BMJ Open

A 180° view on general practitioners' leadership: comparison of leader and practice team assessments using data from the IMPROVEjob study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2022-066298
Article Type:	Original research
Date Submitted by the Author:	07-Jul-2022
Complete List of Authors:	Schmidt, Manuela; University Hospital Bonn, Institute of General Practice and Family Medicine Seifried-Dübon, Tanja; University Hospital Tuebingen Osianderstraße 5, Department of Psychosomatic Medicine and Psychotherapy Göbel, Julian; University Hospital Bonn, Institute of General Practice and Family Medicine Degen, Lukas; University Hospital Bonn, Institute of General Practice and Family Medicine Werners, Brigitte; Ruhr University Bochum Grot, Matthias; Ruhr University Bochum Rind, Esther; Institute for Occupational Medicine, Social Medicine and Health Services Research , University Hospital Tuebingen Pieper, Claudia; University of Duisburg-Essen Jöckel, Karl-Heinz; Institute of Medical Informatics, Biometry and Epidemiology, University Duisburg- Essen Linden, Karen; University Hospital Bonn, Institute of General Practice and Family Medicine Rieger, Monika A.; Institute for Occupational Medicine, Social Medicine and Health Services Research , University Hospital Tuebingen Weltermann, Birgitta; University Hospital Bonn, Institute of General Practice and Family Medicine
Keywords:	PRIMARY CARE, Human resource management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

A 180° view on general practitioners' leadership: comparison of leader and practice team assessments using data from the IMPROVE*job* study

Authors: Manuela Schmidt[1]*, Tanja Seifried-Dübon[2], Julian Göbel[1], Lukas Degen[1], Brigitte Werners[3], Matthias Grot[3], Esther Rind[4], Claudia Pieper[5], Karl-Heinz Jöckel[5 6], Karen Linden[1], Monika A. Rieger[4], Birgitta M. Weltermann[1] on behalf of the IMPROVE*job* consortium

[1] Institute of General Practice and Family Medicine, University Hospital Bonn, University of Bonn, Venusberg-Campus 1, 53127 Bonn, Germany

[2] Department of Psychosomatic Medicine and Psychotherapy, University Hospital Tuebingen, Osianderstraße 5, 72076 Tuebingen, Germany

[3] Institute of Management, Operations Research, Ruhr University Bochum, Universitätsstr. 150, 44801 Bochum, Germany

[4] Institute of Occupational and Social Medicine and Health Services Research, University Hospital Tuebingen, Wilhelmstr. 27, 72074 Tuebingen, Germany

[5] Institute for Medical Informatics, Biometry and Epidemiology, University Hospital Essen, University of Duisburg-Essen, Hufelandstr. 55, 45147 Essen, Germany

[6] Center for Clinical Trials, University Hospital Essen, University of Duisburg-Essen, Hufelandstr. 55, 45147 Essen, Germany.

*Correspondence to: manuelaschmidt@uni-bonn.de

Abstract

Objectives: This study aims to identify general practitioners' leadership skills using self- and staff ratings based on the Full Range of Leadership Model (FRLM) and the Leader-Member Exchange (LMX).

Setting: Cross-sectional analysis with questionnaires among German general practice staff participating in the IMPROVE*job* trial.

Participants: The study population comprised 60 practices with 366 participants, of which 84 GP leaders and 282 staff members (254 practice assistants and 28 employed physicians).

Primary and secondary outcome measures: Leadership was measured using the Integrative Leadership Questionnaire (German *FIF*) and the Leader-Member Exchange (LMX-7) questionnaire, with leaders rating themselves and practice staff rating their leaders. The thus provided 180° feedback was analyzed by paired mean comparisons at the participant level and Kruskal-Wallis tests at the practice level.

Results: Practice owners' self-ratings were higher than practice assistants' ratings for all leadership dimensions. Interestingly, employed physicians' ratings were higher for the dimensions 'performance development' and 'providing a vision', as well as for transactional leadership. Statistically significant differences were found for transformational leadership ($p<.001$, $d=.41$), especially for the dimensions 'innovation' ($p<.001$, $d=.69$) and 'individuality focus' ($p<.001$, $d=.50$). For transactional leadership, only the dimension 'goal setting' showed significant differences ($p<.01$, $d=.30$). There were no significant differences between single and group practices. The LMX-7 scale (ranging from 7 to 35) showed a high relationship quality for both leaders and staff (26 for practice assistants and 28 for practice owners).

Conclusions: This analysis of GP leaders' leadership with self- and staff ratings showed consistent ratings of good relationships, but also highlighted the potential for leadership interventions to improve goal-setting, innovation, and individuality focus.

Trial registration: German Clinical Trials Register, DRKS00012677. Registered 16 October 2019.

Keywords: leadership, leadership quality, general physician, practice, 180-degree feedback, working conditions

Strenghts & limitations:

- Presentation of an innovative 180° feedback approach in the GP setting which allows for analysis of the different occupational fields in the practice
- The data reflects a typical spectrum of German general practices including teaching and non-teaching practices as well as solo and group practices
- Interpretation of the results is limited by the data's cross-sectional nature therefore cause-effect analyses are not possible

- The results for group practices are limited as staff were asked to rate practice leaders in general, not on individual level. However, this approach reflects leadership in group practices typically executed by a leadership team.

Background

Leadership has become an important topic in the medical field, as associations between poor leadership and various outcomes at leader, staff, and patient levels are described. Leadership issues comprise lack of organization, recognition for work, and opportunities to use one's abilities. Among the related adverse outcomes are low job satisfaction [1, 2], lack of staff well-being [3], and poor quality of patient care [1, 4, 5].

Successful efforts to improve leadership have been undertaken in the hospital [6] and healthcare management settings [7], while leadership in primary care practices remains poorly addressed. However, improving leadership in this setting is important because general practitioners (GPs) and practice assistants were twice as likely to report high chronic stress as the general population [8]. Also, higher rates of burnout were documented, especially for young, female, and part-time employed physicians in GP group practices [9]. So far, there are no studies addressing leadership issues in primary care grounded on current leadership theories and frameworks for action, although theory-based leadership intervention(s) have improved respective skills in other medical [6] and non-medical fields [10, 11].

One of the most significant leadership frameworks is the Full Range of Leadership Model (FRLM), which integrates transactional, transformational, and negative leadership [12, 13]. Transactional leadership describes leaders' structuring of work situations, the exchange of contingent rewards (e.g., work against salary), and management by exception [12–14]. In contrast, transformational leadership moves beyond leaders' and staff's self-interests. It focuses on the staff's attitudes and values regarding overarching goals such as self-actualization, organizational achievements, and the well-being of others and society as a whole [12, 13]. Associated leadership behaviors include inspirational motivation, intellectual stimulation, and individualized consideration. Additional aspects are charismatic relationships based on trust and confidence (so-called attributed idealized behavior) and a collective sense of action (so-called idealized influence behavior) [14]. Although they are described as different leadership behaviors, studies indicate that transactional and transformational leadership are highly interrelated [15]. Transactional leadership is often valued as the leadership foundation, while additional transformational leadership creates an 'augmentation effect' that may stimulate staff's extra efforts and high performance, as shown for health care workers [16]. In various settings, transformational and transactional leadership influenced organizational outcomes positively, e.g., performance indicators and employees' job satisfaction [14, 15]. Transformational leadership is associated with lower job stress and strain, less anxiety, higher well-being, and better outcomes for occupational safety [17]. While transformational and transactional leadership include desired

behaviors, a leader can also behave in a way that is detrimental to employees and the organization as a whole [18]. This negative leadership style includes laissez-faire leadership, which represents a highly passive leadership style where the leader's activities are at a minimum [12, 14], and destructive leadership, where harmful interpersonal behavior takes place which is not related to the leadership task [18].

Another essential leadership theory, the Leader-Member Exchange (LMX), reflects the relationship between leaders and staff. It concentrates on the perceived quality of the dyadic relationship between a staff member and their immediate leader [12, 19]. A high-quality relationship positively influences employees' work-related well-being and is associated with higher job satisfaction for health care workers [16].

The described focus on the relationship between leaders and staff is reflected in current methodological strategies. For example, the Leader-Member Exchange questionnaire 'LMX-7' [19] and the newly developed German questionnaire 'Fragebogen für integrative Führung' (FIF; in English: Questionnaire for Integrative Leadership) (Rowold & Poethke, 2017; Rowold & Schlotz, 2009) allow for a multi-rater perspective: the leaders' and the staff's views on the leaders' behavior are measured and compared providing 180° feedback. Including assessments from different perspectives creates a more comprehensive picture of the leaders' actual skills and performances [20, 21]. While such approaches are increasingly applied in medical education and graduate training [22, 23], they have not been used to evaluate GP leaders.

The publicly funded IMPROVE*job* study aims to improve the job satisfaction of physician leaders and practice personnel of German GP practices focusing on leadership, communication, and work processes [24, 25]. Based on the described leadership concepts, a 180° feedback approach compared GP leaders' self-ratings and their staff's ratings of their leadership skills.

Methods

This analysis draws on the baseline data of the IMPROVE*job* study [25], which is designed as a cluster-randomized controlled trial (cRCT) to improve job satisfaction among practice personnel. The details are described in the study protocol [24].

Participants

A total of 56 GP practices in the North Rhine region in Germany were recruited by the Institute of General Practice and Family Medicine of the University of Bonn, aiming for approximately equal strata

of single and group practices as well as of teaching and non-teaching practices. The study aimed to recruit all practice team staff, including physician leaders, employed physicians, and practice assistants. Practice owners answered a short questionnaire on practice characteristics and the questionnaire for practice leaders. Employed physicians and practice assistants filled out different versions of the same employee questionnaire. For details, see [24].

Patient and public involvement

No patient involved

Ethics

The study was approved first by the Ethics Committee of the Medical Faculty of the University of Bonn (reference number: 057/19, date of approval: 20/02/2019). In addition, the Ethics Committees of the Medical Association North-Rhine (Lfd-Nr.: 2019107) and of the Medical Faculty, University Hospital of Tuebingen (project no.: 446/2019BO2) approved the study protocol. The study was performed in accordance with the Declaration of Helsinki. All participating practice team members received written information and signed informed consent forms.

Measures

This analysis uses the IMPROVE*job* participants' baseline data on sociodemographic, professional, and work-related characteristics [25], as well as the following two leadership scales:

1. Integrative Leadership Questionnaire (FIF)

   Transformational, transactional, and negative leadership were measured using the FIF questionnaire [26]. Its scales' validity and internal consistency are confirmed for different populations [26, 27]. The FIF has been used in non-medical [28] and hospital settings [29], but not in primary care.

   All 40 items of the FIF are answered on a 5-point Likert scale and are worded to reflect either the leaders' or the staff's position [26].

   The measures comprise:

   - the transformational leadership scale, consisting of six dimensions: innovation, team spirit, performance development, individuality focus, providing a vision, and being a role model;
   - the transactional leadership scale with two dimensions: goal setting, and management by exception; and
   - the negative leadership scale with two dimensions: laissez-faire, and destructive leadership.

2. Leader-Member Exchange (LMX-7)

The relationship quality between leaders and staff is measured using the Leader-Member Exchange questionnaire (LMX-7) with seven items (5-point Likert scales), which are also worded to reflect the leader or staff position [19, 30, 31].

180° feedback

The multi-rater, 180° approach is applied to the leadership scales, measuring leadership skills from different perspectives within GP practices to create a more comprehensive picture of the actual performance of the leader in question [20, 21] and improve the accuracy of the assessment [22, 23]. While results of such assessments are usually shared with the ratee [26], previous studies showed mixed reactions in the medical setting [32, 33]. Therefore, the results of this 180° feedback are not shared with the participating practices but are used for research purposes only on an aggregated level.

Statistical analysis

Statistical analyses were conducted with SAS version 9.4 for the sociodemographic characteristics and SPSS on Windows version 26 for additional analyses. All analyses were carried out at the participant and the practice level.

The FIF data were analyzed according to the official manual [26]. Mean scores for transformational, transactional, and negative leadership were summarized both for the respective main scale and the dimensions: for transactional and transformational leadership, they ranged from 1 (worst rating) to 5 (best rating); for negative leadership, they ranged from 1 (best rating) to 5 (worst rating). To allow for a comparison with other settings, scores were standardized using T-scaling tables as defined by Rowold & Poethke [26]. These T-values are based on a normal distribution around 50 (SD 10). Thus, values above 70 only reflect about 2% of the reference population from German-speaking countries [26].

The LMX-7 was analyzed per standard protocol by creating a sum score of all seven items without transformation [31]. Thus, the LMX-7 score may range from 7 to 35, with five standard categories for interpretation: score 7 to 14 = very low; 15 to 19 = low; 20 to 24 = moderate; 25 to 29 = high, 30 to 35 = very high [34]. Inadvertently, question seven was missing on all employed physicians' questionnaires, which reduced the answered questions to six. As the LMX-7 manual does not suggest a standard approach for missing values, we excluded employed physicians from further analyses.

Following the 180° feedback approach, the combined mean scores of employed physicians and practice assistants per practice were compared to the self-assessment of their respective leaders using

paired t-tests, as the data satisfied the condition of a normal distribution with the Kolmogorov-Smirnov test. Cohen's d was applied to determine the effect size of mean comparisons with the following standard interpretations: small effects from d=.2, medium from d=.5, and high from d=.8 [35].

Self-ratings were available for each leader, as they were part of the individual questionnaire, allowing for 84 ratings in 60 practices. In single practices, the staff ratings were compared to the leader's assessment. In practices with more than one owner (group practices), each leader's self-rating was compared with the respective rating of the practice personnel, who were asked to rate leadership in the entire practice, not stratified by individuals.

In addition, the ratings of the transformational and transactional leadership scales were compared by practice type (group vs. single and teaching vs. non-teaching practices) using the Kruskal-Wallis tests because the data for practice comparisons did not satisfy the conditions for parametric tests.

Results

Population

The baseline data of the IMPROVE*job* study [25] included 366 participants from 60 practices, consisting of 84 practice owners, 28 employed physicians, and 254 practice assistants. The mean age of all participants was 44.4 years, with a mean of 54.3 years for practice owners, 44.8 for employed physicians, and 41.0 for practice assistants. Among the practice assistants, 99.6% were female, as were 76.6% of the employed physicians and half of the practice owners (52.4%). Most practice owners worked full-time (90.5%), as did about a quarter of the employed physicians (28.6%) and 41.5% of the practice assistants (see Table 1). For details on the sociodemographic descriptions, see [25].

Table 1: Sociodemographic description of participants at baseline [25]

Variable	Total sample N=366	Practice owners N=84	Employed physicians N=28	Practice assistants N=254
Female, %	87.1	52.4	78.6	99.6
Age in years, mean (SD)	44.4 (12.8)	54.3 (6.2)	44.8 (9.8)	41.0 (13.0)
Years in current practice, mean (SD)	10.0 (9.1)	15.3 (8.4)	3.9 (5.4)	8.8 (8.9)
Working full-time, %	52.0	90.5	28.6	41.5
Persons in household over 18 years, mean (SD)	2.2 (1.0)	2.1 (1.0)	2.0 (0.5)	2.2 (1.1)
Persons in household under 18 years, mean (SD)	1.2 (1.0)	1.3 (1.3)	1.4 (1.0)	1.0 (0.9)
Care for next-of-kin, %	20.8	21.7	0.0	22.9

On average, practice owners had been accredited for 26.6 years and licensed for the statutory health insurance for 16.4 years. Seven (25%) of the employed physicians were in GP training. Practice assistants had graduated on average 19.9 years ago, while 7.5% were still in training.

Of the 60 practices, 21 (35%) were single and 39 (65%) were group practices; of these, 34 were teaching (57%) and 26 (43%) were non-teaching practices. On average, practices had been located in their area for 20.4 years (SD 14 years).

Leadership

The transactional and transformational leadership scales showed a high internal consistency with Cronbach's α = .74 to .93 for the staff members' assessment and Cronbach's α = .72 to .87 for the leaders' assessment. For negative leadership, the scales showed a sufficient internal consistency for staff members' (Cronbach's α = .73 to .80) but not for leaders' assessments (Cronbach's α = .47 to .68).

The mean results of the FIF were within the reference population's average (one standard deviation) range. Based on raw values, employed physicians rated their leaders consistently better than practice assistants and —in part— also better than the leaders themselves. While practice assistants rated their leaders more poorly than the practice owners in raw values, reference T-values showed only minor differences; see Table 2 for details.

The LMX-7 scale showed an internal consistency of Cronbach's α = .88 for staff members (practice assistants) and α = .71 for leaders. All groups showed a high relationship quality, scoring 28 for practice owners and 26 for practice assistants. As the seventh question was missing for employed physicians, they were excluded from the analysis. However, the sum score of the remaining six questions also showed a high score of 24.9 out of 30; see Table 2 for details.

Table 2: Leadership assessment by employment group: main scales (in bold) and dimensions (LMX values can range from 7 to 35, FIF scales from 1 to 5)

	Practice owners (N=84)				**Employed physicians (N=28)**				**Practice assistants (N=254)**			
	M	**SD**	**T***	**n**	**M**	**SD**	**T***	**N**	**M**	**SD**	**T***	**n**
Transformational leadership	**3.9**	**0.6**	**45**	**84**	**3.9**	**0.7**	**56**	**27**	**3.5**	**0.8**	**52**	**237**
Innovation	4.2	0.6	49	84	4.0	1.0	55	28	3.7	0.9	52	247
Team spirit	4.1	0.7	49	84	3.8	1.0	54	28	3.6	1.1	52	251
Performance development	3.6	0.8	44	84	4.1	0.7	57	27	3.5	1.0	51	247
Individuality focus	3.9	0.7	47	84	3.7	1.0	54	28	3.5	1.1	53	249

Providing a vision	3.5	0.9	45	84	3.6	0.9	55	28	3.2	1.1	51	245
Being a role model	4.1	0.6	45	84	4.0	0.8	55	27	3.7	1.0	52	246
Transactional leadership	**3.4**	**0.7**	**47**	**83**	**3.5**	**0.7**	**54**	**27**	**3.2**	**0.8**	**50**	**244**
Goal setting	3.5	0.7	44	83	3.7	0.9	56	27	3.1	1.0	50	246
Management by exception	3.3	0.8	51	83	3.4	0.8	52	27	3.3	0.9	51	245
Negative leadership	**1.5**	**0.5**	**51**	**83**	**1.5**	**0.6**	**45**	**28**	**1.7**	**0.7**	**47**	**248**
Laissez-faire	1.6	0.6	52	83	1.6	0.8	45	28	1.7	0.8	46	249
Destructive	1.4	0.5	51	83	1.4	0.6	46	28	1.6	0.7	48	248
LMX-7	**28.1**	**2.6**	**-**	**81**	**n/a**	**n/a**	**-**	**n/a**	**26.7**	**4.8**	**-**	**222**

*Reference T-values range from 0 to 100, as defined by Rowold & Poethke 2017

180° leadership feedback

Practice owners self-rated their leadership skills slightly better than their staff for all dimensions except for 'management by exception'. There were no statistically significant differences for negative leadership. For transactional leadership, goal setting showed a statistically significant difference with a low effect size (p=.009, d=.30). Leaders' scores on transformational leadership were significantly higher than the scores of the teams, with the dimension for innovation reaching the strongest effect size (p=<.001, d=.69), followed by individuality focus with a medium effect size (p=<0.001, d=.50), and team spirit as well as being a role model with slightly smaller, but significant effect sizes. The main scale for transformational leadership also showed a statistically significant difference with a medium effect size (p=<.001, d=.41). For details, see Table 3.

Table 3: Comparison of leaders' self- and staff ratings (n=84 leader-team pairs): main scales (in bold) and dimensions

	Practice owners		Practice staff		Paired t-test		
	M	**SD**	**M**	**SD**	**t(df)**	**p**	**d**
Transformational leadership	**3.9**	**0.5**	**3.6**	**0.6**	**3.721(82)**	**<.001**	**0.41**
Innovation	4.2	0.6	3.8	0.6	6.359(83)	<.001	0.69
Team spirit	4.1	0.7	3.8	0.7	3.462(82)	.001	0.38
Performance development	3.6	0.8	3.7	0.6	-0.208(83)	.836	-
Individuality focus	3.9	0.7	3.5	0.6	4.633(83)	<.001	0.50
Providing a vision	3.5	0.9	3.3	0.8	1.592(82)	.115	-
Being a role model	4.1	0.6	3.8	0.6	2.833(82)	.006	0.31
Transactional leadership	**3.4**	**0.6**	**3.3**	**0.5**	**1.291(81)**	**.200**	**-**
Goal setting	3.5	0.7	3.2	0.6	2.681(81)	.009	0.30
Management by exception	3.3	0.8	3.4	0.6	-0.470(82)	.640	-
Negative leadership	**1.5**	**0.4**	**1.6**	**0.4**	**-1.744(82)**	**.085**	**-**
Laissez-faire	1.6	0.6	1.7	0.5	-1.563(82)	.122	-
Destructive	1.4	0.5	1.6	0.5	-1.514(82)	.134	-
LMX-7	28.1	2.6	26.8	3.5	3.275(79)	.002	0.37

Transformational and transactional leadership by practice type

The Kruskal-Wallis test was applied to analyze for differences in leadership by practice types. It showed slight but non-significant differences in the raw values between practice types (single vs. group, teaching vs. non-teaching practices), e.g., slightly higher ratings for transformational leadership in single and non-teaching practices. These slight differences persisted when using reference T-values. For details, see Table 4.

Table 4: Comparison of leadership assessments by practice type: single / group practices; teaching / non-teaching practices

	Single (n=21)			Group (n=39)			Non-teaching (n=26)			Teaching (n=34)		
	M	**T**	**n**	**M**	**T**	**N**	**M**	**T**	**n**	**M**	**T**	**n**
Practice owners												
Transformational	4.0	47	21	3.9	45	63	3.8	43	37	4.0	47	47
Transactional	3.4	47	21	3.4	47	62	3.3	45	37	3.4	47	46
Negative	1.5	51	21	1.5	51	62	1.6	53	37	1.5	51	46
LMX-7	28.8	-	20	27.9	-	61	27.5	-	36	28.6	-	45
Practice staff												
Transformational	3.7	54	67	3.6	53	212	3.5	52	117	3.6	53	162
Transactional	3.4	53	67	3.3	51	212	3.3	51	117	3.2	50	162
Negative	1.7	47	70	1.7	47	212	1.7	47	117	1.6	46	165
LMX-7	27.5	-	61	26.3	-	190	25.6	-	105	27.3	-	146

Discussion

To our knowledge, this is the first study on GP practice leaders to assess transactional and transformational leadership using 180° feedback approach. Practice owners and practice personnel showed good relationships and a similar understanding of leadership behaviors regarding low levels of negative leadership and moderate levels of transactional leadership. Yet, potential for optimization was shown for transformational leadership, especially regarding the dimensions innovation, team spirit, and individuality focus.

The 180° feedback approach, also called multi-rater assessment, was shown to provide a more realistic picture of leader-team situations [20, 21]. In international studies, 360° or 180° feedback has been found to be a good predictor for promotions [36] and leadership effectiveness [37]. This type of feedback has the potential to identify differences in leadership expectations on behalf of the staff and the leader. Even though ratings are not low on either side, there is a discrepancy between self- and staff ratings, especially for transformational leadership as a more visionary approach. Such differences in the perceptions of leadership are influenced by mutual expectations. They are associated with job satisfaction, which is higher when the ratings are in agreement [38]. In our study, the gap in the

perception of leadership was larger for practice assistants than for employed physicians. This is likely related to the fact that employed physicians are much closer to their physician leader regarding training, roles, and duties compared to practice assistants. Also, practice assistants do not have the prospects to become physician leaders themselves, which implies a fundamentally different perspective. While such differences cannot be overcome, a better mutual understanding of leaders and staff can improve workplaces. Interestingly, staff members who perceive themselves as more similar to the leader also give better ratings [39]. Rowold and Poethke [26] describe how leaders can learn to adapt when receiving this feedback and recommend implementing changes, such as regular team meetings or improving leadership skills through training. Although multi-rater assessments were shown to provide a higher accuracy of ratings in the medical setting and are applied in various levels of medical training [22, 23], this approach has not been implemented routinely in physician leadership training.

Overall, the self- and staff ratings showed average or moderate levels when compared to a representative sample of German leaders provided by the manual of the FIF questionnaire [26]. Notably, these reference values are already corrected for aspects such as self-protective response and social desirability biases [40, 41]. These results are comparable to international studies on GP leadership, where GPs were aware of their leadership role in general but failed to describe explicit actions and perspectives for everyday practice situations [42]. Spehar et al. suggested that a lack of leadership training and credentials may play a role [43].

In line with a larger study in a German hospital setting [29], practice leaders' perception of their leadership skills was moderate. Yet, the differences between self- and staff ratings were smaller in our study. This might have been due to the sample selection, as leaders and employees in the current study were analyzed at the practice team level while only aggregate analyses of leaders' and staff's assessments were possible in the hospital setting. There were no differences in leaders' and staff's assessments of transactional and transformational leadership between practice types, although differences were observed when looking at mental health outcomes such as burnout [9].

Applying the LMX-7 questionnaire, prior studies on relationship quality in the health sector showed associations with enhanced commitment, reduced staff turnover, and better organizational behavior [44]. Also, positive effects on employees' health and well-being were outlined [45, 46]. Our results show a good relationship quality between practice owners and practice assistants when applying reference values [34].

Strengths and limitations

Our study presents an innovative implementation of the 180° feedback approach in the GP setting. The relatively large sample size and the analysis at practice team levels are noteworthy. The results for group practices with two or more leaders are limited as staff were asked to rate practice leaders as a group, not individually. On the other hand, this approach reflects leadership in group practices typically executed by a leadership team. Given the data's cross-sectional nature, cause-effect analyses are not possible. The negative leadership scale showed a low reliability but was included according to the respective manual [26]. For the small number of employed physicians, complete data on the Leader-Member Exchange questionnaire were missing, yet the analysis of the available six rather than seven questions yielded a high relationship quality similar to that documented for practice assistants.

The stratified randomization, which took into account teaching and non-teaching practices as well as solo and group practices, reflects a typical spectrum of German general practices. However, a selection bias cannot be excluded as participating practice leaders might have had a greater interest in the topic.

Conclusion and practical implications

Compared to the reference populations in the literature, GPs and their practice teams have a good relationship quality. However, transactional and transformational leadership skills show potential for improvement, especially regarding the dimensions goal-setting, innovation, and individuality focus. Recent developments call for leadership workshops at every level of medical training [47] in order to help GP leaders better understand staff's needs. Future results from the IMPROVE*job* study will show if the intervention changed leadership skills and their perception.

Formalities

Acknowledgment

All authors are part of the IMPROVE*job* consortium and owe special thanks to all participating physicians and practice assistants. In addition, special thanks go to the other members of the IMPROVE*job* consortium; current members: M. Brinkmann, S. Burgess, A. Herrmann-Werner, K.-H. Jöckel, F. Junne, A. Siegel, F. Stuber, A. Wagner, and former members: J. Block, J.-M. Bois, A. Dreher, M. Hippler, S. Hartmann, L. Imhoff, C. Kersting, C. Ose and E. Tsarouha. Additionally, we are grateful to all members of the scientific advisory board: Stefanie Esper, Katja Götz, Andre Habrock, Albert Nienhaus, Horst Christian Vollmar, and Andrea Wittich. The work of the Institute of Occupational and Social Medicine and Health Services Research is supported by an unrestricted grant of the Employers' Association of the Metal and Electrical Industry Baden-Wuerttemberg (Suedwestmetall). We acknowledge administrative support by the DLR Project Management Agency.

Contributorship statement

Conceptualization, B.M.W., M.A.R., T.S.-D., C.P., and B.W.; methodology, B.M.W. M.A.R., T.S.-D., M.S., B.W., M.G., E.R., C.P., K.-H.J., L.D., and J.G.; validation, K.-H.J.; formal analysis, M.S. and J.G.; investigation, L.D., K.L., T.S.-D., B.W., E.R., C.P., J.G., M.A.R., and B.M.W.; resources, B.M.W., L.D., K.L., T.S.-D., B.W., M.G., C.P., and E.R..; data curation, K.-H.J.; writing—original draft preparation, M.S. and B.M.W.; writing—review and editing, M.S., T.S.-D., B.W., M.G., E.R., C.P., K.-H.J., L.D., J.G., M.A.R. and B.M.W.; visualization, M.S.; supervision, B.M.W., M.A.R., and B.W.; project administration, K.L., E.R., B.M.W. and M.A.R.; funding acquisition, B.M.W., M.A.R., C.P., and B.W. All authors have read and approved the published version of the manuscript.

Competing interests

The authors declare that they have no competing interests.

Funding

This study is funded by the German Federal Ministry for Education and Research (grant numbers: 01GL1851D, 01GL1751B, 01GL1751A, 01GL1751C). The funder had no role in the study design, the data collection, and analysis, the decision to publish, or the preparation of the manuscript.

Data sharing statement

There are no plans to grant access to the full protocol, participant-level dataset, or statistical code as data contain potentially identifying information.

Ethics approval

The study complies with the ethical principles of the World Medical Association Declaration of Helsinki. The study was approved first by the Ethics Committee of the Medical Faculty of the University of Bonn (reference number: 057/19, date of approval: 20/02/2019). In addition, the Ethics Committees of the Medical Association Nordrhein (Lfd-Nr.: 2019107), and of the Medical Faculty, University Hospital of Tuebingen (project no.: 446/2019BO2) approved the study protocol.

All participating practice team members received written information and signed informed consent forms, which are stored at the Institute for General Practice and Family Medicine, University of Bonn.

References

1 Boamah SA, Spence Laschinger HK, Wong C, et al. Effect of transformational leadership on job satisfaction and patient safety outcomes. *Nurs Outlook* 2018;66(2):180–89. doi:10.1016/j.outlook.2017.10.004 [published Online First: 23 November 2017].

2 Goetz K, Campbell S, Broge B, et al. Job satisfaction of practice assistants in general practice in Germany: an observational study. *Fam Pract* 2013;30(4):411–17. doi:10.1093/fampra/cmt015 [published Online First: 19 April 2013].

3 Kersting C, Zimmer L, Thielmann A, et al. Chronic stress, work-related daily challenges and medicolegal investigations: a cross-sectional study among German general practitioners. *BMC Fam Pract* 2019:143. doi:10.1186/s12875-019-1032-6 [published Online First: 24 October 2019].

4 Wallace JE, Lemaire JB, Ghali WA. Physician wellness: a missing quality indicator. *The Lancet* 2009;374(9702):1714–21.

5 Wong CA, Cummings GG, Ducharme L. The relationship between nursing leadership and patient outcomes: a systematic review update. *J Nurs Manag* 2013;21(5):709–24.

6 Saravo B, Netzel J, Kiesewetter J. The need for strong clinical leaders - Transformational and transactional leadership as a framework for resident leadership training. *PLoS One* 2017;12(8):e0183019.

7 Richter A, Lornudd C, Thiele Schwarz U von, et al. Evaluation of iLead, a generic implementation leadership intervention: mixed-method preintervention-postintervention design. *BMJ Open* 2020;10(1):e033227. doi:10.1136/bmjopen-2019-033227 [published Online First: 12 January 2020].

8 Viehmann A, Kersting C, Thielmann A, et al. Prevalence of chronic stress in general practitioners and practice assistants: Personal, practice and regional characteristics. *PLoS One* 2017;12(5):e0176658.

9 Dreher A, Theune M, Kersting C, et al. Prevalence of burnout among German general practitioners: Comparison of physicians working in solo and group practices. *PLoS One* 2019;14(2):e0211223.

10 Abrell C, Rowold J, Weibler J, et al. Evaluation of a long-term transformational leadership development program. *German Journal of Human Resource Management* 2011;25(3):205–24.

11 Kelloway EK, Barling J, Helleur J. Enhancing transformational leadership: The roles of training and feedback. *Leadership & Organization Development Journal* 2000;21(3):145–49.

12 Bass BM. Two Decades of Research and Development in Transformational Leadership. *European Journal of Work and Organizational Psychology* 1999;8(1):9–32.

13 Poethke U. Leadership, values and communication: a cross-cultural investigation of the extended full-range of leadership behaviors: Doctoral dissertation 2019.

14 Rowold J, Schlotz W. Transformational and transactional leadership and followers' chronic stress. *Leadership Review* 2009;9(1):35–48.

15 Judge TA, Piccolo RF. Transformational and transactional leadership: a meta-analytic test of their relative validity. *J Appl Psychol* 2004;89(5):755–68.

16 Gregersen S, Vincent-Höper S, Nienhaus A. Health-Relevant Leadership Behaviour: A Comparison of Leadership Constructs. *German Journal of Human Resource Management* 2014;28(1-2):117–38.

17 Kelloway EK, Barling J. Leadership development as an intervention in occupational health psychology. *Work & Stress* 2010;24(3):260–79.

18 Schyns B, Schilling J. How bad are the effects of bad leaders? A meta-analysis of destructive leadership and its outcomes. *The Leadership Quarterly* 2013;24(1):138–58.

19 Graen GB, Uhl-Bien M. Relationship-based approach to leadership: Development of leader-member exchange (LMX) theory of leadership over 25 years: Applying a multi-level multi-domain perspective. *Leadership Quarterly* 1995;6(2):219–47.

20 Bracken D, Timmreck C, Church A. The Handbook of Multisource Feedback The Comprehensive Resource for Designing and Implementing MSF Processes 2001.

21 Lockyer J. Multisource feedback in the assessment of physician competencies. *J Contin Educ Health Prof* 2003;23(1):4–12.

22 Davis DA, Mazmanian PE, Fordis M, et al. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA* 2006;296(9):1094–102.

23 Donnon T, Al Ansari A, Al Alawi S, et al. The reliability, validity, and feasibility of multisource feedback physician assessment: a systematic review. *Acad Med* 2014;89(3):511–16.

24 Weltermann BM, Kersting C, Pieper C, et al. IMPROVEjob - Participatory intervention to improve job satisfaction of general practice teams: a model for structural and behavioural prevention in small and medium-sized enterprises - a study protocol of a cluster-randomised controlled trial. *Trials* 2020;21(1):532.

25 Degen L, Linden K, Seifried-Dübon T, et al. Job Satisfaction and Chronic Stress of General Practitioners and Their Teams: Baseline Data of a Cluster-Randomised Trial (IMPROVEjob). *Int J Environ Res Public Health* 2021;18(18). doi:10.3390/ijerph18189458 [published Online First: 8 September 2021].

26 Rowold J, Poethke U. FIF: Fragebogen zur Integrativen Führung Manual, 1st edn. Göttingen: Hogrefe 2017.

27 Ingold PV. Fragebogen zur integrativen Führung. *Zeitschrift für Arbeits- und Organisationspsychologie A&O* 2019;63(2):100–06.

28 Lange S, Bormann KC, Rowold J. Mindful leadership: mindfulness as a new antecedent of destructive and transformational leadership behavior. *Gr Interakt Org* 2018;49(2):139–47.

29 Stuber F, Seifried-Dübon T, Rieger MA, et al. Investigating the Role of Stress-Preventive Leadership in the Workplace Hospital: The Cross-Sectional Determination of Relational Quality by Transformational Leadership. *Front Psychiatry* 2019;10:622. doi:10.3389/fpsyt.2019.00622 [published Online First: 3 September 2019].

30 Schyns B. Überprüfung einer deutschsprachigen Skala zum Leader-Member-Exchange-Ansatz. *Zeitschrift für Differentielle und Diagnostische Psychologie* 2002.

31 Schyns B, Paul T. Leader-Member Exchange Skala (LMX7): ZIS - GESIS Leibniz Institute for the Social Sciences 2014.

32 Sargeant J, Mann K, Ferrier S. Exploring family physicians' reactions to multisource feedback: perceptions of credibility and usefulness. *Med Educ* 2005;39(5):497–504.

33 Hu J, Lee R, Mullin S, et al. How physicians change: Multisource feedback driven intervention improves physician leadership and teamwork. *Surgery* 2020;168(4):714–23. doi:10.1016/j.surg.2020.06.008 [published Online First: 6 August 2020].

34 Northouse PG. Leadership: Theory and practice, 6th edn. Thousand Oaks: Sage Publications 2010.

35 Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd edn. Hoboken: Taylor and Francis 1988.

36 Halverson SK, Tonidandel S, Barlow C, et al. Self-other agreement on a 360-degree leadership evaluation. In: 17th Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, Canada 2002.

37 Furnham A, Stringfield P. Congruence of self and subordinate ratings of managerial practices as a correlate of supervisor evaluation. *Journal of Occupational and Organizational Psychology* 1994;67(1):57–67.

38 Erben J, Schneider FM, Maier M. In the Ear of the Beholder: Self-Other Agreement in Leadership Communication and Its Relationship With Subordinates' Job Satisfaction. *International Journal of Business Communication* 2019;56(4):505–29.

39 Fleenor JW, Smither JW, Atwater LE, et al. Self–other rating agreement in leadership: A review. *The Leadership Quarterly* 2010;21(6):1005–34.

40 Sarros JC, Cooper BK, Hartican AM. Leadership and character. *Leadership & Organization Development Journal* 2006;27(8):682–99.

41 Thornton GC. Psychometric properties of self-appraisals of job performance. *Personnel Psychology* 1980;33(2):263–71.

42 Hana J, Rudebeck CE. Leadership in rural medicine: the organization on thin ice? *Scandinavian Journal of Primary Health Care* 2011;29(2):122–28. doi:10.3109/02813432.2011.577148 [published Online First: 28 April 2011].

43 Spehar I, Sjøvik H, Karevold KI, et al. General practitioners' views on leadership roles and challenges in primary health care: a qualitative study. *Scandinavian Journal of Primary Health Care* 2017;35(1):105–10. doi:10.1080/02813432.2017.1288819 [published Online First: 1 March 2017].

44 Chen C-HV, Wang S-J, Chang W-C, et al. The effect of leader-member exchange, trust, supervisor support on organizational citizenship behavior in nurses. *Journal of Nursing Research* 2008;16(4):321–28.

45 Gregersen S, Vincent-Höper S, Nienhaus A. Job-related resources, leader–member exchange and well-being – a longitudinal study. *Work & Stress* 2016;30(4):356–73.

46 Vincent-Höper S, Stein M, Gregersen S, et al. Messung gesundheitsförderlicher Führung in Gesundheitsdienst und Wohlfahrtspflege. *ASU Arbeitsmed Sozialmed Umweltmed* 2018(53):46–53.

47 Frank JR, Snell L, Sherbino J. CanMEDS 2015 Physician Competency Framework. Ottawa: Royal College of Physicians and Surgeons of Canada 2015.

Additional Files

Additional file 1: CONSORT 2010 checklist of information to include when reporting a randomised trial

CONSORT 2010 checklist of information to include when reporting a randomised trial*

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	n/a
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	Abstract, 2
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	Background, 4-5
	2b	Specific objectives or hypotheses	Background, 4-5
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	Methods, 5
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	n/a
Participants	4a	Eligibility criteria for participants	Methods, 5-6
	4b	Settings and locations where the data were collected	Methods, 5-6
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	n/a
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	Methods, 6-7
	6b	Any changes to trial outcomes after the trial commenced, with reasons	n/a
Sample size	7a	How sample size was determined	Study protocol, Weltermann et al., Trials 2020
	7b	When applicable, explanation of any interim analyses and stopping guidelines	n/a

Randomisation:			
Sequence generation	8a	Method used to generate the random allocation sequence	Details in Study protocol, Weltermann et al., Trials 2020
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	Details in Study protocol, Weltermann et al., Trials 2020
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	n/a
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	Details in Study protocol, Weltermann et al., Trials 2020
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	n/a
	11b	If relevant, description of the similarity of interventions	n/a
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	Methods,7-8
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	n/a
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	n/a
	13b	For each group, losses and exclusions after randomisation, together with reasons	n/a
Recruitment	14a	Dates defining the periods of recruitment and follow-up	Details in Study protocol, Weltermann et al., Trials 2020
	14b	Why the trial ended or was stopped	Details in Study protocol,

			Weltermann et al., Trials 2020
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	Results, table 1, Page 8
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	Results, 9
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	Results, 9-11
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	N/A.
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	n/a
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	n/a
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	Discusssion, Strengths and limitatoins, Page 12-13
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	13
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	Discussion, 11-13
Other information			
Registration	23	Registration number and name of trial registry	Trial registration, 2
Protocol	24	Where the full trial protocol can be accessed, if available	Study protocol, Weltermann et al., Trials 2020
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	Funding, 17-18

*We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org.

BMJ Open

A 180° view on general practitioners' leadership skills: practice-level comparisons of leader and staff assessments using data from the IMPROVEjob study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2022-066298.R1
Article Type:	Original research
Date Submitted by the Author:	20-Dec-2022
Complete List of Authors:	Schmidt, Manuela; University Hospital Bonn, Institute of General Practice and Family Medicine Seifried-Dübon, Tanja; University Hospital Tuebingen Osianderstraße 5, Department of Psychosomatic Medicine and Psychotherapy Göbel, Julian; University Hospital Bonn, Institute of General Practice and Family Medicine Degen, Lukas; University Hospital Bonn, Institute of General Practice and Family Medicine Werners, Brigitte; Ruhr University Bochum Grot, Matthias; Ruhr University Bochum Rind, Esther; Institute for Occupational Medicine, Social Medicine and Health Services Research , University Hospital Tuebingen Pieper, Claudia; University of Duisburg-Essen Jöckel, Karl-Heinz; Institute of Medical Informatics, Biometry and Epidemiology, University Duisburg- Essen Linden, Karen; University Hospital Bonn, Institute of General Practice and Family Medicine Rieger, Monika A.; Institute for Occupational Medicine, Social Medicine and Health Services Research , University Hospital Tuebingen Weltermann, Birgitta; University Hospital Bonn, Institute of General Practice and Family Medicine
Primary Subject Heading:	General practice / Family practice
Secondary Subject Heading:	General practice / Family practice
Keywords:	PRIMARY CARE, Human resource management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

A 180° view on general practitioners' leadership skills: practice-level comparisons of leader and staff assessments using data from the IMPROVE*job* study

Authors: Manuela Schmidt[1]*, Tanja Seifried-Dübon[2], Julian Göbel[1], Lukas Degen[1], Brigitte Werners[3], Matthias Grot[3], Esther Rind[4], Claudia Pieper[5], Karl-Heinz Jöckel[5 6], Karen Linden[1], Monika A. Rieger[4], Birgitta M. Weltermann[1] on behalf of the IMPROVE*job* consortium

[1] Institute of General Practice and Family Medicine, University Hospital Bonn, University of Bonn, Venusberg-Campus 1, 53127 Bonn, Germany

[2] Department of Psychosomatic Medicine and Psychotherapy, University Hospital Tuebingen, Osianderstraße 5, 72076 Tuebingen, Germany

[3] Institute of Management, Operations Research, Ruhr University Bochum, Universitätsstr. 150, 44801 Bochum, Germany

[4] Institute of Occupational and Social Medicine and Health Services Research, University Hospital Tuebingen, Wilhelmstr. 27, 72074 Tuebingen, Germany

[5] Institute for Medical Informatics, Biometry and Epidemiology, University Hospital Essen, University of Duisburg-Essen, Hufelandstr. 55, 45147 Essen, Germany

[6] Center for Clinical Trials, University Hospital Essen, University of Duisburg-Essen, Hufelandstr. 55, 45147 Essen, Germany.

*Correspondence to: julian.goebel@ukbonn.de

Abstract

Objectives: Strong primary care leaders are needed to assure high quality services for patient populations. This study analyzed general practitioners' (GP) leadership skills comparing practice-level self- and staff assessments based on the Full Range of Leadership Model (FRLM) and the Leader-Member Exchange (LMX).

Setting: The questionnaire survey was conducted among German general practice leaders and their staff participating in the IMPROVE*job* trial.

Participants: The study population comprised 60 German general practices with 366 participants: 84 GP practice leaders and 282 employees (28 physicians and 254 practice assistants).

Primary and secondary outcome measures: Leadership skills of the practice leaders were measured using the Integrative Leadership Questionnaire (German *FIF*) and the Leader-Member Exchange (LMX-7) questionnaire. Leaders rated themselves and practice staff rated their leaders. The data was analyzed by paired mean comparisons on the practice level.

Results: For most leadership dimensions, practice leaders rated themselves higher than their employees rated them. Differences were found for transformational leadership ($p<.001$, $d=.41$), especially for the dimensions 'innovation' ($p<.001$, $d=.69$) and 'individuality focus' ($p<.001$, $d=.50$). For transactional leadership, the dimension 'goal setting' differed significantly ($p<.01$, $d=.30$) but not the other dimensions. Scores for negative leadership were low and showed no differences between leaders and employees. Interestingly, employed physicians' rated their practice leaders higher on the two transformational ('performance development', 'providing a vision') and all transactional dimensions. The LMX-7 scale showed high quality relationships between leaders and employees.

Conclusions: This 180° analysis of GPs´ leadership skills with self- and employee ratings indicated good relationships. There is a potential to improve leadership regarding goal-setting, innovation and focusing on individual team members. These results allow for the development of targeted interventions.

Trial registration: German Clinical Trials Register, DRKS00012677. Registered 16 October 2019.

Keywords: leadership, leadership quality, general practitioner, practice staff, 180° feedback

Strenghts & limitations:

- Leadership skills are crucial to secure a strong primary care workforce for patient care.
- High quality leadership can protect and promote the mental health of employees.
- This presentation of an innovative 180° feedback approach in the GP setting allows for a better understanding of leadership as viewed by the different professional groups working in practices.
- The data reflects a typical spectrum of German general practices including teaching and non-teaching practices as well as solo and group practices.
- In group practices, leadership teams, not individual leaders were rated by staff to capture leadership comprehensively on practice level.

Background:

Strong primary care leaders and a strong primary care workforce are important to assure the health of populations and primary care teams [1–3]. A recent systematic review of 20 studies by Meredith et al. showed an association between stronger leadership and less burnout among different medical professionals in the United States [4]. In contrast, poor leadership skills have a negative impact on job satisfaction [5–7], staff well-being [8] and the quality of patient care [5, 9, 10]. A review showed correlations between better leadership and various quality of care indicators, e.g. pain, safety and 30-day-mortality [11]. In addition to individual outcomes, leadership is important to promote organizational changes (e.g., the implementation of IT-supported care) [12].

Scientifically, leadership is conceptualized in several theories. One of the most studied leadership frameworks is the Full Range of Leadership Model (FRLM), which integrates transactional, transformational, and negative leadership [13, 14]. Transactional leadership describes leaders' structuring of work situations, the exchange of contingent rewards (e.g. work against salary), and the management by exception [13–15]. In contrast, transformational leadership moves beyond leaders' and staff's self-interests. It focuses on the staff's attitudes and values regarding overarching goals such as self-actualization, organizational achievements, and the well-being of others and society as a whole [13, 14]. Building on the FRLM, a recent further development, the so-called Implementation Leadership Scale, focusses on the role of leadership for implementation of organizational changes [12]. Another important leadership theory, the Leader-Member Exchange (LMX), specifically addresses the relationship between leaders and staff. It concentrates on the perceived quality of the dyadic relationship between a staff member and the immediate leader [13, 16]. The relationship reflects a dyadic social exchange process ranging from low LMX, described by limited social transactions with more transactional leadership to high LMX, which represents a transformational approach with a high degree of social exchange and a mature leader-member partnership [17]. High-quality relationships positively influence employees' work-related well-being and are associated with higher job satisfaction of health care workers [18].

Based on these theories, various questionnaires were developed, e.g. the Leader-Member Exchange questionnaire 'LMX-7' [16] and the German questionnaire 'Fragebogen für integrative Führung' (FIF; in English: Questionnaire for Integrative Leadership) [15, 19]. These instruments allow for a multi-rater perspective: the leader's and the staff's views on the leader's behavior are measured and compared providing 180° feedback. This method is valuable because assessments from different perspectives create a more comprehensive picture of the leaders' actual skills and performances [20]. Two recent reviews of 60 studies from various medical settings showed that such approaches are increasingly

applied in medical education and graduate training [21, 22], but have not been used to evaluate GP leaders and their teams. Effective interventions to improve leadership were developed and evaluated in the hospital [23] and healthcare management setting [24]. For example, Saravo et al. showed an improvement in transformational and transactional leadership performance of 57 medical residents in hospital rotations after a 4-week intervention [23]. In addition, a 2018 study from Hill et al. highlighted positive effects of a leadership training for surgical residents on teamwork and team involvement in decision-making [25]. However, such interventions have not been implemented in German primary care, although high chronic stress and burnout rates are reported for this workforce [26, 27]. The need is even larger as about half of the German GPs who mainly work in GP-owned private practices [28], will reach retirement age in the next ten years [29]. Based on the leadership frameworks mentioned above, the publicly funded IMPROVE*job* study aimed to improve the job satisfaction of physician leaders and practice personnel of German GP practices focusing on leadership, communication, and work processes [30, 31]. At baseline, GPs´ leadership skills were evaluated comparing GP leaders' self and staff ratings on practice level.

Methods

This analysis draws on the baseline data of the IMPROVE*job* study, which is designed as a cluster-randomized controlled trial (cRCT) to improve job satisfaction among practice personnel. The details are described in the study protocol [30].

In short, a total of 60 GP practices in the North Rhine region in Germany were recruited by the Institute of General Practice and Family Medicine of the University of Bonn. The sample comprised single (owned by one practice leader) and group practices (owned by more than one practice leader), some of which were also involved as teaching practices (affiliated to a university). The study aimed to recruit practice teams, including physician leaders, employed physicians, and practice assistants. A total of 84 GP practice leaders, 28 employed physicians and 254 practice assistants were recruited. In Germany, primary care is typically provided by GP-owned practices with 1 to 3 physicians. For each physician, practices employ about 1 to 2 certified practice assistants who finished a vocational training of 3 years. Similar to other regions worldwide, the size of group practices is increasing.

Patient and public involvement

The study targeted general practice staff. Therefore, GPs and practice assistants were involved in all phases of the study. As the study did not target patients, no patients or members of the public were involved in the design or conduct of the study.

Ethics

The study was approved first by the Ethics Committee of the Medical Faculty of the University of Bonn (reference number: 057/19, date of approval: 20/02/2019). In addition, the Ethics Committees of the Medical Association North-Rhine (Lfd-Nr.: 2019107) and of the Medical Faculty, University Hospital of Tuebingen (project no.: 446/2019BO2) approved the study protocol. The study was performed in accordance with the Declaration of Helsinki. All participating practice team members received a study information and signed informed consent forms.

Measures

Practice leaders answered a short questionnaire on practice characteristics and the questionnaire for practice leaders. Employed physicians and practice assistants completed different versions of the same employee questionnaire. Details of the methods and the characteristics of the study population are published [30, 31].

All participants provided sociodemographic, professional, and work-related characteristics which are published [31]. In addition, GP leaders and practice staff filled the following two leadership questionnaires:

1. Integrative Leadership Questionnaire (FIF)

Transformational, transactional, and negative leadership were measured using the FIF questionnaire. Its scales' validity and internal consistency are confirmed for different populations [19, 32]. The FIF has been used in non-medical and hospital settings [33], but not in primary care.

All 40 items of the FIF are answered on a 5-point Likert scale and are worded to reflect either the leader's or the staff's position.

The measures comprise:

- the transformational leadership scale consisting of six dimensions: innovation, team spirit, performance development, individuality focus, providing a vision, and being a role model;
- the transactional leadership scale with two dimensions: goal setting and management by exception;
- the negative leadership scale with two dimensions: laissez-faire and destructive leadership.

2. Leader-Member Exchange (LMX-7)

The relationship quality between leaders and staff is measured using the Leader-Member Exchange questionnaire (LMX-7) with seven items on a 5-point Likert scale, which are worded to reflect the leader or the staff position [16, 34, 35].

The multi-rater, 180° approach is applied to the two leadership scales. Results of such assessments are usually shared with the ratee, yet previous studies showed mixed reactions in the medical setting [36, 37]. Therefore, the results of the 180° feedback in our study were not shared with the participating practices but are used on an aggregated level for research purposes only.

Statistical analysis

Statistical analyses were conducted using SPSS Statistics 27 (IBM Cooperation, Armonk, Ny, USA, 2020). All analyses were carried out at the participant and the practice level.

The FIF data were analyzed according to the official manual [19]. Mean scores for transformational, transactional, and negative leadership were summarized both for the respective main scale and all dimensions: for transactional and transformational leadership, they ranged from 1 (worst rating) to 5 (best rating); for negative leadership, they ranged from 1 (best rating) to 5 (worst rating). For comparison, scores were standardized using T-scaling tables from reference populations as defined by Rowold & Poethke [19]. These T-values are based on a normal distribution around 50 (SD 10). Thus, values above 70 only reflect about 2% of the reference population from German-speaking countries [19].

The LMX-7 was analyzed per standard protocol by creating a sum score of all seven items without transformation [35]. The LMX-7 score can range from 7 to 35 with five standard categories which were interpreted as follows: score 7 to 14 = very low; 15 to 19 = low; 20 to 24 = moderate; 25 to 29 = high, 30 to 35 = very high [38]. Inadvertently, question seven was missing on all employed physicians' questionnaires, which reduced the answered questions to six. As the LMX-7 manual does not suggest a standard approach for missing values, we excluded employed physicians from the further analyses.

For the 180° feedback approach on practice level, the combined mean scores of employed physicians and practice assistants per practice were compared to the self-assessment of their respective leaders using paired t-tests, as the data satisfied the condition of a normal distribution with the Kolmogorov-Smirnov test. Cohen's d was applied to determine the effect size of mean comparisons with the following standard interpretations: small effects from d=.2, medium from d=.5, and high from d=.8 [39]. In single practices, the staff ratings were compared to the leader's assessment. In practices with more than one owner (group practices), each leader's self-rating was compared with the respective ratings of the practice personnel, who were asked to rate the leadership team of the practice, not stratified by individual leaders. This approach was chosen because practice owners of German practices typically work as a leadership team. In addition, the ratings of the transformational and transactional leadership scales were compared stratified by practice type (single vs. group and teaching vs. non-teaching practices) using the Kruskal-Wallis tests because the data for practice comparisons did not satisfy the conditions for parametric tests.

Results

Population

The baseline data of the IMPROVE*job* study included 366 participants from 60 practices, consisting of 84 practice leaders, 28 employed physicians, and 254 practice assistants. The mean age of the participants was 44.4 years, with a mean of 54.3 years for practice leaders, 44.8 for employed physicians, and 41.0 for practice assistants. Among the practice assistants, 99.6% were female, as were 76.6% of the employed physicians and half of the practice leaders (52.4%). Most practice leaders worked full-time (90.5%), as did about a quarter of the employed physicians (28.6%) and 41.5% of the practice assistants (see Table 1). The details on the sociodemographic descriptions are published [31].

Table 1: Sociodemographic description of participants at baseline [31]

Variable	Total sample N=366	Practice leaders N=84	Employed physicians N=28	Practice assistants N=254
Female, %	87.1	52.4	78.6	99.6
Age in years, mean (SD)	44.4 (12.8)	54.3 (6.2)	44.8 (9.8)	41.0 (13.0)
Years in current practice, mean (SD)	10.0 (9.1)	15.3 (8.4)	3.9 (5.4)	8.8 (8.9)
Working full-time, %	52.0	90.5	28.6	41.5

On average, practice leaders had been accredited for 26.6 years and licensed for the statutory health insurance for 16.4 years. Seven (25%) of the employed physicians were in GP training. Practice assistants had graduated on average 19.9 years ago, while 7.5% were still in training. Of the 60 practices, 21 (35%) were single and 39 (65%) were group practices; of these, 34 were teaching (57%) and 26 (43%) were non-teaching practices. On average, practices were in the same location for 20.4 years (SD 14 years).

Leadership

The transactional and transformational leadership scales showed a high internal consistency with Cronbach's α = .74 to .93 for the staff members' assessment and Cronbach's α = .72 to .87 for the leaders' assessment. For negative leadership, the scales showed a sufficient internal consistency for staff members' (Cronbach's α = .73 to .80) but not for leaders' assessments (Cronbach's α = .47 to .68). The mean results of the FIF were within the one standard deviation range of the reference population [19]. Based on raw values, employed physicians rated their leaders consistently better than practice assistants and better than the leaders themselves for some items. While practice assistants rated their leaders more poorly than the practice leaders in raw values, reference T-values showed only minor differences. The details are outlined in Table 2.

The LMX-7 scale showed an internal consistency of Cronbach's α = .88 for staff members (practice assistants) and α = .71 for leaders. Both groups showed a high relationship quality, scoring 28 for

practice leaders and 26 for practice assistants. As the seventh question was missing for employed physicians, they were excluded from the analysis. However, the sum score of the remaining six questions also showed a high score of 24.9 out of 30. The details are shown in Table 2.

Table 2: Leadership assessment by employment group: main scales (in bold) and dimensions (LMX values can range from 7 to 35, FIF scales from 1 to 5)

	Practice leaders (N=84)				Employed physicians (N=28)				Practice assistants (N=254)			
	M	**SD**	**T***	**n**	**M**	**SD**	**T***	**N**	**M**	**SD**	**T***	**n**
Transformational leadership	**3.9**	**0.6**	**45**	**84**	**3.9**	**0.7**	**56**	**27**	**3.5**	**0.8**	**52**	**237**
Innovation	4.2	0.6	49	84	4.0	1.0	55	28	3.7	0.9	52	247
Team spirit	4.1	0.7	49	84	3.8	1.0	54	28	3.6	1.1	52	251
Performance development	3.6	0.8	44	84	4.1	0.7	57	27	3.5	1.0	51	247
Individuality focus	3.9	0.7	47	84	3.7	1.0	54	28	3.5	1.1	53	249
Providing a vision	3.5	0.9	45	84	3.6	0.9	55	28	3.2	1.1	51	245
Being a role model	4.1	0.6	45	84	4.0	0.8	55	27	3.7	1.0	52	246
Transactional leadership	**3.4**	**0.7**	**47**	**83**	**3.5**	**0.7**	**54**	**27**	**3.2**	**0.8**	**50**	**244**
Goal setting	3.5	0.7	44	83	3.7	0.9	56	27	3.1	1.0	50	246
Management by exception	3.3	0.8	51	83	3.4	0.8	52	27	3.3	0.9	51	245
Negative leadership	**1.5**	**0.5**	**51**	**83**	**1.5**	**0.6**	**45**	**28**	**1.7**	**0.7**	**47**	**248**
Laissez-faire	1.6	0.6	52	83	1.6	0.8	45	28	1.7	0.8	46	249
Destructive	1.4	0.5	51	83	1.4	0.6	46	28	1.6	0.7	48	248
LMX-7	**28.1**	**2.6**	**-**	**81**	**n/a**	**n/a**	**-**	**n/a**	**26.7**	**4.8**	**-**	**222**

*Reference T-values range from 0 to 100, as defined by Rowold & Poethke 2017

180° leadership feedback

Practice leaders self-rated their leadership skills slightly better than their staff for all dimensions except for 'management by exception'. There were no statistically significant differences for negative leadership. For transactional leadership, goal setting differed significantly with a low effect size (p=.009, d=.30). Leaders' scores on transformational leadership were significantly higher than the scores of the teams, with the dimension for innovation reaching the strongest effect size (p=<.001, d=.69), followed by individuality focus with a medium effect size (p=<0.001, d=.50). The scores for team spirit and being a role model were slightly lower, but significant. The main scale for transformational leadership also showed a significant difference with a medium effect size (p=<.001, d=.41). The details are outlined in Table 3.

Table 3: Comparison of leaders' self- and staff ratings (n=84 leader-team pairs): main scales (in bold) and dimensions

	Practice leaders		Practice staff		Paired t-test		
	M	SD	M	SD	t(df)	p	d
Transformational leadership	**3.9**	**0.5**	**3.6**	**0.6**	**3.721(82)**	**<.001**	**0.41**
Innovation	4.2	0.6	3.8	0.6	6.359(83)	<.001	0.69
Team spirit	4.1	0.7	3.8	0.7	3.462(82)	.001	0.38
Performance development	3.6	0.8	3.7	0.6	-0.208(83)	.836	-
Individuality focus	3.9	0.7	3.5	0.6	4.633(83)	<.001	0.50
Providing a vision	3.5	0.9	3.3	0.8	1.592(82)	.115	-
Being a role model	4.1	0.6	3.8	0.6	2.833(82)	.006	0.31
Transactional leadership	**3.4**	**0.6**	**3.3**	**0.5**	**1.291(81)**	**.200**	**-**
Goal setting	3.5	0.7	3.2	0.6	2.681(81)	.009	0.30
Management by exception	3.3	0.8	3.4	0.6	-0.470(82)	.640	-
Negative leadership	**1.5**	**0.4**	**1.6**	**0.4**	**-1.744(82)**	**.085**	**-**
Laissez-faire	1.6	0.6	1.7	0.5	-1.563(82)	.122	-
Destructive	1.4	0.5	1.6	0.5	-1.514(82)	.134	-
LMX-7	28.1	2.6	26.8	3.5	3.275(79)	.002	0.37

Transformational and transactional leadership by practice type

The Kruskal-Wallis test was applied to analyze for differences in leadership by practice types. It showed slight but non-significant differences in the raw values between practice types (single vs. group, teaching vs. non-teaching practices), e.g., slightly higher ratings for transformational leadership in single and non-teaching practices. These slight differences persisted when using reference T-values. For details, see Table 4.

Table 4: Comparison of leadership assessments by practice type: single versus group practices and teaching versus non-teaching practices

	Single (n=21)			Group (n=39)			Non-teaching (n=26)			Teaching (n=34)		
	M	T	n	M	T	N	M	T	n	M	T	n
Practice leaders												
Transformational	4.0	47	21	3.9	45	63	3.8	43	37	4.0	47	47
Transactional	3.4	47	21	3.4	47	62	3.3	45	37	3.4	47	46
Negative	1.5	51	21	1.5	51	62	1.6	53	37	1.5	51	46
LMX-7	28.8	-	20	27.9	-	61	27.5	-	36	28.6	-	45
Practice staff												
Transformational	3.7	54	67	3.6	53	212	3.5	52	117	3.6	53	162
Transactional	3.4	53	67	3.3	51	212	3.3	51	117	3.2	50	162
Negative	1.7	47	70	1.7	47	212	1.7	47	117	1.6	46	165
LMX-7	27.5	-	61	26.3	-	190	25.6	-	105	27.3	-	146

Discussion

Using a 180° feedback approach of leadership in GP practices, this study showed good relationships between leaders and staff with low levels of negative leaderships. Practice staff rated their leaders slightly higher on all transformational and transactional dimensions than the 234 German leaders and 713 employees from the FIF questionnaire reference population [19]. Also, agreement between GP leaders and staff was higher than in a study of 1,137 German hospital employees (315 leaders, 822 staff members) from different occupational groups (e.g., physicians, nurses, administration, information technology), which used the same methodology [33]. Interestingly, hospital and GP leaders rated themselves approximately similar [33].

The benefit of 180° and 360° feedback is shown in studies from various settings. In a sample of more than 2,000 U.S. military leaders, 360° feedback (leaders, subordinates, peers) was identified as a good predictor for promotions [40]. This is in line with a 180° feedback (leaders, employees) study among 396 managers from different departments of an international airline: congruence between managers self-ratings and employees ratings predicted managerial behavior such as innovation, decision making, leading, and motivation [41]. In a sample of 1,190 physicians from the U.S. and Canada, the 180° feedback approach, which is also called multi-rater assessment, provided a more realistic picture of leader-team situations as shown by an improvement in a leadership teamwork index [37]. In our study, leadership ratings of employed physicians were markedly higher in most dimensions than those by non-physician practice personnel. This likely reflects that employed physicians are much closer to their physician leaders regarding training, roles, and duties compared to practice assistants. In addition, practice assistants do not have the perspective to become physician leaders themselves, which implies a fundamentally different perspective. This finding is in line with a 2010 review identifying several studies which showed that staff members who perceive themselves as more similar to the leader give better performance ratings [42]. This effect was shown for example among 406 rater and 396 ratees in an insurance company [43].

Multi-rater assessments can provide the basis for analyzing and at best improving the psychological well-being at workplaces by a better mutual understanding of leaders and staff [7, 42, 44]. A 2016 study of 110 insurance managers and their teams showed higher job satisfaction with higher mutual ratings. Job satisfaction among employees (assessed on a 1 to 5 scale) was lowest when leaders rated their leadership skills higher than their subordinates did (Mean 3.89 of 5 compared to 4.53 of 5 in agreement) [45]. Rowold and Poethke who developed the FIF questionnaire conclude from their studies that leaders can learn to adapt when receiving the leadership ratings as feedback. In addition, they recommend to implement for example regular team meetings and improving leadership skills through training [19]. Results from the DIALHS collaboration from South Africa point at the need for accountability strategies such as standard operation procedures, facility audits and target setting [46]. While other studies followed this approach to share the assessment results with the ratee, we

abstained from this because previous studies in the medical field showed mixed reactions. In a 2005 study, 15 family physicians rated multisource feedback extremely different, from negative to positive. This evaluation was affected by the perceived usefulness, accuracy, and credibility [36].

Using the LMX questionnaire, a 2008 study with 200 nurses from six smaller and larger hospitals showed positive associations of high mutual relationship scores with enhanced commitment, reduced staff turnover, and better organizational behavior [47]. Also, positive effects on employees' health and well-being are described in association with good relationships between leaders and employees. Lower levels of emotional exhaustion were associated with higher leader-member exchange quality in a sample of 343 employees working in the German healthcare sector after 11 months [48]. In addition, a hierarchical regression model showed that the LMX was a good predictor for the health of 412 employees in health and social services in Germany [49]. Compared to the LMX reference values based on 113 participants, our study showed an overall better relationship quality between practice leaders and practice assistants (mean value of 28.1 of 35 for practice leaders and 26.8 for practice assistants vs. 22.9 in the LMX reference population) [35]. Higher scores in the practice setting are likely influenced by the fact that GP leaders recruit personnel themselves, while personnel recruitment and placement in larger institutions is not necessarily in the hands of the direct team leaders.

Strengths and limitations

Novel for the German GP setting, we investigated GP leadership in a large sample with analysis on practice level. Our data provide leadership ratings for each solo practice leaders, but not for each group practice leader, as we had asked staff to rate their leadership team to reflect current small team leadership situations. LMX data were missing for one of seven questions for the small number of employed physicians. However, the analysis of the available data yielded a high relationship quality with leaders like the results for practice assistants. A selection bias cannot be excluded as participating practices might have had a greater interest in the topic.

Conclusion and practical implications

Overall, our data from the IMPROVE*job* study show trustful relationships between GP leaders and their staff. Future GPs´ trainings should enable GP leaders to implement goal-setting, innovation, and individuality focus more effectively. Our results support recent calls for leadership workshops on every level of the medical training for strengthening the GP and other health services workforce.

Acknowledgment

All authors are part of the IMPROVE*job* consortium and owe special thanks to all participating physicians and practice assistants. In addition, special thanks go to the other members of the IMPROVE*job* consortium; current members: M. Brinkmann, S. Burgess, A. Herrmann-Werner, K.-H. Jöckel, F. Junne, A. Siegel, F. Stuber, A. Wagner, and former members: J. Block, J.-M. Bois, A. Dreher, M. Hippler, S. Hartmann, L. Imhoff, C. Kersting, C. Ose and E. Tsarouha. Additionally, we are grateful to all members of the scientific advisory board: Stefanie Esper, Katja Götz, Andre Habrock, Albert Nienhaus, Horst Christian Vollmar, and Andrea Wittich. The work of the Institute of Occupational and Social Medicine and Health Services Research is supported by an unrestricted grant of the Employers' Association of the Metal and Electrical Industry Baden-Wuerttemberg (Suedwestmetall). We acknowledge administrative support by the DLR Project Management Agency.

Contributorship statement

Conceptualization, B.M.W., M.A.R., T.S.-D., C.P., and B.W.; methodology, B.M.W. M.A.R., T.S.-D., M.S., B.W., M.G., E.R., C.P., K.-H.J., L.D., and J.G.; validation, K.-H.J.; formal analysis, M.S. and J.G.; investigation, L.D., K.L., T.S.-D., B.W., E.R., C.P., J.G., M.A.R., and B.M.W.; resources, B.M.W., L.D., K.L., T.S.-D., B.W., M.G., C.P., and E.R..; data curation, K.-H.J.; writing—original draft preparation, M.S. and B.M.W.; writing—review and editing, M.S., T.S.-D., B.W., M.G., E.R., C.P., K.-H.J., L.D., J.G., M.A.R. and B.M.W.; visualization, M.S.; supervision, B.M.W., M.A.R., and B.W.; project administration, K.L., E.R., B.M.W. and M.A.R.; funding acquisition, B.M.W., M.A.R., C.P., and B.W. All authors have read and approved the published version of the manuscript.

Competing interests

The authors declare that they have no competing interests.

Funding

This study is funded by the German Federal Ministry for Education and Research (grant numbers: 01GL1851D, 01GL1751B, 01GL1751A, 01GL1751C). The funder had no role in the study design, the data collection, and analysis, the decision to publish, or the preparation of the manuscript.

Data sharing statement

There are no plans to grant access to the full protocol, participant-level dataset, or statistical code as data contain potentially identifying information.

Ethics approval

The study complies with the ethical principles of the World Medical Association Declaration of Helsinki.

The study was approved first by the Ethics Committee of the Medical Faculty of the University of Bonn (reference number: 057/19, date of approval: 20/02/2019). In addition, the Ethics Committees of the Medical Association Nordrhein (Lfd-Nr.: 2019107), and of the Medical Faculty, University Hospital of Tuebingen (project no.: 446/2019BO2) approved the study protocol.

All participating practice team members received written information and signed informed consent forms, which are stored at the Institute for General Practice and Family Medicine, University of Bonn.

REFERENCES

1 Kringos D, Boerma W, Bourgueil Y, et al. The strength of primary care in Europe: an international comparative study. *British Journal of General Practice* 2013;63(616):e742-e750.

2 Rohde J, Cousens S, Chopra M, et al. 30 years after Alma-Ata: has primary health care worked in countries? *The Lancet* 2008;372(9642):950–61.

3 Swanwick T, Varnam R. Leadership development and primary care. *BMJ Leader* 2019;3(2).

4 Meredith LS, Bouskill K, Chang J, et al. Predictors of burnout among US healthcare providers: a systematic review. *BMJ Open* 2022;12(8):e054243.

5 Boamah SA, Spence Laschinger HK, Wong C, et al. Effect of transformational leadership on job satisfaction and patient safety outcomes. *Nurs Outlook* 2018;66(2):180–89. doi:10.1016/j.outlook.2017.10.004 [published Online First: 23 November 2017].

6 Goetz K, Campbell S, Broge B, et al. Job satisfaction of practice assistants in general practice in Germany: an observational study. *Fam Pract* 2013;30(4):411–17. doi:10.1093/fampra/cmt015 [published Online First: 19 April 2013].

7 i Solà GJ, i Badia JG, Hito PD, et al. Self-perception of leadership styles and behaviour in primary health care. *BMC Health Services Research* 2016;16.

8 Kersting C, Zimmer L, Thielmann A, et al. Chronic stress, work-related daily challenges and medicolegal investigations: a cross-sectional study among German general practitioners. *BMC Fam Pract* 2019:143. doi:10.1186/s12875-019-1032-6 [published Online First: 24 October 2019].

9 Wallace JE, Lemaire JB, Ghali WA. Physician wellness: a missing quality indicator. *The Lancet* 2009;374(9702):1714–21.

10 Wong CA, Cummings GG, Ducharme L. The relationship between nursing leadership and patient outcomes: a systematic review update. *J Nurs Manag* 2013;21(5):709–24.

11 Sfantou DF, Laliotis A, Patelarou AE, et al. Importance of Leadership Style towards Quality of Care Measures in Healthcare Settings: A Systematic Review. *Healthcare (Basel, Switzerland)* 2017;5(4):73.

12 Söling S, Pfaff H, Karbach U, et al. How is leadership behavior associated with organization-related variables? Translation and psychometric evaluation of the implementation leadership scale in German primary healthcare. *BMC Health Services Research* 2022;22(1):1–13.

13 Bass BM. Two Decades of Research and Development in Transformational Leadership. *European Journal of Work and Organizational Psychology* 1999;8(1):9–32.

14 Poethke U. Leadership, values and communication: a cross-cultural investigation of the extended full-range of leadership behaviors: Doctoral dissertation 2019.

15 Rowold J, Schlotz W. Transformational and transactional leadership and followers' chronic stress. *Leadership Review* 2009;9(1):35–48.

16 Graen GB, Uhl-Bien M. Relationship-based approach to leadership: Development of leader-member exchange (LMX) theory of leadership over 25 years: Applying a multi-level multi-domain perspective. *Leadership Quarterly* 1995;6(2):219–47.

17 Graen GB, Uhl-Bien M. Relationship-based approach to leadership: Development of leader-member exchange (LMX) theory of leadership over 25 years: Applying a multi-level multi-domain perspective. *The Leadership Quarterly* 1995;6(2):219–47.

18 Gregersen S, Vincent-Höper S, Nienhaus A. Health-Relevant Leadership Behaviour: A Comparison of Leadership Constructs. *German Journal of Human Resource Management* 2014;28(1-2):117–38.

19 Rowold J, Poethke U. FIF: Fragebogen zur Integrativen Führung Manual, 1st edn. Göttingen: Hogrefe 2017.

20 Lockyer J. Multisource feedback in the assessment of physician competencies. *J Contin Educ Health Prof* 2003;23(1):4–12.

21 Davis DA, Mazmanian PE, Fordis M, et al. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA* 2006;296(9):1094–102.

22 Donnon T, Al Ansari A, Al Alawi S, et al. The reliability, validity, and feasibility of multisource feedback physician assessment: a systematic review. *Acad Med* 2014;89(3):511–16.

23 Saravo B, Netzel J, Kiesewetter J. The need for strong clinical leaders - Transformational and transactional leadership as a framework for resident leadership training. *PLoS One* 2017;12(8):e0183019.

24 Richter A, Lornudd C, Thiele Schwarz U von, et al. Evaluation of iLead, a generic implementation leadership intervention: mixed-method preintervention-postintervention design. *BMJ Open* 2020;10(1):e033227. doi:10.1136/bmjopen-2019-033227 [published Online First: 12 January 2020].

25 Hill DA, Jimenez J-C, Cohn SM, et al. How to be a leader: a course for residents. *Cureus* 2018;10(7).

26 Viehmann A, Kersting C, Thielmann A, et al. Prevalence of chronic stress in general practitioners and practice assistants: Personal, practice and regional characteristics. *PLoS One* 2017;12(5):e0176658.

27 Dreher A, Theune M, Kersting C, et al. Prevalence of burnout among German general practitioners: Comparison of physicians working in solo and group practices. *PLoS One* 2019;14(2):e0211223.

28 Freund T, Everett C, Griffiths P, et al. Skill mix, roles and remuneration in the primary care workforce: who are the healthcare professionals in the primary care teams across the world? *International journal of nursing studies* 2015;52(3):727–43.

29 infas im Auftrag der KBV und des NAV-Virchow-Bund. Tabellenband Ärztemonitor 2018, Ergebnisse für Haus- und Fachärzte. Bonn 2018.

30 Weltermann BM, Kersting C, Pieper C, et al. IMPROVEjob - Participatory intervention to improve job satisfaction of general practice teams: a model for structural and behavioural prevention in small and medium-sized enterprises - a study protocol of a cluster-randomised controlled trial. *Trials* 2020;21(1):532.

31 Degen L, Linden K, Seifried-Dübon T, et al. Job Satisfaction and Chronic Stress of General Practitioners and Their Teams: Baseline Data of a Cluster-Randomised Trial (IMPROVEjob). *Int J Environ Res Public Health* 2021;18(18). doi:10.3390/ijerph18189458 [published Online First: 8 September 2021].

32 Ingold PV. Fragebogen zur integrativen Führung. *Zeitschrift für Arbeits- und Organisationspsychologie A&O* 2019;63(2):100–06.

33 Stuber F, Seifried-Dübon T, Rieger MA, et al. Investigating the Role of Stress-Preventive Leadership in the Workplace Hospital: The Cross-Sectional Determination of Relational Quality by Transformational Leadership. *Front Psychiatry* 2019;10:622. doi:10.3389/fpsyt.2019.00622 [published Online First: 3 September 2019].

34 Schyns B. Überprüfung einer deutschsprachigen Skala zum Leader-Member-Exchange-Ansatz. *Zeitschrift für Differentielle und Diagnostische Psychologie* 2002.

35 Schyns B, Paul T. Leader-Member Exchange Skala (LMX7): ZIS - GESIS Leibniz Institute for the Social Sciences 2014.

36 Sargeant J, Mann K, Ferrier S. Exploring family physicians' reactions to multisource feedback: perceptions of credibility and usefulness. *Med Educ* 2005;39(5):497–504.

37 Hu J, Lee R, Mullin S, et al. How physicians change: Multisource feedback driven intervention improves physician leadership and teamwork. *Surgery* 2020;168(4):714–23. doi:10.1016/j.surg.2020.06.008 [published Online First: 6 August 2020].

38 Northouse PG. Leadership: Theory and practice, 6th edn. Thousand Oaks: Sage Publications 2010.

39 Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd edn. Hoboken: Taylor and Francis 1988.

40 Halverson SK, Tonidandel S, Barlow C, et al. Self-other agreement on a 360-degree leadership evaluation. In: 17th Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, Canada 2002.

41 Furnham A, Stringfield P. Congruence of self and subordinate ratings of managerial practices as a correlate of supervisor evaluation. *Journal of Occupational and Organizational Psychology* 1994;67(1):57–67.

42 Fleenor JW, Smither JW, Atwater LE, et al. Self–other rating agreement in leadership: A review. *The Leadership Quarterly* 2010;21(6):1005–34.

43 Antonioni D, Park H. The effects of personality similarity on peer ratings of contextual work behaviors. *Personnel Psychology* 2001;54(2):331–60.

44 Alrubaysh MA, Alshehri MH, Alsuhaibani EA, et al. The leadership styles of primary healthcare center managers and center performance outcomes in Riyadh, Saudi Arabia: A correlational study. *Journal of Family & Community Medicine* 2022;29(1):56.

45 Erben J, Schneider FM, Maier M. In the Ear of the Beholder: Self-Other Agreement in Leadership Communication and Its Relationship With Subordinates' Job Satisfaction. *International Journal of Business Communication* 2019;56(4):505–29.

46 Cleary S, Du Toit A, Scott V, et al. Enabling relational leadership in primary healthcare settings: lessons from the DIALHS collaboration. *Health Policy and Planning* 2018;33(suppl_2):ii65-ii74.

47 Chen C-HV, Wang S-J, Chang W-C, et al. The effect of leader-member exchange, trust, supervisor support on organizational citizenship behavior in nurses. *Journal of Nursing Research* 2008;16(4):321–28.

48 Gregersen S, Vincent-Höper S, Nienhaus A. Job-related resources, leader–member exchange and well-being – a longitudinal study. *Work & Stress* 2016;30(4):356–73.

49 Vincent-Höper S, Stein M, Gregersen S, et al. Messung gesundheitsförderlicher Führung in Gesundheitsdienst und Wohlfahrtspflege. *ASU Arbeitsmed Sozialmed Umweltmed* 2018(53):46–53.

Additional Files

Additional file 1: CONSORT 2010 checklist of information to include when reporting a randomised trial

CONSORT 2010 checklist of information to include when reporting a randomised trial*

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	n/a
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	Abstract, 2
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	Background, 4-5
	2b	Specific objectives or hypotheses	Background, 4-5
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	Methods, 5
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	n/a
Participants	4a	Eligibility criteria for participants	Methods, 5-6
	4b	Settings and locations where the data were collected	Methods, 5-6
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	n/a
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	Methods, 6-7
	6b	Any changes to trial outcomes after the trial commenced, with reasons	n/a
Sample size	7a	How sample size was determined	Study protocol, Weltermann et al., Trials 2020
	7b	When applicable, explanation of any interim analyses and stopping guidelines	n/a

Randomisation:			
Sequence generation	8a	Method used to generate the random allocation sequence	Details in Study protocol, Weltermann et al., Trials 2020
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	Details in Study protocol, Weltermann et al., Trials 2020
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	n/a
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	Details in Study protocol, Weltermann et al., Trials 2020
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	n/a
	11b	If relevant, description of the similarity of interventions	n/a
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	Methods,7-8
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	n/a
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	n/a
	13b	For each group, losses and exclusions after randomisation, together with reasons	n/a
Recruitment	14a	Dates defining the periods of recruitment and follow-up	Details in Study protocol, Weltermann et al., Trials 2020
	14b	Why the trial ended or was stopped	Details in Study protocol,

			Weltermann et al., Trials 2020
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	Results, table 1, Page 8
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	Results, 9
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	Results, 9-11
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	N/A.
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	n/a
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	n/a
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	Discusssion, Strengths and limitatoins, Page 12-13
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	13
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	Discussion, 11-13
Other information			
Registration	23	Registration number and name of trial registry	Trial registration, 2
Protocol	24	Where the full trial protocol can be accessed, if available	Study protocol, Weltermann et al., Trials 2020
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	Funding, 17-18

*We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org.

BMJ Open

A 180° view on general practitioners' leadership skills: practice-level comparisons of leader and staff assessments using data from the cluster-randomized controlled IMPROVEjob study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2022-066298.R2
Article Type:	Original research
Date Submitted by the Author:	17-Jul-2023
Complete List of Authors:	Schmidt, Manuela; University Hospital Bonn, Institute of General Practice and Family Medicine Seifried-Dübon, Tanja; University Hospital Tuebingen Osianderstraße 5, Department of Psychosomatic Medicine and Psychotherapy Göbel, Julian; University Hospital Bonn, Institute of General Practice and Family Medicine Degen, Lukas; University Hospital Bonn, Institute of General Practice and Family Medicine Werners, Brigitte; Ruhr University Bochum Grot, Matthias; Ruhr University Bochum Rind, Esther; Institute for Occupational Medicine, Social Medicine and Health Services Research , University Hospital Tuebingen Pieper, Claudia; University of Duisburg-Essen Jöckel, Karl-Heinz; Institute of Medical Informatics, Biometry and Epidemiology, University Duisburg- Essen Linden, Karen; University Hospital Bonn, Institute of General Practice and Family Medicine Rieger, Monika A.; Institute for Occupational Medicine, Social Medicine and Health Services Research , University Hospital Tuebingen Weltermann, Birgitta; University Hospital Bonn, Institute of General Practice and Family Medicine
Primary Subject Heading:	General practice / Family practice
Secondary Subject Heading:	General practice / Family practice
Keywords:	PRIMARY CARE, Human resource management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

A 180° view on general practitioners' leadership skills: practice-level comparisons of leader and staff assessments using data from the cluster-randomized controlled IMPROVE*job* study

Authors: Manuela Schmidt[1]*, Tanja Seifried-Dübon[2], Julian Göbel[1], Lukas Degen[1], Brigitte Werners[3], Matthias Grot[3], Esther Rind[4], Claudia Pieper[5], Karl-Heinz Jöckel[5 6], Karen Linden[1], Monika A. Rieger[4], Birgitta M. Weltermann[1] on behalf of the IMPROVE*job* consortium

[1] Institute of General Practice and Family Medicine, University Hospital Bonn, University of Bonn, Venusberg-Campus 1, 53127 Bonn, Germany

[2] Department of Psychosomatic Medicine and Psychotherapy, University Hospital Tuebingen, Osianderstraße 5, 72076 Tuebingen, Germany

[3] Institute of Management, Operations Research, Ruhr University Bochum, Universitätsstr. 150, 44801 Bochum, Germany

[4] Institute of Occupational and Social Medicine and Health Services Research, University Hospital Tuebingen, Wilhelmstr. 27, 72074 Tuebingen, Germany

[5] Institute for Medical Informatics, Biometry and Epidemiology, University Hospital Essen, University of Duisburg-Essen, Hufelandstr. 55, 45147 Essen, Germany

[6] Center for Clinical Trials, University Hospital Essen, University of Duisburg-Essen, Hufelandstr. 55, 45147 Essen, Germany.

*Correspondence to: julian.goebel@ukbonn.de

Abstract

Objectives: Strong primary care leaders are needed to assure high quality services for patient populations. This study analyzed general practitioners' (GP) leadership skills comparing practice-level self- and staff assessments based on the Full Range of Leadership Model (FRLM) and the Leader-Member Exchange (LMX).

Setting: The questionnaire survey was conducted among German general practice leaders and their staff participating in the IMPROVE*job* trial.

Participants: The study population comprised 60 German general practices with 366 participants: 84 GP practice leaders and 282 employees (28 physicians and 254 practice assistants).

Primary and secondary outcome measures: Leadership skills of the practice leaders were measured using the Integrative Leadership Questionnaire (German *FIF*) and the Leader-Member Exchange (LMX-7) questionnaire. Leaders rated themselves and practice staff rated their leaders. The data was analyzed by paired mean comparisons on the practice level.

Results: For most leadership dimensions, practice leaders rated themselves higher than their employees rated them. Differences were found for transformational leadership ($p<.001$, $d=.41$), especially for the dimensions 'innovation' ($p<.001$, $d=.69$) and 'individuality focus' ($p<.001$, $d=.50$). For transactional leadership, the dimension 'goal setting' differed significantly ($p<.01$, $d=.30$) but not the other dimensions. Scores for negative leadership were low and showed no differences between leaders and employees. Interestingly, employed physicians' rated their practice leaders higher on the two transformational ('performance development', 'providing a vision') and all transactional dimensions. The LMX-7 scale showed high quality relationships between leaders and employees.

Conclusions: This 180° analysis of GPs´ leadership skills with self- and employee ratings indicated good relationships. There is a potential to improve leadership regarding goal-setting, innovation and focusing on individual team members. These results allow for the development of targeted interventions.

Trial registration: German Clinical Trials Register, DRKS00012677. Registered 16 October 2019.

Keywords: leadership, leadership quality, general practitioner, practice staff, 180° feedback

Strengths & limitations:

- This 180° feedback approach in the GP setting allows for a better understanding of leadership from the perspective of different practice professionals.
- The data reflect a typical spectrum of German general practices with solo and group practices, but results may differ in other settings.
- Leadership teams, not individual leaders, were rated in group practices to capture leadership at the practice level.

Background:

Strong primary care leaders and a strong primary care workforce are important to assure the health of populations and primary care teams [1–3]. A recent systematic review of 20 studies by Meredith et al. showed an association between stronger leadership and less burnout among different medical professionals in the United States [4]. In contrast, poor leadership skills have a negative impact on job satisfaction [5–7], staff well-being [8] and the quality of patient care [5, 9, 10]. A review showed correlations between better leadership and various quality of care indicators, e.g. pain, safety and 30-day-mortality [11]. In addition to individual outcomes, leadership is important to promote organizational changes (e.g., the implementation of IT-supported care) [12].

Scientifically, leadership is conceptualized in several theories. One of the most studied leadership frameworks is the Full Range of Leadership Model (FRLM), which integrates transactional, transformational, and negative leadership [13, 14]. Transactional leadership describes leaders' structuring of work situations, the exchange of contingent rewards (e.g. work against salary), and the management by exception [13–15]. In contrast, transformational leadership moves beyond leaders' and staff's self-interests. It focuses on the staff's attitudes and values regarding overarching goals such as self-actualization, organizational achievements, and the well-being of others and society as a whole [13, 14]. Building on the FRLM, a recent further development, the so-called Implementation Leadership Scale, focusses on the role of leadership for implementation of organizational changes [12]. Another important leadership theory, the Leader-Member Exchange (LMX), specifically addresses the relationship between leaders and staff. It concentrates on the perceived quality of the dyadic relationship between a staff member and the immediate leader [13, 16]. The relationship reflects a dyadic social exchange process ranging from low LMX, described by limited social transactions with more transactional leadership to high LMX, which represents a transformational approach with a high degree of social exchange and a mature leader-member partnership [17]. High-quality relationships positively influence employees' work-related well-being and are associated with higher job satisfaction of health care workers [18].

Based on these theories, various questionnaires were developed, e.g. the Leader-Member Exchange questionnaire 'LMX-7' [16] and the German questionnaire 'Fragebogen für integrative Führung' (FIF; in English: Questionnaire for Integrative Leadership) [15, 19]. These instruments allow for a multi-rater perspective: the leader's and the staff's views on the leader's behavior are measured and compared providing 180° feedback. This method is valuable because assessments from different perspectives create a more comprehensive picture of the leaders' actual skills and performances [20]. Two recent reviews of 60 studies from various medical settings showed that such approaches are increasingly

applied in medical education and graduate training [21, 22], but have not been used to evaluate GP leaders and their teams. Effective interventions to improve leadership were developed and evaluated in the hospital [23] and healthcare management setting [24]. For example, Saravo et al. showed an improvement in transformational and transactional leadership performance of 57 medical residents in hospital rotations after a 4-week intervention [23]. In addition, a 2018 study from Hill et al. highlighted positive effects of a leadership training for surgical residents on teamwork and team involvement in decision-making [25]. However, such interventions have not been implemented in German primary care, although high chronic stress and burnout rates are reported for this workforce [26, 27]. The need is even larger as about half of the German GPs who mainly work in GP-owned private practices [28], will reach retirement age in the next ten years [29]. Based on the leadership frameworks mentioned above, the publicly funded IMPROVE*job* study aimed to improve the job satisfaction of physician leaders and practice personnel of German GP practices focusing on leadership, communication, and work processes [30, 31]. At baseline, GPs´ leadership skills were evaluated comparing GP leaders' self and staff ratings on practice level.

Methods

This analysis draws on the baseline data of the IMPROVE*job* study, which is designed as a cluster-randomized controlled trial (cRCT) to improve job satisfaction among practice personnel. The details are described in the study protocol [30].

In short, a total of 60 GP practices in the North Rhine region in Germany were recruited by the Institute of General Practice and Family Medicine of the University of Bonn. The sample comprised single (owned by one practice leader) and group practices (owned by more than one practice leader), some of which were also involved as teaching practices (affiliated to a university). The study aimed to recruit practice teams, including physician leaders, employed physicians, and practice assistants. A total of 84 GP practice leaders, 28 employed physicians and 254 practice assistants were recruited. In Germany, primary care is typically provided by GP-owned practices with 1 to 3 physicians. For each physician, practices employ about 1 to 2 certified practice assistants who finished a vocational training of 3 years. Similar to other regions worldwide, the size of group practices is increasing.

Patient and public involvement

The study did not target patients, but general practice personnel. Therefore, no patients or members of the public were involved.

Ethics

The study was approved first by the Ethics Committee of the Medical Faculty of the University of Bonn (reference number: 057/19, date of approval: 20/02/2019). In addition, the Ethics Committees of the Medical Association North-Rhine (Lfd-Nr.: 2019107) and of the Medical Faculty, University Hospital of Tuebingen (project no.: 446/2019BO2) approved the study protocol. The study was performed in accordance with the Declaration of Helsinki. All participating practice team members received a study information and signed informed consent forms.

Measures

Practice leaders answered a short questionnaire on practice characteristics and the questionnaire for practice leaders. Employed physicians and practice assistants completed different versions of the same employee questionnaire. Details of the methods and the characteristics of the study population are published [30, 31].

All participants provided sociodemographic, professional, and work-related characteristics which are published [31]. In addition, GP leaders and practice staff filled the following two leadership questionnaires:

1. Integrative Leadership Questionnaire (FIF)

Transformational, transactional, and negative leadership were measured using the FIF questionnaire. Its scales' validity and internal consistency are confirmed for different populations [19, 32]. The FIF has been used in non-medical and hospital settings [33], but not in primary care.

All 40 items of the FIF are answered on a 5-point Likert scale and are worded to reflect either the leader's or the staff's position.

The measures comprise:

- the transformational leadership scale consisting of six dimensions: innovation, team spirit, performance development, individuality focus, providing a vision, and being a role model;
- the transactional leadership scale with two dimensions: goal setting and management by exception;
- the negative leadership scale with two dimensions: laissez-faire and destructive leadership.

2. Leader-Member Exchange (LMX-7)

The relationship quality between leaders and staff is measured using the Leader-Member Exchange questionnaire (LMX-7) with seven items on a 5-point Likert scale, which are worded to reflect the leader or the staff position [16, 34, 35].

The multi-rater, 180° approach is applied to the two leadership scales. Results of such assessments are usually shared with the ratee, yet previous studies showed mixed reactions in the medical setting [36, 37]. Therefore, the results of the 180° feedback in our study were not shared with the participating practices but are used on an aggregated level for research purposes only.

Statistical analysis

Statistical analyses were conducted using SPSS Statistics 27 (IBM Cooperation, Armonk, Ny, USA, 2020). All analyses were carried out at the participant and the practice level.

The FIF data were analyzed according to the official manual [19]. Mean scores for transformational, transactional, and negative leadership were summarized both for the respective main scale and all dimensions: for transactional and transformational leadership, they ranged from 1 (worst rating) to 5 (best rating); for negative leadership, they ranged from 1 (best rating) to 5 (worst rating). For comparison, scores were standardized using T-scaling tables from reference populations as defined by Rowold & Poethke [19]. These T-values are based on a normal distribution around 50 (SD 10). Thus, values above 70 only reflect about 2% of the reference population from German-speaking countries [19].

The LMX-7 was analyzed per standard protocol by creating a sum score of all seven items without transformation [35]. The LMX-7 score can range from 7 to 35 with five standard categories which were interpreted as follows: score 7 to 14 = very low; 15 to 19 = low; 20 to 24 = moderate; 25 to 29 = high, 30 to 35 = very high [38]. Inadvertently, question seven was missing on all employed physicians' questionnaires, which reduced the answered questions to six. As the LMX-7 manual does not suggest a standard approach for missing values, we excluded employed physicians from the further analyses.

For the 180° feedback approach on practice level, the combined mean scores of employed physicians and practice assistants per practice were compared to the self-assessment of their respective leaders using paired t-tests, as the data satisfied the condition of a normal distribution with the Kolmogorov-Smirnov test. Cohen's d was applied to determine the effect size of mean comparisons with the following standard interpretations: small effects from d=.2, medium from d=.5, and high from d=.8 [39]. In single practices, the staff ratings were compared to the leader's assessment. In practices with more than one owner (group practices), each leader's self-rating was compared with the respective ratings of the practice personnel, who were asked to rate the leadership team of the practice, not stratified by individual leaders. This approach was chosen because practice owners of German practices typically work as a leadership team. In addition, the ratings of the transformational and transactional leadership scales were compared stratified by practice type (single vs. group and teaching vs. non-teaching practices) using the Kruskal-Wallis tests because the data for practice comparisons did not satisfy the conditions for parametric tests.

Results

Population

The baseline data of the IMPROVE*job* study included 366 participants from 60 practices, consisting of 84 practice leaders, 28 employed physicians, and 254 practice assistants. The mean age of the participants was 44.4 years, with a mean of 54.3 years for practice leaders, 44.8 for employed physicians, and 41.0 for practice assistants. Among the practice assistants, 99.6% were female, as were 76.6% of the employed physicians and half of the practice leaders (52.4%). Most practice leaders worked full-time (90.5%), as did about a quarter of the employed physicians (28.6%) and 41.5% of the practice assistants (see Table 1). The details on the sociodemographic descriptions are published [31].

Table 1: Sociodemographic description of participants at baseline [31]

Variable	Total sample N=366	Practice leaders N=84	Employed physicians N=28	Practice assistants N=254
Female, %	87.1	52.4	78.6	99.6
Age in years, mean (SD)	44.4 (12.8)	54.3 (6.2)	44.8 (9.8)	41.0 (13.0)
Years in current practice, mean (SD)	10.0 (9.1)	15.3 (8.4)	3.9 (5.4)	8.8 (8.9)
Working full-time, %	52.0	90.5	28.6	41.5

On average, practice leaders had been accredited for 26.6 years and licensed for the statutory health insurance for 16.4 years. Seven (25%) of the employed physicians were in GP training. Practice assistants had graduated on average 19.9 years ago, while 7.5% were still in training. Of the 60 practices, 21 (35%) were single and 39 (65%) were group practices; of these, 34 were teaching (57%) and 26 (43%) were non-teaching practices. On average, practices were in the same location for 20.4 years (SD 14 years).

Leadership

The transactional and transformational leadership scales showed a high internal consistency with Cronbach's α = .74 to .93 for the staff members' assessment and Cronbach's α = .72 to .87 for the leaders' assessment. For negative leadership, the scales showed a sufficient internal consistency for staff members' (Cronbach's α = .73 to .80) but not for leaders' assessments (Cronbach's α = .47 to .68). The mean results of the FIF were within the one standard deviation range of the reference population [19]. Based on raw values, employed physicians rated their leaders consistently better than practice assistants and better than the leaders themselves for some items. While practice assistants rated their leaders more poorly than the practice leaders in raw values, reference T-values showed only minor differences. The details are outlined in Table 2.

The LMX-7 scale showed an internal consistency of Cronbach's α = .88 for staff members (practice assistants) and α = .71 for leaders. Both groups showed a high relationship quality, scoring 28 for

practice leaders and 26 for practice assistants. As the seventh question was missing for employed physicians, they were excluded from the analysis. However, the sum score of the remaining six questions also showed a high score of 24.9 out of 30. The details are shown in Table 2.

Table 2: Leadership assessment by employment group: main scales (in bold) and dimensions (LMX values can range from 7 to 35, FIF scales from 1 to 5)

	Practice leaders (N=84)				Employed physicians (N=28)				Practice assistants (N=254)			
	M	**SD**	**T***	**n**	**M**	**SD**	**T***	**N**	**M**	**SD**	**T***	**n**
Transformational leadership	**3.9**	**0.6**	**45**	**84**	**3.9**	**0.7**	**56**	**27**	**3.5**	**0.8**	**52**	**237**
Innovation	4.2	0.6	49	84	4.0	1.0	55	28	3.7	0.9	52	247
Team spirit	4.1	0.7	49	84	3.8	1.0	54	28	3.6	1.1	52	251
Performance development	3.6	0.8	44	84	4.1	0.7	57	27	3.5	1.0	51	247
Individuality focus	3.9	0.7	47	84	3.7	1.0	54	28	3.5	1.1	53	249
Providing a vision	3.5	0.9	45	84	3.6	0.9	55	28	3.2	1.1	51	245
Being a role model	4.1	0.6	45	84	4.0	0.8	55	27	3.7	1.0	52	246
Transactional leadership	**3.4**	**0.7**	**47**	**83**	**3.5**	**0.7**	**54**	**27**	**3.2**	**0.8**	**50**	**244**
Goal setting	3.5	0.7	44	83	3.7	0.9	56	27	3.1	1.0	50	246
Management by exception	3.3	0.8	51	83	3.4	0.8	52	27	3.3	0.9	51	245
Negative leadership	**1.5**	**0.5**	**51**	**83**	**1.5**	**0.6**	**45**	**28**	**1.7**	**0.7**	**47**	**248**
Laissez-faire	1.6	0.6	52	83	1.6	0.8	45	28	1.7	0.8	46	249
Destructive	1.4	0.5	51	83	1.4	0.6	46	28	1.6	0.7	48	248
LMX-7	**28.1**	**2.6**	**-**	**81**	**n/a**	**n/a**	**-**	**n/a**	**26.7**	**4.8**	**-**	**222**

*Reference T-values range from 0 to 100, as defined by Rowold & Poethke 2017

180° leadership feedback

Practice leaders self-rated their leadership skills slightly better than their staff for all dimensions except for 'management by exception'. There were no statistically significant differences for negative leadership. For transactional leadership, goal setting differed significantly with a low effect size (p=.009, d=.30). Leaders' scores on transformational leadership were significantly higher than the scores of the teams, with the dimension for innovation reaching the strongest effect size (p=<.001, d=.69), followed by individuality focus with a medium effect size (p=<0.001, d=.50). The scores for team spirit and being a role model were slightly lower, but significant. The main scale for transformational leadership also showed a significant difference with a medium effect size (p=<.001, d=.41). The details are outlined in Table 3.

Table 3: Comparison of leaders' self- and staff ratings (n=84 leader-team pairs): main scales (in bold) and dimensions

	Practice leaders		Practice staff		Paired t-test		
	M	SD	M	SD	t(df)	p	d
Transformational leadership	**3.9**	**0.5**	**3.6**	**0.6**	**3.721(82)**	**<.001**	**0.41**
Innovation	4.2	0.6	3.8	0.6	6.359(83)	<.001	0.69
Team spirit	4.1	0.7	3.8	0.7	3.462(82)	.001	0.38
Performance development	3.6	0.8	3.7	0.6	-0.208(83)	.836	-
Individuality focus	3.9	0.7	3.5	0.6	4.633(83)	<.001	0.50
Providing a vision	3.5	0.9	3.3	0.8	1.592(82)	.115	-
Being a role model	4.1	0.6	3.8	0.6	2.833(82)	.006	0.31
Transactional leadership	**3.4**	**0.6**	**3.3**	**0.5**	**1.291(81)**	**.200**	**-**
Goal setting	3.5	0.7	3.2	0.6	2.681(81)	.009	0.30
Management by exception	3.3	0.8	3.4	0.6	-0.470(82)	.640	-
Negative leadership	**1.5**	**0.4**	**1.6**	**0.4**	**-1.744(82)**	**.085**	**-**
Laissez-faire	1.6	0.6	1.7	0.5	-1.563(82)	.122	-
Destructive	1.4	0.5	1.6	0.5	-1.514(82)	.134	-
LMX-7	28.1	2.6	26.8	3.5	3.275(79)	.002	0.37

Transformational and transactional leadership by practice type

The Kruskal-Wallis test was applied to analyze for differences in leadership by practice types. It showed slight but non-significant differences in the raw values between practice types (single vs. group, teaching vs. non-teaching practices), e.g., slightly higher ratings for transformational leadership in single and non-teaching practices. These slight differences persisted when using reference T-values. For details, see Table 4.

Table 4: Comparison of leadership assessments by practice type: single versus group practices and teaching versus non-teaching practices

	Single (n=21)			Group (n=39)			Non-teaching (n=26)			Teaching (n=34)		
	M	T	n	M	T	N	M	T	n	M	T	n
Practice leaders												
Transformational	4.0	47	21	3.9	45	63	3.8	43	37	4.0	47	47
Transactional	3.4	47	21	3.4	47	62	3.3	45	37	3.4	47	46
Negative	1.5	51	21	1.5	51	62	1.6	53	37	1.5	51	46
LMX-7	28.8	-	20	27.9	-	61	27.5	-	36	28.6	-	45
Practice staff												
Transformational	3.7	54	67	3.6	53	212	3.5	52	117	3.6	53	162
Transactional	3.4	53	67	3.3	51	212	3.3	51	117	3.2	50	162
Negative	1.7	47	70	1.7	47	212	1.7	47	117	1.6	46	165
LMX-7	27.5	-	61	26.3	-	190	25.6	-	105	27.3	-	146

Discussion

Using a 180° feedback approach of leadership in GP practices, this study showed good relationships between leaders and staff with low levels of negative leaderships. Practice staff rated their leaders slightly higher on all transformational and transactional dimensions than the 234 German leaders and 713 employees from the FIF questionnaire reference population [19]. Also, agreement between GP leaders and staff was higher than in a study of 1,137 German hospital employees (315 leaders, 822 staff members) from different occupational groups (e.g., physicians, nurses, administration, information technology), which used the same methodology [33]. Interestingly, hospital and GP leaders rated themselves approximately similar [33].

The benefit of 180° and 360° feedback is shown in studies from various settings. In a sample of more than 2,000 U.S. military leaders, 360° feedback (leaders, subordinates, peers) was identified as a good predictor for promotions [40]. This is in line with a 180° feedback (leaders, employees) study among 396 managers from different departments of an international airline: congruence between managers self-ratings and employees ratings predicted managerial behavior such as innovation, decision making, leading, and motivation [41]. In a sample of 1,190 physicians from the U.S. and Canada, the 180° feedback approach, which is also called multi-rater assessment, provided a more realistic picture of leader-team situations as shown by an improvement in a leadership teamwork index [37]. In our study, leadership ratings of employed physicians were markedly higher in most dimensions than those by non-physician practice personnel. This likely reflects that employed physicians are much closer to their physician leaders regarding training, roles, and duties compared to practice assistants. In addition, practice assistants do not have the perspective to become physician leaders themselves, which implies a fundamentally different perspective. This finding is in line with a 2010 review identifying several studies which showed that staff members who perceive themselves as more similar to the leader give better performance ratings [42]. This effect was shown for example among 406 rater and 396 ratees in an insurance company [43].

Multi-rater assessments can provide the basis for analyzing and at best improving the psychological well-being at workplaces by a better mutual understanding of leaders and staff [7, 42, 44]. A 2016 study of 110 insurance managers and their teams showed higher job satisfaction with higher mutual ratings. Job satisfaction among employees (assessed on a 1 to 5 scale) was lowest when leaders rated their leadership skills higher than their subordinates did (Mean 3.89 of 5 compared to 4.53 of 5 in agreement) [45]. Rowold and Poethke who developed the FIF questionnaire conclude from their studies that leaders can learn to adapt when receiving the leadership ratings as feedback. In addition, they recommend to implement for example regular team meetings and improving leadership skills through training [19]. Results from the DIALHS collaboration from South Africa point at the need for accountability strategies such as standard operation procedures, facility audits and target setting [46]. While other studies followed this approach to share the assessment results with the ratee, we

abstained from this because previous studies in the medical field showed mixed reactions. In a 2005 study, 15 family physicians rated multisource feedback extremely different, from negative to positive. This evaluation was affected by the perceived usefulness, accuracy, and credibility [36].

Using the LMX questionnaire, a 2008 study with 200 nurses from six smaller and larger hospitals showed positive associations of high mutual relationship scores with enhanced commitment, reduced staff turnover, and better organizational behavior [47]. Also, positive effects on employees' health and well-being are described in association with good relationships between leaders and employees. Lower levels of emotional exhaustion were associated with higher leader-member exchange quality in a sample of 343 employees working in the German healthcare sector after 11 months [48]. In addition, a hierarchical regression model showed that the LMX was a good predictor for the health of 412 employees in health and social services in Germany [49]. Compared to the LMX reference values based on 113 participants, our study showed an overall better relationship quality between practice leaders and practice assistants (mean value of 28.1 of 35 for practice leaders and 26.8 for practice assistants vs. 22.9 in the LMX reference population) [35]. Higher scores in the practice setting are likely influenced by the fact that GP leaders recruit personnel themselves, while personnel recruitment and placement in larger institutions is not necessarily in the hands of the direct team leaders.

Strengths and limitations

Novel for the German GP setting, we investigated GP leadership in a large sample with analysis on practice level. Our data provide leadership ratings for each solo practice leaders, but not for each group practice leader, as we had asked staff to rate their leadership team to reflect current small team leadership situations. LMX data were missing for one of seven questions for the small number of employed physicians. However, the analysis of the available data yielded a high relationship quality with leaders like the results for practice assistants. A selection bias cannot be excluded as participating practices might have had a greater interest in the topic.

Conclusion and practical implications

Overall, our data from the IMPROVE*job* study show trustful relationships between GP leaders and their staff. Future GPs´ trainings should enable GP leaders to implement goal-setting, innovation, and individuality focus more effectively. Our results support recent calls for leadership workshops on every level of the medical training for strengthening the GP and other health services workforce.

Acknowledgment

All authors are part of the IMPROVE*job* consortium and owe special thanks to all participating physicians and practice assistants. In addition, special thanks go to the other members of the IMPROVE*job* consortium; current members: M. Brinkmann, S. Burgess, A. Herrmann-Werner, K.-H. Jöckel, F. Junne, A. Siegel, F. Stuber, A. Wagner, and former members: J. Block, J.-M. Bois, A. Dreher, M. Hippler, S. Hartmann, L. Imhoff, C. Kersting, C. Ose and E. Tsarouha. Additionally, we are grateful to all members of the scientific advisory board: Stefanie Esper, Katja Götz, Andre Habrock, Albert Nienhaus, Horst Christian Vollmar, and Andrea Wittich. The work of the Institute of Occupational and Social Medicine and Health Services Research is supported by an unrestricted grant of the Employers' Association of the Metal and Electrical Industry Baden-Wuerttemberg (Suedwestmetall). We acknowledge administrative support by the DLR Project Management Agency.

Contributorship statement

Conceptualization, B.M.W., M.A.R., T.S.-D., C.P., and B.W.; methodology, B.M.W. M.A.R., T.S.-D., M.S., B.W., M.G., E.R., C.P., K.-H.J., L.D., and J.G.; validation, K.-H.J.; formal analysis, M.S. and J.G.; investigation, L.D., K.L., T.S.-D., B.W., E.R., C.P., J.G., M.A.R., and B.M.W.; resources, B.M.W., L.D., K.L., T.S.-D., B.W., M.G., C.P., and E.R..; data curation, K.-H.J.; writing—original draft preparation, M.S. and B.M.W.; writing—review and editing, M.S., T.S.-D., B.W., M.G., E.R., C.P., K.-H.J., L.D., J.G., M.A.R. and B.M.W.; visualization, M.S.; supervision, B.M.W., M.A.R., and B.W.; project administration, K.L., E.R., B.M.W. and M.A.R.; funding acquisition, B.M.W., M.A.R., C.P., and B.W. All authors have read and approved the published version of the manuscript.

Competing interests

The authors declare that they have no competing interests.

Funding

This study is funded by the German Federal Ministry for Education and Research (grant numbers: 01GL1851D, 01GL1751B, 01GL1751A, 01GL1751C). The funder had no role in the study design, the data collection, and analysis, the decision to publish, or the preparation of the manuscript.

Data sharing statement

There are no plans to grant access to the full protocol, participant-level dataset, or statistical code as data contain potentially identifying information.

Ethics approval

The study complies with the ethical principles of the World Medical Association Declaration of Helsinki.

The study was approved first by the Ethics Committee of the Medical Faculty of the University of Bonn (reference number: 057/19, date of approval: 20/02/2019). In addition, the Ethics Committees of the Medical Association Nordrhein (Lfd-Nr.: 2019107), and of the Medical Faculty, University Hospital of Tuebingen (project no.: 446/2019BO2) approved the study protocol.

All participating practice team members received written information and signed informed consent forms, which are stored at the Institute for General Practice and Family Medicine, University of Bonn.

REFERENCES

1 Kringos D, Boerma W, Bourgueil Y, et al. The strength of primary care in Europe: an international comparative study. *British Journal of General Practice* 2013;63(616):e742-e750.

2 Rohde J, Cousens S, Chopra M, et al. 30 years after Alma-Ata: has primary health care worked in countries? *The Lancet* 2008;372(9642):950–61.

3 Swanwick T, Varnam R. Leadership development and primary care. *BMJ Leader* 2019;3(2).

4 Meredith LS, Bouskill K, Chang J, et al. Predictors of burnout among US healthcare providers: a systematic review. *BMJ Open* 2022;12(8):e054243.

5 Boamah SA, Spence Laschinger HK, Wong C, et al. Effect of transformational leadership on job satisfaction and patient safety outcomes. *Nurs Outlook* 2018;66(2):180–89. doi:10.1016/j.outlook.2017.10.004 [published Online First: 23 November 2017].

6 Goetz K, Campbell S, Broge B, et al. Job satisfaction of practice assistants in general practice in Germany: an observational study. *Fam Pract* 2013;30(4):411–17. doi:10.1093/fampra/cmt015 [published Online First: 19 April 2013].

7 i Solà GJ, i Badia JG, Hito PD, et al. Self-perception of leadership styles and behaviour in primary health care. *BMC Health Services Research* 2016;16.

8 Kersting C, Zimmer L, Thielmann A, et al. Chronic stress, work-related daily challenges and medicolegal investigations: a cross-sectional study among German general practitioners. *BMC Fam Pract* 2019:143. doi:10.1186/s12875-019-1032-6 [published Online First: 24 October 2019].

9 Wallace JE, Lemaire JB, Ghali WA. Physician wellness: a missing quality indicator. *The Lancet* 2009;374(9702):1714–21.

10 Wong CA, Cummings GG, Ducharme L. The relationship between nursing leadership and patient outcomes: a systematic review update. *J Nurs Manag* 2013;21(5):709–24.

11 Sfantou DF, Laliotis A, Patelarou AE, et al. Importance of Leadership Style towards Quality of Care Measures in Healthcare Settings: A Systematic Review. *Healthcare (Basel, Switzerland)* 2017;5(4):73.

12 Söling S, Pfaff H, Karbach U, et al. How is leadership behavior associated with organization-related variables? Translation and psychometric evaluation of the implementation leadership scale in German primary healthcare. *BMC Health Services Research* 2022;22(1):1–13.

13 Bass BM. Two Decades of Research and Development in Transformational Leadership. *European Journal of Work and Organizational Psychology* 1999;8(1):9–32.

14 Poethke U. Leadership, values and communication: a cross-cultural investigation of the extended full-range of leadership behaviors: Doctoral dissertation 2019.

15 Rowold J, Schlotz W. Transformational and transactional leadership and followers' chronic stress. *Leadership Review* 2009;9(1):35–48.

16 Graen GB, Uhl-Bien M. Relationship-based approach to leadership: Development of leader-member exchange (LMX) theory of leadership over 25 years: Applying a multi-level multi-domain perspective. *Leadership Quarterly* 1995;6(2):219–47.

17 Graen GB, Uhl-Bien M. Relationship-based approach to leadership: Development of leader-member exchange (LMX) theory of leadership over 25 years: Applying a multi-level multi-domain perspective. *The Leadership Quarterly* 1995;6(2):219–47.

18 Gregersen S, Vincent-Höper S, Nienhaus A. Health-Relevant Leadership Behaviour: A Comparison of Leadership Constructs. *German Journal of Human Resource Management* 2014;28(1-2):117–38.

19 Rowold J, Poethke U. FIF: Fragebogen zur Integrativen Führung Manual, 1st edn. Göttingen: Hogrefe 2017.

20 Lockyer J. Multisource feedback in the assessment of physician competencies. *J Contin Educ Health Prof* 2003;23(1):4–12.

21 Davis DA, Mazmanian PE, Fordis M, et al. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA* 2006;296(9):1094–102.

22 Donnon T, Al Ansari A, Al Alawi S, et al. The reliability, validity, and feasibility of multisource feedback physician assessment: a systematic review. *Acad Med* 2014;89(3):511–16.

23 Saravo B, Netzel J, Kiesewetter J. The need for strong clinical leaders - Transformational and transactional leadership as a framework for resident leadership training. *PLoS One* 2017;12(8):e0183019.

24 Richter A, Lornudd C, Thiele Schwarz U von, et al. Evaluation of iLead, a generic implementation leadership intervention: mixed-method preintervention-postintervention design. *BMJ Open* 2020;10(1):e033227. doi:10.1136/bmjopen-2019-033227 [published Online First: 12 January 2020].

25 Hill DA, Jimenez J-C, Cohn SM, et al. How to be a leader: a course for residents. *Cureus* 2018;10(7).

26 Viehmann A, Kersting C, Thielmann A, et al. Prevalence of chronic stress in general practitioners and practice assistants: Personal, practice and regional characteristics. *PLoS One* 2017;12(5):e0176658.

27 Dreher A, Theune M, Kersting C, et al. Prevalence of burnout among German general practitioners: Comparison of physicians working in solo and group practices. *PLoS One* 2019;14(2):e0211223.

28 Freund T, Everett C, Griffiths P, et al. Skill mix, roles and remuneration in the primary care workforce: who are the healthcare professionals in the primary care teams across the world? *International journal of nursing studies* 2015;52(3):727–43.

29 infas im Auftrag der KBV und des NAV-Virchow-Bund. Tabellenband Ärztemonitor 2018, Ergebnisse für Haus- und Fachärzte. Bonn 2018.

30 Weltermann BM, Kersting C, Pieper C, et al. IMPROVEjob - Participatory intervention to improve job satisfaction of general practice teams: a model for structural and behavioural prevention in small and medium-sized enterprises - a study protocol of a cluster-randomised controlled trial. *Trials* 2020;21(1):532.

31 Degen L, Linden K, Seifried-Dübon T, et al. Job Satisfaction and Chronic Stress of General Practitioners and Their Teams: Baseline Data of a Cluster-Randomised Trial (IMPROVEjob). *Int J Environ Res Public Health* 2021;18(18). doi:10.3390/ijerph18189458 [published Online First: 8 September 2021].

32 Ingold PV. Fragebogen zur integrativen Führung. *Zeitschrift für Arbeits- und Organisationspsychologie A&O* 2019;63(2):100–06.

33 Stuber F, Seifried-Dübon T, Rieger MA, et al. Investigating the Role of Stress-Preventive Leadership in the Workplace Hospital: The Cross-Sectional Determination of Relational Quality by Transformational Leadership. *Front Psychiatry* 2019;10:622. doi:10.3389/fpsyt.2019.00622 [published Online First: 3 September 2019].

34 Schyns B. Überprüfung einer deutschsprachigen Skala zum Leader-Member-Exchange-Ansatz. *Zeitschrift für Differentielle und Diagnostische Psychologie* 2002.

35 Schyns B, Paul T. Leader-Member Exchange Skala (LMX7): ZIS - GESIS Leibniz Institute for the Social Sciences 2014.

36 Sargeant J, Mann K, Ferrier S. Exploring family physicians' reactions to multisource feedback: perceptions of credibility and usefulness. *Med Educ* 2005;39(5):497–504.

37 Hu J, Lee R, Mullin S, et al. How physicians change: Multisource feedback driven intervention improves physician leadership and teamwork. *Surgery* 2020;168(4):714–23. doi:10.1016/j.surg.2020.06.008 [published Online First: 6 August 2020].

38 Northouse PG. Leadership: Theory and practice, 6th edn. Thousand Oaks: Sage Publications 2010.

39 Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd edn. Hoboken: Taylor and Francis 1988.

40 Halverson SK, Tonidandel S, Barlow C, et al. Self-other agreement on a 360-degree leadership evaluation. In: 17th Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, Canada 2002.

41 Furnham A, Stringfield P. Congruence of self and subordinate ratings of managerial practices as a correlate of supervisor evaluation. *Journal of Occupational and Organizational Psychology* 1994;67(1):57–67.

42 Fleenor JW, Smither JW, Atwater LE, et al. Self–other rating agreement in leadership: A review. *The Leadership Quarterly* 2010;21(6):1005–34.

43 Antonioni D, Park H. The effects of personality similarity on peer ratings of contextual work behaviors. *Personnel Psychology* 2001;54(2):331–60.

44 Alrubaysh MA, Alshehri MH, Alsuhaibani EA, et al. The leadership styles of primary healthcare center managers and center performance outcomes in Riyadh, Saudi Arabia: A correlational study. *Journal of Family & Community Medicine* 2022;29(1):56.

45 Erben J, Schneider FM, Maier M. In the Ear of the Beholder: Self-Other Agreement in Leadership Communication and Its Relationship With Subordinates' Job Satisfaction. *International Journal of Business Communication* 2019;56(4):505–29.

46 Cleary S, Du Toit A, Scott V, et al. Enabling relational leadership in primary healthcare settings: lessons from the DIALHS collaboration. *Health Policy and Planning* 2018;33(suppl_2):ii65-ii74.

47 Chen C-HV, Wang S-J, Chang W-C, et al. The effect of leader-member exchange, trust, supervisor support on organizational citizenship behavior in nurses. *Journal of Nursing Research* 2008;16(4):321–28.

48 Gregersen S, Vincent-Höper S, Nienhaus A. Job-related resources, leader–member exchange and well-being – a longitudinal study. *Work & Stress* 2016;30(4):356–73.

49 Vincent-Höper S, Stein M, Gregersen S, et al. Messung gesundheitsförderlicher Führung in Gesundheitsdienst und Wohlfahrtspflege. *ASU Arbeitsmed Sozialmed Umweltmed* 2018(53):46–53.

Additional Files

Additional file 1: CONSORT 2010 checklist of information to include when reporting a randomised trial

CONSORT 2010 checklist of information to include when reporting a randomised trial*

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	n/a
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	Abstract, 2
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	Background, 4-5
	2b	Specific objectives or hypotheses	Background, 4-5
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	Methods, 5
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	n/a
Participants	4a	Eligibility criteria for participants	Methods, 5-6
	4b	Settings and locations where the data were collected	Methods, 5-6
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	n/a
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	Methods, 6-7
	6b	Any changes to trial outcomes after the trial commenced, with reasons	n/a
Sample size	7a	How sample size was determined	Study protocol, Weltermann et al., Trials 2020
	7b	When applicable, explanation of any interim analyses and stopping guidelines	n/a

Randomisation:			
Sequence generation	8a	Method used to generate the random allocation sequence	Details in Study protocol, Weltermann et al., Trials 2020
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	Details in Study protocol, Weltermann et al., Trials 2020
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	n/a
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	Details in Study protocol, Weltermann et al., Trials 2020
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	n/a
	11b	If relevant, description of the similarity of interventions	n/a
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	Methods,7-8
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	n/a
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	n/a
	13b	For each group, losses and exclusions after randomisation, together with reasons	n/a
Recruitment	14a	Dates defining the periods of recruitment and follow-up	Details in Study protocol, Weltermann et al., Trials 2020
	14b	Why the trial ended or was stopped	Details in Study protocol,

			Weltermann et al., Trials 2020
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	Results, table 1, Page 8
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	Results, 9
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	Results, 9-11
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	N/A.
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	n/a
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	n/a
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	Discusssion, Strengths and limitatoins, Page 12-13
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	13
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	Discussion, 11-13
Other information			
Registration	23	Registration number and name of trial registry	Trial registration, 2
Protocol	24	Where the full trial protocol can be accessed, if available	Study protocol, Weltermann et al., Trials 2020
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	Funding, 17-18

*We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org.