METHODS ARTICLE

Development and psychometric assessment of a novel survey to measure care coordination from the specialist's perspective

Varsha G. Vimalananda MD, MPH^{1,2} | Benjamin Graeme Fincke MD^{1,3} | Shirley Qian MS¹ | Molly E. Waring PhD^{1,4} | Ryan G. Seibert MD, MS⁵ | Mark Meterko PhD^{3,6}

¹Center for Healthcare Organization and Implementation Research (CHOIR), Edith Nourse Rogers Memorial VA Medical Center, Bedford, Massachusetts

²Section of Endocrinology, Diabetes and Metabolism, Boston University School of Medicine, Boston, Massachusetts

³Department of Health Law, Policy and Management, Boston University School of Public Health, Boston, Massachusetts

⁴Department of Allied Health Sciences, College of Agriculture, Health, and Natural Resources, University of Connecticut, Storrs, Connecticut

⁵Section of General Internal Medicine, Boston University School of Medicine, Boston, Massachusetts

⁶VHA Office of Reporting, Analytics, Performance, Improvement and Deployment (RAPID – 10EA), Field-based at the Edith Nourse Rogers Memorial VA Medical Center, Bedford, Massachusetts

Correspondence

Varsha G. Vimalananda, MD, MPH, Center for Healthcare Organization and Implementation Research (CHOIR), Edith Nourse Rogers Memorial VA Medical Center, Bedford, MA.
Email: Varsha.vimalananda@va.gov

Funding information

This work was funded by a Career Development Award (CDA 15-070-3) from the Department of Veterans Affairs, Health Services Research & Development Service (V.V.).

Objective: To develop an online survey of care coordination with primary care providers as experienced by medical specialists, evaluate its psychometric properties, and test its construct validity.

Data Sources: Physicians (N = 633) from 13 medical specialties across the Veterans Health Administration.

Study Design: We developed the survey based on prior work (literature review, specialist interviews) and by adapting existing measures and developing new items. Multitrait scaling analysis and confirmatory factor analysis were used to assess scale structure. We used multiple linear regression to examine the relationship of the final coordination scales to specialists' overall experience of care coordination.

Data Collection: November 2016–December 2016.

Principal Findings: Results suggest a 13-item, four-factor survey [Relationships ($k = 4$), Roles and Responsibilities ($k = 4$), Communication ($k = 3$), and Data Transfer ($k = 2$)] that measures the medical specialist experience of coordination with good internal consistency reliability, convergent validity, discriminant validity, and goodness of fit. Together, the four scales explained nearly 50 percent of the variance in specialists' overall experience of care coordination.

Conclusions: The 13-item Coordination of Specialty Care—Specialist Survey (CSC-Specialist) is the first of its kind. It can be used alone or embedded in other surveys to measure four domains of care coordination as experienced by medical specialists.

KEYWORDS

consultation, coordinated care, psychometrics, specialty care, survey research

1 | INTRODUCTION

Over 100 million specialty care visits occur yearly in the United States.¹ They commonly result in fragmented health care involving missed and unmet needs,² duplicated tests,^{3,4} medication errors,⁵ and patient confusion.⁶ Risks increase exponentially when patients have more sources of medical care,⁷ putting sicker patients

at greatest risk and increasing costs.² Preventing these outcomes through care coordination is a cornerstone of many efforts to improve the quality of health care.

Care coordination is defined as “the deliberate organization of patient care between two or more participants... to facilitate the appropriate delivery of health care services and account for each other's actions.”^{8,9} Organizing care in this way involves marshaling

personnel and other resources, and managing the exchange of information among participants responsible for different aspects of care.

Health care providers' reports of care coordination have been empirically related to variations in patient outcomes. Better coordination among inpatient multidisciplinary care teams as measured by provider surveys is related to reduced postoperative pain.¹⁰ Nursing home aide-reported coordination is associated with better nursing home resident quality of life,¹¹ registered nurse (RN)-reported coordination is associated with fewer medical errors and hospital-acquired infections,¹² and better coordination among surgeons, medical specialists, and allied health professionals is associated with reduced lower extremity amputation rates among patients with diabetes.¹³ Patient-reported problems with coordination are associated with greater joint pain after knee replacement,¹⁴ while better patient-reported coordination between primary care providers (PCPs) and specialists is linked to lower emergency department and outpatient care utilization.¹⁵ These links between the experience of coordination and meaningful clinical and cost outcomes make a compelling case for the importance of valid and reliable measures of coordination.

The relationship of provider-reported specialty care coordination to outcomes has not been well described. To examine these Relationships, appropriate measures are needed. In the case of specialty care referrals, important information flows in both directions between all three participants in the "specialty care triad": specialist, PCP, and patient. Each member of this triad has different needs and priorities for coordination success. Therefore, measures of coordination must account for the perspective of each triad member. There are existing coordination measures that focus on the PCP or patient perspective, but there is no valid and reliable measure of care coordination from the medical specialist perspective. Furthermore, existing measures omit some concepts that are both unique and central to specialty care coordination. For example, referrals have been called "the link between primary and specialty care"¹⁶ but measures of care coordination do not include assessments of referral completeness or appropriateness. These shortcomings in existing measures limit the scope and usefulness of current evaluations of specialty care coordination.

The present study had two major goals: (a) to adapt content from existing measures and develop new items as needed to create a new and comprehensive measure of care coordination that is particular to the experience of the medical specialist; and (b) to evaluate the measure's psychometric properties, including a preliminary assessment of construct validity. We conducted this study in the Veterans Health Administration (VA), the largest integrated health care system in the United States. VA providers all share a single electronic medical record nationwide, but this advantage has by no means resolved all the challenges to VA primary care-specialty care coordination^{17,18} and those that remain closely mirror those in nonintegrated health care settings.^{2,16,19-24}

2 | METHODS

The Bedford VA Medical Center Institutional Review Board approved this study.

2.1 | Instrument development

2.1.1 | Conceptual framework

To guide our work, we adapted the model developed by McDonald et al²⁵ that was adopted by the Agency for Healthcare Research and Quality (AHRQ) and is prominent in care coordination research. A central principle of that model is that the impacts of any efforts to achieve care coordination will differ depending on the perspective of each party experiencing those effects. The McDonald model describes the perspectives of patient, PCP, and system. Based on our prior qualitative research,²⁶ we modified this framework by replacing "system" with "specialist" and thereby focused the model on considerations specific to the coordination of specialty care.

To then identify the key features of coordination from the medical specialist perspective, we used the AHRQ definition of coordination, literature review, and findings from our previous qualitative study that included interviews with specialists as members of the specialty care triad.²⁶ This extensive preliminary work resulted

Construct	Characteristics
Strong Relationships with referring primary care providers	Specialists and PCPs have mutually respectful Relationships and collaborate well.
Clear Roles and Responsibilities	Specialists and PCPs are clear on and agree on their Responsibilities in the referral process. PCPs provide clear, complete consult requests, complete adequate workup before placing referral, ensure patients know the reason for referral, and follow through on specialists' recommendations.
Clear and timely Communication	Specialists are able to reach PCPs and their team in a timely manner for help.
Timely and accurate Data Transfer	Specialists have access to relevant and correct data from the PCP at the time of the patient visit.

TABLE 1 Conceptual framework of specialty care coordination from the medical specialist perspective

in the identification of four constructs that provided the conceptual framework for further survey development (Table 1). These constructs are as follows: strong Relationships, clear Roles and Responsibilities, clear and timely Communication, and timely and accurate Data Transfer.

2.1.2 | Establishing a gap in existing surveys

To determine whether any existing surveys contained items and/or scales that might be used to measure coordination from the specialist perspective, we examined coordination surveys listed in the 2010⁸ (and the 2014 updated)²⁷ AHRQ Care Coordination Atlas, a series of publications from the National Quality Forum,²⁸ the Consumer Assessment of Healthcare Providers and Systems family of surveys,²⁹ the American College of Physicians Neighbor³⁰ and National Committee on Quality Assurance tools,³¹ and a 2013 systematic review of care coordination surveys.³² Some surveys were included in more than one resource.

We focused on surveys that approached coordination from the physician perspective and: (a) were applicable to adult outpatient care, (b) were independent of a patient's diagnosis, (c) evaluated whether the respondent experienced coordination to be successful, and (d) provided data on reliability and validity. Of the 101 surveys we evaluated, we excluded from further review 56 that took a patient or system perspective on coordination, 27 that were not applicable to adult outpatient care, six that were not independent of a patient's diagnosis, four that measured the presence of mechanisms intended to improve care rather than the success of such mechanisms, and three that did not have evidence of reliability or validity testing. This resulted in five surveys from the PCP perspective. No surveys from the specialist perspective met our four criteria, thereby confirming a gap in available instruments for the assessment of specialty care coordination. The five surveys from the PCP perspective that met our criteria provided a total of 279 candidate items for our new measure.^{10,33-36}

Four of the five surveys were developed to assess multiple aspects of coordination relevant to primary care providers; however, many of these are not relevant to specialist-PCP coordination. For example, the PCAT includes 153 items. PCAT topics include directly what services are available at the PCP practice site, preventive care issues discussed, family centeredness, and community orientation. Across the five surveys, we identified 78 items potentially relevant to specialty care coordination.

2.1.3 | Selection and modification of candidate items and scales

Whenever possible, we aimed to populate the four constructs of our conceptual framework by adapting items from the five PCP surveys. One investigator (V.V.) used qualitative data analysis software (NVIVO³⁷) to map relevant candidate items from each of the five PCP surveys to the constructs of our framework. Of the 78 potentially relevant items, 59 mapped to one or more of those constructs.

The 19 candidate items which could not be assigned to a construct in our conceptual framework were excluded from further consideration. Three investigators (V.V., G.F., and M.M.) independently reviewed the 59 mapped candidate items (and companion items if they were part of a multi-item scale in the original survey). The investigators rated the clarity of each item and, separately, its relevance to its assigned construct (1-5 scales). As a group, we reached consensus on the best candidate items for each construct. No existing scales were applicable in their entirety to the specialist perspective and therefore we retained the best individual items, adapting language to the specialist perspective.

At the end of this process, some framework constructs were represented by only one or two items adapted from existing instruments. We developed new items as necessary to assure there were at least four items for each construct. In aiming for at least four items per construct, our goal was to balance brevity in the final measure with comprehensive coverage of the construct as well as having sufficient candidate items to form multi-item scales even if psychometric analyses suggesting dropping some items. The response scale for all items was 1: Never; 2: Rarely—less than 10 percent of the time; 3: Occasionally—about 30 percent of the time; 4: Sometimes—about 50 percent of the time; 5: Frequently—about 70 percent of the time; 6: Usually—about 90 percent of the time; and 7: Always.³⁸

2.1.4 | Item refinement

A nine-member expert panel, which included VA specialists and PCPs, reviewed the candidate items and response scales. They independently rated the clarity and, separately, the relevance of each item to its assigned construct. Panel members also rated the appropriateness of response options for each item. For each criterion (clarity, relevance, and appropriateness), the rating options were "Acceptable" or "Needs Improvement." When "Needs Improvement" was selected, panel members were asked to provide details about the nature of the shortcoming and how it might be addressed. The candidate items were then revised based on this feedback and then re-reviewed by the expert panel to confirm that they met our three criteria for inclusion. To ensure face validity, applicability to multiple medical specialties, and future adaptability for use in non-VA settings, 12 VA and non-VA specialists independently reviewed the draft measure for item relevance and coverage of issues related to coordination for their respective fields. This process included four VA specialists (Neurology, Pulmonology, Infectious Diseases, and Endocrinology) and eight non-VA specialists (Cardiology, Dermatology, Endocrinology, Gastroenterology, Hematology/Oncology, Infectious Diseases, Nephrology, and Rheumatology). No new issues were raised.

We then conducted two rounds of cognitive interviews with four VA and two non-VA medical specialists using a standardized protocol that took a retrospective approach with targeted probes.^{39,40} Specifically, informants were asked to read a hard copy version of the survey, provide their interpretation of selected items, and comment on any items that were unclear. Revised and new items were

tested for clarity and appropriate response scales in a second round of cognitive interviews.

The resulting draft Coordination of Specialty Care—Specialist Survey (CSC-Specialist) included 14 items about different aspects of coordination. Our hypothesized scale structure included all 14 items and comprised 4 multi-item scales measuring specialist-PCP coordination.

2.2 | Survey administration

We developed a survey that included the 14 candidate items for the CSC-Specialist, demographic and practice characteristics, two questions each about organizational aspects of coordination and coordination with non-VA specialists, two questions about overall coordination, and an item to capture the degree to which referrals to the respondent relate to procedures. We also included a three-item measure of burnout and a three-item measure of job satisfaction⁴¹ because they represented factors hypothesized to be related to specialist's perceptions of coordination. Finally, we included several items about specific mechanisms to coordinate care and specialists' use and perception of those mechanisms as helpful for coordination. This work resulted in a 51-item survey that included the 14-item draft CSC-Specialist.

2.2.1 | Pilot administration

We conducted a pilot study to examine patterns of item nonresponse and identify problems with item clarity, instructions, or logistical issues. The survey was administered using SurveyMonkey (SurveyMonkey Inc., San Mateo, CA).

We used a three-pronged approach to specialist recruitment. First, we obtained letters of support from five specialty section chiefs known to the investigators, representing 54 medical specialists from five different VA facilities. Chiefs provided an up-to-date list of email addresses and sent informational emails about the survey to their section members that encouraged them to consider responding. We followed this email within 36 hours with a personalized survey invitation and link. Three reminders were sent, one every 3-5 days. Second, we obtained a list of specialists from VA Managerial Cost Accounting (MCA). We selected a random sample of 200 email addresses and directly emailed those providers with a personalized survey invitation and link. Reminders were sent as for the letter of support approach. Third, a VA dermatologist posted information about the survey on the VA dermatology listserv, which has about 138 subscribers. We followed this information with a post that included the survey invitation and link that was not personalized. There were no reminders posted in this third approach.

All email addresses used in the pilot administration were removed from the list of potential participants in the large-scale administration. We also excluded all dermatologist email addresses from the list of potential participants in the large-scale administration. These steps ensured that no email address would be sent a survey invitation twice.

2.2.2 | Large-scale administration

We administered the online survey in two phases to a national sample of VA medical specialists in November-December 2016. First, we used the letter of support approach in collaboration with 33 section chiefs (representing 261 specialists). We then randomly sampled 1388 specialists (excluding those contacted with the letter of support approach) from Veterans Integrated Service Networks (VISNs) where we had few or no respondents to date, with the goal of maximizing geographic variation in the sample. The survey software removes identifiers to create an anonymous analytic dataset.

For the multitrait analysis (MTA), 180 completed surveys were required to achieve a power of 0.80 to identify a medium effect (difference of 0.30 between item-to-hypothesized scale correlations and the item-to-other scale correlations) at an alpha of 0.05. As we anticipated that not all respondents would provide complete data, the target sample size was 500 respondents.

2.3 | Statistical analysis

Analyses were conducted using SAS 9.4 (SAS Institute, Cary, NC).

2.3.1 | Methodological assessment

First, we examined the response rates associated with the different recruitment strategies used, the time required to complete the 51-item questionnaire, and various indicators of data quality including item missing data rates, and score distributions as described by measures of central tendency, floor and ceiling effects, and skew and kurtosis, from the pilot study and large-scale administration separately.

2.3.2 | Psychometric assessment

Next, we combined data from the pilot study and large-scale administration. We examined item descriptive statistics in the entire eligible sample. For the psychometric assessment, we checked our hypothesized scale structure using multitrait scaling analysis (MTA) and followed this with a confirmatory factor analysis (CFA) to obtain more specific fit statistics.

We conducted the MTA using our hypothesized four-scale structure. Following the recommendations of Ware, we used only the subset of eligible respondents who had complete data on at least half of the items in the four hypothesized scales (half-scale criterion), and estimated values for unanswered items based on the within-subject average of answered items on that scale.⁴² The MTA yields Cronbach's alpha coefficient as an estimate of scale reliability, item-to-hypothesized scale (HS) correlations as estimates of convergent validity, and comparisons of each item's HS correlation to its correlations with all other scales (OS) as indications of discriminant validity. Convergent validity evaluates the degree to which items are correlated with their assigned scale, and ≥ 0.40 is evidence of good convergence. Discriminant validity is a measure of the degree to which

items are significantly more strongly correlated with their assigned versus any other scale, with a higher percent of items meeting this criterion being more desirable. Based on empirical findings and conceptual considerations, we reassigned and dropped items to improve the psychometric properties and conceptual clarity of the scales. We then repeated the MTA on the revised item-to-scale model in the same sample.

Next, we conducted a CFA to obtain more precise empirical estimates of the robustness of the proposed refined item-to-scale structure. We used only the sample of respondents that had complete data on all items in the scales. Following recommended practice,⁴³ we evaluated model fit by examining the preponderance of evidence across multiple measures. We computed (a) the overall chi-square as an index of absolute fit; (b) the standardized root mean residual (SRMR), another absolute measure of goodness of fit; (c) the root mean square error of approximation (RMSEA), which provides a goodness-of-fit measure rewarding parsimony; and (d) the comparative fit index (CFI) and Tucker-Lewis fit index (TLI), both measures of comparative fit. The SRMR estimates the average discrepancy between the observed and predicted model covariances, with a possible range of 0.0-1.0 and a value of <0.08 indicating good fit.⁴⁴ The RMSEA estimates model residuals, with a value of ≤ 0.08 indicating reasonable fit and <0.06 indicating good fit.⁴⁴ The CFI and TLI metrics both compare the observed pattern of item Relationships to models assuming that the items are not correlated. For both statistics, ≥ 0.95 indicates good fit and 0.90-0.95 indicates reasonable fit.

As a preliminary examination of construct validity, we conducted a multiple linear regression analysis using the final proposed CSC-Specialist scales as the predictors of overall specialist-PCP coordination. We restricted the sample to respondents with complete data on all variables in the model. The outcome measure was provided by a single survey item which asked, "All things considered, how would you rate the overall quality of specialty care coordination between you and referring primary care providers over the last 3 months?" Response options for this item ranged from 0 (worst coordination possible) to 10 (best coordination possible). The mean response for overall coordination was 6.3 (standard deviation = 1.9) with a generally normal distribution, so we modeled it as a linear outcome variable.

Given that data on both the predictors and the outcome were collected with the same instrument in a cross-sectional design, common method variance could bias the observed Relationships between the variables. To minimize this effect, we first controlled for other factors that might either (a) affect perceptions of overall coordination (eg, certain organizational characteristics and overall job satisfaction) or (b) affect survey response patterns in general (eg, age or gender), or (c) both. Unlike objective measures such as age, gender, and clinic sessions, job satisfaction and burnout are measures of attitudes that may be confounded with perceptions of coordination as a consequence of their shared measurement methodology and/or a halo effect from the respondent's global affect (positive or negative) toward their work situation. Our premise was that, if, after controlling for satisfaction and burnout, the

coordination scales were still significantly related to the outcome of overall coordination, then it would suggest that the scales possess considerable signal strength over and above whatever noise may be associated with collinearity from various sources among the predictors. We did not have further specific theory-based hypotheses about the Relationships of these predictors to the outcome, and therefore, the direction of causality was not a factor in our choice of predictors.

In our final model, we included demographics (gender, age, years in VA); practice characteristics (number of VA half-day outpatient clinic sessions weekly, percent of referrals related to procedures); job satisfaction (degree of satisfaction with people worked within the clinical setting; degree of satisfaction with organization); and burnout (frequency of feeling burned out from work, frequency of accomplishing many worthwhile things in job). We took a sequential approach to building the final model, adding potential predictors in blocks: demographics, then practice characteristics, job satisfaction, and finally burnout. We then added the four coordination scales, allowing SAS to select the order in which they were included in the model on the basis of their strength of association with the outcome.

2.3.3 | Reliability assessment

To assess the extent to which our measures could inform facility-level comparisons, we calculated one-way random intraclass correlation coefficients [ICC(1)s] to obtain an estimate of the reliability of facility-level means for each of the four scales. We then used the Spearman-Brown prophecy formula to solve for the number of respondents needed to obtain facility-level reliability of 0.70 for each scale.

3 | RESULTS

3.1 | Methodological assessment

During the pilot administration, the response rate varied by recruitment strategy. The letter of support approach netted a 48 percent response rate (26/54). The random sampling approach resulted in an 18 percent response rate (36/200), similar to the 17 percent response rate in a recent web survey of VA PCPs that used a similar recruitment approach.⁴⁵ The colleague referral approach resulted in a 27 percent response rate (37/138).

The median time to complete the full 51-item survey was 12 minutes (range 5-22 minutes); 90 percent of respondents completed all items. We observed low item missing rates, no problems with ceiling or floor effects, and values for skew and kurtosis were not high with the exception of two items on which some skew was anticipated (both related to having access to patient data, which is facilitated in VA by the universal EMR). These results did not indicate the need to modify survey content in preparation for the broad administration.

In the large-scale administration, the response rate using letters of support was 46 percent (120/261). Random sampling netted a 22 percent (414/1900) response rate.

3.2 | Sample characteristics

Overall, 633 of 2533 specialists responded (25 percent). Of those, 24 were disqualified automatically by skip logic in the survey before answering any coordination questions based on reporting a professional discipline of "Other," or a primary specialty of "Primary Care" or "Other." Of the remaining 609, 31 were excluded from the analytic sample due to missing data for professional discipline or specialty, or reporting no VA outpatient specialty clinic duties. This resulted in an analytic sample of 578 providers from 13 specialties (Table 2).

3.3 | Psychometric assessment

Item descriptive statistics are reported in Table 3. For item Q12, the score was reversed so that a higher score indicated a higher level of care coordination, consistent with the other items. Responses showed adequate variation; the top box percent ranged from 7.6 percent to 20.0 percent among the three items with the highest such percentage. Missing data followed predictable patterns related to screener questions.

For the MTA, we began with the 4 hypothesized scales related to specialist-PCP coordination: Relationships, Roles and Responsibilities, Communication, and Data Transfer. Applying the half-scale criterion resulted in an analytic sample of 453 (78 percent of the 578 eligible respondents). The MTA identified one item originally assigned to the Roles and Responsibilities scale (frequency with which PCPs follow through on specialist recommendations) that was significantly more highly correlated with the Relationships scale, and that item was reassigned accordingly. A second item originally assigned to the Roles and Responsibilities scale (frequency with which specialists notify PCPs prior to placing further consults for referred patients) was found to have no significant correlations with any of the scales and was omitted from the next round of MTA. We tested the refined 4-scale model in a repeat MTA in the same sample. Final item-to-scale correlations for this model are in Table 3.

We found strong support for the refined scale structure in the patterns of convergent and divergent correlations. One hundred percent of items met criteria for convergent validity. One hundred percent of items were more highly correlated with their assigned than with other scales; for 90 percent of items, the difference was statistically significant. All 4 scales demonstrated adequate internal consistency reliability with Cronbach's alphas of ≥ 0.80 , which is above conventional standards of >0.70 .

Table 4 reports basic descriptive statistics for the four scales, including their means, standard deviations, percent at ceiling (mean score ≥ 6.5), and the observed range of responses. The mean score and percent at ceiling were lowest for Roles (4.95 ± 0.96 ; 2.6 percent). As expected in a health care system with a shared EMR, the mean score and percent ceiling (scale score ≥ 6.5) were highest for Data Transfer (5.57 ± 1.04 ; 18.7 percent).

Table 5 includes the scale internal consistency reliabilities (Cronbach's alphas) in the diagonal entries and the correlations

between scales in the off-diagonal entries. For each scale, the alpha coefficients were higher than the interscale correlations, supporting the coherence and integrity of the proposed scales. The highest interscale correlation was only 0.66, which suggests about 44 percent shared variance and is substantially lower than the 0.80 correlation criterion for substantial overlap.⁴⁶ Overall, this pattern suggests that the proposed scales measure related but distinguishable factors.

To further assess the robustness of the final model, we conducted a CFA, which yields more empirical measures of goodness of fit than does MTA. We included the 273 specialists who provided complete data on all the items being tested. The preponderance of evidence from the CFA strongly supported the proposed four-scale model. As is often the case,⁴⁷ the chi-square test was statistically significant, rejecting the strict null hypothesis of exact fit (chi-square = 134, $P = <0.0001$). However, other key fit statistics consistently supported the proposed model: SRMR = 0.05, RMSEA = 0.069 (95% confidence limit 0.053-0.084), CFI = 0.95, TLI = 0.94. A post hoc assessment based on the confidence interval of the RMSEA⁴⁸ indicated power of 0.94 at $\alpha = 0.05$.

As a preliminary assessment of construct validity, we examined the association between the CSC-Specialist and overall PCP-specialist coordination among respondents with complete data on all predictors ($N = 361$). With demographics, practice characteristics, job satisfaction, and burnout included, the model explained 22 percent of the variance in overall coordination (see online Appendix S1: Table 6, titled "Predictors of overall coordination: standardized regression coefficients and 95% confidence intervals from sequential block regression models"). After adding the four coordination scales, the percent of variance in overall specialist-PCP coordination explained increased to 70 percent ($F(13, 347) = 64.24$, $P < 0.0001$). Three of the new coordination scales were significant predictors with standardized betas of Relationships, 0.40; Roles and Responsibilities, 0.25; and Communication, 0.20; all $P < 0.0001$. These latter two scales were significant after controlling for the other coordination scales previously entered into the model, further supporting their representation of separate aspects of coordination-related experience. Data Transfer was not significant ($b = 0.07$, $P = 0.07$).

We calculated one-way random intraclass correlation coefficients to obtain an estimate of the reliability of facility-level means for each of the four scales. For this calculation, we used data from the 20 facilities from which we received at least eight independent responses. This analysis revealed modest but significant facility-level reliability estimates for the Relationships [ICC(1) 0.59, $P = 0.002$], Roles and Responsibilities [ICC(1) 0.46, $P = 0.02$], and Communications scales [ICC(1) 0.48, $P = 0.02$]. The ICC for the Data Transfer scale was lower and not significant [ICC(1) 0.17, $P = .26$]. Using the Spearman-Brown prophecy formula to solve for the number of respondents needed to obtain facility-level reliability of 0.70, we found the following required sample size estimates: Relationships, $n = 13$; Roles and Responsibilities, $n = 22$; Communication, $n = 20$; and Data Transfer, $n = 91$.

TABLE 2 Characteristics of VA specialist providers (N = 578)^a

	N (%)
Professional discipline (N = 578)	
Physician	557 (96.4)
Physician Assistant	14 (2.4)
Nurse Practitioner	7 (1.2)
Specialty (N = 578)	
Allergy and/or Immunology	9 (1.6)
Cardiology	77 (13.3)
Dermatology	47 (8.1)
Endocrinology	41 (7.1)
Gastroenterology	68 (11.7)
Geriatrics	10 (1.7)
Hematology and/or Oncology	40 (6.9)
Infectious Disease	50 (8.6)
Nephrology	39 (6.7)
Neurology	74 (12.8)
Pulmonology and/or Critical Care	64 (11.1)
Rheumatology	36 (6.2)
Sleep Medicine	23 (4.0)
Half-day VA outpatient clinics weekly (N = 576)	
1-5	459 (79.7)
6-10	117 (20.3)
Hours per week salaried to work at VA (N = 541)	
Up to 10 h (up to 2/8ths)	25 (4.6)
From 10 up to 20 h (2/8ths up to 4/8ths)	32 (5.9)
From 20 up to 30 h (4/8ths up to 6/8ths)	81 (15.0)
From 30 up to 40 h (6/8ths up to 8/8ths)	399 (73.8)
No VA hours (contract, fee basis, other)	4 (0.7)
Percent of new consults related to procedures (performing, scheduling, or decision making) (N = 535)	
None	99 (18.5)
1%-25%	215 (40.1)
26%-50%	111 (20.7)
51%-75%	78 (14.6)
75%-100%	32 (6.0)
Number of other physicians, NPs, or PAs who see patients in your specialty department (N = 539)	
No other providers	29 (5.4)
1 other provider	55 (10.2)
2-4 other providers	199 (36.9)
5 or more other providers	256 (47.5)
Gender (N = 534)	
Female	210 (39.3)
Male	324 (60.7)
Age (N = 517)	
Less than 30 y	3 (0.6)
30-39 y	102 (19.7)

(Continues)

TABLE 2 (Continued)

	N (%)
40-49 y	146 (28.2)
50-59 y	124 (24.0)
60 y or older	142 (27.5)
Years worked as specialist in VA, including as a trainee (N = 540)	
Less than 1 y	16 (3.0)
Between 1 and 5 y	151 (28.0)
Between 5 and 10 y	124 (23.0)
More than 10 y	249 (46.1)

^aN for each item varies due to missing data.

The CSC-Specialist is available online as Appendix S2; scoring instructions are available online as Appendix S3. Additional information is available by contacting the first author.

4 | DISCUSSION

We developed and assessed the psychometric properties of a novel measure of medical specialists' perspective on care coordination with PCPs. The Coordination of Specialty Care—Specialist Survey (CSC-Specialist) fills a gap in available coordination measures. The CSC-Specialist can be used by researchers and managers to capture the specialist's assessment of four central elements of care coordination, assess the extent to which these elements are influenced by context and impact outcomes, and guide efforts to improve delivery of specialty care.

We found strong empirical support for a four-dimensional model of coordination. The final scale structure is highly consistent with existing literature on specialty care coordination and maps very closely to the a priori hypothesized scale structure. The scales measure the interactions between PCPs and their teams with specialists with respect to Relationships (four items), Roles and Responsibilities (four items), Communication (three items), and Data Transfer (two items). Nearly all measures of internal consistency reliability, discriminant validity, and goodness of fit exceeded conventional standards.

Strong evidence for construct validity of the final four scales was provided by estimating a multivariate linear regression model with an overall measure of specialist-PCP coordination as an outcome. The Relationships, Roles and Responsibilities, and Communication scales were each significantly associated with overall coordination, and their standardized regression coefficients were much higher than any of the estimates for demographics, practice characteristics, job satisfaction, or burnout. The variance explained by the model increased from 22 percent to over 70 percent when the coordination scales were added to the other predictors. The Data Transfer scale demonstrated the weakest association with overall coordination. We anticipate that this scale will demonstrate greater variance and predictive power in a setting where PCPs and specialists do not have

TABLE 3 Multitrait analysis of the Coordination of Specialty Care—Specialist Survey (CSC-Specialist): item descriptive statistics and item-to-scale correlations^{a,b}

	N	Mean	SD	Percentage top box	Relationships	Roles	Communication	Data Transfer
Relationships								
Q12. How often did you feel that you and the referring PCP had conflicting expectations of each other's Roles during the course of the specialty care you provided? (score reversed)	549	5.04	1.23	5.1	(0.62)	0.51	0.42	0.41
Q15. How often did you feel that you and the referring PCP worked well together in caring for patients?	549	4.83	1.16	5.3	(0.74)	0.61	0.70	0.47
Q16. How often did you feel that the referring PCP valued your contribution in caring for patients?	547	5.29	1.13	9.9	(0.62)	0.45	0.57	0.39
Q10. When you saw patients for follow up visits, how often had the referring PCP followed through on your recommendations for interim care?	467	4.50	1.19	1.3	(0.66)	0.52	0.48	0.42
Roles and Responsibilities								
Q5. How often was the reason for the consult request sufficiently clear, such that you understood what the referring PCP was asking of you?	573	4.85	1.10	2.8	0.45	(0.61)	0.32	0.37
Q6. How often did the consult request itself include sufficient clinical history and other information to meet your immediate needs?	569	3.65	1.28	1.1	0.40	(0.56)	0.32	0.23
Q7. How often did the referring PCP adequately evaluate the patient's condition prior to requesting the consult?	566	3.68	1.11	0.2	0.54	(0.60)	0.39	0.37
Q9. How often did consult requests reflect an understanding on the part of the PCP about what constitutes an appropriate referral to your specialty clinic?	562	4.53	1.06	1.1	0.65	(0.69)	0.43	0.53
[If answered yes to a screener question] Q18. How often did you try to discuss the further consult with the referring PCP before placing the consult request? ^c	500	2.75	1.49	1.3	0.09	0.10	0.11	-0.04
Communication								
Q13. When you tried to communicate directly with the referring PCP, how often could you reach the PCP in a timely manner?	480	4.58	1.44	5.4	0.56	0.40	(0.73)	0.37
Q14. How often was the PCP helpful in providing you further information or other assistance when you requested it?	415	4.64	1.39	7.6	0.62	0.43	(0.72)	0.42
Q19. When you needed help from a primary care team member other than the referring PCP, how often were you able to get the help you needed in a timely manner?	376	4.15	1.47	3.7	0.57	0.38	(0.74)	0.44
Data Transfer								
Q8. At the time of the patient's first visit with you, how often did you have access to the most recent relevant information from the PCP's practice?	568	5.46	1.17	15.2	0.46	0.47	0.40	(0.73)
Q11. At the time of the patient's follow-up visit with you, how often did you have access to the most recent relevant information from the PCP's practice?	531	5.71	1.05	20.0	0.51	0.40	0.46	(0.73)

^aResponse scale for all items is 1: Never; 2: Rarely—less than 10% of the time; 3: Occasionally—about 30% of the time; 4: Sometimes—about 50% of the time; 5: Frequently—about 70% of the time; 6: Usually—about 90% of the time; and 7: Always.

^bValues in parentheses are item-to-hypothesized scale Pearson correlation coefficients.

^cThis item was dropped as it had no correlation >0.40 with any scale.

TABLE 4 Descriptive statistics for final scales of the Coordination of Specialty Care—Specialist Survey (CSC-Specialist)

	Number of items per scale	N ^a	Mean (SD)	Percent at ceiling ^b (%)	Observed range ^c
Relationships	4	552	4.95 (0.96)	4.5	1.75-7.00
Roles and Responsibilities	4	573	4.18 (0.90)	0.3	2.00-6.75
Communication	3	453	4.48 (1.26)	4.6	1.00-7.00
Data Transfer	2	568	5.57 (1.04)	18.7	2.00-7.00

^aNumber who were nonmissing on at least half the items in each scale.^bCeiling = scale score ≥ 6.5 .^cTheoretical range = 1-7.**TABLE 5** Correlations among scales of the Coordination of Specialty Care—Specialist Survey (CSC-Specialist)^a

	Relations	Roles	Communication	Data Transfer
Relationships	(0.83)			
Roles and Responsibilities	0.64	(0.80)		
Communication	0.66	0.46	(0.86)	
Data Transfer	0.52	0.47	0.46	(0.84)

^aInternal consistency reliability in parentheses in the diagonal.

access to a common medical record. The utility of the Data Transfer scale would likely be enhanced in such contexts if more items were added to the current two, and we recommend the development and testing of an expanded version of the scale. Future research should also examine the reliability estimates for this scale (and all scales) after modification for use in settings which do not offer a common medical record.

The CSC-Specialist can facilitate identification of the causes and consequences of good specialty care coordination, including those aspects of coordination that represent the stress points for interventions to improve high-priority outcomes. The CSC-Specialist can also be used to target and evaluate quality improvement efforts. We developed the survey with the input of physicians from multiple medical specialties to ensure its broad applicability. The coordination scales make up a brief measure of only 13 items, with scales ranging from 2 to 4 items. The CSC-Specialist is therefore practical to administer and could also be embedded into other surveys to enrich the evaluation of the quality of specialty care.

Our study has some limitations. First, we administered the survey in the VA setting, which may limit generalizability. However, throughout instrument development we included steps to maximize its relevance to a broad range of delivery settings and increase the likelihood of validity outside of the VA. Second, we developed the instrument with the input of medical specialists and tested it among medical specialists only. Future work should include a determination of what, if any, changes need to be made in order to ensure the measure is applicable to other types of specialties (eg, surgical, and psychiatric). Third, our response rate was 25 percent across all recruitment methods, raising the possibility of nonresponse bias. However, if our respondents were consistently more highly satisfied (or dissatisfied) with their experience of coordination than the

nonrespondents, that bias would have likely been manifest in a high level of uniformity of responses (either positive or negative), which in turn would have made it difficult to differentiate among clusters of items. We did not observe this in our data.

Fourth, we tested construct validity using an outcome measure of overall coordination derived from the same data source as the predictors. This raised the possibility of common method bias, because our measures of both the predictors (coordination scales) and the outcome (overall assessment of coordination) were obtained from the same survey and using similar Likert-type items. However, we endeavored to minimize the possible effect of such spurious associations due to method-specific variance by controlling for several factors that might affect perceptions of overall coordination or reflect survey response patterns. Even after doing so, we found that three of the coordination scales emerged as much stronger predictors of overall coordination than the other variables in the model.

It is also possible that variation in responses to the evaluation of overall coordination was affected by disagreement over what the question means, or that responses to that question reflected an overall tendency to give positive or negative responses on surveys. However, if these factors were the primary drivers of respondents' judgments regarding overall coordination, then one would expect to observe a pattern of relatively weak and/or uniform correlations between each of the scales and the measure of overall coordination. Instead, we observed wide variation in the strength of correlations, with standardized betas ranging from 0.07 (Data Transfer) to 0.40 (Relationships). Nonetheless, we recommend the addition of one or two items about overall coordination to create a more robust multi-item scale to mitigate the possible effects of variation in the interpretation of this single question.

An important limitation of our questionnaire is that it evaluates coordination only as experienced by the specialist. A comprehensive assessment of specialty care coordination must also include the experience of the other stakeholders most closely involved—primary care providers and patients. We ask specialists, for example, how often PCPs follow through on their recommendations. Such a question assumes that all such recommendations should be acted upon. However from the perspective of the PCP, some or all recommendations may be judged to be inappropriate. Our intent in phrasing the CSC-Specialist survey items from the perspective of the specialist was not to generate a rating of PCP performance, but to flag coordination problems as experienced by the specialist. If scores on a given scale are low, all possibilities for the low score could and should be explored—including, to continue our example, the possibility that specialists are making unreasonable recommendations given the constraints of the primary care context at a given location. Further, we would suggest that a complete understanding of coordination requires assessment of the PCP and patient perspectives as well. For this reason, we are in the process of developing questionnaires for the other two members of this triad.

Additional research is needed to further validate the CSC-Specialist. Tests of criterion validity would enable assessment of the degree to which the measure and its individual scales predict important outcomes. An assessment of sensitivity to change is needed in order to understand the degree to which the measure can be used to evaluate interventions in relation to each other, over time, and at different facilities. The Data Transfer construct should be further developed, as it may have greater relevance to settings without a shared EMR. Finally, the measure should be administered and its psychometric characteristics assessed in other health care organizations.

Care coordination is critical to promoting reduced costs, improved patient satisfaction, and optimal health outcomes. To properly evaluate and improve coordination for specialty care referrals, a measure is needed that can assess the specialist perspective. The CSC-Specialist is a brief, multidimensional measure of care coordination from the specialist's perspective that fills this gap.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: The authors would like to thank Jolie Wormwood, PhD, for her assistance in revising the manuscript.

CONFLICT OF INTEREST

The authors have no conflicts of interest to disclose.

DISCLAIMER

The contents do not represent the views of the U.S. Department of Veterans Affairs or the U.S. Government.

ORCID

Varsha G. Vimalananda <https://orcid.org/0000-0003-0738-672X>

REFERENCES

1. Barnett ML, Song Z, Landon BE. Trends in physician referrals in the United States, 1999-2009. *Arch Intern Med.* 2012;172(2):163-170.
2. Committee on Quality Health Care in America IoM. *Crossing the Quality Chasm: A New Health System for the 21st Century.* Washington, DC: National Academy Press; 2001.
3. Anderson R, Barbara A, Feldman S. What patients want: A content analysis of key qualities that influence patient satisfaction. *J Med Pract Manage.* 2007;22(5):255-261.
4. Stille CJ, Jerant A, Bell D, Meltzer D, Elmore JG. Coordinating care across diseases, settings, and clinicians: a key role for the generalist in practice. *Ann Intern Med.* 2005;142(8):700-708.
5. Fialova D, Onder G. Medication errors in elderly people: contributing factors and future perspectives. *Br J Clin Pharmacol.* 2009;67(6):641-645.
6. Group Health's MacColl Institute for Healthcare Innovation. Reducing care fragmentation: a toolkit for coordinating care. 2011. http://www.improvingchroniccare.org/downloads/reducing_care_fragmentation.pdf. Accessed November 28, 2016.
7. Schoen C, Osborn R, How SK, Doty MM, Peugh J. In chronic condition: experiences of patients with complex health care needs, in eight countries, 2008. *Health Aff (Millwood).* 2009;28(1):w1-w16.
8. McDonald KM, Schultz EM, Albin L, et al. Care Coordination Atlas Version 3 (Prepared by Stanford University under subcontract to Battelle on Contract No. 290-04-0020). AHRQ Publication No 11-0023-EF; 2010.
9. Schultz EM, McDonald KM. What is care coordination? *Int J Care Coord.* 2014;17(1-2):5-24.
10. Gittel JH, Fairfield KM, Bierbaum B, et al. Impact of relational coordination on quality of care, postoperative pain and functioning, and length of stay: a nine-hospital study of surgical patients. *Med Care.* 2000;38(8):807-819.
11. Gittel JH, Weinberg D, Pfefferle S, Bishop C. Impact of relational coordination on job satisfaction and quality outcomes: a study of nursing homes. *Hum Resour Manage J.* 2008;18(2):154-170.
12. Havens DS, Vasey J, Gittel JH, Lin WT. Relational coordination among nurses and other providers: impact on the quality of patient care. *J Nurs Manag.* 2010;18(8):926-937.
13. Wrobel JS, Charns MP, Diehr P, et al. The relationship between provider coordination and diabetes-related foot outcomes. *Diabetes Care.* 2003;26(11):3042-3047.
14. Weinberg DB, Gittel JH, Lusenhop RW, Kautz CM, Wright J. Beyond our walls: impact of patient and provider coordination across the continuum on outcomes for surgical patients. *Health Serv Res.* 2007;42(1 Pt 1):7-24.
15. Fryer A-K, Friedberg MW, Thompson RW, Singer SJ. Patient perceptions of integrated care and their relationship to utilization of emergency, inpatient and outpatient services. Paper presented at: Healthcare; 2017.
16. Mehrotra A, Forrest CB, Lin CY. Dropping the baton: specialty referrals in the United States. *Milbank Q.* 2011;89(1):39-68.
17. Hysong SJ, Esquivel A, Sittig DF, et al. Towards successful coordination of electronic health record based-referrals: a qualitative analysis. *Implement Sci.* 2011;6:84.
18. Zuchowski JL, Rose DE, Hamilton AB, et al. Challenges in referral communication between VHA Primary Care and Specialty Care. *J Gen Intern Med.* 2015;30(3):305-311.
19. Bodenheimer T. Coordinating care—a perilous journey through the health care system. *N Engl J Med.* 2008;358(10):1064-1071.

20. Forrest CB, Glade GB, Baker AE, Bocian A, von Schrader S, Starfield B. Coordination of specialty referrals and physician satisfaction with referral care. *Arch Pediatr Adolesc Med*. 2000;154(5):499-506.
21. Gandhi TK, Sittig DF, Franklin M, Sussman AJ, Fairchild DG, Bates DW. Communication breakdown in the outpatient referral process. *J Gen Intern Med*. 2000;15(9):626-631.
22. Greenberg JO, Barnett ML, Spinks MA, Dudley JC, Frolkis JP. The "medical neighborhood": integrating primary and specialty care for ambulatory patients. *JAMA Intern Med*. 2014;174(3):454-457.
23. O'Malley AS, Reschovsky JD. Referral and consultation communication between primary care and specialist physicians: finding common ground. *Arch Intern Med*. 2011;171(1):56-65.
24. Stille CJ, Primack WA. Interspecialty communication: old problem, new hope? *Arch Intern Med*. 2011;171(14):1300.
25. McDonald KM, Sundaram V, Bravata DM, et al. *Care coordination*. Stanford, CA: Stanford-UCSF Evidence-based Practice Center; 2007.
26. Vimalananda VG, Dvorin K, Fincke BG, Tardiff N, Bokhour BG. Patient, primary care provider, and specialist perspectives on specialty care coordination in an integrated health care system. *J Ambul Care Manage*. 2018;41(1):15-24.
27. McDonald K, Schultz EM, Albin L, et al. Care coordination measures atlas update. AHRQ Publication No 14-0037-EF; 2014.
28. National Quality Forum. Effective Communication and Care Coordination. http://www.qualityforum.org/Topics/Effective_Communication_and_Care_Coordination.aspx. Accessed May 22, 2015.
29. Agency for Health Research and Quality. CAHPS surveys and tools. <https://www.cahps.ahrq.gov/default.asp>. Accessed July 15, 2013.
30. American College of Physicians. The patient-centered medical home neighbor. The interface of the patient-centered medical home with specialty/subspecialty practices. 2010. https://www.acponline.org/advocacy/current_policy_papers/assets/pcmh_neighbors.pdf. Accessed April 13, 2015.
31. National Committee on Quality Assurance. Standards and Guidelines for NCQA's Patient-Centered Medical Home (PCMH). 2011.
32. Schultz EM, Pineda N, Lonhart J, Davies SM, McDonald KM. A systematic review of the care coordination measurement landscape. *BMC Health Serv Res*. 2013;13:119.
33. Zillich AJ, Doucette WR, Carter BL, Kreiter CD. Development and initial validation of an instrument to measure physician-pharmacist collaboration from the physician perspective. *Value Health*. 2005;8(1):59-66.
34. Shi LY, Starfield B, Xu J. Validating the adult primary care assessment tool. *J Fam Pract*. 2000;50:161.
35. Birnberg JM, Drum ML, Huang ES, et al. Development of a safety net medical home scale for clinics. *J Gen Intern Med*. 2011;26(12):1418-1425.
36. Hess BJ, Lynn LA, Holmboe ES, Lipner RS. Toward better care coordination through improved communication with referring physicians. *Acad Med*. 2009;84(10):S109-S112.
37. Version 10: QSR International Pty Ltd.; 2012.
38. Diefenbach MA, Weinstein ND, O'reilly J. Scales for assessing perceptions of health hazard susceptibility. *Health Educ Res*. 1993;8(2):181-192.
39. Fowler FJ. *Improving Survey Questions: Design and Evaluation*, vol. 38. Thousand Oaks, CA: Sage; 1995.
40. Dillman D, Groves B. Internet, mail and mixed-mode surveys: the tailored design method, 3rd ed. *Survey Res*. 2011;34(833):635.
41. Osatuke K, Draime J, Moore SC, et al. Organization development in the Department of Veterans Affairs. In: Miller T, ed. *The Praeger Handbook of Veterans Health: History, Challenges, Issues and Developments. Volume IV: Future Directions in Veterans Healthcare*. Santa Barbara, CA: Praeger; 2012:21-76.
42. Ware J, Harris W, Gandek B, Rogers B, Reese P. *MAP-R for Windows: Multitrait/Multi-item Analysis Program—Revised User's Guide*, vol. 1. Boston, MA: Health Assessment Lab; 1997:997.
43. Brown Timothy A. *Confirmatory Factor Analysis for Applied Research*. New York, NY: Guilford; 2006.
44. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Eq Model Multi J*. 1999;6(1):1-55.
45. Linsky A, Simon SR, Stolzmann K, Bokhour B, Meterko M. Prescribers' perceptions of medication discontinuation: survey instrument development and validation. *Am J Manag Care*. 2016;22:747-754.
46. Midi H, Sarkar S, Rana S. Collinearity diagnostics of binary logistic regression model. *J Interdisciplin Math*. 2010;13(3):253-267.
47. Brown TA. *Confirmatory Factor Analysis for Applied Research*. New York, NY: Guilford Press; 2006.
48. O'Rourke N, Hatcher L. *A Step-by-step Approach to using SAS for Factor Analysis and Structural Equation Modeling*. Cary, NC: Sas Institute; 2013.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Vimalananda VG, Fincke BG, Qian S, Waring ME, Seibert RG, Meterko M. Development and psychometric assessment of a novel survey to measure care coordination from the specialist's perspective. *Health Serv Res*. 2019;54:689–699. <https://doi.org/10.1111/1475-6773.13148>