Excellent Quality	57.1	63.0	58.6	63.8	61.1	62.1	54.8	73.5	55	70.3
<i>p</i> value	0.642		0.650		0.942		0.092		0.167	
I am confident about being able to choose the best coping responses for hard situations										
% I can't do this or I can do this a little bit	43.6	23.2	37.5	23.3	42.4	25.2	38.2	16.9	42.1	14.7
% I can do this pretty well/extremely well	56.4	76.8	62.5	76.7	57.6	74.8	61.8	83.1	57.9	85.3
<i>p</i> value	0.018*		0.067		0.058		0.005*		<0.001*	
I can come up with emotionally balanced thoughts even during negative times										
% I can't do this or I can do this a little bit	55.3	29.6	48.2	29.1	62.5	29.2	46.7	23.4	46	26.5
% I can do this pretty well/extremely well	44.7	70.4	51.9	70.9	37.5	70.8	53.3	76.6	54.1	73.5
<i>p</i> value	0.005*		0.022*		0.001*		0.004*		0.016*	
I can ask people in my life for support or assistance whenever I need it										
% I can't do this or I can do this a little bit	66.7	43.4	57.1	43	60.6	46.7	63.2	30.8	67.1	27.9
% I can do this pretty well/extremely well	33.3	56.6	42.9	57	39.4	53.3	36.8	69.2	32.9	72.1
<i>p</i> value	0.014*		0.100		0.163		<0.001*		<0.001*	
Rate your experiences this past year with reflecting on your own performance										
% poor/fair	60.5	27	50	26.4	57.6	29	44.7	23.4	41.3	26.9
% good/excellent	39.5	73	50	73.6	42.4	71	55.3	76.6	58.7	73.1
<i>p</i> value	<0.001*		0.004*		0.003*		0.008*		0.070	
Rate your experiences this past year with receiving feedback about your skills/performance as a physician										
% poor/fair	69.2	56	65.5	54	78.8	53.4	63.2	52.3	59.2	58.2
% good/excellent	30.8	44	34.5	46	21.2	46.3	36.8	47.7	40.8	41.8
<i>p</i> value	0.153		0.178		0.010*		0.193		0.903	

while 24% had only 1 coachee ($p=0.013$). The length of the quarterly meetings and quality of communications in the meetings did not show any significant relationships with the primary outcomes.

Being able to come up with emotionally balanced thoughts during negative times (vs. unable to do so) was associated with a significantly higher proportion of coaches with low burnout (70 vs. 30%, $p=0.005$), low WE (71 vs. 29%, $p=0.022$), low IPD (71 vs. 29%, $p=0.001$), high Professional Fulfillment (77 vs. 23%, $p=0.004$) and the top 2 quartiles for PERMA (74 vs. 27%, $p=0.016$). Being able to ask people for support (vs. unable to do so) was associated with a significantly higher proportion of coaches who had low burnout (57 vs. 43%, $p=0.014$), low Professional Fulfillment 69 vs. 31%, $p<0.001$), and in the top 2 quartiles of PERMA (72 vs. 28%, $p<0.001$). Good or excellent experiences with reflection (vs. poor or fair experiences) were associated with a significantly higher proportion of coaches who had low burnout (73 vs. 27%, $p<0.001$), low WE (74 vs. 26%, $p=0.004$), low IPD (71 vs. 29%, $p=0.003$), and high Professional Fulfillment (77 vs. 23%, $p=0.008$).

Discussion

We evaluated the experience of academic physician coaches participating in a longitudinal coaching program. Coaches reported several benefits related to their experience, such as being able to stay emotionally balanced and using their coaching skills in non-coaching relationships. In our adjusted analyses, burnout, interpersonal disengagement, and well-being decreased from baseline. However, these changes do not correlate with time spent coaching and appeared to be more a reflection of the coaches' baseline state rather than an effect of program participation itself. On further analyses, better experiences and comfort with the skills emphasized in the PDCP were significantly associated with better outcomes on burnout, professional fulfillment, and well-being.

Coaches reported benefits from coaching in the PDCP. More coaches felt confident about staying emotionally balanced during difficult times after coaching compared to baseline. There was also an upward trend in the proportion of coaches who had good/excellent experiences with reflection and receiving feedback, though these were not significant. A significantly higher proportion of coaches reported using skills gained through coaching in other relationships after coaching compared to baseline. Given that baseline was measured in between completion of training and onset of coaching, it is likely that use of these skills with

non-coachees was not necessarily motivated by the training, but by using them with the coachees.

Based on the available literature, we hypothesized that burnout might increase over time in our coach population. In our bivariate analyses, both the mean scores and proportions of coaches with high burnout and WE trended downward over time, but this change was not significant. At the 3 time points, only 16.1 to 32.1% of our coaches had high burnout, with mean scores ranging from 2.48 to 2.74. This is similar to reports of non-coach physicians in other studies measuring burnout using the PFI. A 2020 study of 7364 physicians across 9 institutions had a mean burnout of 2.8 [14]. The PERMA scores in our study ranged in the three time points from 54.7 to 57.3, which is lower than the mean PERMA of 154 medicine faculty (60.9) who participated in PDCPs across multiple institutions. This suggests that our faculty in this study had slightly lower PERMA than prior cohorts [11]. We had also hypothesized that professional fulfillment might increase due to participation as a coach. We did not see an improvement in professional fulfillment in our coaches, and noted that the professional fulfillment means of our coaches was similar to the means of the population in the aforementioned 2020 study (6.5–6.96 compared to 6.61, respectively) [14]. Given the stressors that academic faculty physicians experience, a decline in professional fulfillment over time might be expected and the stability in that outcome that we see in our volunteer faculty coaches could be a “positive” finding. We are unable to explore this further without a control group of non-coach physicians; however, this is a notable result to explore in future studies.

It is challenging to compare the baseline values of our physician coaches to the non-coaching physicians given the lack of a control arm. However, published and internal data on burnout can provide some context. Rao et al. 2019 surveyed physicians at our institution from all specialties and found a burnout rate in 2017 of 45% on the Maslach Burnout Index (MBI) [15]. While this is a different burnout scale, Trockel et al. 2018 showed that MBI and Stanford PFI correlate strongly ($r=0.71$). Furthermore, internal survey data from this institution in 2019 found similar rates of burnout in physicians across all specialties on the Stanford PF and the MBI (~43%). Given that these scales are comparable, it is worth noting that the physician coaches in our study who started in 2017 (from select specialties) had a burnout rate of 35% at their baseline, slightly lower than the overall physician population at our institution.

Comparison of these findings with studies of other coaches is challenging given the heterogeneity of coaching programs and the few studies that have been performed. Elster et al.'s previously mentioned mixed methods analyses found that nearly 2/3 of academic physicians who were

trained and paid to coach medical students met criteria for burnout on the MBI. It is unclear why rates of burnout in Elster's study are different from ours, though there are likely factors specific to the institution or setting [7].

Both Elster et al.'s and Brooks et al.'s studies of medical student coaches revealed similar themes that could contribute to a coach's professional satisfaction or fulfillment, such as guiding struggling trainees, building a longitudinal relationship, and observing their successes [6, 7]. Coaches in both studies were challenged by the fragmentation of one's professional identity due to the coaching role and the expectations and burden this carried. Our measure of Professional Fulfillment asks about a global view of the work experience, of which coaching is simply one part. Given that the professional benefits reported by coaches in prior studies may not be captured in this Stanford Professional Fulfillment metric, this could help explain why we do not observe our hypothesized improvement in Professional Fulfillment. Future studies should investigate effects of coaching on both global- and role-specific professional fulfillment.

We adjusted our analyses of primary outcomes vs. time spent coaching with covariates representing who the coaches were at baseline including the value of that primary outcome at baseline, gender, race, and department. The covariate with the largest effect size was how the coach scored on that outcome at baseline. After adjusting for baseline value, gender, race, and department, we found that burnout, IPD, and PERMA did decline from baseline. However, for those who coached for 2 years, none of the outcomes significantly changed between EOY-1 and EOY-2. Thus, this was likely not an accumulative, dose-dependent effect from coaching. These cohorts also had a very high retention rate, so this lack of change between EOY-1 and EOY-2 was likely not due to self-selecting attrition. Without a comparison arm, it is hard to interpret how much of these changes are due to time spent being a physician. Ultimately, the level of burnout, well-being, and professional fulfillment of the physicians at baseline influenced outcomes more significantly than the time spent coaching. Lastly, while the trends in burnout and IPD (improving) and well-being on PERMA (worsening) may seem contradictory, it is worth emphasizing that burnout and well-being are not mirror opposite constructs. They capture different elements of a subject's emotional health and recent work by this group has even shown that PERMA and burnout (on the MBI) are not correlated [11].

To further investigate if elements of the coaching experience were associated with our primary outcomes, we performed bivariate analyses of the outcomes vs. characteristics of the coaching relationships and skills emphasized in

the program. Having more than one coachee was significantly associated with scoring in the bottom 2 quartiles on PERMA, suggesting a potential impact of increased burden on the coach's time. Good or excellent experiences with reflection, as well as coping skills like staying emotionally balanced and asking for support, were associated with a higher proportion of coaches who had low burnout, high professional fulfillment, and scoring in the top 2 quartiles for PERMA. This suggests that these skills may be protective in these outcomes and could be emphasized further in training.

This study has several limitations to consider. There was a relatively low response rate when using paired data for analyses, which could indicate a non-response bias. Because we did not gather data on the non-respondents, we cannot draw inferences on how that population might differ from our respondents. The relatively low response rate also led to challenges in our analyses. Constructs such as burnout and well-being are complex and multi-factorial. While we attempted to control for some of these factors in our multivariate analyses, we were limited by our relatively low sample size. Furthermore, we did not have a comparison arm of non-coach physicians, which makes it harder to distinguish how much of the effects are due to being a coach vs. being a physician. Without a comparison arm, it also makes it harder to understand what "no change" may mean in a particular metric, as other coaching studies with comparison groups have shown a decline in many metrics within the comparison arm. This suggests a potentially supportive and stabilizing role that coaching may play in physician resilience and well-being. Finally, this study took place in a single institution with heterogeneity in coachee population, prior experience, and many other factors.

While this variability does make it harder to generalize this study to other academic physician coaches, our findings create an initial, albeit incomplete, profile of burnout, professional fulfillment, and well-being of the physician faculty who volunteer to be novice coaches. These findings demonstrate changes in these outcomes that warrant further investigation and provide potential elements and skills in coaching that can be emphasized to improve these outcomes. Future studies with larger populations and comparison groups will allow for further exploration of the role of coaching on the coach.

Appendix

See Tables 6, 7, 8, 9, 10.

Table 6 Adjusted analyses of PFI burnout through multivariate linear regression

Predictors	PFI burnout (0–10)		
	Coefficient	Std. error	<i>p</i> value
Burnout at baseline: low vs. high (ref)	– 1.30	0.36	<0.001
Race: non-white vs. white (ref)	– 0.378	0.313	0.230
Gender: female vs. male (ref)	0.231	0.253	0.364
Dept.: non-surgical vs. surgical subspecialties (ref)	– 0.167	0.152	0.270
Length: 1 year of coaching vs. baseline (ref)	0.780	0.373	0.039
Length: 2 years of coaching vs. baseline (ref)	0.871	0.418	0.039
Length: 2 years of coaching vs. 1 year (ref)	0.091	0.328	0.781
	<i>F</i> (6, 118) = 3.23, <i>R</i> -sq: 0.141		

Table 7 Adjusted analyses of workplace exhaustion subscale through multivariate linear regression

Predictors	WE: workplace exhaustion subscale (0–10)		
	Coefficient	Std. error	<i>p</i> value
WE at baseline: low vs. high (ref)	– 1.353	0.402	0.001
Race: non-white vs. white (ref)	– 0.292	0.394	0.458
Gender: female vs. male (ref)	0.423	0.319	0.187
Dept.: non-surgical vs. surgical subspecialties (ref)	– 0.300	0.190	0.118
Length: 1 year of coaching vs. baseline (ref)	0.469	0.408	0.253
Length: 2 years of coaching vs. baseline (ref)	0.3466	0.464	0.458
Length: 2 years of coaching vs. 1 year (ref)	– 0.123	0.406	0.762
	<i>F</i> (6, 124) = 3.50, <i>R</i> -sq: 0.145		

Table 8 Adjusted analyses of Interpersonal Disengagement subscale through multivariate linear regression

Predictors	IPD: interpersonal disengagement subscale (0–10)		
	Coefficient	Std. error	<i>p</i> value
IPD at baseline: low vs. high (ref)	– 1.29	0.386	0.001
Race: non-white vs. white (ref)	– 0.377	0.331	0.257
Gender: female vs. male (ref)	– 0.063	0.265	0.812
Dept.: non-surgical vs. surgical subspecialties (ref)	– 0.135	0.159	0.395
Length: 1 year of coaching vs. baseline (ref)	1.01	0.415	0.017
Length: 2 years of coaching vs. baseline (ref)	1.13	0.436	0.011
Length: 2 years of coaching vs. 1 year (ref)	0.126	0.340	0.711
	<i>F</i> (6, 122) = 2.35, <i>R</i> -sq: 0.104		

Table 9 Adjusted analyses of Professional Fulfillment through multivariate linear regression

Predictors	Professional Fulfillment (0–10)		
	Coefficient	Std. error	<i>p</i> value
Professional Fulfillment at baseline: high vs. low (ref)	2.05	0.361	<0.001
Race: non-white vs. white (ref)	– 0.144	0.373	0.700
Gender: female vs. male (ref)	– 0.627	0.288	0.032
Dept.: non-surgical vs. surgical subspecialties (ref)	0.361	0.176	0.043
Length: 1 year of coaching vs. baseline (ref)	– 0.699	0.358	0.053
Length: 2 years of coaching vs. baseline (ref)	– 0.760	0.398	0.059
Length: 2 years of coaching vs. 1 year (ref)	– 0.061	0.369	0.869
	<i>F</i> (6, 122) = 7.55, <i>R</i> -sq: 0.271		

Table 10 Adjusted analyses of PERMA (Well-Being) through multivariate linear regression

Predictors	PERMA (well-being) (15–75)		
	Coefficient	Std. error	<i>p</i> value
PERMA at baseline: top 2 quartiles vs. bottom 2 quartiles (ref)	10.0	1.72	<0.001
Race: non-white vs. white (ref)	– 1.94	1.72	0.263
Gender: female vs. male (ref)	– 1.64	1.38	0.235
Dept.: non-surgical vs. surgical subspecialties (ref)	– 0.0260	0.84	0.976
Length: 1 year of coaching vs. baseline (ref)	– 4.47	1.73	0.011
Length: 2 years of coaching vs. baseline (ref)	– 7.01	1.89	<0.001
Length: 2 years of coaching vs. 1 year (ref)	– 2.54	1.89	0.146
	<i>F</i> (6, 129) = 7.09, <i>R</i> -sq: 0.248		

Acknowledgements Thank you to Mike Healy, PhD and Yoon Soo Park, PhD for their input on the statistical analyses of the work. Thank you to Emil Petrusa, PhD for his help with the editing. Finally, the authors wish to acknowledge the Institute of Coaching, and in particular Carol Kauffman, PhD for the support in designing the original MGH Professional Development Coaching Program, which this intervention was based on.

Funding This work was supported by funding from the MGH Physicians Foundation (Grant number: 2850944).

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Palamara K, Kauffman C, Chang Y, Barreto EA, Yu L, Bazari H, et al. Professional development coaching for residents: results of a 3-year positive psychology coaching intervention. *J Gen Intern Med* [Internet]. 2018;33(11):1842 (PMCID6206332).
2. Dyrbye LN, Shanafelt TD, Gill PR, Satele DV, West CP. Effect of a professional coaching intervention on the well-being and distress of physicians: a pilot randomized clinical trial. *JAMA Intern Med* [Internet]. 2019;179(10):1406 (PMCID6686971).
3. Müller AA, Kotte S, Möller H. Coach and no regrets about it: on the life satisfaction, work-related mental strain, and use of supervision of workplace coaches. *Coaching* [Internet]. 2019;13(1):16–29. <https://doi.org/10.1080/17521882.2019.1636841>.
4. Schermuly C. Negative effects of coaching for coaches—an explorative study [Internet]. *Int Coach Psychol Rev*. 2014. <https://doi.org/10.1037/t42021-000>.
5. Hamidi MS, Bohman B, Sandborg C, Smith-Coggins R, De Vries P, Albert MS, et al. Estimating institutional physician turnover attributable to self-reported burnout and associated financial burden: a case study. *BMC Health Serv Res* [Internet]. 2018;18(1):1–8. <https://doi.org/10.1186/s12913-018-3663-z>.
6. Brooks JV, Ista K, Barth BE. Becoming a coach: experiences of faculty educators learning to coach medical students. *BMC Med Educ* [Internet]. 2020;20(1):1–7. <https://doi.org/10.1186/s12909-020-02119-z>.
7. Elster MJ, O’Sullivan PS, Muller-Juge V, Sheu L, Kaiser SV, Hauer KE. Does being a coach benefit clinician-educators? A mixed methods study of faculty self-efficacy, job satisfaction and burnout. *Perspect Med Educ* [Internet]. 2022;11(1):45–9 (PMCID8371581).
8. Palamara K, Kauffman C, Stone VE, Bazari H, Donelan K. Promoting success: a professional development coaching program for interns in medicine. *J Grad Med Educ* [Internet]. 2015;7(4):630–7 (PMCID4675202).
9. Palamara K, McKinley SK, Chu JT, Merrill A, Yu L, Parangi S, et al. Impact of a virtual professional development coaching program on the professional fulfillment and well-being of women surgery residents: a randomized controlled trial. *Ann Surg*. 2022. <https://doi.org/10.1097/sla.0000000000005562>.
10. Stein SL, Chu JT, Merrill A, McKinley SA, Parengi S, Donelan K, et al. Implementation of a Coaching Program by a National Surgical Association. *J Surg Educ* [Internet]. 2022. Available from: <https://pubmed.ncbi.nlm.nih.gov/35995676/>. Cited 9 Sep 2022.
11. Larsen D, Chu JT, Yu L, Chang Y, Donelan K, Palamara K. Correlating burnout and well-being in a multisite study of internal medicine residents and faculty. *J Gen Intern Med* [Internet]. 2021;36(5):1422–6.
12. Palamara K, Chu JT, Chang Y, Yu L, Cosco D, Higgins S, et al. Who benefits most? a multisite study of coaching and resident well-being. *J Gen Intern Med* [Internet]. 2022;37(3):539–47. <https://doi.org/10.1007/s11606-021-06903-5>.
13. Trockel MT, Hamidi MS, Menon NK, Rowe SG, Dudley JC, Stewart MT, et al. Self-valuation: attending to the most important instrument in the practice of medicine. *Mayo Clin Proc* [Internet]. 2019;94(10):2022–31.
14. Trockel MT, Menon NK, Rowe SG, Stewart MT, Smith R, Lu M, et al. Assessment of physician sleep and wellness, burnout, and clinically significant medical errors. *JAMA Netw Open* [Internet]. 2020;3(12):e2028111–e2028111.
15. Rao SK, Ferris TG, Hidrue MK, Lehrhoff SR, Lenz S, Heffernan J, et al. Physician burnout, engagement and career satisfaction in a large academic medical practice. *Clin Med Res* [Internet]. 2020;18(1):3–10.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.