HEALTHCARE
POLICY
Politiques de Santé

Health Services, Management and Policy Research
Services de santé, gestion et recherche de politique

Volume 7 • Special Issue

Measurement of Primary Healthcare Attributes from the Patient Perspective

Mesure des caractéristiques des soins de santé primaires du point de vue du patient

A LONGWOODS PUBLICATION   Fonds   WWW.HEALTHCAREPOLICY.NET

HEALTHCARE QUARTERLY: Best practices, policy and innovations in the administration of healthcare. For administrators, academics, insurers, suppliers and policy leaders. *Edited by* Dr. Peggy Leatt, University of North Carolina, Chapel Hill. ✦ **CANADIAN JOURNAL OF NURSING LEADERSHIP**: Covering politics, policy, theory and innovations that contribute to leadership in nursing administration, practice, teaching and research. Peer reviewed. *Edited by* Dr. Lynn Nagle, University of Toronto, Toronto. ✦ **HEALTHCARE PAPERS**: Review of new models in healthcare. Bridging the gap between the world of academia and the world of healthcare management and policy. Authors explore the potential of new ideas. *Edited by* Dr. Peggy Leatt, University of North Carolina, Chapel Hill. ✦ **HEALTHCARE POLICY**: Healthcare policy research and translation. Peer reviewed. For health system managers, practitioners, politicians and their administrators, and educators and academics. Authors come from a broad range of disciplines including social sciences, humanities, ethics, law, management sciences and knowledge translation. *Edited by* Dr. Jennifer Zelmer, Adjunct Faculty, University of Victoria, Victoria. ✦ **ELECTRONIC HEALTHCARE**: Best practices, policy and innovations exploring e-models, e-practices and e-products for e-health. For administrators, academics, insurers, suppliers and policy pundits. *Edited by* Dr. Michael Guerriere, University of Toronto, Toronto, and Denis Protti, University of Victoria, Victoria. ✦ **LAW & GOVERNANCE**: Within the framework of the law and the role of governance providing policies, programs, practices and opinions for the providers, administrators and insurers of healthcare services. *Editorial Chair,* Dr. Kevin Smith, McMaster University, Hamilton. ✦ **HRRESOURCES**: Cases, commentary and policy reviews for healthcare clinicians, human resources managers and the policy leaders, insurers, academics, administrators, boards and advisors of all healthcare organizations. *Editorial Chair,* Dr. Louise Lemieux-Charles, University of Toronto, Toronto. ✦ **WORLD HEALTH & POPULATION**: Best practices, policy and innovations in the administration of healthcare in developing communities and countries. For administrators, academics, researchers and policy leaders. Includes peer reviewed research papers. *Edited by* Dr. John Paul, University of North Carolina, Chapel Hill. ✦ **LONGWOODS.COM**: Enabling excellence in healthcare. Providing electronic access to news, information, career opportunities, conference schedules, research, case studies, policy reviews and commentary that cover politics, policy, theory, best practices and innovations in healthcare.

Longwoods Publishing | *Enabling Excellence*

HEALTHCARE
POLICY
Politiques de Santé

Health Services, Management and Policy Research
Services de santé, gestion et recherche de politique

VOLUME 7 SPECIAL ISSUE • DECEMBER 2011

Healthcare Policy/Politiques de Santé seeks to bridge the worlds of research and decision-making by presenting research, analysis and information that speak to both audiences. Accordingly, our manuscript review and editorial processes include researchers and decision-makers.

We publish original scholarly and research papers that support health policy development and decision-making in spheres ranging from governance, organization and service delivery to financing, funding and resource allocation. The journal welcomes submissions from researchers across a broad spectrum of disciplines in health sciences, social sciences, management and the humanities and from interdisciplinary research teams. We encourage submissions from decision-makers or researcher–decision-maker collaborations that address knowledge application and exchange.

While *Healthcare Policy/Politiques de Santé* encourages submissions that are theoretically grounded and methodologically innovative, we emphasize applied research rather than theoretical work and methods development. The journal maintains a distinctly Canadian flavour by focusing on Canadian health services and policy issues. We also publish research and analysis involving international comparisons or set in other jurisdictions that are relevant to the Canadian context.

* * *

Politiques de Santé/Healthcare Policy cherche à rapprocher le monde de la recherche et celui des décideurs en présentant des travaux de recherche, des analyses et des renseignements qui s'adressent aux deux auditoires. Ainsi donc, nos processus rédactionnel et d'examen des manuscrits font intervenir à la fois des chercheurs et des décideurs.

Nous publions des articles savants et des rapports de recherche qui appuient l'élaboration de politiques et le processus décisionnel dans le domaine de la santé et qui abordent des aspects aussi variés que la gouvernance, l'organisation et la prestation des services, le financement et la répartition des ressources. La revue accueille favorablement les articles rédigés par des chercheurs provenant d'un large éventail de disciplines dans les sciences de la santé, les sciences sociales et la gestion, et par des équipes de recherche interdisciplinaires. Nous invitons également les décideurs ou les membres d'équipes formées de chercheurs et de décideurs à nous envoyer des articles qui traitent de l'échange et de l'application des connaissances.

Bien que *Politiques de Santé/Healthcare Policy* encourage l'envoi d'articles ayant un solide fondement théorique et innovateurs sur le plan méthodologique, nous privilégions la recherche appliquée plutôt que les travaux théoriques et l'élaboration de méthodes. La revue veut maintenir une saveur distinctement canadienne en mettant l'accent sur les questions liées aux services et aux politiques de santé au Canada. Nous publions aussi des travaux de recherche et des analyses présentant des comparaisons internationales qui sont pertinentes pour le contexte canadien.

CONTENTS

Guest Editorial

Overview and Commentary

Research Papers

Peer Reviewed

* * *

TABLE DES MATIÈRES

Éditorial – collaboration spéciale

Présentation et commentaire

Rapports de recherche

Examen par les pairs

HEALTHCARE
POLICY
Politiques de Santé

EDITOR-IN-CHIEF

JENNIFER ZELMER, BSC, MA, PHD
Adjunct Faculty, University of Victoria

SENIOR EDITOR

FRANÇOIS BÉLAND, PHD
Professor, Department of Health Administration, Faculté de médecine, Université de Montréal, Member, Groupe de recherche interdisciplinaire en santé (GRIS), Co-Director, Groupe de recherche Université de Montréal-Université McGill sur les personnes âgées, Montréal, QC

EDITORS

ROGER CHAFE, PHD
Director of Pediatric Research and Assistant Professor, Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL

RAISA DEBER, PHD
Professor, Department of Health Policy, Management and Evaluation, Faculty of Medicine, University of Toronto, Toronto, ON

MARK DOBROW, PHD
Director, Analysis and Reporting, Health Council of Canada
Assistant Professor, Department of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON

ERIC LATIMER, PHD
Researcher, Douglas Institute
Associate Professor, Department of Psychiatry, McGill University
Associate Member, Department of Epidemiology, Biostatistics, and Occupational Health, McGill University
Montreal, QC

JOEL LEXCHIN, MSC, MD
Professor and Associate Chair, School of Health Policy and Management, Faculty of Health, York University, Emergency Department, University Health Network, Toronto, ON

PATRICIA J. MARTENS, PHD
Professor, Department of Community Health Sciences, Faculty of Medicine, University of Manitoba
Director, Manitoba Centre for Health Policy, University of Manitoba, Winnipeg, MB

CLAUDE SICOTTE, PHD
Professor, Department of Health Administration, Faculty of medicine, University of Montreal
Researcher, Groupe de recherche interdisciplinaire en santé (GRIS), Montréal, QC

CONTRIBUTING EDITOR

STEVEN LEWIS
President, Access Consulting Ltd., Saskatoon, SK
Adjunct Professor of Health Policy,
University of Calgary & Simon Fraser University

EDITORIAL ADVISORY BOARD

TONI ASHTON
Associate Professor Health Economics, School of Population Health, The University of Auckland, Auckland, NZ

LUC BOILEAU, MD, MSC, FRCPC
President and Chief Executive Officer, Agence de la santé et des services sociaux de la Montérégie, Montréal, QC

PHILIP DAVIES
Government Social Research Unit, London, UK

MICHAEL DECTER
Founding and Former Chair, Health Council of Canada, Toronto, ON

ROBERT G. EVANS
Professor, Department of Economics, University of British Columbia, Member, Centre for Health Services and Policy Research, University of British Columbia, Vancouver, BC

KENNETH FYKE
Victoria, BC

STEFAN GREß
Department of Health Sciences, University of Applied Sciences Fulda, Germany

CHRIS HAM
Professor of Health Policy and Management, Health Services Management Centre, The University of Birmingham, Birmingham, UK

PAUL LAMARCHE
Professor, Departments of Health Administration & Social and Preventive Medicine, Director, GRIS, Faculté de médecine, Université de Montréal, Montréal, QC

DAVID LEVINE
Président directeur général, Agence de développement de réseaux locaux de services de santé et de services sociaux de Montréal-Centre, Montréal, QC

CHRIS LOVELACE
Senior Manager, World Bank, Kyrgyz Republic Country Office, Central Asia Human Development, Bishkek, Kyrgyz Republic

THEODORE R. MARMOR
Professor of Public Policy and Management, Professor of Political Science, Yale School of Management, New Haven, CT

VICENTE ORTÚN
Economics and Business Department and Research Center on Health and Economics (CRES), Pompeu Fabra University, Barcelona, Spain

ROBIN OSBORN
Vice President and Director, International Program in Health Policy and Practice, Commonwealth Fund, New York, USA

DOROTHY PRINGLE
Professor Emeritus and Dean Emeritus, Faculty of Nursing, University of Toronto, Toronto, ON

MARC RENAUD
Lisbon, Portugal (on sabbatical)

JEAN ROCHON
Expert associé, Systèmes de soins et services, Institut national de santé publique du Québec, Sainte-Foy, QC

NORALOU P. ROOS
Manitoba Centre for Health Policy
Professor, Community Health Sciences
University of Manitoba, Winnipeg, MB

RICHARD SALTMAN
Professor of Health Policy and Management, Rollins School of Public Health, Emory University, Atlanta, GA

HON. HUGH D. SEGAL, CM
Senator, Kingston-Frontenac-Leeds, Ottawa, ON

ALAN WOLFSON
South Africa

GUEST EDITOR
JACK WILLIAMS, PHD, FCES
Health Services Research and Epidemiology, Scientist Emeritus, Institute for Clinical Evaluative Sciences and Clinical Epidemiology, Sunnybrook Health Sciences Centre, Toronto, ON

MANAGING EDITOR
ANIA BOGACKA
abogacka@longwoods.com

EDITORIAL DIRECTOR
DIANNE FOSTER-KENT
dkent@longwoods.com

COPY EDITOR
FRANCINE GERACI

TRANSLATOR
ÉRIC BERGERON

PROOFREADER
NATHALIE LEGROS

DESIGN AND PRODUCTION
YVONNE KOO
ykoo@longwoods.com

JONATHAN WHITEHEAD
jwhitehead@longwoods.com

PUBLISHER
ANTON HART
ahart@longwoods.com

ASSOCIATE PUBLISHER
REBECCA HART
rhart@longwoods.com

ASSOCIATE PUBLISHER
SUSAN HALE
shale@longwoods.com

ASSOCIATE PUBLISHER
MATTHEW HART
mhart@longwoods.com

ASSOCIATE PUBLISHER/ADMINISTRATION
BARBARA MARSHALL
bmarshall@longwoods.com

HOW TO REACH THE EDITORS AND PUBLISHER
Telephone: 416-864-9667 Fax: 416-368-4443

ADDRESSES
All mail should go to: Longwoods Publishing Corporation, 260 Adelaide Street East, No. 8, Toronto, Ontario M5A 1N1, Canada.

For deliveries to our studio: 54 Berkeley St., Suite 305, Toronto, Ontario M5A 2W4, Canada.

SUBSCRIPTIONS
Individual subscription rates for one year are [C] $110 for online only and [C] $163 for print + online. For individual subscriptions contact Barbara Marshall at telephone 416-864-9667, ext. 100 or by e-mail at bmarshall@longwoods.com.

Institutional subscription rates are [C] $483 for online only and [C] $590 for print + online. For institutional subscriptions, please contact Susan Hale at telephone 416-864-9667, ext. 104 or by e-mail at shale@longwoods.com.

Subscriptions must be paid in advance. An additional tax (GST/HST) is payable on all Canadian transactions. Rates outside of Canada are in US dollars. Our GST/HST number is R138513668.

SUBSCRIBE ONLINE
Go to www.healthcarepolicy.net and click on "Subscribe."

REPRINTS/SINGLE ISSUES
Single issues are available at $45. Includes shipping and handling. Reprints can be ordered in lots of 100 or more. For reprint information call Barbara Marshall at 416-864-9667 or fax 416-368-4443 or e-mail to bmarshall@longwoods.com.

Return undeliverable Canadian addresses to: Circulation Department, Longwoods Publishing Corporation, 260 Adelaide Street East, No. 8, Toronto, Ontario M5A 1N1, Canada.

EDITORIAL
To submit material or talk to our editors please contact Ania Bogacka at 416-864-9667, ext. 108 or by e-mail at abogacka@longwoods.com. Author guidelines are available online at http://www.longwoods.com/pages/hpl-for-authors.

ADVERTISING
For advertising rates and inquiries, please contact Matthew Hart at 416-864-9667, ext. 113 or by e-mail at mhart@longwoods.com.

PUBLISHING
To discuss supplements or other publishing issues contact Anton Hart at 416-864-9667, ext. 109 or by e-mail at ahart@longwoods.com.

Healthcare Policy/Politiques de Santé is published four times per year by Longwoods Publishing Corp., 260 Adelaide St. East, No. 8, Toronto, ON M5A 1N1, Canada. The journal is published with support from the Canadian Institutes of Health Research's Institute of Health Services and Policy Research. Manuscripts are reviewed by the editors and a panel of peers appointed by the editors. Information contained in this publication has been compiled from sources believed to be reliable. While every effort has been made to ensure accuracy and completeness, these are not guaranteed. The views and opinions expressed are those of the individual contributors and do not necessarily represent an official opinion of *Healthcare Policy* or Longwoods Publishing Corporation. Readers are urged to consult their professional advisers prior to acting on the basis of material in this journal.

Healthcare Policy/Politiques de Santé is indexed in the following: PubMed Central, CINAHL, CSA (Cambridge), Ulrich's, Embase, IndexCopernicus, Scopus and is a partner of HINARI

No liability for this journal's content shall be incurred by Longwoods Publishing Corporation, the editors, the editorial advisory board or any contributors. ISSN No. 1715-6572

Publications Mail Agreement No. 40069375

EDITORIAL

The Continuing Quest for Primary Healthcare Reform: Measuring Performance

PRIMARY HEALTHCARE (PHC) REFORM CONTINUES AS A DEMANDING, IF NOT VEXing, challenge. The last decade has seen a cascade of initiatives to advance PHC. In 2000, the government of Canada established the $800-million Primary Health Care Transition Fund in agreement with the First Ministers who were seeking improvements to PHC that were deemed crucial to the renewal of health services. The report by Health Canada (2007) summarized the initiatives that were undertaken during the six years of the transition fund.

The Canadian Institute for Health Information (CIHI) is continuing to track changes in PHC. The CIHI-PHC (Webster 2009) initiative includes collaborations with Statistics Canada, the Health Council of Canada and health administrative data from across the provinces. The projects and major reports include the *Primary Health Care Indicators Chartbook*, the 2008 Canadian Survey of Experiences with Primary Health Care, the Canadian Community Health Survey (CCHS) (including the diabetes care module) and the 2010 National Physician Survey. For international comparisons, CIHI-PHC incorporates findings from the Commonwealth Fund and the Organisation for Economic Co-operation and Development (OECD).

The numbers of physicians in Canada remained at 2.1 per 1,000 population from 1987 to 2004, but by 2009 the number of physicians increased to 2.4 per 1,000 population, a 19% increase (OECD 2011). One can assume that some of the increase involves PHC, but we do not know for certain how many physicians are in PHC, nor how they are distributed across fee-for-service models (solo practice, walk-in clinics, family health teams), family health groups paid by capitation, community health centres (CHCs) or centres locaux de services communautaires (CLSCs in Quebec). In any case, Statistics Canada (2010) reports that the percentages of persons without PHC physicians and those having problems in accessing care have remained relatively unchanged over the last decade. As CIHI noted in its 2009 report on health during the past 10 years, the effectiveness of PHC is largely unknown.

The contexts for PHC are changing across Canada, and the changes in Ontario are but one example. The government is increasing the number of nurse practitioners in PHC, either through multidisciplinary family practices or by establishing independent nurse practitioner clinics (DiCenso et al. 2010).

Measuring Performance of PHC Systems – Filling in the Gaps

In 2003 the British Columbia Ministry of Health Services contracted the Centre for Health

Services and Policy Research at the University of British Columbia to establish an information system that could be used to describe the PHC sector from temporal, geographic, population and provider perspectives. As Watson (2009a,b) has noted, a comprehensive PHC information system requires data from all four domains.

Improving the measurement of PHC performance calls for a conceptual model for developing indicators, measures and performance data (Hogg and Dyke 2011). Assessing the perspectives of individuals across communities requires well-defined concepts and constructs and reliable and valid measures of the attributes. The lack of information and guidance about measures prompted Jeannie Haggerty and her colleagues to define PHC attributes from the patient's perspective and to critique and assess measures related to those attributes. Their study was funded by the Canadian Institutes of Health Research.

Haggerty and her team started with the basics and worked forward. They consulted with PHC experts to define the key attributes to be operationally defined—accessibility, interpersonal communication, relational continuity, comprehensiveness, management continuity and respectfulness. Secondly, they mapped the definitions to widely known PHC measures and chose six of them for the research project. Thirdly, 645 individuals from Quebec and Nova Scotia living in urban and rural locations with varying educational backgrounds participated in the surveys. The research team held 13 discussion groups with sub-samples of the respondents to obtain the assessments and criticisms of the French and English versions of the measures. Lastly, the team critically analyzed the responses to the individual items and the reliability and construct validity of how well the responses mapped onto the attributes of interest. The results are summarized in the papers in this special issue of *Healthcare Policy/Politiques de Santé*. Full results are available on the website: http://www.longwoods.com/publications/healthcare-policy/22634.

This special issue is a superb resource for rethinking the Canadian population's perspectives on PHC.

JACK WILLIAMS
Guest Editor

REFERENCES

Canadian Institute for Health Information (CIHI). 2009 (July). "Experiences with Primary Health Care in Canada – Analysis in Brief." Retrieved July 20, 2011. <http://secure.cihi.ca/cihiweb/products/cse_phc_aib_en.pdf>.

DiCenso, A., I. Bourgeault, J. Abelson, R. Martin-Misener, S. Kaasalainen, et al. 2010. "Utilization of Nurse Practitioners to Increase Patient Access to Primary Healthcare in Canada – Thinking Outside the Box." *Canadian Journal of Nursing Leadership* 23 (Special Issue, December): 239–59.

Haggerty, J.L. 2011. "Measurement of Primary Healthcare Attributes from the Patient Perspective." *Healthcare Policy* 7 (Special Issue): 13–20.

Health Canada. 2007. *Primary Health Care Transition Fund: Summary of Initiatives – Final Edition.* Retrieved July 20, 2011. <http://www.hc-sc.gc.ca/hcs-sss/pubs/prim/2007-initiatives/index-eng.php>.

Hogg, W. and E. Dyke. 2011. "Improving Measurement of Primary Care System Performance." *Canadian Family Physician* 57: 758–60.

Organisation for Economic Co-operation and Development (OECD). 2011. *OECD Health Data 2011.* "Frequently Requested Data." Retrieved July 20, 2011. <http://www.oecd.org/health/healthdata>.

Statistics Canada. 2010. *Access to a Regular Medical Doctor, 2010.* Retrieved July 20, 2011. <http://www.statcan.gc.ca/pub/82-625-x/2011001/article/11456-eng.htm>.

Watson, D.E. 2009a. "The Development of a Primary Healthcare Information System to Support Performance Measurement and Research in British Columbia." *Healthcare Policy* 5 (Special Issue): 16–22.

Watson, D.E. 2009b. "For Discussion: A Roadmap for Population-Based Information Systems to Enhance Primary Healthcare in Canada." *Healthcare Policy* 5 (Special Issue): 105–20.

Webster, G. 2009 (May 13). "Primary Health Care Indicators and CIHI's PHC Information Program." Presentation to the Canadian Association of Health Services and Policy Research. Retrieved July 13, 2011. <http://www.f2fe.com/CAHSPR/2009/docs/E1/Greg%20Webster.pdf>.

ÉDITORIAL

La quête continuelle pour la réforme des soins de santé primaires : mesure de la performance

La réforme des soins de santé primaires (SSP) continue de présenter des défis exigeants, sinon exaspérants. Au cours des dix dernières années, il y a eu une myriade d'initiatives visant l'avancement des SSP. En 2000, le gouvernement du Canada créait le Fonds pour l'adaptation des soins de santé primaires, doté de 800 millions de dollars, en accord avec les premiers ministres qui voulaient apporter des améliorations aux SSP, jugés essentiels pour le renouveau des services de santé. Le rapport de Santé Canada (2007) présente un sommaire des initiatives mises en place pendant les six ans de vie du Fonds.

L'Institut canadien d'information sur la santé (ICIS) continue de surveiller les changements dans les SSP. L'initiative de l'ICIS en matière de SSP (Webster 2009) comprend des collaborations avec Statistique Canada, le Conseil canadien de la santé et des bases de données administratives provinciales sur la santé. Les principaux projets et rapports sont, notamment, le *Recueil de graphiques sur les indicateurs de soins de santé primaires*, l'Enquête canadienne sur l'expérience des soins de santé primaires (2008), l'Enquête sur la santé dans les collectivités canadiennes (ESCC) (y compris le module des soins pour le diabète) et le Sondage national auprès des médecins (2010). À des fins de comparaison internationale, l'initiative de l'ICIS en matière de SSP incorpore des résultats provenant du Fonds du Commonwealth et de l'Organisation de coopération et de développement économiques (OCDE).

Entre 1987 et 2004, le nombre de médecins au Canada est demeuré à 2,1 par 1000 habitants, mais en 2009, ce chiffre s'est élevé à 2,4, soit une augmentation de 19 % (OCDE 2011). On peut penser qu'une partie de cette augmentation touche aux SSP, mais on ne connaît pas exactement combien de médecins exercent dans les SSP, ni quelle est leur distribution au sein des modèles de paiement à l'acte (pratique en solo, cliniques sans rendez-vous, équipes de médecins de famille), des groupes de santé familiale financés par capitation, des centres de santé communautaires ou des centres locaux de services communautaires (CLSC, au Québec). Quoi qu'il en soit, Statistique Canada (2010) indique que le pourcentage de personnes sans médecin de SSP et le pourcentage de gens qui connaissent des problèmes d'accès aux services de santé sont restés relativement les mêmes au cours de la dernière décennie. Tel que mentionné dans le rapport de l'ICIS (2009) sur la santé au cours des dix dernières années, l'efficacité des SSP est en grande partie inconnue.

Le contexte des SSP connaît des changements partout au Canada, et l'Ontario n'est qu'un exemple du changement. Le gouvernement accroît le nombre d'infirmières praticiennes dans les SSP, soit au sein de cliniques familiales multidisciplinaires ou en établissant des cliniques indépendantes d'infirmières praticiennes (DiCenso et al. 2010).

Mesure du rendement des systèmes de SSP – combler les lacunes

En 2003, le ministère britanno-colombien des Services de Santé retenait les services du Centre for Health Services and Policy Research de l'Université de la Colombie-Britannique pour la création d'un système d'information qui pourrait servir à décrire le secteur des SSP du point de vue temporel, géographique, de la population et des fournisseurs. Tel que l'indique Watson (2009a,b), pour être complet, un tel système doit comporter des données dans ces quatre domaines.

L'amélioration des mesures du rendement des SSP est tributaire d'un modèle conceptuel qui recueille des données sur les indicateurs, les mesures et le rendement (Hogg et Dyke 2011). L'évaluation du point de vue d'individus dans les collectivités requiert des concepts et des constructs clairement définis et fiables ainsi que des mesures valides des caractéristiques. Le manque d'information et d'orientation pour les mesures ont poussé Jeannie Haggerty et ses collègues à définir les caractéristiques des SSP du point de vue du patient et à faire une évaluation critique des mesures liées à ces caractéristiques. Leur travail a été financé par les Instituts de recherche en santé du Canada.

Mme Haggerty et son équipe ont commencé leur travail à partir de la base. En premier lieu, ils ont consulté des experts des SSP pour déterminer quelles caractéristiques devaient recevoir une définition opérationnelle : l'accessibilité, la communication interpersonnelle, la continuité relationnelle, l'intégralité, la continuité d'approche et le respect. En deuxième lieu, ils ont mis en correspondance ces définitions avec les instruments de mesure des SSP les plus connus, pour en choisir six aux fins du projet de recherche. En troisième lieu, 645 personnes du Québec et de la Nouvelle-Écosse, vivant en milieu urbain ou rural et ayant différents niveaux de scolarisation, ont participé aux sondages. L'équipe de recherche a tenu 13 groupes de discussion avec des sous-échantillons de répondants pour obtenir des évaluations et des critiques des versions française et anglaise des instruments de mesure. Pour terminer, l'équipe a fait l'analyse critique des réponses aux items et de la fiabilité et la validité conceptuelle de ces réponses en fonction desdites caractéristiques. Un sommaire de leurs résultats est présenté dans ce numéro spécial de *Politiques de Santé/Healthcare Policy*. Les résultats intégraux sont disponibles sur le site Web suivant : http://www.longwoods.com/publications/healthcare-policy/22634.

Ce numéro spécial constitue une ressource formidable pour reconsidérer le point de vue de la population canadienne sur les SSP.

JACK WILLIAMS
Éditorialiste de collaboration spéciale

RÉFÉRENCES

DiCenso, A., I. Bourgeault, J. Abelson, R. Martin-Misener, S. Kaasalainen, et al. 2010. "Utilization of Nurse Practitioners to Increase Patient Access to Primary Healthcare in Canada – Thinking Outside the Box." *Canadian Journal of Nursing Leadership* 23 (Special Issue, December): 239–59.

Haggerty, J.L. 2011. « Mesure des caractéristiques des soins de santé primaires du point de vue du patient » *Politiques de Santé* 7 (numéro spécial): 13–20.

Hogg, W. et E. Dyke. 2011. "Improving Measurement of Primary Care System Performance." *Canadian Family Physician* 57: 758–60.

Institut canadien d'information sur la santé (ICIS). 2009 (juillet). « Expériences vécues en soins de santé primaires au Canada – Analyse en bref. » Consulté le 20 juillet 2011. <http://secure.cihi.ca/cihiweb/products/cse_phc_aib_fr.pdf>.

Organisation de coopération et de développement économiques (OCDE). 2011. *Base de données de l'OCDE sur la santé 2011.* « Données fréquemment demandées. » Consulté le 20 juillet 2011. <http://www.oecd.org/health/healthdata>.

Santé Canada. 2007. *Fonds pour l'adaptation des soins de santé primaires : Résumé des initiatives : Dernière édition.* Consulté le 20 juillet 2011. <http://www.hc-sc.gc.ca/hcs-sss/pubs/prim/2007-initiatives/index-fra.php>.

Statistique Canada. 2010. *Médecin régulier, 2010.* Consulté le 20 juillet 2011. <http://www.statcan.gc.ca/pub/82-625-x/2011001/article/11456-fra.htm>.

Watson, D.E. 2009a. « Création d'un système d'information pour les soins de santé primaires afin d'appuyer la mesure du rendement et la recherche en Colombie-Britannique. » *Politiques de Santé* 5 (numéro spécial): 16–22.

Watson, D.E. 2009b. « Pour discussion : feuille de route pour les systèmes d'information démographique afin de renforcer les soins de santé primaires au Canada. » *Politiques de Santé* 5 (numéro spécial): 105–20.

Webster, G. 2009 (13 mai). "Primary Health Care Indicators and CIHI's PHC Information Program." Présentation à l'Association canadienne pour la recherche sur les services et les politiques de la santé. Consulté le 13 juillet 2011. <http://www.f2fe.com/CAHSPR/2009/docs/E1/Greg%20Webster.pdf>.

OVERVIEW AND COMMENTARY

Measurement of Primary Healthcare Attributes from the Patient Perspective

Mesure des caractéristiques des soins de santé primaires du point de vue du patient

JEANNIE L. HAGGERTY, PHD
Department of Family Medicine, McGill University
Montreal, QC

THIS SPECIAL ISSUE OF *HEALTHCARE POLICY/POLITIQUES DE SANTÉ* PRESENTS A series of papers reporting on the concurrent validation of instruments that assess primary healthcare (PHC) delivery from the patient's perspective. The study was funded in 2004 by the Canadian Institutes of Health Research (CIHR) at the height of the Primary Health Care Transition Fund, an $8-million investment by Health Canada to catalyze a renewal of the PHC system in Canada. In planning the evaluation for initiatives, it became evident that program evaluators and researchers had little guidance for selecting among available instruments. For example, although various instruments purported to measure accessibility, it was not obvious that all measured the same underlying construct. Nor was it clear how to compare results collected with different instruments in different jurisdictions or at different times.

This research program consists of three studies. The first is a consensus consultation of PHC experts across Canada to formulate operational definitions of attributes to be evaluated (Haggerty et al. 2007). Second, we mapped the operational definitions to validated instruments (available at www.programmeprecise.ca/en/publications). Third, we administered six instruments back-to-back in Nova Scotia and Quebec to adults with a regular source of care to examine and compare how well the instruments measure essential attributes of primary healthcare.

We selected six instruments in the public domain that assess usual care rather than a single visit, that are generic (not limited to a specific patient group or dimension of care) and that had been most proposed or used in Canada:

1. Primary Care Assessment Survey (PCAS) (Safran et al. 1998);
2. Primary Care Assessment Tool – Short Form, adult (PCAT-S) (Shi et al. 2001);
3. Components of Primary Care Index (CPCI) (Flocke 1997);
4. First version of the European general practice evaluation instrument (EUROPEP-I) (Grol et al. 2000);
5. Interpersonal Processes of Care – 18-item version (IPC-II) survey (Stewart et al. 2007);
6. Veterans Affairs National Outpatient Customer Satisfaction Survey (VANOCSS) (Borowsky et al. 2002).

Overview of the Findings

The first paper in this special issue reports how primary healthcare experts prioritized attributes for different primary healthcare models (Lévesque, Haggerty et al. 2011). Some attributes were identified as essential to any primary healthcare model; others assumed greater or lesser importance if the model was professional or community-oriented.

The methods article (Haggerty, Burge, Beaulieu et al. 2011) describes the sampling, recruitment and descriptive results of the concurrent instrument administration. Most instrument subscales discriminate among poor, average and excellent overall experience, but the most discriminating attributes are interpersonal communication and respectfulness. Inclusion of these attributes underlines their importance to patients and the crucial value of good measures.

Our sample was balanced by French/English language, high/low education, urban/rural location and overall healthcare experience, so we had sufficient statistical power to assess whether instruments perform differentially (are biased) by these categories. Because we administered French instruments in one province and English in another, we used advanced statistics to determine whether observed differences were due to language or the health system. The paper on differential item functioning (Haggerty, Bouharaoui et al. 2011) presents an overview of this technique and demonstrates that differential functioning was prevalent between French- and English-language instruments. Using only unbiased questions, we found that primary healthcare experience was consistently better in Nova Scotia than in Quebec. Also, with unbiased questions, we reversed an initial finding and found instead that first-contact accessibility was significantly worse for rural than urban respondents.

We learned qualitatively through 13 discussion groups (Haggerty, Beaulieu, Lawson et al. 2011) that patients overwhelmingly prefer Likert response scales with labels adapted to the context of the question. They do not like agree–disagree options to elicit frequency. They prefer long and clear formats to those that are short and crowded. Patients want to be good respondents and admit to guessing when they cannot evaluate directly or don't understand the question. They were disappointed that the instruments did not allow them to report on problems with the interface between different providers in the system.

The core question in our study, however, was how well different instruments measure the attributes essential to primary healthcare: accessibility, interpersonal communication, comprehensiveness, relational continuity, management continuity and respectfulness. All the study instruments have passed standards of reliability and validity, but we went further by comparing subscale values using a common metric, conducting factor analysis of items from different instruments, and then examining how well individual items measured the common construct that emerged across instruments. These findings are outlined and explained in the analytic overview paper (Santor et al. 2011). Pooling items allowed us to discern different dimensions within an attribute, some within the same instrument subscale.

Accessibility has been identified as a weakness in Canadian health systems in sequential international surveys (Schoen et al. 2004, 2007). It is therefore good to know that two subscales perform well (Haggerty, Lévesque et al. 2011). The PCAT-S First-Contact Access, despite some measurement problems, is the best and only measure of patients' confidence in being seen rapidly. The PCAS Organizational Accessibility has good metric properties and measures accommodation rather than rapid access.

Interpersonal communication – the heart of patient-centred care and foundational to establishing a therapeutic alliance – can suffer in team-based and shared care (Rodriguez et al. 2007; Safran 2003), so monitoring is critical. We discerned sub-dimensions of eliciting, explaining and shared decision-making that are captured in the PCAS Communication and the EUROPEP-I Clinical Behaviour subscales, with the former showing better measurement properties (Beaulieu et al. 2011). Three subscales in the IPC-II measure these dimensions specifically, but the response options or scoring could be adjusted to permit better discriminability.

Relational continuity is valued highly in family medicine as having therapeutic potential in itself, and is a potential victim of reforms towards team-based care. The assumption that continuity is achieved through concentrating care in a single physician is reflected in the measures we studied (Burge et al. 2011), though scores and percentage of visits with doctor do not correlate highly. This attribute is inferred from accumulated and comprehensive knowledge of the patient, captured in two subscales with similar content, the CPCI Accumulated Knowledge and the PCAS Contextual Knowledge, the latter with better metric properties. These may also measure an aspect of whole-person care.

Comprehensiveness is one of the most invoked qualifiers of good primary healthcare, but the lack of definitional clarity complicates measurement (Haggerty, Beaulieu, Pineault et al. 2011). Comprehensiveness as whole-person care is missing from these measures and needs development. The CPCI Comprehensive Care measures the patient's confidence in the physician's capacity to care for a range of health problems but may not reflect the actual range of services. The PCAT-S Comprehensive Services Available elicits the range of services, but measurement is limited by patient knowledge, needs or both. Range of services is probably best assessed by providers.

Management continuity is the experience of care coordination. Only some provider efforts to link and coordinate care are visible to patients, a fact that may explain why patient assessments of their primary care physician's coordination actions with the PCAT-S Coordination and the PCAS Integration are predominantly positive, even when problems are reported (Haggerty, Burge, Pineault et al. 2011). Participants in discussion groups wanted to report on their care experience across the entire system (Haggerty, Beaulieu et al. 2011), and the VANOCSS Overall Coordination subscale is the best tool to capture this experience.

Respectfulness – attention to dignity, interpersonal treatment and adequate privacy – is a new dimension, important to patients and their confidence in the health system. Respectfulness is the principal way patients experience responsiveness, defined as a fundamental dimension of health system performance by the World Health Organization and in which Canada showed room for improvement (WHO 2000). Though the paper by Lévesque, Pineault and colleagues (2011) is more exploratory than others in this series, they found respectfulness items in various instruments: addressed explicitly in IPC-II subscales but also by the PCAS Interpersonal Treatment. The measures apply to organizational processes as well as the clinical encounter.

Across attributes, patient assessments of care typically have a skewed distribution, with the vast majority of patients endorsing the more positive response options. This deviation from a normal distribution compromises the robustness of some psychometric measures, but most problematically, it reduces the capacity to discriminate between different levels of positive experience and, hence, to detect improvements. Patients are reluctant to evaluate providers negatively unless they know that the provider is responsible for the negative experience (Collins and O'Cathain 2003), suggesting that negative assessments are true negatives, whereas some positive assessments will be false positives. Indeed, item response analysis showed enhanced discriminability and information yield in the negative zone of assessments. This finding has implications for reporting of measures: rather than averaging item values, it may be more meaningful and informative to dichotomize assessments and to report the percentage with negative evaluations.

Patient assessments may be "contaminated" by the regard – usually positive – for the physician, and we may need new approaches to isolate experience of the attributes themselves. Instruments such as the VANOCSS, inspired by the Picker Institute suite of tools, elicit patient reports of incidents and preclude patients' judging their provider. This approach may help provide more information about positive experience. However, because it does not lend itself to classic psychometric analysis, we may need a new approach to assessing these instruments' validity and reliability.

Discussion

The results reported in the various papers in this special issue will provide guidance in the selection of instruments to be used in evaluation of primary healthcare and in the refinement and development of other instruments. Despite the end of the Primary Health Care Transition Fund in 2006, the results are still relevant because PHC renewal continues.

In fact, with the withdrawal of special funding, it is more important than ever to base resource allocations on sound evaluations and evidence of impact. These results are relevant not only for Canada, but also internationally – for example, to evaluate the patient-centred medical home (PCMH) in the United States and to inform the surveys of patient experience that are part of the pay-for-performance Quality and Outcomes Framework in the United Kingdom.

Ultimately, however, the enduring relevance of this study comes from the patient's being the *raison d'être* of the healthcare enterprise. Patient-centredness is espoused as a core value, but without mechanisms to continually remind us of how patients experience care, care delivery too easily becomes provider-centred. Patient surveys are critical to continuous quality improvement and the greater health system accountability promised to Canadians in the 2003 Health Accord.

In these papers, we refer unabashedly to "patients" rather than "consumers" or "clients," even though use of surveys is most often associated with the market concerns that these latter terms evoke. The etymological root for "patient" is the Latin *pati*: "to suffer." To "endure in suffering" constitutes the moral virtue of patience, one that is too often required in the healthcare delivery process as well as on the journey from illness to recovery. It is our sincere hope that this study will help evaluators institute mechanisms to listen effectively to patients' voices and adapt to their concerns.

Limitations of our study

This unique study gave us the rare privilege of comparing measures directly. However, there are some important limitations and things we might do differently were we to repeat the study.

First, few of the findings are generalizable beyond the six instruments included in our study. Resource limitations and the burden of response imposed the sometimes painful necessity of selecting instruments we believed to be most relevant and those that measured the usual care experience. Consequently, we selected the EUROPEP-I over the Consumer Assessment of Health Plan Survey (CAHPS, www.cahps.ahrq.gov), which is widely used in the United States, because the former fits our tax-based healthcare system. Likewise, we did not include any visit-based measures that, in retrospect, may be the most precise way to measure accessibility, interpersonal communication or respectfulness.

Second, we did not measure indicators of intermediate outcomes. Thus, we cannot estimate the relative association between, say, accessibility and unmet needs for care or between management continuity and gaps in information transfer.

Third, while most reliability and validity parameters of the instruments are similar in the Canadian context to the context in which they were developed, there are still unacceptably high levels of missing values in some items and in some instruments. Some of these are due to offering patients "not applicable" response options, which were appreciated by respondents but are essentially non-informative. These missing values limited our statistical power despite our having a robust sample size of 645 respondents. Evaluators should be alert to this pitfall in selecting instruments, as it could compromise the information yield of administered instruments.

Recommendations for researchers, program evaluators and policy makers

We did not aim to evaluate the instruments per se, but because our opinions are frequently solicited, I venture some observations.

The PCAS seems to have the best measurement properties overall: the response scales behave as expected, and the formatting and readability enhance ease of response. However, the questions pertain to the "regular personal doctor" and are not specific to primary healthcare – indeed, the content could well apply to specialist care. In contrast, the PCAT-S and the CPCI were specifically designed for and apply to North American primary healthcare; both have measurement problems that could be ameliorated easily by formatting to ease response and by adjusting the response options to better fit the questions. The EUROPEP-I, designed for European general practice, is easy to answer and its broad use allows for benchmarking; but it has only two subscales, which capture generic experience but present difficulties for monitoring specific attributes. The IPC-II measures dimensions that are important to patients and can be focused on primary healthcare, but it was not designed specifically for this; its response scale is not fully exploited, and the scoring may mask negative experiences.

Researchers need to develop measures for such attributes as whole-person care, shared decision-making, team relational continuity, information management and cultural sensitivity. Studies need to link patient assessments of attributes to intermediate and health outcomes. Finally, some fundamental methodological development is needed to assess the metric performance of instruments that elicit reports of critical incidents as a way of inferring quality of care.

We are confident in endorsing the instruments in our study for the Canadian context, despite their having been developed elsewhere. Program evaluators should, however, be cautious about subscales having high (over 4%) missing values, as these will compromise the capacity to detect intended effects. Our study identified sub-dimensions within the attributes being measured; these would be a consideration in selecting instruments. Data from providers is needed for such attributes as comprehensive services, integration and quality of care. Finally, we caution against using our results to select individual items for use in an isolated manner; we affirm the integrity of subscales as constructed by the developers.

Finally, policy makers may be assured that we have adequate measurement instruments to consistently and routinely measure patient perceptions of such attributes as accessibility, interpersonal communication, relational continuity and respectfulness. These attributes are important to Canadians and are predictive of their overall experience of care. Even if different instruments are used across time or across jurisdictions, our study demonstrates how a common metric can be established to compare results.

ACKNOWLEDGEMENTS

This research was funded by the Canadian Institutes of Health Research. This special supplement was sponsored by: a dissemination grant from the Canadian Institutes for Health Research; funds allocated to Jeannie Haggerty's Canada Research Chair by the Université de Sherbrooke and the Centre de recherche de l'Hôpital Charles LeMoyne; the Réseau de recherche en santé de populations du Québec and the Canadian Institute for Health Information. The production of an extended report was made possible by contributions from the following authors' institutions: Centre de recherche du Centre hospitalier de l'Université de Montréal; the McGill Research Chair in Family & Community Medicine at St. Mary's Hospital; Dalhousie University's Department of Family Medicine; University of Ottawa's School of Psychology and the Département de santé publique de l'Agence de santé et services sociaux de Montréal-Centre.

Correspondence may be directed to: Jeannie L. Haggerty, Associate Professor, Department of Family Medicine, McGill University, St. Mary's Research Centre, Hayes Pavilion – Suite 3734, 3830 Lacombe Ave., Montreal QC H3T 1M5; tel.: 514-345-3511 ext. 6332; fax: 514-734-2652; e-mail: jeannie.haggerty@mcgill.ca.

REFERENCES

Beaulieu, M.-D., J.L. Haggerty, C. Beaulieu, F. Bouharaoui, J.-F. Lévesque, R. Pineault et al. 2011. "Interpersonal Communication from the Patient Perspective: Comparison of Primary Healthcare Evaluation Instruments." *Healthcare Policy* 7 (Special Issue): 108–23.

Borowsky, S.J., D.B. Nelson, J.C. Fortney, A.N. Hedeen, J.L. Bradley and M.K. Chapko. 2002. "VA Community-Based Outpatient Clinics: Performance Measures Based on Patient Perceptions of Care." *Medical Care* 40(7): 578–86.

Burge, F., J.L. Haggerty, R. Pineault, M.-D. Beaulieu, J.-F. Lévesque, C. Beaulieu et al. 2011. "Relational Continuity from the Patient Perspective: Comparison of Primary Healthcare Evaluation Instruments." *Healthcare Policy* 7 (Special Issue): 124–38.

Collins, K. and A. O'Cathain. 2003 "The Continuum of Patient Satisfaction – From Satisfied to Very Satisfied." *Social Science and Medicine* 57(12): 2465–70.

Flocke, S. 1997. "Measuring Attributes of Primary Care: Development of a New Instrument." *Journal of Family Practice* 45(1): 64–74.

Grol, R., M. Wensing and Task Force on Patient Evaluations of General Practice. 2000. "Patients Evaluate General/Family Practice: The EUROPEP Instrument." Nijmegen, Netherlands: Centre for Quality of Care Research, Raboud University.

Haggerty, J.L., M.-D. Beaulieu, R. Pineault, F. Burge, J.-F. Lévesque, D. A. Santor et al. 2011. "Comprehensiveness of Care from the Patient Perspective: Comparison of Primary Healthcare Evaluation Instruments." *Healthcare Policy* 7 (Special Issue): 154–66.

Haggerty, J.L., C. Beaulieu, B. Lawson, D.A. Santor, M. Fournier and F. Burge. 2011. "What Patients Tell Us about Primary Healthcare Evaluation Instruments: Response Formats, Bad Questions and Missing Pieces." *Healthcare Policy* 7 (Special Issue): 66–78.

Haggerty, J.L., F. Bouharaoui and D.A Santor. 2011. "Differential Item Functioning in Primary Healthcare Evaluation Instruments by French/English Version, Educational Level and Urban/Rural Location." *Healthcare Policy* 7 (Special Issue): 47–65.

Haggerty, J.L., F. Burge, M.-D. Beaulieu, R. Pineault, C. Beaulieu et al. 2011. "Validation of Instruments to Evaluate Primary Healthcare from the Patient Perspective: Overview of the Method." *Healthcare Policy* 7 (Special Issue): 31–46.

Haggerty, J.L., F. Burge, J.-F. Lévesque, D. Gass, R. Pineault, M. Beaulieu et al. 2007. "Operational Definitions of Attributes of Primary Health Care: Consensus among Canadian Experts." *Annals of Family Medicine* 5(4): 336–44.

Haggerty, J.L., F. Burge, R. Pineault, M.-D. Beaulieu, F. Bouharaoui, C. Beaulieu et al. 2011. "Management Continuity from the Patient Perspective: Comparison of Primary Healthcare Evaluation Instruments." *Healthcare Policy* 7 (Special Issue): 139–53.

Haggerty, J.L., J.-F. Lévesque, D.A Santor, F. Burge, C. Beaulieu, F. Bouharaoui et al. 2011. "Accessibility from the Patient Perspective: Comparison of Primary Healthcare Evaluation Instruments." *Healthcare Policy* 7 (Special Issue): 94–107.

Lévesque, J.-F., J.L. Haggerty, F. Burge, M.-D. Beaulieu, D. Gass, R. Pineault et al. 2011. "Canadian Experts' Views on the Importance of Attributes within Professional and Community-Oriented Primary Healthcare Models." *Healthcare Policy* 7 (Special Issue): 21–30.

Lévesque, J.-F., R. Pineault, J.L. Haggerty, F. Burge, M.-D. Beaulieu, D. Gass et al. 2011. "Respectfulness from the Patient Perspective: Comparison of Primary Healthcare Evaluation Instruments." *Healthcare Policy* 7 (Special Issue): 167–79.

Rodriguez, H.P., W.H. Rogers, R.E. Marshall and D.G. Safran. 2007. "Multidisciplinary Primary Care Teams. Effects on the Quality of Clinician–Patient Interactions and Organizational Features of Care." *Medical Care* 45(1): 19–27.

Safran, D.G. 2003. "Defining the Future of Primary Care: What Can We Learn from Patients?" *Annals of Internal Medicine* 138(3): 248–55.

Safran, D.G., J. Kosinski, A.R. Tarlov, W.H. Rogers, D.A. Taira, N. Lieberman and J.E. Ware. 1998. "The Primary Care Assessment Survey: Tests of Data Quality and Measurement Performance." *Medical Care* 36(5): 728–39.

Santor, D.A., J.L. Haggerty, J.-F. Lévesque, F. Burge, M.-D. Beaulieu, D. Gass et al. 2011. "An Overview of Confirmatory Factor Analysis and Item Response Analysis Applied to Instruments to Evaluate Primary Healthcare." *Healthcare Policy* 7 (Special Issue): 79–92.

Schoen, C., R. Osborn, M. Doty, M. Bishop, J. Peugh and N. Murukutla. 2007. "Toward Higher-Performance Health Systems: Adults' Health Care Experiences in Seven Countries." *Health Affairs* 26(6): 717–34.

Schoen, C., R. Osborn, P.T. Huynh, M. Doty, K. Davis, K. Zapert et al. 2004. "Primary Care and Health System Performance: Adults' Experiences in Five Countries." *Health Affairs* (Millwood), Suppl. Web Exclusives: W4–503.

Shi, L., B. Starfield and J. Xu. 2001. "Validating the Adult Primary Care Assessment Tool." *Journal of Family Practice* 50(2): n161w–n171w.

Stewart, A.L., A.M. Nápoles-Springer, S.E. Gregorich and J. Santoyo-Olsson. 2007. "Interpersonal Processes of Care Survey: Patient-Reported Measures for Diverse Groups." *Health Services Research* 42(3 Pt. 1): 1235–56.

World Health Organization (WHO). 2000. *The World Health Report 2000 – Health Systems: Improving Performance.* Geneva: Author.

RESEARCH PAPER

Canadian Experts' Views on the Importance of Attributes within Professional and Community-Oriented Primary Healthcare Models

Points de vue d'experts canadiens sur l'importance des caractéristiques au sein des modèles de soins primaires professionnels et axés sur la communauté

JEAN-FRÉDÉRIC LÉVESQUE, MD, PHD
Centre de recherche du Centre hospitalier de l'Université de Montréal
Montréal, QC

JEANNIE L. HAGGERTY, PHD
Department of Family Medicine, McGill University
Montreal, QC

FREDERICK BURGE, MD, MSC
Department of Family Medicine, Dalhousie University
Halifax, NS

MARIE-DOMINIQUE BEAULIEU, MD, MSC
Chaire Dr Sadok Besrour en médecine familiale
Centre de recherche du Centre hospitalier de l'Université de Montréal
Montréal, QC

DAVID GASS, MD
Department of Family Medicine, Dalhousie University
Halifax, NS

RAYNALD PINEAULT, MD, PHD
Centre de recherche du Centre hospitalier de l'Université de Montréal
Montréal, QC

DARCY A. SANTOR, PHD
School of Psychology, University of Ottawa
Ottawa, ON

Abstract

Purpose: The aim of this study was to rate the importance of primary healthcare (PHC) attributes in evaluations of PHC organizational models in Canada.
Methods: Using the Delphi process, we conducted a consensus consultation with 20 persons recognized by peers as Canadian PHC experts, who rated the importance of PHC attributes within professional and community-oriented models of PHC.
Results: Attributes rated as essential to all models were designated core attributes: first-contact accessibility, comprehensiveness of services, relational continuity, coordination (management) continuity, interpersonal communication, technical quality of clinical care and clinical information management. Overall, while all were important, non-core attributes – except efficiency/productivity – were rated as more important in community-oriented than in professional models. Attributes rated as essential for community-oriented models were equity, client/community participation, population orientation, cultural sensitivity and multidisciplinary teams.
Conclusion: Evaluation tools should address core attributes and be customized in accordance with the specific organizational models being evaluated to guide health reforms.

Résumé

Objet : L'objectif de cette étude était de classer selon leur importance différentes caractéristiques liées à l'évaluation des modèles organisationnels de soins de santé primaires (SSP) au Canada.
Méthode : Au moyen du procédé Delphi, nous avons mené une consultation de consensus auprès de 20 personnes reconnues par leurs pairs comme des experts canadiens des SSP. Ils ont classé selon leur importance des caractéristiques des SSP au sein des modèles de SSP professionnels et axés sur la communauté.
Résultats : Les caractéristiques classées comme essentielles pour tous les modèles ont été désignées comme caractéristiques centrales : accessibilité de premier contact, intégralité des services, continuité relationnelle, continuité d'approche (gestion), communication interpersonnelle, qualité technique des soins cliniques et gestion de l'information clinique. En général, bien qu'elles soient toutes importantes, les caractéristiques non centrales (à l'exception de l'efficience et de la productivité) ont été classées plus importantes pour les modèles axés sur la communauté que pour les modèles professionnels. Les caractéristiques classées essentielles pour les modèles axés sur la communauté sont : l'équité, la participation des clients et de la communauté, l'approche axée sur la population, la sensibilité aux aspects culturels et les équipes multidisciplinaires.
Conclusion : Les outils d'évaluation devraient tenir compte des caractéristiques centrales et être adaptés en fonction des modèles organisationnels évalués, afin d'orienter les réformes de la santé.

* * *

PRIMARY HEALTHCARE (PHC) HAS RECENTLY RECEIVED SUBSTANTIAL ATTENTION in many industrialized countries. In Canada, national and provincial commissions have identified problems in how PHC is organized (Romanow 2002; Kirby and LeBreton 2002; Clair 2000; Government of Ontario 2000; Government of Saskatchewan 2001; Government of Alberta 2001), and have suggested reform of the PHC system as a means of solving problems affecting the healthcare system as a whole. In particular, the accessibility and comprehensiveness of PHC services should be strengthened to better fulfill their mandate and maximize their contribution to the health of Canadians.

Most Canadian provinces have undertaken PHC reforms. Although most efforts focus on encouraging physicians and other health professionals to work collaboratively, in various jurisdictions reforms have involved specific organizational innovations in PHC models. It is thus crucial to develop frameworks and tools for ongoing evaluation of reforms that take into account the heterogeneity of organizational forms present in the health system.

This paper is part of a series of studies aimed at providing such a framework and tools, one of which was a consultation with Canadian PHC experts to establish operational (i.e., measurable) definitions of the attributes that should be evaluated in current and proposed models in Canada (Haggerty et al. 2007). In addition, experts rated the importance of different attributes within different models of PHC. In this paper, we report the results of this process.

We use "primary healthcare" to refer minimally both to services serving as first contact with the healthcare system and comprehensively to the place where a broad range of health-related needs are managed. We consider PHC organizational models to be those that provide at least general medical services by family physicians or generalists, who may or may not work with other health and social service professionals. Furthermore, we use a taxonomy (Lamarche et al. 2003) to classify PHC models as two archetypes, either professional or community-oriented. These two classes emerge based on an analysis of the vision, resources, governance structures and practices of PHC models worldwide, and they correspond roughly to what are often identified as "primary care" and "primary health care," respectively.

Professional models are generally managed autonomously by family physicians, funded as small businesses through physician compensation. They provide general medical services, and their aim is to respond to medical needs of patients who present to, or have registered with, the PHC provider. This is the predominant model in Canada, but it encompasses a variety of sub-models such as single-physician practices, walk-in clinics and family medicine group practices. Recent reforms have seen a move towards the integration of primary care nurses and other disciplines and towards physician remuneration methods other than fee-for-service. Examples are family health teams (FHTs, Ontario), primary care networks (PCNs, Alberta) and primary health care organizations (PHCOs, British Columbia).

Community-oriented PHC models aim at improving the health of populations and are organized to meet health needs in the broader sense than just medical care. They are generally managed by public or community administrations and usually include a wide range of professionals who deliver a broad spectrum of health and social services. This model is less prevalent in Canada, although there is a long history with local community service centres (CLSCs) in Quebec and community health centres (CHCs) in Ontario and Saskatchewan. Recent reforms saw the establishment of similar models in New Brunswick and British Columbia.

Methods

We used the Delphi method for consensus-building. In this iterative process, documents pass through a series of revisions by a small group of experts, whose feedback each time is incorporated into successive rounds until consensus is achieved. This method has been described elsewhere (Haggerty et al. 2007) and is summarized here.

We used a snowball sampling technique to identify Canadian PHC experts. These were persons identified by at least two persons as having accumulated significant knowledge about PHC through clinical, managerial or academic activities. We identified 26 such experts, equally balanced among clinicians, academics and decision-makers from all regions in the country, of whom 20 were successfully contacted and 18 participated in at least two rounds. We conducted four rounds between June and October 2004, averaging 13 participants each.

TABLE 1. Operational definitions of attributes of primary healthcare to be evaluated and consensus on best data source for evaluation

Core Attributes	Best Data Source
First-Contact Accessibility: The ease with which a person can obtain needed care (including advice and support) from the practitioner of choice within a time frame appropriate to the urgency of the problem.	Patient
Comprehensiveness of Services: The provision, either directly or indirectly, of a full range of services to meet patients' healthcare needs. This includes health promotion, prevention, diagnosis and treatment of common conditions, referral to other providers, management of chronic conditions, rehabilitation, palliative care and, in some models, social services.	Patient, provider, administrative data
Relational Continuity: A therapeutic relationship between a patient and one or more providers that spans various healthcare events and results in accumulated knowledge of the patient and care consistent with the patient's needs.	Patient
Coordination (Management) Continuity: The delivery of services by different providers in a timely and complementary manner such that care is connected and coherent.	Patient
Interpersonal Communication: The ability of the provider to elicit and understand patient concerns, to explain healthcare issues and engage in shared decision-making, if desired.	Patient
Technical Quality of Clinical Care: The degree to which clinical procedures reflect current research evidence and/or meet commonly accepted standards for technical content or skill.	Provider, chart audit
Clinical Information Management: The adequacy of methods and systems to capture, update, retrieve and monitor patient data in a timely, pertinent and confidential manner.	Provider
Person-Oriented Dimensions	
Advocacy: The extent to which providers represent the best interests of individual patients and patient groups in matters of health (including broad determinants) and healthcare.	Patient
Cultural Sensitivity: The extent to which a provider integrates cultural considerations into communication, assessment, diagnosis and treatment planning.	Patient
Family-Centred Care: The extent to which the provider considers the family (in all its expressions), understands its influence on a person's health and engages it as a partner in ongoing healthcare.	Patient
Respectfulness: The extent to which health professionals and support staff meet users' expectations about interpersonal treatment, demonstrate respect for the dignity of patients and provide adequate privacy.	Patient

TABLE 1. Continued

Core Attributes	Best Data Source
Whole-Person Care: The extent to which a provider elicits and considers the physical, emotional and social aspects of a patient's health and considers the community context in the patient's care.	Patient
Community-Oriented Dimensions	
Client/Community Participation: The involvement of clients and community members in decisions regarding the structure of the practice and services provided (e.g., advisory committees, community governance).	Patient, provider
Equity: The extent to which access to healthcare and good-quality services is provided on the basis of health needs, without systematic differences on the basis of individual or social characteristics.	All
Intersectoral Team: The extent to which the primary care provider collaborates with practitioners from non-health sectors in providing services that influence health. (Note: This dimension is relevant only to community models of primary care.)	Provider
Population Orientation: The extent to which primary care providers assess and respond to the health needs of the population they serve. (In professional models, the population is the patient population served; in community models, it is defined by geography or social characteristics.)	Patient, provider
Structural Dimensions	
Accessibility–Accommodation: The way primary healthcare resources are organized to accommodate a wide range of patients' abilities to contact healthcare providers and reach healthcare services (e.g., the organization of characteristics such as telephone services, flexible appointment systems, hours of operation and walk-in periods).	Patient
Informational Continuity*: The extent to which information about past care is used to make current care appropriate to the patient.	(Not assessed)
Multidisciplinary Team: Practitioners from various health disciplines collaborate in providing ongoing healthcare.	Provider
Quality Improvement Process: The institutionalization of policies and procedures that provide feedback about structures and practices and that lead to improvements in clinical quality of care and provide assurance of safety.	Provider
System Integration: The extent to which the healthcare unit organization has established and maintains linkages with other parts of the healthcare and social services system to facilitate transfer of care and coordinate concurrent care among different healthcare organizations.	Provider
System Performance	
Accountability: The extent to which the responsibilities of professionals, management and governance structures are defined, their performance is monitored and appropriate information on results is made available to stakeholders.	Provider
Availability: The fit between the number and type of human and physical resources and the volume and types of care required by the catchment population served in a defined period of time.	Administrative data
Efficiency/Productivity: Achieving the desired results with the most cost-effective use of resources. (This definition is non-operational.)	Administrative data

* This definition and best data source were not submitted to the consensus process but are included for completeness by general agreement of the research team.

In addition to identifying attributes and proposing operational definitions for them, the experts were asked to rank each attribute's importance for the two archetypes of PHC models. In the first round, they were asked to score the importance of attributes as (1) essential to core functions of PHC, (2) important but not essential or (3) relevant only to some forms of PHC organization. Attributes unanimously scored as 1 were designated core attributes and were no longer submitted to discussion.

In subsequent rounds, we asked experts to rank the importance of remaining attributes, this time within professional models and community-oriented PHC models. They were asked to score the importance of attributes for each model type as (1) essential to its core function,

(2) important but not essential to its core function or (3) not relevant to this model. Again, attributes unanimously scored as 1 were considered core attributes for that model.

Based on the importance of the scores of the fourth round, we calculated the means and standard deviations of the scores. These were used to rank attributes within organizational models. In this way, we identified attributes that were essential (scores of 1.0–1.1), essential–important (1.2–1.3), very important (1.4–1.8), important (1.9) and somewhat important (2.0+).

Results

The Delphi process identified and defined 24 attributes of primary care (Haggerty et al. 2007). In the first round, seven attributes were scored unanimously as essential to core function and were thus identified as core attributes for all types of primary care organization. These are: first-contact accessibility; comprehensiveness of services; relational continuity; coordination (management) continuity; interpersonal communication; technical quality of clinical care; and clinical information management. Table 1 presents the 24 attributes and their operational definitions; core attributes are grouped together and listed first. The table also shows the best identified data source for evaluation. The remaining attributes are grouped within person-oriented, community-oriented, structural and system performance categories.

FIGURE 1. Synthesis of ranking scores of non-core attributes across models

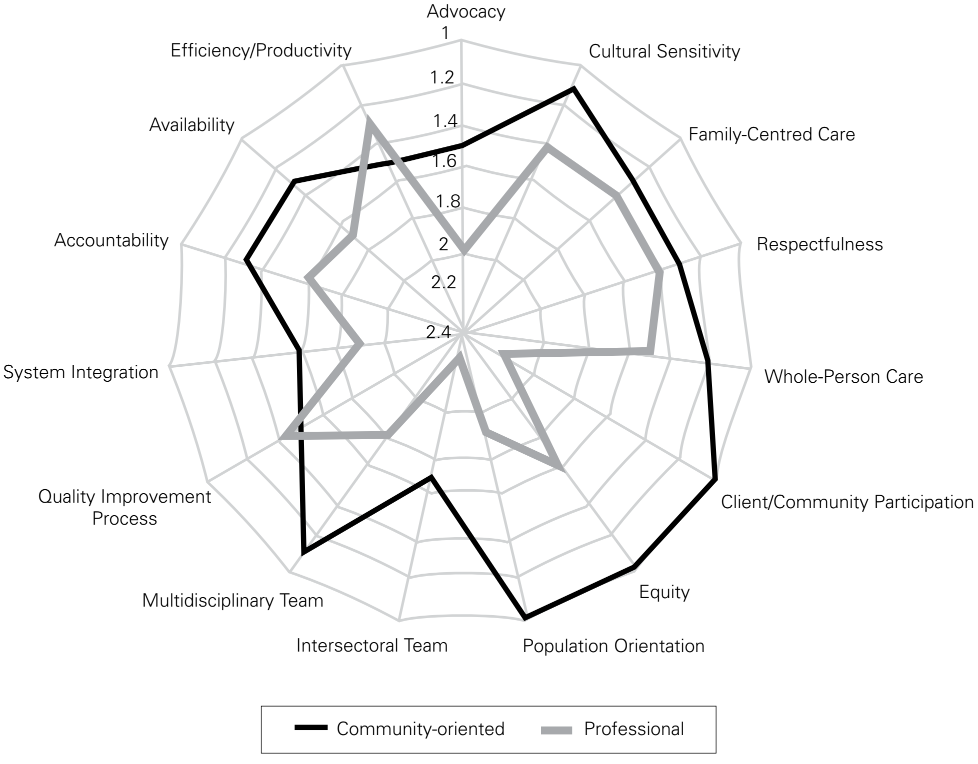


Figure 1 represents schematically the mean importance scores on non-core attributes for professional and community-oriented models (all core attributes are ranked as 1). Overall, attributes are assigned greater importance in the community-oriented than in the professional model. Ten attributes are classified as essential (rank 1.0–1.3) in the community-oriented model, compared to only one in the professional model (efficiency/productivity). No attributes ranked below important in community-oriented models, while three were judged only somewhat important for professional models (advocacy, intersectoral team and client/community participation). This is not to say that attributes did not rank high in the professional model. Nine were considered very important, but even those were rated as more important or essential to community models' performance. Two attributes (accessibility–accommodation and informational continuity) were added at the final face-to-face meeting with a subgroup of experts, and consequently were not scored for importance within organizational models of PHC. The others are ranked in order of importance first in the community-oriented model, then in the professional model. Three attributes – population orientation, client/community participation and equity – were unanimously ranked as essential in the community-oriented model. These would be considered core attributes of this model, with equity being very important also in professional models.

Table 2 highlights, for each of the attributes, the difference in importance between community-oriented and professional models. Attributes are ordered by the extent to which each was considered more important in the community-oriented model (determined by simple subtraction of the mean importance scores). A larger positive difference indicates that an attribute is relevant principally in community-oriented models, while negative values identify those more relevant in professional models; a score of zero indicates equal relevance in both (e.g., core attributes). From Table 2, we see that some attributes (such as client/community participation, population orientation, multidisciplinary team, equity, intersectoral team and advocacy) are relevant and of utmost importance in community-oriented models, and others (efficiency/productivity, quality improvement process) in professional models. Other attributes (such as system integration) are equally important in both.

TABLE 2. Clustering of attributes according to discriminating power and relevance to both organizational models

	Ranking Difference between Models	Relevance to Both Models
Client/community participation	1.1	–
Population orientation	0.9	–
Multidisciplinary team	0.7	+
Equity	0.6	+
Intersectoral team	0.6	–
Advocacy	0.5	– –
Availability	0.4	+
Cultural sensitivity	0.3	++
Accountability	0.3	++

TABLE 2. Continued

	Ranking Difference between Models	Relevance to Both Models
Whole-person care	0.3	++
System integration	0.3	–
Family-centred care	0.1	++
Respectfulness	0.1	++
Quality improvement process	– 0.1	+
Efficiency/productivity	– 0.2	++
First-contact accessibility	0	+++
Clinical information management	0	+++
Technical quality of clinical care	0	+++
Comprehensiveness of services	0	+++
Continuity – coordination	0	+++
Continuity – relational	0	+++
Interpersonal communication	0	+++

Discussion and Conclusion

The experts identified seven core attributes that must be present in any model of primary care: first-contact accessibility, comprehensiveness of services, relational continuity, coordination (management) continuity, interpersonal communication, technical quality of clinical care and clinical information management. From an evaluation standpoint, it is critical to ensure we have appropriate tools to assess PHC performance in these core dimensions, which should always be assessed whatever model is under study or attribute is of interest. If an intervention's focus is, say, on multidisciplinary teams or accessibility–accommodation, then, in addition to examining improvement in that attribute, the evaluation process must ensure that this improvement does not occur at the expense of core attributes.

Our findings on the scoring of non-core attributes suggest that more is expected of community-oriented than professional models. Community-oriented models are expected to provide both a large scope of services and care that addresses health determinants and equity. Some person-oriented attributes (such as cultural sensitivity and whole-person care) and community-oriented attributes (such as equity and population orientation) ranked as truly essential to this model.

Interestingly, efficiency was not considered an essential attribute of community-oriented models. This is interesting because "[services] provided at a cost that communities can afford" – one of the most-repeated phrases from the Alma Ata Declaration on Primary Health Care (WHO 1978) – is the inspiration for the community-oriented model. This result may reflect the Canadian experience with community health centres, which is that whatever they may achieve in community-oriented attributes, their productivity with respect to medical services is generally much lower than that of professional models (Pineault et al. 2009). On the other hand, a direct efficiency comparison with professional models may be inappropriate, given the community-oriented models' fundamentally different driving vision of achieving population health benefits rather than focusing on general medical services.

System performance attributes are very important in both models. Interestingly, accountability and availability of resources are seen as essential to community-oriented models, while efficiency/productivity and quality improvement process are more important in professional models.

The implication for evaluation of community-oriented PHC organizations is that a broad scope of measures is required to evaluate the full array of functions considered essential. Evaluation of professional models can address a narrower spectrum of attributes. This distinction further suggests that different standards of performance might be expected from community-oriented and professional models, and evaluative efforts should compare similar models on specific attributes that are most relevant to the model family, using specifically tailored measures. However, all models should be evaluated for core attributes. We mapped the operational definition of all 24 attributes to subscales in 13 unique instruments that evaluate care from the patient perspective (report available from the authors). We found that most core attributes – except technical quality of care – were well covered by these instruments, and this special issue of *Healthcare Policy* provides evaluators with valuable information to guide their choices between instruments.

Measuring achieved outcomes for expected attributes – and not measuring non-expected attributes – would provide an appropriate evaluation of performance. More specific attributes should be better at discriminating among organizations within each model, while attributes with similar rankings of relevance should be better at discriminating among models.

This work also highlights attributes that can be monitored as indicators of change during the process of shifting from one PHC model to another, as is currently occurring in many jurisdictions. This suggests that the aims of PHC reforms should inform the choice of attributes to measure. Attributes not traditionally seen as relevant to an organizational model should be measured if reforms involve moving away from traditional ways of organizing PHC.

We found that attributes, and consequently indicators, are not equally relevant for different organizational models. Therefore, current evaluative tools should be reviewed in accordance with the specific organizational models being evaluated to provide more relevant information to guide health reforms. In addition, targeting operational attributes of PHC that are core to all models, as well as at those that are specific to some models or that are the focus of model transformations or innovations, should be a guiding principle in the design of evaluation and performance assessment strategies.

ACKNOWLEDGEMENTS

This paper is based on a presentation made during the North American Primary Care Research Group Conference in Quebec City on October 16, 2005. The authors, who accept full responsibility for the content of this paper, would like to thank anonymous participants for their comments and suggestions. Special thanks to Christine Beaulieu for research support throughout this survey.

This study was supported by a grant from the Canadian Institutes of Health Research (CIHR).

Correspondence may be directed to: Jean-Frédéric Lévesque, Direction de santé publique de Montréal/Institut national de santé publique du Québec, 1301 Sherbrooke East, Montreal, QC H2L 1M3; tel.: 514-528-2400 ext. 3216; fax: 514-528-2470; e-mail: jflevesq@santepub-mtl.qc.ca.

REFERENCES

Clair, M. 2000. *Les Solutions émergentes: Rapport et recommandations.* Quebec, QC: Government of Quebec, Commission d'étude sur les services de santé et les services sociaux.

Government of Alberta. 2001. *A Framework for Reform: Report on the Premier's Advisory Council on Health.* Don Mazankowski, PC, OC, ed. ISBN 0-7785-1547-8 Report; ISBN 0-7785-1548-6 Appendices. 2001. Retrieved May 9, 2011. <http://www.assembly.ab.ca/lao/library/egovdocs/alpm/2001/132279.pdf>.

Government of Ontario. 2000. *Looking Back, Looking Forward: The Ontario Health Services Restructuring Commission (1996–2000). A Legacy Report.* Toronto: Ontario Health Services Restructuring Commission.

Government of Saskatchewan. 2001. *Caring for Medicare: Sustaining a Quality System.* Regina: Saskatchewan Commission on Medicare.

Haggerty, J., F. Burge, J.-F. Lévesque, D. Gass, R. Pineault, M.-D. Beaulieu and D. Santor. 2007. "Operational Definitions of Attributes of Primary Health Care: Consensus among Canadian Experts." *Annals of Family Medicine* 5: 336–44.

Kirby, J.L. and M. LeBreton. 2002. *The Health of Canadians – The Federal Role: Volume 6: Recommendations for Reform.* Ottawa: Government of Canada, Standing Senate Committee on Social Affairs, Science and Technology.

Lamarche, P.-A., M.-D. Beaulieu, R. Pineault, A.-P. Contandriopoulos, J.-L. Denis and J. Haggerty. 2003. *Choices for Change: The Path for Restructuring Primary Healthcare Services in Canada.* Ottawa: Canadian Health Services Research Foundation.

Pineault, R., J.-F. Lévesque, D. Roberge, M. Hamel, P. Lamarche and J. Haggerty. 2009. *Accessibility and Continuity of Care: A Study of Primary Healthcare in Quebec. Research Report.* Montreal: Institut national de santé publique du Québec, Public Health Department of the Agence de la santé et des services sociaux de Montréal and Charles LeMoyne Hospital Research Centre.

Romanow, R.J. 2002. *Building on Values – The Future of Health Care in Canada. Final Report.* Ottawa: Government of Canada, Commission on the Future of Health Care in Canada.

World Health Organization (WHO). 1978. *Primary Health Care: Report of the International Conference on Primary Health Care, Alma-Ata, USSR, 6–12 September 1978.* Geneva: Author.

RESEARCH PAPER

Validation of Instruments to Evaluate Primary Healthcare from the Patient Perspective: Overview of the Method

Validation des instruments d'évaluation des soins primaires du point de vue des patients : aperçu de la méthode

JEANNIE L. HAGGERTY, PHD
Department of Family Medicine, McGill University
Montreal, QC

FREDERICK BURGE, MD, MSC
Department of Family Medicine, Dalhousie University
Halifax, NS

MARIE-DOMINIQUE BEAULIEU, MD, MSC
Chaire Dr Sadok Besrour en médecine familiale
Centre de recherche du Centre hospitalier de l'Université de Montréal
Montréal, QC

RAYNALD PINEAULT, MD, PHD
Centre de recherche du Centre hospitalier de l'Université de Montréal
Montréal, QC

CHRISTINE BEAULIEU, MSC
St. Mary's Research Centre, St. Mary's Hospital Center
Montreal, QC

JEAN-FRÉDÉRIC LÉVESQUE, MD, PHD
Centre de recherche du Centre hospitalier de l'Université de Montréal
Montréal, QC

DARCY A. SANTOR, PHD
School of Psychology, University of Ottawa
Ottawa, ON

DAVID GASS, MD
Department of Family Medicine, Dalhousie University
Halifax, NS

BEVERLEY LAWSON, MSC
Department of Family Medicine, Dalhousie University
Halifax, NS

Abstract

Patient evaluations are an important part of monitoring primary healthcare reforms, but there is little comparative information available to guide evaluators in the choice of instruments or to determine their relevance for Canada.

Objective: To compare values and the psychometric performances of validated instruments thought to be most pertinent to the Canadian context for evaluating core attributes of primary healthcare.

Method: Among validated instruments in the public domain, we selected six: the Primary Care Assessment Survey (PCAS); the Primary Care Assessment Tool – Short Form (PCAT-S); the Components of Primary Care Index (CPCI); the first version of the EUROPEP (EUROPEP-I); the Interpersonal Processes of Care Survey, version II (IPC-II); and part of the Veterans Affairs National Outpatient Customer Satisfaction Survey (VANOCSS). We mapped subscales to operational definitions of attributes. All were administered to a sample of adult service users balanced by English/French language (in Nova Scotia and Quebec, respectively), urban/rural residency, high/low education and overall care experience. The sample was recruited from previous survey respondents, newspaper advertisements and community posters. We used common factor analysis to compare our factor resolution for each instrument to that of the developers.

Results: Our sample of 645 respondents was approximately balanced by design variables, but considerable effort was required to recruit low-education and poor-experience respondents. Subscale scores are statistically different by excellent, average and poor overall experience, but interpersonal communication and respectfulness scores were the most discriminating of overall experience. We found fewer factors than did the developers, but when constrained to the number of expected factors, our item loadings were largely similar to those found by developers. Subscale reliability was equivalent to or higher than that reported by developers.

Conclusion: These instruments perform similarly in the Canadian context to their original development context, and can be used with confidence. Interpersonal and respectfulness scores are most discriminating of excellent, average or poor overall experience and are crucial dimensions of patient evaluations.

Résumé

L'évaluation des soins de santé primaires par les patients constituent un aspect important du suivi des reformes des soins de santé primaires, mais il existe peu d'information comparative pour orienter les évaluateurs dans le choix d'instruments ou pour déterminer leur pertinence au Canada.

Objectif : Comparer la valeur et la performance psychométrique des instruments valides et considérés comme les plus pertinents au contexte canadien pour l'évaluation des caractéristiques centrales des soins de santé primaires.
Méthode : Nous avons choisi six instruments validés du domaine public, soit : Primary Care Assessment Survey (PCAS); Primary Care Assessment Tool – version courte (PCAT-S); Components of Primary Care Index (CPCI); la première version de l'EUROPEP (EUROPEP-I); Interpersonal Processes of Care Survey, version II (IPC-II); et une partie du Veterans Affairs National Outpatient Customer Satisfaction Survey (VANOCSS). Les sous-échelles ont été mises en relation avec les définitions opérationnelles des caractéristiques. Tous ces instruments ont été appliqués a un échantillon d'adultes utilisateurs de services qui était équilibré en fonction de la langue (anglais : Nouvelle-Écosse ou français : Québec), du lieu de résidence urbain ou rural, du niveau de scolarité élevé ou faible et de l'expérience générale des soins. L'échantillon provenait de répondants recrutés grâce à un sondage antérieur, à des annonces dans les journaux et à des affiches dans des lieux communautaires. Nous avons procédé a une analyse factorielle normale pour comparer nos structures factorielles à celles des concepteurs, et ce, pour chacun des instruments.
Résultats : L'échantillon de 645 répondants était globalement équilibré en fonction des paramètres du devis, mais des efforts considérables ont été nécessaires pour trouver des répondants avec un faible niveau de scolarité et une faible expérience des soins. Les scores des sous-échelles sont statistiquement différents en fonction d'une expérience générale excellente, moyenne ou faible, mais les scores présentant la discrimination la plus forte pour l'expérience générale étaient ceux de la communication interpersonnelleé. Nous avons observé moins de facteurs que ne l'avaient fait les concepteurs, mais en forçant le nombre de facteurs au nombre attendu, la correspondance de nos items était très similaire à celle observée par les concepteurs. La fiabilité des sous-échelles était équivalente ou plus élevée que celle indiquée par les concepteurs.
Conclusion : Le rendement de ces instruments dans le contexte canadien est similaire à celui obtenu de leur contexte original, et ils peuvent être utilisés en toute confiance. Les scores de la communication interpersonnelle et du respect sont les plus discriminants d'une expérience générale excellente, moyenne ou faible, et ils constituent des aspects importants de l'évaluation par les patients.

* * *

INSTRUMENTS TO EVALUATE CARE FROM THE PATIENT'S PERSPECTIVE HAVE BEEN developed and validated elsewhere, but no comparative information is available on their performance in the Canadian context to guide researchers and policy makers in selecting one instrument over another. Our objective was to compare validated instruments thought to be most pertinent to the Canadian context. Specifically, we aimed to compare subscales from different instruments for the same attribute of care and to ensure that the instruments' reported psychometric properties were similar in the Canadian context. Program evaluators could then be confident of the tools' applicability to that context, and if different instruments were used, either at different times or in different jurisdictions, our results would provide a common benchmark for comparing relative scores. In this paper, we report in detail on how we selected the instruments, identified and recruited the study sample and administered the instruments. We provide general descriptive results and compare psychometric properties with those reported by the instrument developers.

Method

Ethical approval for this study was obtained from the Research Centre of the Université de Montréal Hospital (Quebec) and the Capital Health Research Ethics Board (Nova Scotia).

Identification and selection of instruments

We conducted an electronic search of the Medline and CINAHL databases in spring 2004 using as keywords "primary healthcare," "outcome and process assessment," "questionnaires" and "psychometrics." From identified instruments, we eliminated those used to screen for illnesses, functional health status or perceived outcomes of care for specific conditions (migraines, mental healthcare). We identified additional instruments by consulting with colleagues

TABLE 1. Subscales selected from six instruments retained for the study and their correspondence to attributes of PHC, in the order used in the study questionnaire, showing subscale as named by the instrument developer (number of items shown in parentheses). The last row names the scales excluded from this study.

Attribute of Care to which Subscale was Mapped	Primary Care Assessment Survey PCAS	Primary Care Assessment Tool PCAT	EUROPEP	Components of Primary Care Index CPCI	Interpersonal Processes of Care IPC	Veterans Affairs National Outpatient Community Services Survey VANOCSS
Accessibility	Organizational access (6)	First-contact access (4); First-contact utilization (3)	Organization of care (7)			
Relational Continuity	Contextual knowledge of patient (5); Visit-based continuity (2)	Ongoing care (4)		Accumulated knowledge (8); Preference for regular physician (5)		
Interpersonal Communication	Communication (6); Trust (8)		Clinical behaviour (16)	Interpersonal communication (6)	Elicitation, responsiveness, explanations (6); Patient-centred decision-making (4)	
Respectfulness	Interpersonal treatment (5)				Emotional support (4); Non-hurried, attentive (6); Perceived discrimination (4); Respectfulness (4); Respectfulness of office staff (4)	
Comprehensiveness of Services		Comprehensiveness (services available) (4)		Comprehensive care (6)		

TABLE 1. Continued

Attribute of Care to which Subscale was Mapped	Primary Care Assessment Survey PCAS	Primary Care Assessment Tool PCAT	EUROPEP	Components of Primary Care Index CPCI	Interpersonal Processes of Care IPC	Veterans Affairs National Outpatient Community Services Survey VANOCSS
Whole-Person Care (Community-Oriented Care)		Community orientation (3)		Community context (2)		
Management Continuity (Coordination)	Integration (6)	Coordination (4)		Coordination of care (8)		Overall coordination of care (6); Specialty provider access (4)
Subscales excluded from the Study	**Financial access (2); Physical examination (1); Preventive counselling (7)**	**Culturally competent (3); Coordination (information systems) (3); Family-centredness (3)**		**Advocacy (9); Family context (3)**	**Cultural sensitivity (2); Doctor's sensitivity to language (3); Office staff's sensitivity to language (2); Empowerment (3); Explain medications (2); Self-care (2)**	**Visit-based scales: Access/timeliness (7); Coordination of care at visit (5); Courtesy (2); Emotional support (4); Patient education information (7); Preferences (5)**

and scanning reference lists in published papers. When several instruments were derived from or inspired by a common instrument, for example, the General Practice Assessment Questionnaire derived from the Primary Care Assessment Survey, we retained only the parent instrument. We identified 13 unique validated instruments, on which we then obtained psychometric information from available publications or from the instrument developers.

Three instruments were visit-based, and the other 10 were retrospective, addressing usual care. We excluded the visit-based instruments and one that focused exclusively on satisfaction with all healthcare received – the Patient Satisfaction Questionnaire (PSQ-18) (Marshall and Hays 1994). Each researcher independently ranked the remaining nine instruments according to their relevance in the Canadian context, and we retained the six highest-ranked instruments: the Primary Care Assessment Survey (PCAS) (Safran et al. 1998); the Primary Care Assessment Tool – Short Form (PCAT-S) (Shi et al. 2001); the Components of Primary Care Index (CPCI) (Flocke 1997); the first version of the European general practice evaluation instrument EUROPEP-I (Grol et al. 2000); the Interpersonal Processes of Care Survey – 18-item version (IPC-II) (Stewart et al. 2007); and the Veterans Affairs National Outpatient Customer Satisfaction Survey (VANOCSS) (Borowsky et al. 2002). Permission to use the instruments was obtained from all instrument developers.

Because our objective was to compare measures by attribute of care, we further retained only the subscales of attributes measured in more than one instrument. For example, we dropped the Advocacy subscale from the CPCI because it is measured only in this instrument. The six instruments and the subscales retained for this study are listed in Table 1.

Study population

Our target population was English- and French-speaking adult primary healthcare (PHC) users in Canada, undifferentiated by age, health condition, geographic location or level of functional literacy. Eligible subjects were adults (≥18 years) with a regular source of PHC that they had consulted in the previous 12 months. We maximized the statistical efficiency for conducting subgroup comparisons by balancing our study sample by English/French language, urban/rural location and educational level. We also stratified by excellent, average and poor primary care experience based on a single screening question: "Overall, has your experience of care from your regular family doctor or medical clinic been excellent, poor or average?" The sample size was designed to provide statistical power for factor analysis of up to 150 items with 25 subjects in each sampling cell.

We used educational achievement as a proxy for literacy above and below high-school reading level. Because the association between literacy and education varies as a function of age, we used an age-sensitive cut-off for high school: completed high school, if under age 45; completed 10 years, for ages 45 to 55; and less than eight years, if over age 55 (Smith and Haggerty 2003). Urban location was defined as residing in a census metropolitan area; rural, in areas more than one hour's travel from a metropolitan area; and remote (Quebec only), in areas more than four hours' travel from the nearest metropolitan area. Questionnaires were administered exclusively in English in Nova Scotia and in French in Quebec.

Subjects were recruited by various means. Our goal was to achieve representativeness of the sampling strata, not of the population as a whole. We initially used a sampling frame of persons from previous PHC surveys who had agreed to future contact: 647 from a 2002 clinic-based survey in Quebec (Haggerty et al. 2007) and 1,247 from a 2005 telephone survey in Nova Scotia. Eligibility for different strata was determined from screening questions administered by telephone or e-mail.

Owing to difficulties in recruiting low-literacy participants and those with poor experience of care, we obtained ethical approval in Quebec only to expand recruitment strategies to newspaper advertisements, then community posters in laundromats, grocery stores, recreation centres and health centres and, finally, word of mouth. All participants were offered compensation for completing the questionnaire.

Data collection

The study questionnaire consisted of the retained subscales from the six selected instruments (153 items, 28 specific to care from multiple providers), as well as socio-demographic and utilization information (total, 198 questions). Instruments were formatted in their original form, and subscales were grouped by instrument family in the sequence shown in Table 1. The VANOCSS was placed last because it was specific to those who had seen multiple providers.

Participants were offered either paper-based or online response modalities. To maximize response, we used a protocol of two reminder postcards or e-mails at two-week intervals, followed by a second posting of the questionnaire, then phone calls. A subset of participants completed the questionnaire in a group setting where they could be observed directly and then participated in a 30- to 45-minute discussion; these results are reported elsewhere (Haggerty, Beaulieu et al. 2011). Data were collected between February and July 2005.

Analysis

We analyzed our recruitment descriptively by substrata in terms of the success of different recruitment strategies and differential response rates.

The details of individual subscales are presented in the attribute-specific papers elsewhere in this special issue of the journal. We expressed the value of each subscale as the mean of the values of the items. Thus, the mean of several items with a 1-to-5 Likert response scale varied between 1 and 5 regardless of the number of items in the subscale. We calculated the internal reliability of each subscale using Cronbach's alpha.

We used one-way ANOVA tests to determine whether subscale means differed significantly among respondents with poor, average and excellent overall experience of care; we used discriminant analysis to examine the magnitude of the Fisher linear discriminant (F-test) as an indicator of which subscale score best differentiates between the groups. We also conducted exploratory factor analysis using common factor analysis for each instrument, to determine how many factors emerged with an eigenvalue >1. We repeated the analysis forcing the number of factors found by instrument developers, then examined whether item loading accorded with that identified by the developer. Factor analysis used only observations with no missing values on any item (listwise missing), but we repeated the analyses, imputing for missing values by using either maximum likelihood within the subscale (Jöreskog and Sörbom 1996) or the developer's suggested imputation algorithm.

Results

Recruitment of study population

Of the 647 Quebec residents in the initial sampling frame, the first 208 who met the eligibility criteria for specific strata were selected for telephone contact; 168 had still-active telephone numbers, and 38% (62/168) agreed to participate. Of these, 85% (53/62) returned the questionnaire. Of the 1,247 persons in Nova Scotia, 290 had provided e-mail addresses and were invited by e-mail without being pre-screened, of which 112 (38.5%) responded to the questionnaire. The final overall response rates were similar. While the telephone strategy (Quebec) was more resource-intensive, the resulting sample corresponded more closely to the desired design; the e-mail strategy (Nova Scotia) oversampled high-education respondents (91% vs. the 50% desired).

We had difficulty recruiting eligible subjects with low education and/or poor experience of care from their regular provider. Advertising in local newspapers (Quebec only) was most cost-effective in urban areas. Posters in laundromats, grocery stores, community recreation centres and credit unions were effective for reaching low-education participants in urban areas, but not rural areas. This method required few resources but provided a steady trickle of responses. In both provinces, peer recruitment by word of mouth (snowball sampling) was the most effective strategy for targeted recruitment in rural areas and among people with low educational attainment.

Table 2 presents the final sample size and distribution by sampling design variables. The sample distribution was more balanced in the design variables in Quebec (French) than in Nova Scotia (English). Additionally, the Nova Scotia sample was in better health than the Quebec sample and more likely to be affiliated to a family doctor and for a longer time, to concentrate care among fewer unique family physicians and to have shorter waits for care (details presented elsewhere in this special issue of the journal, Haggerty, Bouharaoui et al. 2011).

TABLE 2. Final recruitment of study subjects by design variables; original aim was for 25 subjects per cell

Prior Experience with Primary Care	French (n=302, 46%)				English (n=343, 53%)				Total
	Urban (n=148, 49%)		Rural (n=154, 51%)		Urban (n=203, 59%)		Rural (n=140, 41%)		
	Low Education	High Education	Low Education	High Education	Low Education	High Education	Low Education	High Education	
Excellent	31	31	28	32	24	66	11	41	264 (41%)
Average	22	31	28	31	14	57	11	39	233 (36%)
Poor	9	24	17	18	10	32	16	22	148 (23%)
Total	62 (21%)	86 (28%)	73 (24%)	81 (27%)	48 (14%)	155 (45%)	38 (11%)	102 (30%)	

Of the 645 respondents, 130 (20.2%) responded to the online version of the questionnaire: 25% in urban areas and 14% in rural areas ($\chi^2=11.6$, $p=.0007$). Of the high-education participants, 26.9% responded online, compared to 7.2% of low-education participants ($\chi^2=34.9$, $p<.0001$). There was no difference in subscale scores by response modality after controlling for language, geographic location and educational status.

Table 3 presents the sample characteristics and compares them with respect to their reported overall experience of care. Compared to those participants with just average or poor experience, those with excellent experience are more likely to be in better health, to be affiliated with a physician rather than a clinic (and with longer affiliations), to have seen fewer unique physicians in the year and to report shorter waits for appointments.

TABLE 3. Characteristics of the study sample and comparison of subjects by overall experience of care

Characteristic	Total (n=645)	Overall Experience of Care			Test for Difference
		Excellent (n=264)	Average (n=232)	Poor (n=149)	
Personal Characteristics					
Average age	48.0 (14.9)	48.4 (14.9)	47.6 (14.3)	47.8 (15.8)	NS
Per cent female	64.6 (414)	63.7 (167)	65.8 (152)	64.6 (95)	NS
Per cent indicating health status as good or excellent	37.8 (241)	43.0 (113)	37.3 (85)	29.9 (43)	$x^2=6.9$; *df* 2 $p=.03$

TABLE 3. Continued

		Overall Experience of Care			
Characteristic	**Total (n=645)**	**Excellent (n=264)**	**Average (n=232)**	**Poor (n=149)**	**Test for Difference**
Per cent with disability	31.6 (200)	29.5 (77)	32.4 (73)	33.8 (50)	NS
Per cent with chronic health problem*	61.6 (392)	61.1 (160)	60.5 (138)	64.4 (94)	NS
Healthcare Use					
Regular provider: Physician	94.1 (607)	97.4 (257)	92.7 (215)	90.6 (135)	$x^2=9.2$; *df* 2
Clinic only	5.9 (38)	2.7 (7)	7.3 (17)	9.4 (14)	$p=.01$
Mean number of years of affiliation	11.2 (9.0)	11.9 (10)	11.3 (8.5)	9.7 (7.8)	NS
Mean number of primary care visits in last 12 months	6.3 (7.0)	7.1 (8.3)	4.9 (4.6)	7.1 (7.3)	$F=6.9$; *df* 2 $p=.001$
Mean number of unique general or family physicians seen	2.0 (1.3)	1.8 (1.1)	2.0 (1.5)	2.3 (1.5)	$F=8.3$; *df* 2 $p=.003$
Usual wait time for appointment					
Less than 2 days	35.2 (220)	47.3 (123)	30.6 (68)	20.3 (29)	
2 to 7 days	32.6 (204)	28.5 (74)	37.4 (83)	32.9 (47)	$x^2=45$; *df* 8
7 days to 2 weeks	11.8 (74)	9.2 (24)	9.0 (20)	21.0 (30)	$p<.0001$
2 weeks to 4 weeks	9.3 (58)	5.8 (15)	11.7 (26)	11.9 (17)	
More than 4 weeks	11.0 (69)	9.2 (24)	11.3 (25)	14.0 (20)	
Usual wait time in waiting room before clinical visit					
Less than 15 minutes	34.7 (218)	37.6 (99)	38.7 (87)	22.7 (32)	
15 to 29 minutes	38.8 (244)	39.2 (103)	35.6 (80)	43.3 (61)	$x^2=15.8$; *df* 6
30 to 59 minutes	19.9 (125)	19.0 (50)	18.7 (42)	23.4 (33)	$p=.02$
More than an hour	6.7 (42)	4.2 (11)	7.1 (16)	10.6 (15)	

* Percentage indicating they had been told by a doctor that they had any of the following: high blood pressure, diabetes, cancer, depression, arthritis, respiratory disease, heart disease.

Comparison of instrument scores

The subscale scores grouped by PHC attribute are presented in Table 4. Several points are noteworthy. First, with few exceptions, the score distributions are skewed; the median is higher than the mean, indicating that the mass of the distribution is in the positive zone of assessment. Second, the back-to-back placement of scores demonstrates the challenge of comparing scores even within the same group of respondents, let alone between groups or jurisdictions. Third, the subscale means differ significantly by levels of overall experience, as shown in the last two columns of Table 4. All subscales scores, except the VANOCSS Specialty Provider Access, distinguish between poor and excellent care; the vast majority, between poor and average and between average and excellent care. The magnitude of the Fisher test shows that subscales for interpersonal communication and respectfulness provide the greatest discrimination among poor, average and excellent overall experience of care. Average Fisher test values are 66.5 and 55.8, respectively, compared to average values in the 20s and 30s for other attribute families.

Psychometric properties

In Table 5, the subscales are grouped within their instrument families. Note that the Cronbach's alphas reported by the developers are similar to those observed in our sample. For factor analysis, with the exception of the EUROPEP, the number of factors observed by common factor analysis was approximately half that expected (item loading available on request). When we constrained the factor resolution to the number of factors found by the instrument developer, the item loading corresponded generally to that identified by the developer. The observed factor solutions deviated most from the expected for the CPCI and PCAT-S instruments. The deviation for the CPCI may be explained by halo effects related to the instrument's format and response scale, and for the PCAT-S, by problems related to missing values – a case that merits additional explanation.

TABLE 4. Subscale values, grouped by attribute of care, showing comparison of statistically significant differences in mean values by overall experience of care

				Raw Values			Heathcare Experience*	*F*-Test of Discrimination
Instrument	**Developer's Subscale Name**	**# Items**	**Likert Response Range**	**Mean**	**Median**	**SD**	**Poor Average Excellent**	
Accessibility								
PCAS	Organizational Access	6	1 to 6	3.97	4.00	0.92	4,85 5,73 6,74	62.6
PCAT	First-Contact Accessibility	4	1 to 4	2.68	2.75	0.78	4,51 5,32 6,45	11.8
PCAT	First-Contact Utilization	3	1 to 4	3.73	4.00	0.48	8,71 8,94 9,44	13.8
EUROPEP	Organization of Care	7	1 to 5	3.61	3.71	0.90	5,01 6,27 7,59	29.9
Comprehensiveness								
PCAT	Comprehensiveness (services available)	4	1 to 4	3.32	3.50	0.74	7,35 7,52 8,08	4.9
CPCI	Comprehensive Care	6	1 to 6	4.86	5.00	1.10	6,73 7,54 8,43	35.5
Interpersonal Communication								
PCAS	Communication	6	1 to 6	4.66	4.83	1.05	5,85 6,98 8,45	101.4
PCAS	Trust	8	1 to 5	4.01	4.13	0.71	6,28 7,28 8,46	96.9
CPCI	Interpersonal Communication	6	1 to 6	4.59	4.83	1.16	5,87 6,90 8,16	57.2
EUROPEP	Clinical Behaviour	16	1 to 5	4.14	4.33	0.83	6,48 7,61 8,86	79.9

TABLE 4. Continued

				Raw Values			Heathcare Experience*	*F*-Test of Discrimination
Instrument	**Developer's Subscale Name**	**# Items**	**Likert Response Range**	**Mean**	**Median**	**SD**	**Poor Average Excellent**	
IPC-II	Communication (elicited concerns, responded)	3	1 to 5	4.12	4.33	0.87	6,55 7,54 8,77	61.2
IPC-II	Communication (explained results, medications)	4	1 to 5	3.96	4.25	1.00	6,24 7,03 8,39	42.9
IPC-II	Decision-Making (patient-centred decision-making)	4	1 to 5	3.17	3.25	1.26	4,71 4,89 6,30	16.8
Management Continuity								
PCAS	Integration	6	1 to 6	4.45	4.67	1.00	5,74 6,64 7,80	42.7
PCAT	Coordination	4	1 to 4	3.27	3.50	0.80	6,61 7,38 8,30	18.0
CPCI	Coordination of Care	8	1 to 6	4.30	4.38	1.00	5,80 6,27 7,34	34.4
VANOCSS	Coordination of Care (overall), number of problems	6	0 to 6	2.51	2.00	1.88	5,05 5,66 6,47	4.7
VANOCSS	Specialty Provider Access (number of problems)	4	0 to 4	0.62	0.00	0.90	–	
Relational Continuity								
PCAS	Visit-Based Continuity	2	1 to 6	5.17	5.50	1.05	7,54 8,44 8,72	15.5
PCAS	Contextual Knowledge	5	1 to 6	3.96	4.10	1.14	4,67 5,55 6,93	60.5
PCAT	Ongoing Care	4	1 to 4	3.15	3.25	0.70	5,94 6,89 8,06	46.1
CPCI	Accumulated Knowledge	8	1 to 6	4.50	4.75	1.24	5,84 6,58 7,99	45.1
CPCI	Patient Preference for Regular Physician	5	1 to 6	4.84	5.00	1.00	6,86 7,53 8,27	29.3
Respectfulness								
PCAS	Interpersonal Treatment	5	1 to 6	4.72	4.90	1.08	5,90 7,14 8,56	97.3
IPC-II	Hurried Communication	5	1 to 5	4.20	4.37	0.71	7,01 7,90 8,68	48.1

TABLE 4. Continued

				Raw Values			Heathcare Experience*	*F*-Test of Discrimination
Instrument	**Developer's Subscale Name**	**# Items**	**Likert Response Range**	**Mean**	**Median**	**SD**	**Poor Average Excellent**	
IPC-II	Interpersonal Style (compassionate, respectful)	5	1 to 5	4.21	4.60	0.9	6,57 7,86 9,01	60.4
IPC-II	Interpersonal Style (respectful office staff**)	4	1 to 5	4.51	5.00	0.73	8,05 8,84 9,14	17.4
Whole-Person Care – Community Context								
PCAT	Community Orientation	3	1 to 4	2.47	2.50	0.86	3,75 4,60 5,79	24.8
CPCI	Community Context	2	1 to 6	4.23	4.50	1.56	5,07 6,21 7,46	30.2

* Means by group only presented where difference statistically significant at $p<.01$.

** Subscale reversed as well as normalized; raw value indicates frequency of disrespectful behaviour. Consequently, the normalized score of 10 = never disrespectful, 0 = always disrespectful.

The PCAT-S offers five response options to desirable characteristics in PHC: 1 = definitely not; 2 = probably not; 3 = probably; 4 = definitely, and "don't know/not sure." Processed classically, the "don't know" response counts as a missing value, yielding us only 146/645 valid observations for factor analysis. The developer suggests replacing this latter response with a value of 2 (probably not) for respondents with at least 50% true values within the subscale, based on the logic that when patients are unsure of service attributes at the clinic, this reflects negatively on the provider. Using the developer's replacement algorithm yielded 470 observations, and the factor resolution corresponded more closely to that of the developer, although the grouping of items in factors 3 and 6 (Table 5) persisted, suggesting a construct overlap between first-contact accessibility and community orientation, and between first-contact accessibility and ongoing care (details available on request).

TABLE 5. Reported and observed internal consistency (Cronbach alpha) and factor resolution by instrument, showing observed factors with eigenvalue > 1 and factor solution when constrained to expected number

Instrument and Subscale (number items)	Mapped Attributes	Reported Alpha	Observed Alpha	Solution of Expected Number of Factors (eigen) Subscales
Primary Care Assessment Survey (PCAS)				*Expected=6, Observed=4 (n=377)*
Organizational Access (6)	Accessibility	.84	.83	(17.45) Communication + Interpersonal Treatment
Visit-Based Continuity (2)	Relational Continuity	—	.69	(1.98) Contextual Knowledge
Contextual Knowledge (5)	Relational Continuity	.92	.91	(1.48) Integration

TABLE 5. Continued

Instrument and Subscale (number items)	Mapped Attributes	Reported Alpha	Observed Alpha	Solution of Expected Number of Factors (eigen) Subscales
Communication (6)	Interpersonal Communication	.95	.96	(1.06) Organizational Access
Trust (8)	Interpersonal Communication	.86	.88	(0.90) 4/8 Trust
Interpersonal Treatment (5)	Respectfulness	.95	.96	(0.65) 4/8 Trust
Integration (6)	Management Continuity	.92	.93	(0.51) Visit-based continuity
Primary Care Assessment Tool (PCAT)				*Expected=6, Proposed=3 (n=470)*
First-Contact Utilization (3)	Accessibility / Comprehensiveness	TBD	.68	(5.01) Coordination
First-Contact Access	Accessibility		.72	(1.40) 3/4 Ongoing Care
Comprehensiveness (services available) (4)	Comprehensiveness of Services		.72	(0.86) Comprehensive Services
Ongoing Care (4)	Relational Continuity		.73	(0.63) 2/4 First-Contact Access + 1/4 Ongoing Care (telephone)
Coordination (4)	Management Continuity		.76	(0.51) Community Orientation + 2/4 First-Contact Access
Community Orientation (3)	Whole-Person Care		.65	(0.40) First-Contact Utilization
Components of Primary Care Instrument (CPCI)				*Expected=6, Proposed=3 (N=487)*
Comprehensive Care (6)	Comprehensiveness of Services	.79	.83	(13.75) Community Context + 6/8 Coordination + 1/5 Preference
Accumulated Knowledge (8)	Relational Continuity	.88	.91	(1.29) 7/8 Accumulated Knowledge + 1/6 Communication
Preference for Regular Physician (5)	Relational Continuity	.71	.68	(1.15) 5/6 Comprehensive
Interpersonal Communication (6)	Interpersonal Communication	.75	.83	(0.93) 5/6 Communication
Coordination of Care (8)	Management Continuity	.92	.74	(0.85) 4/5 Preference + 1/8 Coordination
Community Context (2)	Whole-Person Care	—	.82	(0.51) 2/8 Coordination
EUROPEP				*Expected=2, Proposed=2 (n=355)*
Organization of Care (7)	Accessibility	.87	.89	(13.62) Clinical Behaviour
Clinical Behaviour (16)	Interpersonal Communication	.96	.97	(1.56) Organization of Care
Interpersonal Processes of Care (IPC-II)				*Expected=6, Proposed=3 (n=536)*
Elicit Concerns, Respond (3)	Interpersonal Communication	.80	.86	(11.92) Compassionate + (3/5) Non-Hurried, Attentive
Explain Results, Medications (4)	Interpersonal Communication	.81	.88	(2.61) Decision-Making
Decision-Making (4)	Interpersonal Communication	.75	.91	(1.36) Respectful Office Staff
Non-Hurried, Attentive (5)	Respectfulness	.65	.95	(0.79) Explain Results
Compassionate, Respectful (5)	Respectfulness	.71	.95	(0.57) Non-Hurried, Attentive (3/5 load equally with factor 1)
Respectful Office Staff (5)	Respectfulness	.90	.93	(0.39) Elicit Concerns
Veterans Affairs Outpatient Community Services Survey				*NB: Dichotomous scoring of items, factor analysis not applicable*
Management Continuity (6)	Overall Coordination of Care		NA	
Management Continuity (4)	Specialty Provider Access		NA	

Discussion and Conclusion

We found that relevant subscales from generic PHC evaluation instruments demonstrate general psychometric properties in a Canadian sample that are similar to those observed in the United States and Europe, where the instruments were developed. Despite important differences in PHC organization among countries, our results suggest that Canadian program evaluators and researchers can confidently rely on the reported psychometric properties of these instruments for evaluating PHC attributes.

Almost all the subscales demonstrate skewed distribution, regardless of whether the response type is reporting or rating. We would expect the skewing to be even more extreme in a representative sample of the population that was not selected to balance the sample by overall experience of care, as ours was. This skewing has been demonstrated consistently in other studies (Crow et al. 2002) and is a major challenge in program evaluation. Qualitative studies suggest that patients are reluctant to report negative assessments of care even when not entirely satisfied, unless clear responsibility can be attributed to the source of the negative experience (Collins and O'Cathain 2003). This means that positive assessments will reflect a mix of experiences ranging from only adequate to excellent, and therefore have low sensitivity and specificity. Negative assessments, by contrast, tend to be true negatives, indicating good specificity of negative scores. Thus, in reports to decision-makers about PHC performance, it may be more informative and accurate to report the percentage of less-than-positive scores, rather than masking the negative scores within generally positive average scores.

Our recruitment experience illustrates the difficulty of including low-literacy subjects in surveys of healthcare experience. These patients are not reached easily by written material. Yet their participation in evaluations is important, because low literacy is an independent health risk (Smith and Haggerty 2003), and subjects will be more dependent than high-literacy subjects on their doctors' actions and advice (Bostick et al. 1994; Fiscella et al. 2002; Breitkopf et al. 2004; Willems et al. 2005). We found that for the most part, these instruments function equivalently in low-literacy and high-literacy responders (Haggerty, Bouharaoui et al. 2011), further highlighting the importance of reaching these patient groups.

All the instrument subscales distinguish between different levels of overall experience of care, but interpersonal communication and respectfulness are the most discriminating. This finding has important policy implications. The implication for policy makers is that public support for proposed healthcare innovations will suffer if reforms interfere with providers' capacity to attend to interpersonal communication and respectfulness. These attributes were not targeted for accountability within the Health Accords (Health Canada 2003) nor for renewal in the Primary Health Care Transition Fund (Health Canada 2007), but they are of critical importance to patients, and it is crucial to ensure that reforms not be implemented at their expense.

ACKNOWLEDGEMENTS

This research was funded by the Canadian Institute for Health Research. During this study Jeannie L. Haggerty held a Canada Research Chair in Population Impacts of Healthcare at the Université de Sherbrooke. The authors wish to thank Beverley Lawson for conducting the survey in Nova Scotia and Christine Beaulieu in Quebec, and Donna Riley for support in preparation and editing of the manuscript.

Correspondence may be directed to: Jeannie L. Haggerty, Associate Professor, Department of Family Medicine, McGill University, St. Mary's Research Centre, Hayes Pavilion – Suite 3734, 3830 Lacombe Ave., Montreal QC H3T 1M5; tel.: 514-345-3511 ext. 6332; fax: 514-734-2652; E-mail: jeannie.haggerty@mcgill.ca.

REFERENCES

Borowsky, S.J., D.B. Nelson, J.C. Fortney, A.N. Hedeen, J.L. Bradley and M.K. Chapko. 2002. "VA Community-Based Outpatient Clinics: Performance Measures Based on Patient Perceptions of Care." *Medical Care* 40(7): 578–86.

Bostick, R.M., J.M. Sprafka, B.A. Virnig and B.A. Potter. 1994. "Predictors of Cancer Prevention Attitudes and Participation in Cancer Screening Examinations." *Preventive Medicine* 23: 816–26.

Breitkopf, C.R., J. Catero, J. Jaccard and A.B. Berenson. 2004. "Psychological and Sociocultural Perspectives on Follow-up of Abnormal Papanicolaou Results." *Obstetrics and Gynecology* 104: 1347–54.

Collins, K. and A. O'Cathain. 2003. "The Continuum of Patient Satisfaction – from Satisfied to Very Satisfied." *Social Science and Medicine* 57: 2465–70.

Crow, R., H. Gage, S. Hampson, J. Hart, A. Kimber, L. Storey et al. 2002. "The Measurement of Satisfaction with Healthcare: Implications for Practice from a Systematic Review of the Literature." *Health Technology Assessment* 6: 1–244.

Fiscella, K., M.A. Goodwin and K.C. Stange. 2002. "Does Patient Educational Level Affect Office Visits to Family Physicians?" *Journal of the National Medical Association* 94: 157–65.

Flocke, S. 1997. "Measuring Attributes of Primary Care: Development of a New Instrument." *Journal of Family Practice* 45(1): 64–74.

Grol, R., M. Wensing and Task Force on Patient Evaluations of General Practice. 2000. *Patients Evaluate General/Family Practice: The EUROPEP Instrument.* Nijmegen, Netherlands: Centre for Quality of Care Research, Raboud University.

Haggerty, J.L., R. Pineault, M.-D. Beaulieu, Y. Brunelle, J. Gauthier, F. Goulet et al. 2007. "Room for Improvement: Patient Experience of Primary Care in Quebec Prior to Major Reforms." *Canadian Family Physician* 53: 1056–57.

Haggerty, J.L., C. Beaulieu, B. Lawson, D.A. Santor, M. Fournier and F. Burge. 2011. "What Patients Tell Us about Primary Healthcare Evaluation Instruments: Response Formats, Bad Questions and Missing Pieces." *Healthcare Policy* 7 (Special Issue): 66–78.

Haggerty, J.L., F. Bouharaoui and D.A. Santor. 2011. "Differential Item Functioning in Primary Healthcare Evaluation Instruments by French/English Version, Educational Level and Urban/Rural Location." *Healthcare Policy* 7 (Special Issue): 47–65.

Health Canada. 2003. *First Ministers' Health Accord on Health Care Renewal.* Ottawa: Author.

Health Canada. 2007. *Primary Health Care Transition Fund.* Retrieved May 10, 2011. <http://www.hc-sc.gc.ca/hcs-sss/prim/phctf-fassp/index-eng.php>.

Jöreskog, K.G. and D. Sörbom. 1996. *LISREL 8: User's Reference Guide.* Lincolnwood, IL: Scientific Software International.

Marshall, G.N. and R.D. Hays. 1994. *The Patient Satisfaction Questionnaire – Short Form (PSQ-18).* Report no. P-7865. Santa Monica, CA: Rand.

Safran, D.G., J. Kosinski, A.R. Tarlov, W.H. Rogers, D.A. Taira, N. Lieberman and J.E. Ware. 1998. "The Primary Care Assessment Survey: Tests of Data Quality and Measurement Performance." *Medical Care* 36(5): 728–39.

Shi, L., B. Starfield and J. Xu. 2001. "Validating the Adult Primary Care Assessment Tool." *Journal of Family Practice* 50(2): n161w–n171w.

Smith, J.L. and J. Haggerty. 2003. "Literacy in Primary Care Populations: Is It a Problem?" *Canadian Journal of Public Health* 94: 408–12.

Stewart, A.L., A.M. Nápoles-Springer, S.E. Gregorich and J. Santoyo-Olsson. 2007. "Interpersonal Processes of Care Survey: Patient-Reported Measures for Diverse Groups." *Health Services Research* 42(3 Pt. 1): 1235–56.

Willems, S., S. De Maesschaick, M. Deveugele, A. Derese and J. De Maeseneer. 2005. "Socio-Economic Status of the Patient and Doctor–Patient Communication: Does It Make a Difference?" *Patient Education and Counselling* 56: 139–46.

RESEARCH PAPER

Differential Item Functioning in Primary Healthcare Evaluation Instruments by French/English Version, Educational Level and Urban/Rural Location

Le fonctionnement différentiel des items dans les instruments d'évaluation des soins de santé primaires en fonction des versions française ou anglaise, du niveau de scolarisation et du lieu de résidence urbain ou rural

JEANNIE L. HAGGERTY, PHD
Department of Family Medicine, McGill University
Montreal, QC

FATIMA BOUHARAOUI, MSC
St. Mary's Research Centre, St. Mary's Hospital Center
Montreal, QC

DARCY A. SANTOR, PHD
School of Psychology, University of Ottawa
Ottawa, ON

Abstract

Evaluating the extent to which groups or subgroups of individuals differ with respect to primary healthcare experience depends on first ruling out the possibility of bias.

Objective: To determine whether item or subscale performance differs systematically between French/English, high/low education subgroups and urban/rural residency.

Method: A sample of 645 adult users balanced by French/English language (in Quebec and Nova Scotia, respectively), high/low education and urban/rural residency responded to six validated instruments: the Primary Care Assessment Survey (PCAS); the Primary Care Assessment Tool – Short Form (PCAT-S); the Components of Primary Care Index (CPCI); the first version of the EUROPEP (EUROPEP-I); the Interpersonal Processes of Care Survey, version II (IPC-II); and part of the Veterans Affairs National Outpatient Customer Satisfaction Survey (VANOCSS). We normalized subscale scores to a 0-to-10 scale and tested for between-group differences using ANOVA tests. We used a parametric item response model to test for differences between subgroups in item discriminability and item difficulty. We re-examined group differences after removing items with differential item functioning.
Results: Experience of care was assessed more positively in the English-speaking (Nova Scotia) than in the French-speaking (Quebec) respondents. We found differential English/French item functioning in 48% of the 153 items: discriminability in 20% and differential difficulty in 28%. English items were more discriminating generally than the French. Removing problematic items did not change the differences in French/English assessments. Differential item functioning by high/low education status affected 27% of items, with items being generally more discriminating in high-education groups. Between-group comparisons were unchanged. In contrast, only 9% of items showed differential item functioning by geography, affecting principally the accessibility attribute. Removing problematic items reversed a previously non-significant finding, revealing poorer first-contact access in rural than in urban areas.
Conclusion: Differential item functioning does not bias or invalidate French/English comparisons on subscales, but additional development is required to make French and English items equivalent. These instruments are relatively robust by educational status and geography, but results suggest potential differences in the underlying construct in low-education and rural respondents.

Résumé

Afin d'évaluer à quel point des groupes ou sous-groupes d'individus divergent quant à leur expérience en matière de soins de santé primaires, il faut d'abord éliminer les possibilités de biais.
Objectif : Déterminer si la performance d'un item ou d'une sous-échelle diffère systématiquement en fonction de la langue (français/anglais), des sous-groupes de scolarisation (élevée/faible) et du lieu de résidence (urbain/rural).
Méthode : Un échantillon de 645 adultes utilisateurs, équilibré en fonction de la langue (français : Québec et anglais : Nouvelle-Écosse), du niveau de scolarisation (élevé/faible) et du lieu de résidence (urbain/rural), a répondu aux six instruments validés suivants : Primary Care Assessment Survey (PCAS); Primary Care Assessment Tool – version courte (PCAT-S); Components of Primary Care Index (CPCI); la première version de l'EUROPEP (EUROPEP-I); Interpersonal Processes of Care Survey, version II (IPC-II); et une partie du Veterans Affairs National Outpatient Customer Satisfaction Survey (VANOCSS). Nous avons normalisé les scores des sous-échelles selon une échelle de 0 à 10 et nous avons vérifié les différences entre les sous-groupes au moyen de tests ANOVA. Nous avons utilisé un

modèle paramétrique de la théorie des réponses par items (IRT) pour tester les différences entre les sous-groupes selon le pouvoir discriminant des items et leur niveau de difficulté. Puis, nous avons réexaminé les différences entre les groupes après avoir retiré les items qui présentaient un fonctionnement différentiel (DIF).

Résultats : L'expérience de soins a été évaluée plus positivement au sein du groupe anglophone (Nouvelle-Écosse) par rapport au groupe francophone (Québec). Nous avons observé un fonctionnement différentiel d'item selon la langue anglais/français dans 48 % des 153 items : une discrimination différentielle dans 20 % des cas et une difficulté différentielle dans 28 % des cas. Les items anglais étaient généralement plus discriminants que les items français. Il n'y a pas eu de changement des différences français/anglais observées après le retrait des items problématiques. Le fonctionnement différentiel des items selon le niveau de scolarisation (élevé/faible) affectait 27 % des items, qui étaient généralement plus discriminants pour les groupes de scolarisation élevée. Les comparaisons entre les groupes n'ont pas montré de changement. Par contre, seulement 9 % des items montraient un fonctionnement différentiel en fonction du lieu géographique, affectant principalement l'accessibilité. Le retrait des items problématiques a provoqué le renversement d'un résultat préalablement non significatif, révélant un plus faible accès de premier contact dans les zones rurales par rapport aux zones urbaines.

Conclusion : Le fonctionnement différentiel des items ne cause pas de biais ou n'invalide pas les comparaisons français/anglais par sous-échelle, mais une adaptation supplémentaire est nécessaire pour développer des items équivalents en français et en anglais. Ces instruments sont relativement robustes en fonction du niveau de scolarisation et du lieu géographique, mais les résultats suggèrent des différences potentielles dans le construit sous-jacent, pour les répondants de niveau de scolarisation plus faible et des zones rurales.

* * *

EXAMINING GROUP DIFFERENCES IN HEALTHCARE EXPERIENCE, WHETHER ACROSS geographic locations or linguistic/ethnic groups, is essential to ensuring that healthcare is delivered as equitably and effectively as possible. However, observed differences between two groups do not necessarily imply true differences unless it can be demonstrated that the evaluation scales and measures function similarly in both groups. To interpret group differences, we must first rule out any bias in how individuals answer questions.

Differential item functioning (sometimes called item bias) occurs when, at the same level of the underlying construct, responses differ significantly by group membership. If several items in a subscale demonstrate differential item functioning, this may adversely affect the conclusions of between-group comparisons by creating a false difference or failing to detect a true difference.

Among instruments developed to measure the quality of primary healthcare from the patient's perspective, we identified six in the public domain that appeared of greatest relevance for Canada: the Primary Care Assessment Survey (PCAS) (Safran et al. 1998); the adult Primary Care Assessment Tool – Short form (PCAT-S) (Shi et al. 2001); the Components of Primary Care Index (CPCI) (Flocke 1997); the first version of the EUROPEP (EUROPEP-I) (Wensing et al. 2000); the Interpersonal Processes of Care Survey, version II (IPC-II) (Stewart et al. 2007); and the Veterans Affairs National Outpatient Customer Satisfaction Survey (VANOCSS) (Borowsky et al. 2002).

We wanted to determine whether French- and English-language versions of the instruments were equivalent and whether item or subscale performance differed systematically by high/low educational status or by urban/rural location.

Specific research questions

All the instruments used in our study were originally developed in English, and thus equivalence with French versions was a major concern. In translation to French, some phrases proved problematic. For example, for rating response options, the European French version of the EUROPEP translated "poor" as "médiocre" (second-rate), whereas Quebec translators rendered it as "mauvais" (bad). The English question "How often…," followed by frequency response options "always," "usually" and "sometimes," was translated in French as "Combien de fois…" (How many times), which naturally elicits a count rather than a frequency response. Finally, one instrument used the term "primary care provider" to refer to both person and place, for which there is no single French equivalent. Consequently, the French term varied by the context of the question specifying as "source habituelle de soins" (usual source of care), "clinique" (clinic) or "médecin" (physician).

Our concern about differential functioning by geographic area arose from previous studies in which rural residents reported better accessibility than did residents of metropolitan areas (Haggerty et al. 2007). We hypothesized that measures of accessibility may function differently by context. We had no a priori concerns regarding educational achievement, but we wanted to ensure that all instruments performed equally well in low-literacy groups, because we found considerable variation in readability among instruments.

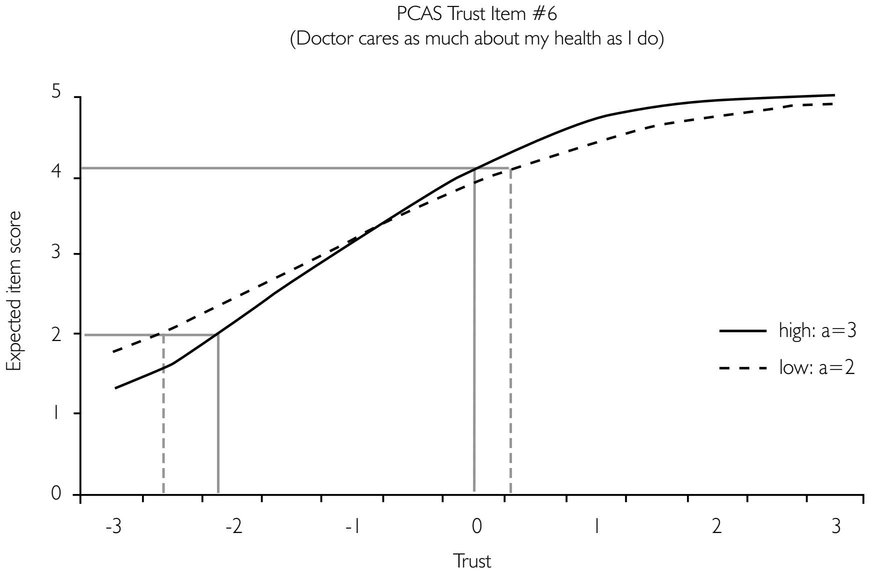
Overview of differential functioning

The language of differential item functioning analysis reflects its origins in educational assessment. The method, developed to assess the performance of questions that estimate a student's understanding of a topic, evaluates performance in two ways – first in terms of *discriminability*, i.e., how well the item can differentiate between individuals with different levels of ability, and then in terms of *difficulty*, i.e., how hard it is to answer correctly a question at different levels of student ability.

For example, if the probability of answering correctly changes depending on the student's level of ability and can detect even a small difference in ability between two individuals, then the question has good discriminability. If a student has a 50% probability of responding correctly only in the high range of ability, then the question or item is considered difficult; if 50% probability is achieved in the low range of ability, then it is considered easy. A good instrument includes questions with difficulty thresholds across the entire ability range, each with good discriminability.

This approach has also been used to evaluate item performance of attitudinal surveys. Discriminability is an item's sensitivity to differences between individuals on the construct being measured (e.g., trust in the provider) and is represented with a slope in item response models. The steeper the slope, the more discriminating the item, with slopes ≥1 (the "a" parameter) considered appropriate; i.e., each unit increase in the item predicts a unit increase in the underlying construct. Ideally, the item's slope should not differ among subgroups; if it does, the item has differential discriminability. Figure 1 illustrates differential discriminability by educational level for an item in the Trust subscale from the Primary Care Assessment Survey (PCAS), showing higher discrimination in high- than in low-education respondents. Differences in item discrimination indicate that the question is understood or interpreted differently by each subgroup. This would occur, for example, when the French translation is not equivalent in meaning to the original English version.

FIGURE 1. Differential discriminability between high- and low-education respondents for item in the PCAS Trust subscale

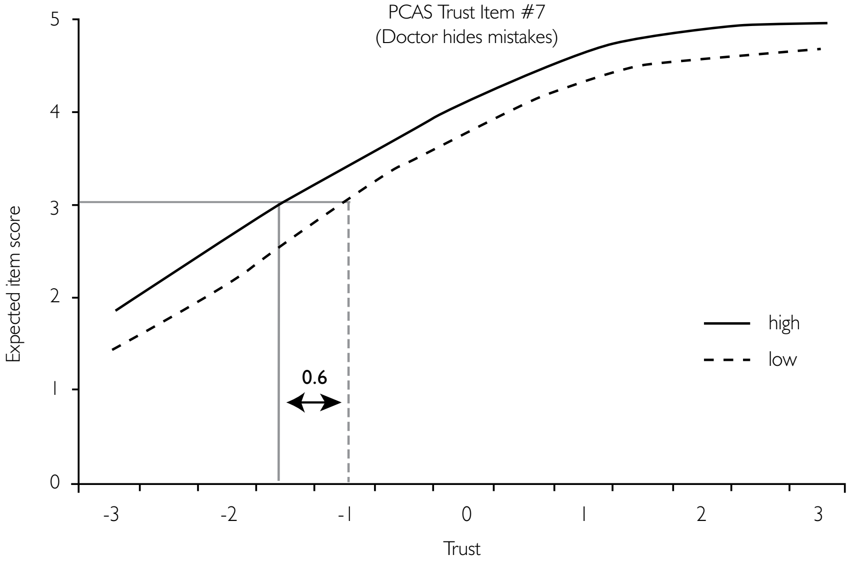


Note: Difficulty threshold is lower (easy) in low- than in high-education respondents at low levels of trust, but higher (more difficult) at high levels of trust.

Difficulty in attitudinal surveys refers to the probability of endorsing a specific response option for a given level of the construct being measured. When an item's difficulty threshold varies by group membership, the item is said to exhibit differential difficulty functioning, e.g., in the PCAS Trust subscale that elicits agreement with statements using a five-point Likert scale of "1 = strongly disagree" to "5 = strongly agree." Figure 2 illustrates the differential difficulty threshold between high- and low-education respondents. Note that for the same expected item score of 3, low-education respondents will have 0.6 higher level of trust on the standardized trust score (i.e., the item is more difficult) than will high-education respondents. The difficulty differential is uniform across all levels of trust, whereas in Figure 1, showing differential discriminability, the difficulty differential is not uniform across levels of trust. Consistent differences in difficulty thresholds in a scale's items may point to differences in how response options are interpreted.

The potential impact of differential item functioning is assessed by removing problematic items from the subscale or instrument, recalculating the scores using only the purified scale (non-problematic or "anchor" items) and comparing the group values again. If the between-group comparison using the purified scale reaches a different conclusion, differential item functioning is said to have an impact, and using the original scale could give biased measures. If the comparison remains essentially unchanged (typically, when differences are minor or in different directions), differential item functioning is said to have no impact.

FIGURE 2. Differential difficulty between high- and low-education respondents for item in the PCAS Trust subscale, showing equal discriminability (a=1.64) and uniform difficulty threshold (b=0.6) across all levels of trust



Method

Study population

The target population for this study was adult, Canadian, primary healthcare users, undifferentiated by age or health condition. The sample was selected to be balanced by French/English language, high/low educational level and urban/rural location. We also stratified by excellent, average and poor primary care experience based on a single screening question: "Overall, has your experience of care from your regular family doctor or medical clinic been excellent, poor or average?"

Participants responded to all six instruments and provided socio-demographic and utilization information. Data were collected between February and July 2005. English-language questionnaires were administered in Nova Scotia and French-language questionnaires in Quebec.

Urban location was defined as census metropolitan areas; rural, as more than one hour's travel from a metropolitan area; remote (Quebec only), more than four hours' travel. We used an age-sensitive cut-off to denote educational achievement as a proxy for reading level. Subjects were considered to have a high-school reading level or lower if they had (a) completed only high school and were under 45 years old, (b) completed 10 years of school and were 45 to 55 years old or (c) completed less than eight years and were over 55 years old (Smith and Haggerty 2003).

Analysis

We examined the distribution of missing values by language, educational achievement and geography. The score for each subscale, calculated as the mean of items, was normalized to a 0-to-10 scale to permit comparisons on a common metric (formula, Table 2). We compared normalized subscale scores by language, education and geography using regression modelling controlling for the other design variables as well as for overall experience, using $\alpha=.05$ to denote statistical significance, despite multiple testing, to maintain a high sensitivity to potential differences. We conducted exploratory factor analysis to examine whether factor resolution for the subscales was the same by language, education and geography.

All methods of assessing differential item functioning consist of examining the distribution of responses in the subgroups of interest when they are conditioned on the same level of the underlying construct or latent variable (Santor and Ramsay 1998; Kristjansson et al. 2005; Reeve 2006; Teresi and Fleishman 2007). In this study, we used parametric item response analysis using Multilog software (Du Toit 2003) to test for differential discrimination and difficulty across all options within an item using a chi-square test. The latent variable was the total score of the subscale. We first assessed significant between-group differences in the discrimination parameter. If none was found, we fixed the discrimination parameter to be equal between groups and tested for uniform and non-uniform differences in the difficulty threshold across response options. We retested the discrimination parameter after removing problematic items from the latent variable and repeated the process until we found no differentially functioning items. We used a critical value of $\alpha=.01$ to indicate statistical significance because lower values detected trivial differences.

Finally, we re-examined group differences with a series of standard ANOVA tests using subscale scores based on the subset of items found to be free of differential item functioning.

Results

The six instruments contained 153 validated items. Despite attempts to balance the sample equally by French/English language, high/low education and urban/rural location, the 645 respondents were not equally distributed. The English-language group was more urban (59% vs. 49%, $\chi^2=6.7$, $p<.096$), more likely to have a high-school reading level (75% vs. 55%, $\chi^2=27.5$, *p*<.001) and also more likely to perceive their health as good or excellent, to be affiliated with a family doctor rather than a clinic and to have longer affiliations (Table 1). Wait times for appointments were better among English-speaking, high-education and urban respondents.

TABLE 1. Characteristics of the study sample compared by language, geography and education (only statistically significant differences are shown*)

Characteristic	Total (n = 645) % (n)	Language % (n)		Geography % (n)		Education % (n)	
		English (n=343)	French (n=302)	Urban (n=351)	Rural (n=294)	High (n=424)	Low (n=221)
Overall experience of care							
Poor	23.1 (149)	–		–		–	
Average	36.0 (232)						
Excellent	40.9 (264)						
Mean age in years (SD)	48.0 (14.9)	–		46.4 (15.0)	49.8 (14.4)**	49.5 (13.9)	45.1 (13.2)**
Per cent female	64.7 (414)	–		–		–	
Mean years of education (SD)	13.0 (3.4)	13.5 (3.0)	12.4 (3.6) ***	13.5 (3.4)	12.4 (3.3) ***	14.4 (2.8)	10.3 (2.6) ***
Regular provider							
Physician	94.1 (607)	96.8 (332)	91.1 (275) **	–		96.2 (408)	90.1 (199) **
Clinic only	5.9 (38)	3.2 (11)	8.9 (27)			3.8 (16)	9.9 (22)
Per cent indicating health status as good or excellent	38.0 (241)	45.6 (155)	29.2 (86) ***	–		42.5 (179)	29.0 (62) ***
Per cent with chronic health problem	61.6 (392)	–		–		–	
Mean usual wait time for appointment			***		***		*
Less than 2 days	35.2 (220)	53.2 (177)	14.7 (43)	43.7 (149)	25.0 (71)	38.5 (158)	28.8 (62)
2 to 7 days	32.6 (204)	36.3 (121)	28.4 (83)	33.4 114)	31.7 (90)	32.9 (135)	32.1 (69)
7 days to 2 weeks	11.8 (74)	8.7 (29)	15.4 (45)	11.1 (38)	12.7 (36)	12.0 (49)	11.6 (25)
2 weeks to 4 weeks	9.3 (58)	1.2 (4)	18.5 (54)	5.6 (19)	13.7 (39)	7.6 (31)	12.6 (27)
More than 4 weeks	11.0 (69)	0.6 (2)	23.0 (67)	6.2 (21)	16.9 (48)	9.0 (37)	14.9 (32)
Mean number of visits to other providers (SD)							
Specialist	4.2 (5.1)	–		–		–	
Other providers	9.4 (19.2)	–		–		–	

* p≤.05

** p≤.01

*** p≤.001

Descriptive results

The number of missing values was not systematically higher by language or educational achievement, but was higher in rural than in urban respondents for five items, but they were not consistently in one instrument or attribute. English-speaking respondents showed a higher tendency than the French speakers to select the "don't know/not applicable" option when offered.

The normalized subscale scores, grouped by primary care attribute, are compared by language, education and geography in Table 2. Strikingly, subscale scores are systematically higher (more positive assessment) in the English than in the French subgroup, with the exception of the CPCI Coordination of Care subscale. The study design does not allow us to determine whether the difference is due to true differences between Nova Scotia and Quebec or differential functioning between English- and French-language versions.

TABLE 2. Comparison of mean normalized subscale scores by language, geography, education and total sample (only statistically significant differences are shown***)

Questionnaire	Developer's Subscale Name	# of Items in Subscale	Language English (SD) French (SD)	Education Low (SD) High (SD)	Geography Rural (SD) Urban (SD)	Overall Mean (SD)
Accessibility						
PCAS	Organizational Access	6	6.31 (1.73) 5.52 (1.85)***	–	–	5.94 (1.83)
PCAT	First-Contact Utilization	3	9.26 (1.36) 8.89 (1.80)**	–	9.25 (1.35) 8.95 (1.76)*	9.10 (1.60)
PCAT	First-Contact Access	4	6.36 (2.43) 4.71 (2.54)***	5.31 (2.66) 5.83 (2.56)*	–	5.60 (2.60)
EUROPEP	Organization of Care	7	7.05 (2.09) 5.92 (2.50)***	–	–	6.51 (2.36)
Comprehensiveness of Services						
PCAT	Comprehensiveness (services available)	4	8.54 (1.73) 6.71 (2.84)***	–	–	7.72 (2.47)
CPCI	Comprehensive Care	6	8.16 (1.85) 7.23 (2.24)***	–	–	7.72 (2.09)
Relational Continuity						
PCAS	Visit-Based Continuity	2	8.78 (1.60) 7.86 (2.48)***	–	–	8.35 (2.11)
PCAS	Contextual Knowledge	5	6.09 (2.24) 5.73 (2.31)*	–	–	5.92 (2.28)
PCAT	Ongoing Care	4	7.38 (2.13) 6.89 (2.54)**	–	–	7.15 (2.34)
CPCI	Accumulated Knowledge	8	7.21 (2.42) 6.75 (2.54)*	–	–	6.99 (2.49)
CPCI	Patient Preference for Regular Physician	5	8.27 (1.72) 7.02 (2.12)***	–	–	7.68 (2.01)

TABLE 2. Continued

Questionnaire	Developer's Subscale Name	# of Items in Subscale	Language English (SD) French (SD)	Education Low (SD) High (SD)	Geography Rural (SD) Urban (SD)	Overall Mean (SD)
Management Continuity						
PCAS	Integration	6	7.28 (1.94) 6.41 (2.14)***	–	–	6.90 (2.07)
PCAT	Coordination	4	7.92 (2.36) 7.13 (2.86)***	–	–	7.57 (2.62)
CPCI	Coordination of Care	8	6.37 (1.33) 6.86 (2.59)**	–	–	6.60 (2.03)
VANOCSS§	Coordination of Care (overall), number of problems	6	2.33 (1.81) 2.80 (1.91)**	–	–	2.51 (1.88)
VANOCSS§	Specialty Provider Access (number of problems)	4	0.42 (0.78) 0.78 (0.97)**	0.79 (0.94) 0.52 (0.88)*	–	0.62 (0.91)
Interpersonal Communication						
PCAS	Communication	6	7.71 (1.92) 6.90 (2.22)***	–	–	7.33 (2.11)
PCAS	Trust	8	7.86 (1.66) 7.17 (1.83)***	–	7.36 (1.74) 7.68 (1.80)*	7.53 (1.78)
CPCI	Interpersonal Communication	6	7.60 (2.28) 6.72 (2.27)***	–	–	7.19 (2.32)
EUROPEP	Clinical Behaviour	16	8.14 (1.84) 7.54 (2.26)***	–	–	7.85 (2.07)
IPC-II	Communication (elicited concerns, responded)	3	8.40 (1.79) 7.15 (2.35)***	–	–	7.81 (2.16)
IPC-II	Communication (explained results, medications)	4	7.66 (2.37) 7.12 (2.59)**	–	–	7.40 (2.49)
IPC-II	Decision-Making (patient-centred decision-making)	4	5.82 (3.06) 4.97 (3.19)***	–	–	5.41 (3.15)
Respectfulness						
PCAS	Interpersonal Treatment	5	7.83 (2.01) 7.00 (2.26)***	–	–	7.44 (2.17)
IPC-II	Hurried Communication	5	8.44 (1.54) 7.52 (1.88)***	–	–	8.01 (1.77)
IPC-II	Interpersonal Style (compassionate, respectful)	5	8.35 (2.16) 7.66 (2.46)***	–	–	8.02 (2.33)
IPC-II	Interpersonal Style (respectful office staff)	4	9.05 (1.72) 8.47 (1.91)***	–	–	8.78 (1.83)

TABLE 2. Continued

Questionnaire	Developer's Subscale Name	# of Items in Subscale	Language English (SD) French (SD)	Education Low (SD) High (SD)	Geography Rural (SD) Urban (SD)	Overall Mean (SD)
Whole-Person Care						
PCAT	Community Orientation	3	5.31 (2.75) 4.44 (2.93)***	–	–	4.88 (2.87)
CPCI	Community Context	2	7.28 (2.71) 5.55 (3.32)***	–	–	6.47 (3.13)

* $p \leq .05$
** $p \leq .01$
*** $p \leq .001$
§ The VANOCSS scores are not normalized; the score represents the number of problems reported.

Only one subscale differs by education, the PCAT-S First-Contact Access, with fewer positive assessments in low- than in high-education groups. Rural respondents indicate more positive assessments than urban respondents in PCAT-S First-Contact Utilization and fewer positive assessments in PCAS Trust.

Most subscales had similar factor resolution by subgroup. Three subscales found two factors (eigenvalue >1) in one group and the expected single factor in the other: the CPCI Coordination of Care (management continuity) subscale had two factors in English; the CPCI Preference for Regular Physician (relational continuity) had two in rural; and the PCAS Trust (interpersonal communication) subscale had two in French-speaking and in low-education respondents.

Differential functioning

Because of space constraints, we report only summary results at a subscale level; item-specific results are available upon request. The discriminability of individual items is reported in the attribute-specific papers elsewhere in this special issue of the journal. Table 3 shows the number of items within each subscale that are free of differential functioning and would be considered pure or anchor items for making valid comparisons between subgroups.

The French/English comparison exhibited the most differential item functioning and urban/rural, the least. Of the 153 items, only 80 (52%) were free of French/English differential functioning, compared to 111 (73%) in high/low education and 139 (91%) in urban/rural location.

Of the items with differential French/English functioning, one-third (24/73) were important differences in discriminability or difficulty. Overall, 41% (30/73) of items showed differences in discriminatory capacity, but only 13 of these had discriminability differentials greater than 1 (Figure 1 demonstrates a discriminability differential of 1). English items tended to be more discriminating, but only four items discriminated adequately in English and poorly in French, all from the CPCI. For example, agreement with the statement, "If I am sick I would always contact this doctor first" (CPCI Preference for Regular Physician) had a discrimination value of 1.63 in English and 0.87 in French.

TABLE 3. Number of items free from differential item functioning (discrimination or difficulty) within each validated subscale by language, geography and education

Developer's Subscale Name	Number of Items without Differential Item Functioning		
	Language (Province) English/French	Education Low/High	Geography Urban/Rural
Accessibility			
PCAS Organizational Access	4/6 (67%)	6/6 (100%)	5/6 (83%)
PCAT First-Contact Utilization	2/3 (67%)	3/3 (100%)	3/3 (100%)
PCAT First-Contact Accessibility	1/4 (25%)	3/4 (75%)	2/4 (50%)
EUROPEP Organization of Care	3/7 (43%)	7/7 (100%)	5/7 (71%)
Subtotal	*10/20 (50%)*	*19/20 (95%)*	*15/20 (75%)*
Comprehensiveness of Services			
PCAT Comprehensiveness (services available)	2/4 (50%)	4/4 (100%)	4/4 (100%)
CPCI Comprehensive Care	3/6 (50%)	4/6 (67%)	6/6 (100%)
Subtotal	*5/10 (50%)*	*8/10 (80%)*	*10/10 (100%)*
Relational Continuity			
PCAS Visit-Based Continuity	0/2 (0%)	0/2 (0%)	2/2 (100%)
PCAS Contextual Knowledge	3/5 (60%)	5/5 (100%)	4/5 (80%)
PCAT Ongoing Care	1/4 (25%)	4/4 (100%)	3/4 (75%)
CPCI Accumulated Knowledge	1/8 (13%)	4/8 (50%)	7/8 (88%)
CPCI Patient Preference for Regular Physician	2/5 (40%)	4/5 (80%)	4/5 (80%)
Subtotal	*7/24 (29%)*	*17/24 (71%)*	*20/24 (83%)*
Management Continuity			
PCAS Integration	3/6 (50%)	5/6 (83%)	6/6 (100%)
PCAT Coordination	2/4 (50%)	4/4 (100%)	3/4 (75%)
CPCI Coordination of Care	3/8 (38%)	5/8 (63%)	6/8 (75%)
VANOCSS Coordination of Care (overall), number of problems	6/6 (100%)	6/6 (100%)	6/6 (100%)
VANOCSS Specialty Provider Access, number of problems	4/4 (100%)	4/4 (100%)	4/4 (100%)
Subtotal	*18/28 (64%)*	*24/28 (86%)*	*25/28 (89%)*
Interpersonal Communication			
PCAS Communication	5/6 (83%)	5/6 (83%)	6/6 (100%)
PCAS Trust	1/8 (13%)	1/8 (13%)	8/8 (100%)
CPCI Interpersonal Communication	1/6 (17%)	2/6 (33%)	6/6 (100%)
EUROPEP Clinical Behaviour	10/16 (63%)	15/16 (94%)	16/16 (100%)

TABLE 3. Continued

Developer's Subscale Name	Number of Items without Differential Item Functioning		
	Language (Province) English/French	Education Low/High	Geography Urban/Rural
IPC-II Communication (elicited concerns, responded)	3/3 (100%)	0/3 (0%)	2/3 (67%)
IPC-II Communication (explained results, medications)	4/4 (100%)	2/4 (50%)	4/4 (100%)
IPC-II Decision-Making (patient-centred decision-making)	4/4 (100%)	3/4 (75%)	4/4 (100%)
Subtotal	*28/47 (60%)*	*28/47 (60%)*	*46/47 (98%)*
Respectfulness			
PCAS Interpersonal Treatment	4/5 (80%)	0/5 (0%)	5/5 (100%)
IPC-II Hurried Communication	4/5 (80%)	5/5 (100%)	5/5 (100%)
IPC-II Interpersonal Style (compassionate, respectful)	1/5 (20%)	5/5 (100%)	5/5 (100%)
IPC-II Interpersonal Style (respectful office staff)	2/4 (50%)	2/4 (50%)	3/4 (75%)
Subtotal	*11/19 (58%)*	*12/19 (63%)*	*18/19 (95%)*
Whole-Person Care – Community Orientation			
PCAT Community Orientation	1/3 (33%)	3/3 (100%)	3/3 (100%)
CPCI Community Context	0/2 (0%)	0/2 (0%)	2/2 (100%)
Subtotal	*1/5 (20%)*	*3/5 (60%)*	*5/5 (100%)*
Number of subscales with no differential item functioning	5/29 (17%)	12/29 (41%)	18/29 (62%)
Number of subscales where ≥50% of items exhibit differential functioning	12/29 (41%)	6/29 (21%)	0/29 (0%)

Of the 43 items with differential French/English difficulty, only 11 had differentials over 0.5, which is approximately the magnitude illustrated in Figure 2. The pattern of differences does not support a systematic difference between English and French when "poor" is translated as "médiocre" versus "mauvais," and it appears that frequency response scales were understood equivalently in both French and English. However, the difficulty threshold for the "fortement en désaccord" option is consistently more positive than for "strongly disagree" across several subscales and two instruments (CPCI and PCAS Trust). The response option "strongly disagree" seems to be more negative than "fortement en désaccord." We found no systematic direction of difficulty differences for "strongly agree."

By education, 43% (18/42) of differentially functioning items were due to differential discriminability, with seven being differentials >1. Items tended to have higher discrimination values in high-education respondents, although the reverse was seen for respectfulness. Only 12 of the remaining 24 items had difficulty differentials >0.5. The items tend to be more discriminating and difficult in the high-education than in the low-education groups; specifically, low-education respondents have a higher probability of responding positively at lower levels of the construct of interest (communication, respectfulness). One of the largest observed difficulty differentials was in the PCAS Interpersonal Treatment (respectfulness) subscale, where all items had differential functioning, with an average difficulty threshold being 0.8 higher for high-education than low-education respondents.

By geography, there were only 14 differentially functioning items, with four out of seven discriminability differentials being >0.5. All were in accessibility and relational continuity. All items were more discriminating in urban than in rural groups.

Table 4 compares the subscale scores by language, education and geography after we removed items with differential functioning. Of the 29 subscales, only five (17%) are free from French/English differential functioning, compared to 12 (41%) in education and 18 (62%) in geography. Valid comparison by language was impossible for subscales with no remaining non-problematic items: PCAS Visit-Based Continuity, PCAT-S Community Orientation and CPCI Community Context. Comparisons based on less than 50% of the original items must be interpreted cautiously; this affects 12 (41%) subscales by language, six (21%) by education and none by geography. However, the results show that, for language, the conclusions are essentially unchanged from those of Table 2: assessments for all attributes remain more positive in English (Nova Scotia) than in French (Quebec). The previous more positive French scores on CPCI Coordination of Care disappear in the purified subscale.

TABLE 4. Subscale comparisons by language, geography and education using purified subscale scores (free of items with differential item functioning)

Developer's Subscale Name	Language (Province) French/English	Education High/Low	Geography Urban/Rural
Accessibility			
PCAS Organizational Access	+	NS	NS
PCAT First-Contact Utilization	+	NS	+
PCAT First-Contact Access	+	+ / NS	NS / +
EUROPEP Organization of Care	+	NS	NS
Comprehensiveness of Services			
PCAT Comprehensiveness (services available)	+	NS	NS
CPCI Comprehensive Care	+	NS	NS
Relational Continuity			
PCAS Visit-Based Continuity	+ / 0	NS / 0	NS
PCAS Contextual Knowledge	+	NS	NS
PCAT Ongoing Care	+ / NS	NS	NS
CPCI Accumulated Knowledge	+ / NS	NS	NS

TABLE 4. Continued

Developer's Subscale Name	Language (Province) French/English	Education High/Low	Geography Urban/Rural
CPCI Patient Preference for Regular Physician	+	NS	NS
Management Continuity			
PCAS Integration	+	NS	NS
PCAT Coordination	+	NS	NS
CPCI Coordination of Care	+ / –	NS	NS
VANOCSS Coordination of Care (overall), number of problems	+	NS	NS
VANOCSS Specialty Provider Access, number of problems	+	+	NS
Interpersonal Communication			
PCAS Communication	+	NS	NS
PCAS Trust	+ / NS	NS	+
CPCI Interpersonal Communication	+	NS	NS
EUROPEP Clinical Behaviour	+	NS	NS
IPC-II Communication (elicited concerns, responded)	+	NS / 0	NS
IPC-II Communication (explained results, medications)	+	NS	NS
IPC-II Decision-Making (patient-centred decision-making)	+	NS	NS
Respectfulness			
PCAS Interpersonal Treatment	+	NS / 0	NS
IPC-II Hurried Communication	+	NS	NS
IPC-II Interpersonal Style (compassionate, respectful)	+	NS	NS
IPC-II Interpersonal Style (respectful office staff)	+	NS	NS
Whole-Person Care			
PCAT Community Orientation	+	NS	NS
CPCI Community Context	+ / 0	NS / 0	NS

The symbol "+" indicates that previous positive differences between categories remain positive; "NS" indicates that previously non-significant differences remain non-significant. Symbols separated by "/" indicate a change between original and purified subscale results; "–" indicates negative and "0" indicates that no items remained on which to test the purified result.

For the high/low education comparison, the previous difference on PCAT-S First-Contact Access disappears, and no other scores are statistically different. However, it is difficult to conclude that non-significant differences by education are valid on the IPC Communication and PCAS Interpersonal Treatment (respectfulness) subscales, because no items were free from differential functioning. The difficulty differential may be such that non-significant difference in Table 2 may be masking an actual difference for these subscales.

When the urban and rural groups are compared using the purified subscales, rural scores become significantly lower than urban scores for PCAT-S First-Contact Access (likelihood of obtaining same-day needed care from regular provider), but the higher rural score persists in PCAT-S First-Contact Utilization (tendency to contact the regular provider first). The previous difference with respect to PCAS Trust disappears.

Discussion and Conclusion

We found that assessments of primary healthcare attributes were systematically more positive by English- than French-speaking respondents despite an a priori expectation of equivalency. Without analyzing differential functioning, it is difficult to determine whether this difference is due to differences in the Quebec and Nova Scotia healthcare systems or to problems with measurement equivalency of the French and English versions. The answer seems to be both. We found substantial differential item functioning between English and French versions. However, the systematically more positive assessments in Nova Scotia persist even after removing problematic items. The differences in wait times and proportion having a regular physician also support the existence of a real difference.

These results suggest that continued refinement is needed to ensure that French-language versions are equivalent to the original English-language instruments, but that most of the differential functioning is minor and has minimal impact on comparisons at the subscale level. The parametric item response models detected differences as small as 0.4 in discriminability and small differences in difficulty; rarely was discriminability compromised in French and adequate in English. Rather, the differences meant that an item showed good discrimination in one group and slightly better in the other, so that overall, the functioning of the items and scales was acceptable despite differential functioning.

In some cases, our results helped us detect slight shifts in meaning in French translations. For example, the English word "ability" in the PCAS Organizational Accessibility subscale was translated as "facilité," "to get through to the practice by telephone" and as "possibilité," "to talk to the doctor by telephone." The former resulted in differential discriminatory capacity, but not the latter, suggesting that "possibilité" is a more equivalent translation for "ability" in this context than is "facilité." Likewise, the varied translation of "primary care provider" in the PCAT-S instrument may have introduced differential functioning by creating specific, limited terms in the French-language versions while retaining a broad and flexible term in English. In other cases, we could not identify the source of non-equivalence, suggesting differences in cultural interpretation or in interacting with the healthcare system.

We did not detect systematic patterns in difficulty differences that would suggest difference in how response options or scales function in these groups, with the exception of the agree/disagree response scale by French/English. The observed difference suggests that "disagree" may not be equivalent in sense and meaning to "désaccord." In French, "désaccord" seems to be a slightly different concept from, rather than the opposite of, "accord" (agree). It may be analogous to the finding that "dissatisfied" is not the same construct as "not satisfied" (Eriksen 1995; Coyle 1999). This difference explains the high level of differential functioning in the CPCI, which uses a disagree/agree response format. We recommend that French-language versions of "disagree" response options be reformulated as extremes of agreement such that "pas du tout d'accord" (not at all in agreement) is the equivalent of "strongly disagree."

In this study, we assumed that the original English-language version is the gold standard and that French-language versions must be modified to achieve equivalence. However, results from discussion groups also suggest that some original English statements should be modified to be more valid or precise. For example, English-speaking respondents expressed confusion about the meaning of "primary care provider" (Haggerty et al. 2011), and we believe that the specificity that was required for the French translation resulted in a more precise measure.

It is a tribute to the instrument developers that the instruments and subscales mostly perform equivalently across high- and low-education groups. However, differences in difficulty thresholds, especially in attributes such as respectfulness and interpersonal communication, suggest some measures may systematically underdetect true differences in experience between high- and low-education patients. Higher difficulty thresholds in high-education patients would be consistent with higher expectations among these respondents, a finding that has been repeatedly observed in studies of satisfaction (Crow et al. 2002).

Differential item functioning by urban/rural residence specifically affected the attribute of accessibility. The finding that rural residents are less likely than urban residents to obtain same-day care from their provider when they are sick (PCAT-S First-Contact Access) becomes evident only when differentially functioning items are removed. This is a concern because urban/rural comparisons of accessibility have important implications for health planners' decisions on health services location to optimize equity of access.

The strength of this study is that the same questionnaires were administered to each subject, so that the underlying construct can be directly compared across groups rather than relying on model assumptions of equivalence. However, some differences we found may be spurious owing to multiple testing, and the analytic software we used was highly sensitive to even small differences in difficulty threshold, so that some of the statistically significant differences may not be meaningful. Furthermore, removing problematic items from subscales may compromise construct representation so that comparing subscale scores before and after removal of problematic items is no longer meaningful.

We feel comfortable recommending the use of the French-language versions, while continuing to refine them. At a subscale level, the differential functioning did not introduce bias because the conclusions were largely unchanged when problematic items were removed. We recommend that the original versions be reviewed where translation has posed a problem. We recommend caution in interpreting rural and urban comparisons of access, and urge the development of unbiased measures. We found little evidence of bias by educational status and are confident in recommending these instruments for a broad educational spectrum of patients.

ACKNOWLEDGEMENTS

This research was funded by the Canadian Institute for Health Research. During this study, Jeannie L. Haggerty held a Canada Research Chair in Population Impacts of Healthcare at the Université de Sherbrooke. The authors wish to thank Beverley Lawson for conducting the survey in Nova Scotia and Christine Beaulieu in Quebec, and Donna Riley for support in preparation and editing of the manuscript.

Correspondence may be directed to: Jeannie L. Haggerty, Associate Professor, Department of Family Medicine, McGill University, St. Mary's Research Centre, Hayes Pavilion – Suite 3734, 3830 Lacombe Ave., Montreal QC H3T 1M5; tel.: 514-345-3511 ext. 6332; fax: 514-734-2652; e-mail: jeannie.haggerty@mcgill.ca.

REFERENCES

Borowsky, S.J., D.B. Nelson, J.C. Fortney, A.N. Hedeen, J.L. Bradley and M.K. Chapko. 2002. "VA Community-Based Outpatient Clinics: Performance Measures Based on Patient Perceptions of Care." *Medical Care* 40(7): 578–86.

Coyle, J. 1999. "Exploring the Meaning of 'Dissatisfaction' with Health Care: The Importance of 'Personal Identity Threat.'" *Sociology of Health and Illness* 21: 95–124.

Crow, R., H. Gage, S. Hampson, J. Hart, A. Kimber, L. Storey and H. Thomas. 2002. "The Measurement of Satisfaction with Healthcare: Implications for Practice from a Systematic Review of the Literature." *Health Technology Assessment* 6: 1–244.

Du Toit, M. 2003. *IRT from SSI: Bilog-mg, Multilog, Parscale, Testfact.* Lincolnwood, IL: Scientific Software International.

Eriksen, L.R. 1995. "Patient Satisfaction with Nursing Care: Concept Clarification." *Journal of Nursing Measurement* 3: 59–76.

Flocke, S. 1997. "Measuring Attributes of Primary Care: Development of a New Instrument." *Journal of Family Practice* 45(1): 64–74.

Haggerty, J.L., R. Pineault, M.-D. Beaulieu, Y. Brunelle, J. Gauthier, F. Goulet and J. Rodrigue. 2007. "Room for Improvement: Patients' Experiences of Primary Care in Quebec Before Major Reforms." *Canadian Family Physician* 53: 1056–57.

Haggerty, J.L., C. Beaulieu, B. Lawson, D.A. Santor, M. Fournier and F. Burge. 2011. "What Patients Tell Us about Primary Healthcare Evaluation Instruments: Response Formats, Bad Questions and Missing Pieces." *Healthcare Policy* 7 (Special Issue): 66–78.

Kristjansson, E., R. Aylesworth, I. McDowell and B.D. Zumbo. 2005. "A Comparison of Four Methods for Detecting Differential Item Functioning in Ordered Response Items." *Educational and Psychological Measurement* 65: 935–53.

Reeve, B. 2006. *An Introduction to Modern Measurement Theory.* Rockville, MD: National Cancer Institute. Retrieved May 12, 2011. <http://appliedresearch.cancer.gov/areas/cognitive/immt.pdf>.

Safran, D.G., J. Kosinski, A.R. Tarlov, W.H. Rogers, D.A. Taira, N. Lieberman and J.E. Ware. 1998. "The Primary Care Assessment Survey: Tests of Data Quality and Measurement Performance." *Medical Care* 36(5): 728–39.

Santor, D.A. and J.O. Ramsay. 1998. "Progress in the Technology of Measurement: Applications of Item Response Models." *Psychological Assessment* 10: 345–59.

Shi, L., B. Starfield and J. Xu. 2001. "Validating the Adult Primary Care Assessment Tool." *Journal of Family Practice* 50(2): n161w–n171w.

Smith, J.L. and J. Haggerty. 2003. "Literacy in Primary Care Populations: Is It a Problem?" *Canadian Journal of Public Health* 94: 408–12.

Stewart, A.L., A.M. Nápoles-Springer, S.E. Gregorich and J. Santoyo-Olsson. 2007. "Interpersonal Processes of Care Survey: Patient-Reported Measures for Diverse Groups." *Health Services Research* 42(3 Pt. 1): 1235–56.

Teresi, J.A. and J.A. Fleishman. 2007. "Differential Item Functioning and Health Assessment." *Quality of Life Research* 16(Suppl. 1): 33–42.

Wensing, M., J. Mainz and R. Grol. 2000. "A Standardised Instrument for Patient Evaluations of General Practice Care in Europe." *European Journal of General Practice* 6: 82–87.

RESEARCH PAPER

What Patients Tell Us about Primary Healthcare Evaluation Instruments: Response Formats, Bad Questions and Missing Pieces

Ce que disent les patients au sujet des instruments d'évaluation des soins de santé primaires : formats des réponses, mauvaises questions et pièces manquantes

JEANNIE L. HAGGERTY
Department of Family Medicine, McGill University
Montreal, QC

CHRISTINE BEAULIEU
St.Mary's Research Centre, St.Mary's Hospital Center
Montreal, QC

BEVERLY LAWSON
Department of Family Medicine, Dalhousie University
Halifax, NS

DARCY A. SANTOR
School of Psychology, University of Ottawa
Ottawa, ON

MARTINE FOURNIER
Centre de recherche CHU Sainte-Justine
Montreal, QC

FREDERICK BURGE
Department of Family Medicine, Dalhousie University
Halifax, NS

Abstract

Instruments have been developed that measure consumer evaluations of primary healthcare using different approaches, formats and questions to measure similar attributes. In 2004 we concurrently administered six validated instruments to adults and conducted discussion groups to explore how well the instruments allowed patients to express their healthcare experience and to get their feedback about questions and formats.

Method: We held 13 discussion groups (n=110 participants): nine in metropolitan, rural and remote areas of Quebec; four in metropolitan and rural Nova Scotia. Participants noted critical incidents in their healthcare experience over the previous year, then responded to all six instruments under direct observation and finally participated in guided discussions for 30 to 40 minutes. The instruments were: the Primary Care Assessment Survey; the Primary Care Assessment Tool; the Components of Primary Care Index; the EUROPEP; the Interpersonal Processes of Care Survey; and part of the Veterans Affairs National Outpatient Customer Satisfaction Survey. Two team members analyzed discussion transcripts for content.

Results: While respondents appreciated consistency in response options, they preferred options that vary to fit the question. Likert response scales functioned best; agreement scales were least appreciated. Questions that average experience over various providers or over many events diluted the capacity to detect critical negative or positive incidents. Respondents tried to answer all questions but stressed that they were not able to report accurately on elements outside their direct experience or in the provider's world. They liked short questions and instruments, except where these compromise clarity or result in crowded formatting. All the instruments were limited in their capacity to report on the interface with other levels of care.

Conclusion: Each instrument has strengths and weaknesses and could be marginally improved, but respondents accurately detected their intent and use. Their feedback offers insight for instrument development.

Résumé

Des instruments ont été conçus pour mesurer l'évaluation des soins de santé primaires par les patients, et ce, par diverses démarches, formats et questions qui mesurent des caractéristiques similaires. En 2004, nous avons administré simultanément à des adultes six instruments validés et avons méne des groupes de discussion pour voir, d'une part, à quel point ces instruments permettent aux patients d'exprimer leur expérience de soins et pour obtenir, d'autre part, leurs commentaires sur les questions et les formats.

Méthodologie : Nous avons organisé 13 groupes de discussion (n=110 participants) : neuf dans des régions métropolitaines, rurales et éloignées du Québec; quatre dans des régions métropolitaines et rurales de la Nouvelle-Écosse. Les participants ont noté des incidents critiques dans le cadre de leur expérience de soins au cours de l'année précédente, puis ont répondu aux six instruments sous observation directe, et finalement ont participé à des échanges dirigés de 30 à 40 minutes. Les instruments suivants ont été utilisés : Primary Care Assessment Survey; Primary Care Assessment Tool; Components of Primary Care Index;

EUROPEP; Interpersonal Processes of Care Survey; et une partie du Veterans Affairs National Outpatient Customer Satisfaction Survey. Deux membres de l'équipe ont analysé le contenu des transcriptions des discussions.

Résultats : Bien que les répondants apprécient la constance dans les choix de réponses offerts, ils préfèrent les choix qui varient de façon à s'ajuster aux questions. Les échelles de réponse de type Likert fonctionnent le mieux; les « échelles d'accord » sont les moins appréciées. Les questions qui tracent une moyenne de l'expérience pour plusieurs fournisseurs de services ou pour plusieurs événements diluent la capacité de détecter des incidents critiques négatifs ou positifs. Les répondants ont tenté de répondre à toutes les questions, mais ont indiqué qu'ils étaient incapables de rapporter avec précision les éléments externes à leur expérience directe ou ceux qui faisaient partie de l'univers du fournisseur de soins. Ils ont apprécié les questions courtes et les instruments courts, à l'exception des cas où un format compact compromettait la clarté. Tous les instruments présentaient des limites en terme de capacité pour faire état de l'interface avec les autres niveaux de soins.

Conclusion : Tous les instruments présentent des points forts et des points faibles, et pourraient donc être légèrement amélioré, cependant les répondants ont détecté avec précision l'intention et l'emploi de chacun d'eux. Les commentaires des répondants ont donné des pistes pour le développement d'instruments.

* * *

In 2004, we concurrently administered six instruments that assess primary healthcare from the patient's perspective to the same group of respondents. We report here on qualitative insights gained from debriefing a subgroup of respondents who responded to the questionnaires in a group setting, then shared their reactions and discussed how well the instruments allowed them to express their primary healthcare experience. In this paper, we describe each instrument briefly and present the reactions that respondents most frequently expressed. Then we present findings that emerge across instruments.

Method

The method, described in detail elsewhere in this special issue of the journal (Haggerty et al. 2011), consisted of administering validated and widely used evaluation instruments to healthcare users in Quebec and Nova Scotia, in French and English, respectively.

The six study instruments compared were (in order of presentation in the questionnaire): (1) the Primary Care Assessment Survey (PCAS); (2) the Primary Care Assessment Tool (PCAT-S, short, adult version); (3) the Components of Primary Care Instrument (CPCI); (4) the EUROPEP-I; (5) the Interpersonal Processes of Care (IPC-II) Survey; and (6) the Veterans Affairs National Outpatient Customer Satisfaction Survey (VANOCSS). We retained only subscales of attributes addressed in more than one instrument. We depicted each instrument as closely as possible to the original version in format, font and instructions to participants.

Thirteen discussion groups were held: three each in metropolitan, rural and remote areas in Quebec and two each in metropolitan and rural areas of Nova Scotia. Participants first had a few minutes to make notes about their most important healthcare experience over the previous year, then they responded to the instruments while their reactions and time were directly observed. They then participated in a recorded, guided group discussion lasting 30 to 40 minutes. The discussion guide, the same for all groups, asked which instruments were most and least liked, and which questions were confusing or difficult and how well the questions allowed them to express the essence of their initial critical incidents. The group facilitators made briefing notes, and recordings were transcribed and analyzed independently for content by two members of the research team who did not participate in the discussions (MF, JH). Issues raised in most group discussions were retained as important.

Results

In total, 110 subjects participated in 13 focus groups (average, nine per group); 64.6% were female, average age was 49.7 (SD: 12.8) years and 67% had a post-secondary education. Overall experience of care was excellent for 49%, average for 38% and poor for 13%, as assessed by the question: "In general, would you say that the care you receive at your regular clinic is excellent, average or poor?"

Description and reaction by instrument

Here, each instrument is described, followed by a synopsis of respondents' reactions. The figures depict a salient sample of items from each instrument showing the formatting actually used.

1. PRIMARY CARE ASSESSMENT SURVEY (PCAS)

The PCAS (Safran et al. 1998) (Figure 1) was developed by Dana Safran to measure the achievement of the Institute of Medicine's definition of primary care: "the provision of integrated, accessible healthcare services by clinicians who are accountable for addressing a large majority of personal healthcare needs, developing a sustained partnership with patients, and practicing in the context of family and community." The instrument applies only to respondents with a "regular personal doctor," and asks respondents to judge or rate the acceptability of different aspects of care. It has a large, easy-to-read font; readability is at grade 4 level. The Likert scale labels change by question context, as does item presentation.

This instrument was most consistently preferred by respondents. It is easy to read, and the adaptation of response scale options to the nature of the questions adds variety to the questionnaire – "made me more alert to answering" [NS] – without feeling "tricky." However, participants did find the Trust subscale questions irrelevant or requiring too much guessing (item 1d).

FIGURE 1. Primary Care Assessment Survey: Sample items from different subscales demonstrating the structure of Likert-type ratings on a 6-point scale

How would you rate the usual **wait** for an appointment when you are sick and call the doctor's office asking **to be seen**?

	☐	☐	☐	☐	☐	☐
1a.	Very poor	Poor	Fair	Good	Very good	Excellent
	[1]	[2]	[3]	[4]	[5]	[6]

Thinking about **talking** with your regular doctor, how would you rate the following:

		[1]	[2]	[3]	[4]	[5]	[6]
		Very poor	Poor	Fair	Good	Very good	Excellent
1b.	Doctor's **explanations** of your health problems or treatments that you need	☐	☐	☐	☐	☐	☐

How often do you leave your doctor's office with **unanswered questions**?

	☐	☐	☐	☐	☐	☐
1c.	Always	Almost always	A lot of the time	Some of the time	Almost never	Never
	[1]	[2]	[3]	[4]	[5]	[6]

Thinking about how much you **TRUST** your doctor, how strongly do you **agree** or **disagree** with the following statements:

		[1]	[2]	[3]	[4]	[5]
		Strongly agree	Agree	Not sure	Disagree	Strongly disagree
1d.	My doctor sometimes **pretends** to know things when he/she is really not sure	☐	☐	☐	☐	☐
1e.	I completely trust my doctor's **judgments** about my medical care	☐	☐	☐	☐	☐

2. PRIMARY CARE ASSESSMENT TOOL (PCAT)

The PCAT (Figure 2) was developed by Barbara Starfield and colleagues (1998; Cassady et al. 2000) to assess the extent to which care is consistent with four unique and three essential attributes of primary care. Initially a paediatric tool, it was adapted and validated for adult care (Shi et al. 2001) and is available in short form (PCAT-S). The items refer to the "Primary Care Provider," and respondents (and evaluators) are guided to identify this as the regular source of care and/or provider who knows the patient best and/or takes responsibility for most care (Strength of Affiliation scale). It has been translated into several languages.

"Primary care provider" is used to refer to both a person and a site. Translating this phrase into French proved too confusing in our cognitive tests, so the French version differentiated "usual doctor" (referring to a healthcare professional) and "clinic" (for site), depending on the context of the question, to achieve appropriate subject–verb agreement. However, even English-language respondents found this term confusing. One wondered "how my doctor could be open or closed" [NS, rural] (e.g., items 2a and 2b). Many found the font too small and the matrix formatting difficult to follow, but various focus groups felt this instrument covered more ground or delved deeper than others.

The response scale elicits probability of occurrence. Some questions are posed hypothetically (e.g., Figure 2, item a), whereas others relate to frequency of occurrence (e.g., Figure 2, item c). The probability response options for questions eliciting frequency were particularly problematic in French, where poor syntax concordance created confusion (e.g., Figure 2, item c). The two extremes of response options were clear – "definitely not" means no or never, "definitely" means yes or always – but the intermediate options were neither clear nor equivalent among respondents.

> Looking at my answers again, I realize that I answered those in a speculative way, which is not an accurate experience. So I think the word "probably" is more of a speculative nature, and perhaps that should be changed. [NS[1], rural]

> For me, "probably," "probably not" and "not sure" all mean the same thing. [QC, remote]

FIGURE 2. Primary Care Assessment Tool: Sample items from two subscales showing a 4-point Likert response scale with the same option labels for all items

	FIRST CONTACT – ACCESS					
	Please check the **one** best answer.	Definitely	Probably	Probably not	Definitely not	Not sure/ Don't remember
2a.	When your Primary Care Provider is *open* and you get sick, would someone from there see you the same day?	4 ☐	3 ☐	2 ☐	1 ☐	9 ☐
2b.	When your Primary Care Provider is *closed*, is there a phone number you can call when you get sick?	4 ☐	3 ☐	2 ☐	1 ☐	9 ☐
	ONGOING CARE					
	Please check the **one** best answer.	Definitely	Probably	Probably not	Definitely not	Not sure/ Don't remember
2c.	When you go to your Primary Care Provider, are you taken care of by the *same* doctor or nurse each time?	4 ☐	3 ☐	2 ☐	1 ☐	9 ☐
2d.	If you have a question, can you call and talk to *the doctor or nurse who knows you best?*	4 ☐	3 ☐	2 ☐	1 ☐	9 ☐
2e.	Does your Primary Care Provider know you very well as a *person*, rather than as someone with a medical problem?	4 ☐	3 ☐	2 ☐	1 ☐	9 ☐

Note: Italics denote emphasis and contrast between similarly worded items in a subscale. Respondents are instructed at the beginning that the questionnaire is about their "Primary Care Provider."

Most problematic for evaluators is the "not sure/don't remember" response option. Although respondents appreciate having this option rather than having to guess, it counts as a missing value in analysis and leads to information loss. For unusual scenarios, such as having a phone number to call when the office is closed (Figure 2, item b), a large proportion of respondents endorsed this response.

3. COMPONENTS OF PRIMARY CARE INSTRUMENT (CPCI)

The 20-item CPCI (Flocke 1997) (Figure 3) was developed based on direct observations of care processes in 138 primary healthcare clinics. Although items are grouped approximately within constructs, no grouping is identifiable in the instrument formatting. It uses a disagree/agree semantic differential response scale with "strongly agree/disagree" labels attached only to the opposite extremes of a set of six categories.

FIGURE 3. Components of Primary Care Instrument: Sample items demonstrating a 6-point semantic differential response scale of the reporting type, with occasional reverse wording

Mark the response that best describes your **regular doctor**

		Strongly disagree [1]	[2]	[3]	[4]	[5]	Strongly agree [6]
3a.	I go to this doctor for almost all of my medical care.	○	○	○	○	○	○
3b.	This doctor handles emergencies.	○	○	○	○	○	○
3c.	This doctor can take care of almost any medical problem I might have.	○	○	○	○	○	○
3d.	I could go to this doctor for help with a personal or emotional problem.	○	○	○	○	○	○
3e.	This doctor does not know my medical history very well.	○	○	○	○	○	○
3f.	This doctor knows a lot about my family medical history.	○	○	○	○	○	○
3g.	This doctor and I have been through a lot together.	○	○	○	○	○	○

Respondents appreciated this instrument's brevity and simple language, but its format was the least liked. Brevity is achieved at the expense of a cramped format, and respondents had difficulty linking statements to the correct response line. Many were observed physically tracking the statements to the response scale. Reverse-worded items were seen as "tricky" (item 3e).

Respondents did not like the semantic differential response scale as much as the Likert scales of other instruments, and they did not like responding to frequency statements with an agree/disagree response (item 3a). Combined with the crowded formatting, they tended to endorse extreme responses rather than use the full response scale. This produced significant halo effects. For instance, many respondents strongly agreed with both positively and negatively worded items in the same construct (e.g., item 3e and 3f).

4. EUROPEP

The EUROPEP (Figure 4) was developed by Richard Grol and colleagues (2000) to compare the performance of general practice in different European countries and to incorporate patient perspectives in care improvement initiatives. The 23-item instrument uses a rating approach to assess (1) Clinical Behaviour, which includes interpersonal communication and technical aspects of care and (2) Organization of Care, which principally addresses issues around accessing care. The instrument has been translated into 15 languages.

FIGURE 4. EUROPEP: Sample items demonstrating 5-point semantic differential response scale of the rating type that is the same for all items, with a "not applicable" option

	How would you rate the following care provided by your general practitioner in the last 12 months?	**1=Poor** **1**	**2**	**3**	**5=Excellent** **4**	**5**	**NA (does not apply)**
4a.	Making you feel you had time during consultations	☐	☐	☐	☐	☐	☐
4b.	Keeping your records and data confidential	☐	☐	☐	☐	☐	☐
4c.	Quick relief of your symptoms	☐	☐	☐	☐	☐	☐
4d.	Thoroughness	☐	☐	☐	☐	☐	☐
4e.	Help in dealing with emotional problems related to your symptoms	☐	☐	☐	☐	☐	☐
4f.	The helpfulness of staff (other than the doctor)	☐	☐	☐	☐	☐	☐
4g.	Getting an appointment to suit you	☐	☐	☐	☐	☐	☐
4h.	Getting through to the practice on the phone	☐	☐	☐	☐	☐	☐
4i.	Providing quick services for urgent problems	☐	☐	☐	☐	☐	☐

The EUROPEP uses a semantic differential response scale ("poor" and "excellent" on a five-point scale). Like the PCAT-S, the EUROPEP offered a "not applicable" response option. It is short with clear formatting, which respondents found easy to read and to answer. Some item statements may be too short (e.g., item 4d, "Thoroughness"), with non-equivalent interpretations among respondents.

> It was very quick and easy to fill out. But I'm not sure that it really describes my experience with my doctor. [NS, rural]

5. INTERPERSONAL PROCESSES OF CARE (IPC)

The IPC (Figure 5) was developed by Anita Stewart and colleagues (1999) to measure interpersonal aspects of quality of care as a complement to technical components. The 45-item instrument was designed to be particularly sensitive to issues of equity and discrimination against patients with limited language proficiency or capacity to advocate strongly in their own care. It has been mostly administered by telephone, so typeface and formatting were discretionary. The questions use a frequency of response scale for "How often did doctors…" – referring to care received from all doctors, not only those in primary care. The items are regularly separated by phrases that frame the next set of questions.

One strength of this instrument is that it includes a separate subscale for assessing office staff behaviour (e.g., items 5e and 5f), whereas others such as the EUROPEP (Figure 4, item f) embed the question among items about the doctor or clinic. Patients often have clearly differentiated experiences between staff and doctors.

> The staff has a great impact on your entire experience. If the receptionist is not someone you feel you can approach, you feel blocked from your Doctor. [NS, rural]

Respondents found the content of the questions highly relevant to critical incidents in their healthcare experience. The major difficulty with this instrument was the frequency response scale, which applied to all providers and all visits, diminishing the importance of individual, negative incidents. The vast majority of respondents selected the two most positive categories of experience. It may be more informative to score dichotomously (see VANOCSS, below), whereby any category other than "always" for positive events or "never" for negative events would be interpreted as a problem.

FIGURE 5. Interpersonal Processes of Care: Sample items demonstrating that the 5-point Likert response scale of the reporting type using a frequency scale is the same for all options (format used was discretionary)

The following questions ask about your experience with your regular doctor or doctors at your regular clinic over the past 12 months.						
		Never [1]	Rarely [2]	Sometimes [3]	Usually [4]	Always [5]
5a.	How often did the doctor(s) speak too fast?	☐	☐	☐	☐	☐
5b.	How often did the doctor(s) let you say what you thought was important?	☐	☐	☐	☐	☐
5c.	How often did the doctor(s) seem bothered if you asked several questions?	☐	☐	☐	☐	☐
The next questions ask about how you and your regular doctor or doctors at your regular clinic decide about your healthcare. Over the past 12 months …						
		Never [1]	Rarely [2]	Sometimes [3]	Usually [4]	Always [5]
5d.	How often did you and your doctor(s) work out a treatment plan together?	☐	☐	☐	☐	☐
The next four questions ask about the doctor's front office staff, meaning the receptionist or the person you talk to on the phone to make an appointment. In the past 12 months…						
		Never [1]	Rarely [2]	Sometimes [3]	Usually [4]	Always [5]
5e.	How often were office staff rude to you?	☐	☐	☐	☐	☐
5f.	How often did office staff have a negative attitude toward you?	☐	☐	☐	☐	☐

6. VETERANS AFFAIRS NATIONAL OUTPATIENT CUSTOMER SATISFACTION SURVEY (VANOCSS)

The VANOCSS (Borowsky et al. 2002) (Figure 6) is based on the Picker–Commonwealth Fund approach to performance assessment (Gerteis et al. 1993). Most items pertain to the last visit, and were therefore excluded from our study; we retained only the usual-care subscales of Overall Coordination and Specialist Access. The instrument elicits the frequency of critical incidents but applies a binary scoring to each item to detect problems. Because the occurrence of at least one problem is considered problematic, this instrument is very sensitive to negative experiences.

FIGURE 6. Veterans Affairs Outpatient Customer Satisfaction Survey (VANOCSS): Sample of items demonstrating a response scale that varies across items

6a. Were the providers who cared for you always familiar with your most recent medical history?
☐ [1] No*
☐ [2] Yes, sometimes*
☐ [3] Yes, always

6b. Were there times when one of your providers did not know about changes in your treatment that another provider recommended?
☐ [1] No
☐ [2] Yes*
☐ [3] No changes in the past 12 months

6c. Were there times when you were confused because different providers told you different things?
☐ [1] No
☐ [2] Yes*

These next questions are about visits in the past 12 months to specialists that your regular doctor knows about or has recommended you see.
IF YOU HAVE NOT SEEN A SPECIALIST IN THE PAST 12 MONTHS, PLEASE GO TO…
6d. How often did you get to see specialists when you thought you needed to?
☐ [1] Never*
☐ [2] Sometimes*
☐ [3] Usually
☐ [4] Often
☐ [5] Didn't need any specialists

6e. How often did your specialists have the information they needed from your medical records?
☐ [1] Never*
☐ [2] Sometimes*
☐ [3] Usually
☐ [4] Often
☐ [5] Didn't need any specialists

* Indicates response options that are considered problematic for dichotomous scoring (not indicated on the respondent version).

Because the instrument applies only to those who had seen providers other than the regular doctor, we placed it last in the questionnaire. Not all respondents answered it, and we obtained little qualitative information. However, it addresses whole-system issues that respondents identified as important.

Response formats

Among the various instruments and response scale formats, respondents consistently preferred Likert response scales, sans serif fonts in large sizes, squares as opposed to circles for responses, questions that were grouped under headings, and varying response scales and formats throughout the questionnaire. They found matrix formatting (e.g., PCAT-S and some PCAS) difficult. Respondents appreciated being able to indicate that an item was not applicable or that they did not know, as in the PCAT-S and EUROPEP. Several expressed confusion about scoring information indicated in response options (e.g., "Always 5").

Bad questions

When asked about questions they did not like, respondents consistently stated they did not like assessing processes occurring in the provider's mind or that they did not directly observe or experience. Almost all questionnaires contained such questions. For example, the PCAS

Trust subscale asked about the doctor pretending to "know things" (Figure 1, item d), and the EUROPEP asked about the confidential treatment of the medical record (Figure 4, item b). Respondents gave the benefit of the doubt and recognized that their best guess would not be an accurate assessment.

> One respondent said: "There aren't many people who know what their doctor [knows] … do you all know what your doctor thinks?" Another respondent stated: "Probably not." [QC, urban]

Items about the depth of the physician–patient relationship provoked divergent and strong reactions. Some participants had few expectations that the physician would know anything beyond their medical history; others attested to the importance of whole-person knowledge in appropriate care:

> They ask questions that don't make any sense: "this doctor and I have been through a lot together," "I could consult this doctor for a personal or emotional problem." Come on. You're not close to your doctor, to consult for a personal problem. You go … to see the doctor. [QC, rural]

> I find the more your doctor knows you and your particulars, the better he's going to be able to adjust things. [NS, urban]

Likewise, despite the importance of respectfulness in defining respondents' experience, the notion of equality was problematic. One item in particular evoked comment from several groups: "How often did doctors treat you as an equal?" (IPC)

> I want [a professional doctor–patient relationship], I don't want to be treated as an equal, I don't want him to talk to me like I'm an equal because I don't understand – I want him to talk to me in plain language but to talk to me with respect and caring and understanding. [NS, rural]

Missing pieces across instruments

Most respondents said the instruments elicited general aspects of their healthcare experience that they considered important. Several groups felt 12 months was too short a reference because it encompassed only one or two visits, whereas their experience was informed by a longer period.

Some missing pieces related to reporting on the system as a whole and others to the regular physician. Respondents consistently expressed frustration at being limited to evaluating their regular doctor because many of the most important critical incidents occur at other levels of the system or outside the clinical encounter. They wanted more questions on delays for access to tests and specialists and elicitation of technical mistakes.

Respondents clearly welcomed an opportunity to give anonymous feedback, and they expected this information to go directly to their physician. They also wanted to communicate to their physicians their desire for more input into decision-making regarding choices of specialist or treatment options.

Respectfulness is critical to patients, but respondents felt there were not enough questions addressing aspects of respectfulness such as respecting the appointment time, responding to voicemail messages, recognizing personal worth and provision of privacy and confidentiality by office staff. Respondents wanted more questions about the physical set-up of the waiting room (privacy, toys for children), cleanliness of the bathroom or the clinic in general and wheelchair accessibility.

Discussion and Conclusion

Every instrument has features that respondents appreciated: the readability of the PCAS; the PCAT-S's comprehensiveness and "don't know" response option (although problematic analytically); the CPCI's clear language; the EUROPEP's conciseness; the IPC's focus on respectfulness; and the VANOCSS's whole-system perspective. The PCAS was considered the most readable and applicable overall, although its placement as first in the questionnaire may have affected this perception. All instruments include questions that patients found difficult to evaluate: things not directly observed or occurring in the provider's mind. Patients can accurately report critical incidents, such as breaches in confidentiality or trust, but if they do not do so, it does not mean these incidents have not occurred. Offering a "don't know" option gives patients more options but results in missing values for analysis. Each instrument has its strength, but this study also provides pointers for continued refinement.

Our results reinforce recommended good practice about questionnaire formatting such as large typeface, instructions placed close to responses, presenting questions together by thematic grouping and minimizing response format changes between sections (Dillman 2007). Additionally, we recommend removing numeric scoring information from response options, using matrix formatting sparingly and using sans serif fonts.

Respondents not only tolerated variation in response scales but appreciated variety. The PCAS is exemplary in this regard; the response scale changes across dimensions while remaining consistent within a block of questions, facilitating the response task. Likewise, validity may be compromised by using a graded response scale for occurrences that are rare, when a binary (yes/no) option would be better. The VANOCSS offers an interesting combination of graded reporting of events with binary scoring, an approach that may be relevant for instruments with highly skewed responses.

Finally, respondents affirmed the importance they accord to evaluating healthcare services and their expectation that this information be used to communicate suggestions for improvement anonymously and clearly to their providers. They desire to be "good respondents" and will respond to questions even when they are not sure of the answer or of what the question means.

The information in this study can be used to ensure that evaluators and clinicians select instruments that not only demonstrate acceptable psychometric properties, but are also well accepted by patients.

NOTE

[1] Citations are from Nova Scotia [NS] or Quebec [QC]; Quebec citations have been translated from French.

ACKNOWLEDGEMENTS

This research was funded by the Canadian Institute for Health Research (CIHR). The authors wish to thank Donna Riley for support in preparation and editing of the manuscript.

Correspondence may be directed to: Jeannie L. Haggerty, Associate Professor, Department of Family Medicine, McGill University, St. Mary's Research Centre, Hayes Pavilion – Suite 3734, 3830 Lacombe Ave., Montreal QC H3T 1M5; tel.: 514-345-3511 ext. 6332; fax: 514-734-2652; e-mail: jeannie.haggerty@mcgill.ca.

REFERENCES

Borowsky, S.J., D.B. Nelson, J.C. Fortney, A.N. Hedeen, J.L. Bradley and M.K. Chapko. 2002. "VA Community-Based Outpatient Clinics: Performance Measures Based on Patient Perceptions of Care." *Medical Care* 40: 578–86.

Cassady, C.E., B. Starfield, M.P. Hurtado, R.A. Berk, J.P. Nanda and L.A. Friedenberg. 2000. "Measuring Consumer Experiences with Primary Care." *Pediatrics* 105(4): 998–1003.

Dillman, D.A. 2007. *Mail and Internet Surveys: The Tailored Design Method* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Flocke, S. 1997. "Measuring Attributes of Primary Care: Development of a New Instrument." *Journal of Family Practice* 45(1): 64–74.

Gerteis, M., S. Edgman-Levitan, J. Daley and T.L. Delbanco. 1993. *Through the Patient's Eyes: Understanding and Promoting Patient-Centered Care.* San Francisco: Jossey-Bass.

Grol, R., M. Wensing and Task Force on Patient Evaluations of General Practice. 2000. *Patients Evaluate General/Family Practice: The EUROPEP Instrument.* Nijmegen, Netherlands: Centre for Research on Quality in Family Practice, University of Nijmegen.

Haggerty, J.L., F. Burge, M.-D. Beaulieu, R. Pineault, C. Beaulieu, J.-F. Lévesque et al. 2011. "Validation of Instruments to Evaluate Primary Healthcare from the Patient Perspective: Overview of the Method" *Healthcare Policy* 7 (Special Issue): 31–46.

Safran, D.G., J. Kosinski, A.R. Tarlov, W.H. Rogers, D.A. Taira, N. Leiberman and J.E. Ware. 1998. "The Primary Care Assessment Survey: Tests of Data Quality and Measurement Performance." *Medical Care* 36(5): 728–39.

Shi, L., B. Starfield and J. Xu. 2001. "Validating the Adult Primary Care Assessment Tool." *Journal of Family Practice* 50(2): 161–71.

Starfield, B., C. Cassady, J. Nanda, C.B. Forrest and R. Berk. 1998. "Consumer Experiences and Provider Perceptions of the Quality of Primary Care: Implications for Managed Care." *Journal of Family Practice* 46: 216–26.

Stewart, A.L., A. Nápoles-Springer and E.J. Pérez-Stable. 1999. "Interpersonal Processes of Care in Diverse Populations." *Milbank Quarterly* 77(3): 305–39, 274.

RESEARCH PAPER

An Overview of Confirmatory Factor Analysis and Item Response Analysis Applied to Instruments to Evaluate Primary Healthcare

Aperçu de l'analyse factorielle confirmatoire et de l'analyse de réponse par item appliquées aux instruments d'évaluation des soins primaires

DARCY A. SANTOR, PHD
School of Psychology, University of Ottawa
Ottawa, ON

JEANNIE L. HAGGERTY, PHD
Department of Family Medicine, McGill University
Montreal, QC

JEAN-FRÉDÉRIC LÉVESQUE, MD, PHD
Centre de recherche du Centre hospitalier de l'Université de Montréal
Montréal, QC

FREDERICK BURGE, MD, MSC
Department of Family Medicine, Dalhousie University
Halifax, NS

MARIE-DOMINIQUE BEAULIEU, MD, MSC
Chaire Dr Sadok Besrour en médecine familiale
Centre de recherche du Centre hospitalier de l'Université de Montréal,
Montréal, QC

DAVID GASS, MD
Department of Family Medicine, Dalhousie University
Halifax, NS

RAYNALD PINEAULT, MD, PHD
Centre de recherche du Centre hospitalier de l'Université de Montréal,
Montréal, QC

Abstract

This paper presents an overview of the analytic approaches that we used to assess the performance and structure of measures that evaluate primary healthcare; six instruments were administered concurrently to the same set of patients. The purpose is (a) to provide clinicians, researchers and policy makers with an overview of the psychometric methods used in this series of papers to assess instrument performance and (b) to articulate briefly the rationale, the criteria used and the ways in which results can be interpreted. For illustration, we use the case of instrument subscales evaluating accessibility. We discuss (1) distribution of items, including treatment of missing values, (2) exploratory and confirmatory factor analysis to identify how items from different subscales relate to a single underlying construct or sub-dimension and (3) item response theory analysis to examine whether items can discriminate differences between individuals with high and low scores, and whether the response options work well. Any conclusion about the relative performance of instruments or items will depend on the type of analytic technique used. Our study design and analytic methods allow us to compare instrument subscales, discern common constructs and identify potentially problematic items.

Résumé

Cet article présente un aperçu des approches analytiques que nous avons utilisées pour évaluer le rendement et la structure des mesures qui servent à évaluer les soins de santé primaires : six instruments ont été appliqués simultanément au même groupe de patients. L'objectif est (a) de fournir aux cliniciens, aux chercheurs et aux responsables de politiques, un aperçu des méthodes psychométriques utilisées dans cette série pour évaluer le rendement de l'instrument et (b) d'articuler brièvement l'analyse raisonnée, les critères employés et les façons dont peuvent être interprétés les résultats. À titre d'exemple, nous avons utilisé le cas des sous-échelles qui servent à évaluer l'accessibilité. Nous discutons (1) la distribution des items, y compris le traitement des valeurs manquantes, (2) les analyses factorielles exploratoires et confirmatoires afin de voir comment les items de différentes sous-échelles sont liés à un seul construit (ou sous-dimension) sous-jacent et (3) l'analyse de réponse par item pour voir si les items permettent de discriminer les différences entre les unités qui présentent des scores élevés et faibles, et pour voir si les choix de réponses fonctionnent bien. Toute conclusion sur le rendement relatif des instruments ou des items dépend du type de technique analytique employé. La conception et les méthodes analytiques de cette étude permettent de comparer les sous-échelles des instruments, de discerner les construit communs et de repérer les items potentiellement problématiques.

* * *

PSYCHOMETRIC SCALES AND INSTRUMENTS HAVE BEEN USED TO ASSESS VIRTUALLY every component of healthcare, whether to identify gaps in service, to assess special needs of patients, or to evaluate the performance and efficiencies of programs, organizations or entire healthcare systems. In this special issue of *Healthcare Policy*, we examine the performance of several instruments that assess different attributes of primary healthcare from the patient's perspective. Any conclusion about the relative performance of instruments and items from those instruments will depend on the type of analytic technique used to assess performance.

The purpose of this paper is to describe the analytic approach and psychometric methods that we used to assess and compare the performance of the instrument subscales. We articulate the rationale of different approaches, the criteria we used and how results can be interpreted. For illustration, we use the case of instrument subscales evaluating accessibility, described in detail elsewhere in this special issue (Haggerty, Lévesque et al. 2011).

Data Sources

All the instruments used in this study have previously been validated and meet standard criteria for validity and reliability. The goal of the study was to extend this process of validation and compare the performance of the six instruments for measuring core attributes of primary healthcare for the Canadian context. Our intention is not to recommend one instrument over another, but to provide insight into how different subscales measure various primary healthcare attributes, such as accessibility and care.

We administered the six instruments to 645 health service users in Nova Scotia and Quebec: the Primary Care Assessment Survey (PCAS, Safran et al. 1998); the adult version of the Primary Care Assessment Tool – Short (PCAT-S, Shi et al. 2001); the Components of Primary Care Index (CPCI, Flocke 1997); the first version of the EUROPEP (EUROPEP-I, Grol et al. 2000); the Interpersonal Processes of Care – 18-item version (IPC-II, Stewart et al. 2007); and the Veterans Affairs National Outpatient Customer Satisfaction Survey (VANOCSS, Borowsky et al. 2002). The sample was balanced by overall rating of primary healthcare, high and low level of education, rural and urban location, and English and French language (Haggerty, Burge et al. 2011).

Distribution of Responses

The first step was to examine the distribution of the responses, flagging as problematic items where a high proportion of respondents select the most negative (floor effect) or the most positive (ceiling effect) response options, or have missing values. Missing values, where respondents failed to respond or wrote in other answers, may indicate questions that are not clear or are difficult to understand. We expected low rates of true missing values because truly problematic items would be eliminated during initial validation by the instrument developers, but remained sensitive to items that are problematic in the Canadian context.

However, some instruments offer response options such as "not applicable" (EUROPEP-I) or "not sure" (PCAT-S), which count as missing values in analysis because they cannot be interpreted as part of the ordinal response scale. They represent a loss of information as they cannot be interpreted.

Missing information, for whatever reason, is problematic for our study given that missing information on any item means that data for the entire participant is excluded from factor analysis (listwise missing), compromising statistical power and potentially introducing bias.

In the case of accessibility subscales, 340 of the 645 respondents (53%) were excluded from factor analysis because of missing values, of which 267 (79%) were for selecting "not sure" or "not applicable" options. We examined the potential for bias by testing for differences between included and excluded respondents on all relevant demographic and healthcare use variables. We also imputed values for most of the missing values using maximum likelihood imputation (Jöreskog and Sörbom 1996), which uses the subject's responses to other items and characteristics to impute a likely value. Then, we repeated all the factor analyses to ensure that our conclusions and interpretations remained unchanged, and reducing the possibility of bias. Nonetheless, the high proportion of missing values in some instances is an important limitation of our study, and needs to be considered in the selection of instruments.

Subscale Scores

Next, we examined the performance by subscale. Subscale scores were mostly calculated as the mean of item values if over 50% of the items were complete. This score was not affected by the number of items, and it reflects the response options. For example, a subscale score of 3.9 in Organizational Access on the PCAS corresponds approximately to the "4=good" option on the response scale of 1 (very poor) to 6 (excellent). But it is difficult to know how this compares to the score of 3.6 on the EUROPEP-I Organization of Care for a similar dimension of accessibility from 1 (poor) to 5 (excellent). To compare the subscales between different response scales, we normalized scores to a common 0-to-10 metric using the following formula:

$$\text{New score} = \{(\text{raw score} - \text{minimum possible}) / (\text{raw maximum} - \text{raw minimum})\} * 10$$

So, the normalized mean for PCAS Organizational Access, 5.9, is seen to be considerably lower than 6.5 on the EUROPEP-I Organization of Care, and the PCAS variance is lower than the EUROPEP-I (1.8 vs. 2.4). Thus, if accessibility were measured in one population using the PCAS and in another using the EUROPEP-I, the scores of the EUROPEP-I would be expected to be higher than the PCAS, even if there were no difference in accessibility between the two populations.

Reliabilities

The reliability of each subscale was evaluated using Cronbach's coefficient α, which estimates how much each item functions as a parallel, though correlated, test of the underlying construct. Cronbach's α ranges from 0 (items completely uncorrelated, all variance is random) to 1 (each item yields identical information), with the convention of .70 indicating a minimally reliable subscale. The subscales included in our study all reported adequate to good internal consistency. The coefficient α is sensitive to sample variation as well as the number and quality of items. Given that our study sample was selected to overrepresent the extremes of poor and excellent experience with primary healthcare relative to a randomly selected sample, we expected our alpha estimates to meet or exceed the reported values.

We then calculated Pearson correlations among the subscale scores, controlling for educational level, geography and language (partial correlation coefficients) to account for slight deviations from our original balanced sampling design. We expect high correlations between subscales mapped to the same attribute (convergent validity) and lower correlations with subscales from other attributes (some degree of divergent validity), indicating that the items in the subscales are indeed specific to that attribute and that respondents appropriately distinguish between attributes. Pearson correlation coefficients indicate expected relationships observed in factor analysis. If we observe high correlations within an attribute, then we would expect all the items from those subscales to "load" on a common factor. So, for example, after the correlation analysis for accessibility, we observed that the PCAT-S First-Contact Utilization subscale correlated less strongly with the other accessibility subscales than it did with relational continuity. We had high expectations that its items would not only form a separate factor, but also that it would relate poorly to accessibility as a whole.

Exploratory and Confirmatory Factor Analysis

Subscales from different instruments that were designed to assess the same primary healthcare attribute should relate to a single underlying factor or construct. We used both exploratory and confirmatory factor analysis to examine this premise, as well as to determine whether items across subscales related to sub-dimensions within the attribute.

Exploratory analyses

Exploratory factor analysis is a descriptive technique that can detect an overarching structure that explains the relationships between items in a parsimonious way. We used the common factor analysis procedure computed in SAS v. 9.1 (SAS 2003). This procedure identifies how much items can be represented by a smaller group of variables (i.e., common factors) that account for as much of the variability in the data as possible. The procedure assigns an eigenvalue to each factor that corresponds to the total variance in item responses that can be explained by the factor. Typically, factors with eigenvalues greater than 1.0 are retained.

This procedure also computes how strongly each individual item maps on to each factor. "Factor loadings" range from −1.0 to 1.0 and can be interpreted much like a correlation coefficient. These indicate (a) the extent to which all items relate to one or more distinctive factors, (b) how strongly each item is related to each factor (and whether the item should be retained or eliminated within a factor) and (c) how much variation in responses to items can be accounted for by each factor or subgroup. We considered items with factor loadings $\geq | .4 |$ as strongly related to the underlying factor. It is important to note that common factor analysis assumes a normal distribution; items with highly skewed distributions will affect both the loadings and the extent to which factors can be easily interpreted (Gorsuch 1983).

Results of an exploratory factor analysis for items in subscales from three different instruments assessing accessibility are presented in Table 1. Factor loadings are presented for each item showing how each item is related to three distinct factors. The first factor has a large eigenvalue (7.84) and accounts for approximately 41% of variance in the responses given to items, compared to just 6% for the second factor (eigenvalue=1.19) and less than 1% for the third. As a result, only two factors would be considered worth interpreting. This confirms our expectation based on the correlation analysis that the PCAT-S First-Contact Utilization subscale might not fit with other accessibility subscales. The two important underlying factors could be characterized as timeliness and accommodation (Haggerty, Lévesque et al. 2011).

TABLE 1. Factor loadings from an oblique exploratory principal component analysis for accessibility items drawn from four measures of accessibility

		Factor 1 Eigen=7.84	Factor 2 Eigen=1.19	Factor 3 Eigen=0.77
PCAS: Accessibility (Organizational Access)				
PS_oa1	How would you rate the convenience of your regular doctor's office **location**?	—	—	—
PS_oa2	How would you rate the **hours** that your doctor's office is open for medical appointments?	—	—	—
PS_oa3	How would you rate the usual **wait** for an appointment when you are sick and call the doctor's office asking to be seen?	0.41	0.49	—
PS_oa4	How would you rate the amount of time you **wait** at your doctor's office for your appointment to **start**?	—	0.91	—
PS_oa5	Thinking about the times you have needed to **see or talk to** your doctor, how would you rate the following: ability to **get through to** the doctor's office by phone?	—	0.51	—
PS_oa6	Thinking about the times you have needed to **see or talk to** your doctor, how would you rate the following: ability to **speak to** your doctor by phone when you have a question or need medical advice?	—	0.55	—
PCAT-S: Accessibility (First-Contact Utilization)				
PT_fcu1	When you need a regular general checkup, do you go to your Primary Care Provider before going somewhere else?	—	—	0.78
PT_fcu2	When you have a new health problem, do you go to your Primary Care Provider before going somewhere else?	—	—	0.74
PT_fcu3	When you have to see a specialist, does your Primary Care Provider have to approve or give you a referral?	—	—	—
PCAT-S: Accessibility (First-Contact Accessibility)				
PT_fca1	When your Primary Care Provider is **open** and you get sick, would someone from there see you the same day?	0.61	—	—
PT_fca2	When your Primary Care Provider is **open**, can you get advice quickly over the phone if you need it?	0.64	—	—
PT_fca3	When your Primary Care Provider is **closed**, is there a phone number you can call when you get sick?	0.73	—	—
PT_fca4	When your Primary Care Provider is **closed** and you get sick **during the night**, would someone from there see you that night?	0.65	—	—
EUROPEP-I: Accessibility (Organization of Care)				
EU_oc1	Preparing you for what to expect from specialist or hospital care	0.51	—	—
EU_oc2	The helpfulness of staff (other than the doctor)	0.41	—	—
EU_oc3	Getting an appointment to suit you	0.51	—	—
EU_oc4	Getting through to the practice on the phone	0.42	—	—

TABLE 1. Continued

		Factor 1 Eigen=7.84	Factor 2 Eigen=1.19	Factor 3 Eigen=0.77
EU_oc5	Being able to speak to the general practitioner on the telephone	0.68	—	—
EU_oc6	Waiting time in the waiting room	—	0.87	—
EU_oc7	Providing quick services for urgent health problems	0.64	—	—

Note: Factor loadings smaller than 0.40 have not been presented.

Confirmatory analyses

Confirmatory factor analysis differs from exploratory factor analysis by allowing the investigator to impose a structure or model on the data and test how well that model "fits." The "model" is a hypothesis about (a) the number of factors, (b) whether the factors are correlated or uncorrelated and (c) how items are associated with the factor. Models with different configurations are compared using structural equation modelling. Statistical software packages produce various "goodness-of-fit" statistics that capture how well the implied variance–covariance matrix of the proposed model corresponds to the observed variance–covariance matrix (i.e., how items from the instrument actually correlate). Confirmatory factor analysis attempts to account for the covariation among items (ignoring error variance), whereas common factor analysis accounts for the "common variance" shared among items, differentiating variance attributable to an underlying factor and error variance. Although similar in spirit, factor loadings are computed in fundamentally different ways from common factor analysis and should be interpreted differently.

Our premise was that different measures of an attribute can still be viewed as indicators (i.e., items) assessing the same underlying construct despite being drawn from different instruments employing different phrasing and response scales. Testing this hypothesis allows researchers and policy makers to view similarly the results obtained from different measures of, say, accessibility.

Figure 1 presents the results of a confirmatory factor analysis for accessibility subscales. Figure 1A depicts a standard unidimensional model, where every item is linked to the same, single underlying construct, namely accessibility. Constructs (designated with ellipses) are linked to (designated with arrows with loading coefficients) individual items (designated with rectangular boxes). The model shows that most items load strongly (i.e., factor loading greater than .90) on the latent construct called Access, but that some items do not (e.g., loadings of .71 and .78) or they have high residual error, shown to the right of the item.

The performance of the model is evaluated by examining the ensemble of "goodness-of-fit" statistics, such as the comparative fit index (CFI), the normed fit index (NFI), consistent Akaike's information criterion (CAIC) or the root mean squared error of approximation (RMSEA). They all assess in different ways the discrepancy between the pattern of variances and covariances implied by the model and the actual pattern of variances and covariances observed in the data (see Kline 1998 for an in-depth review of basic issues in structural equation modelling). If the implied pattern is close to what is observed in the data, then the model is said to fit – it accurately accounts for the manner in which items are interrelated.

Fit statistics for the model in Figure 1A were all good. Unlike the usual interpretation of significance, lower chi-squared (χ^2) values suggest better fit. The χ^2 value was 649 with 152 degrees of freedom and was significant, which might indicate the model does not fit well (though χ^2 is sensitive to large samples such as ours). However, other fit statistics, such as NFI, CAIC and GFI (results not shown), which take into consideration both the sample size and model complexity, were all well above the conventional criterion of .90 of "good fit." The RMSEA of .104 is higher than the .05 criterion indicating good fit, but is still reasonable. Altogether, these results suggest that although items were drawn from distinct subscales, response to questions can be accounted for by a single underlying construct, namely, accessibility.

However, we might hypothesize that because items were drawn from subscales with different numbers of response options and formats, the pattern of responses would be even better explained by a model that explicitly locates individual items with the subscale from their parent instrument (first-order factor) and then links these first-order factors within the general construct of accessibility (a second-order factor). Figure 1B depicts this second-order, multidimensional model. Fit statistics for this model were also extremely good. Again, NFI, CAIC and GFI were all .98. The χ^2 value was 514 (with 148 degrees of freedom).

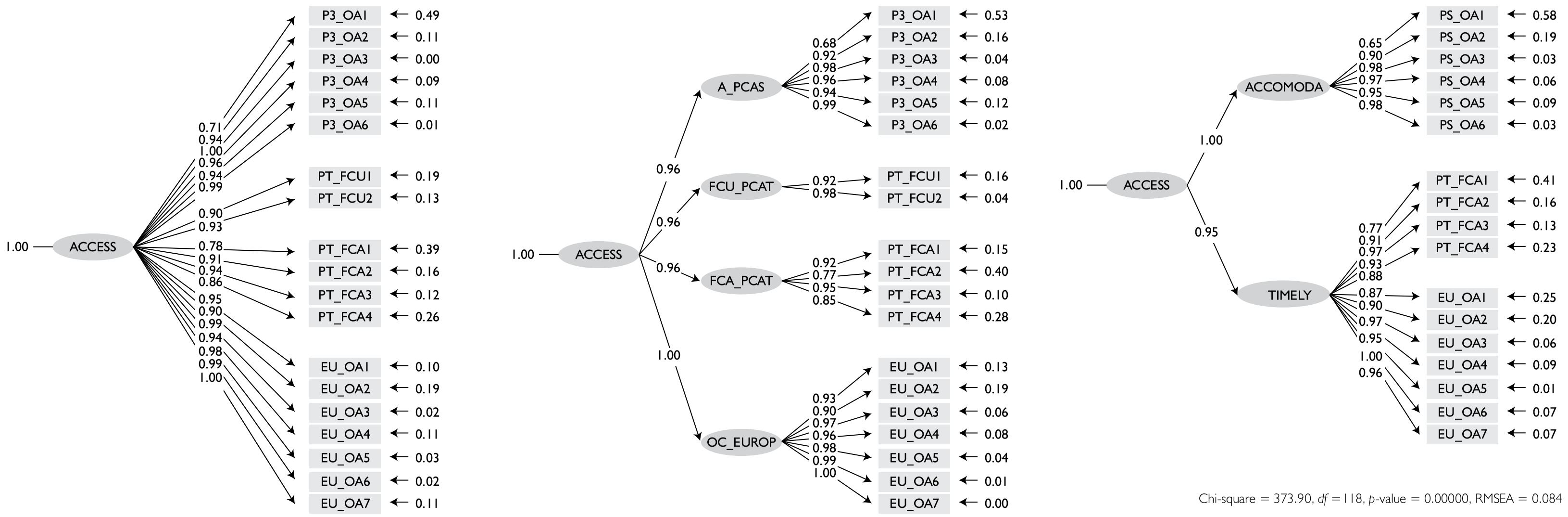
Comparing models

One of the strengths of a confirmatory factor analysis is the ability to compare "nested" models, where one model is a simpler version of a more complex model. Because these models differ only in the number of paths that are being estimated, χ^2 values for one model can be subtracted from the other and the significance of the difference evaluated. The χ^2 difference between the simple model in Figure 1A and the more complex model in Figure 1B (649 – 514 = 135, 4 *df*) is statistically significant, so we can infer that the complex model is more valid. Some of the variability in how individuals respond to questions is not just determined by the underlying construct of accessibility, but also by the specific measure from which the question is drawn.

We also test a model that groups items within sub-dimensions of accessibility. Figure 1C depicts a second-order, multidimensional model in which items are grouped within two first-order factors, namely, the timeliness of service and the extent to which patients' access barriers are accommodated, which are themselves part of a broader, second-order factor: accessibility. This model says there are two components of the more general construct of accessibility, and that these transcend specific instrument subscales.

It is important to note that not all models can be compared directly. The model in Figure 1C does not include the items from the PCAT First-Contact Utilization subscale; it differs from the model in Figure 1A by more than the number of paths. To test the validity of this model, we compared its own restricted version rather than the model depicted in Figure 1A and found that grouping items within sub-factors of timeliness and accommodation is superior to the one-dimensional model (χ^2 426 – 364 = 52, 3 degrees of freedom).

FIGURE 1. Results of a confirmatory factor analysis for accessibility subscales



1A. Standard unidimensional model

Chi-square = 649.37, *df* =152, *p*-value = 0.00000, RMSEA = 0.104

Note: This type of model says that variability in response to individual items can be explained by a single underlying factor, namely accessibility, irrespective of the scale from which items are drawn.

1B. Second-order, multidimensional model

Chi-square = 514.03, *df* =148, *p*-value = 0.00000, RMSEA = 0.090

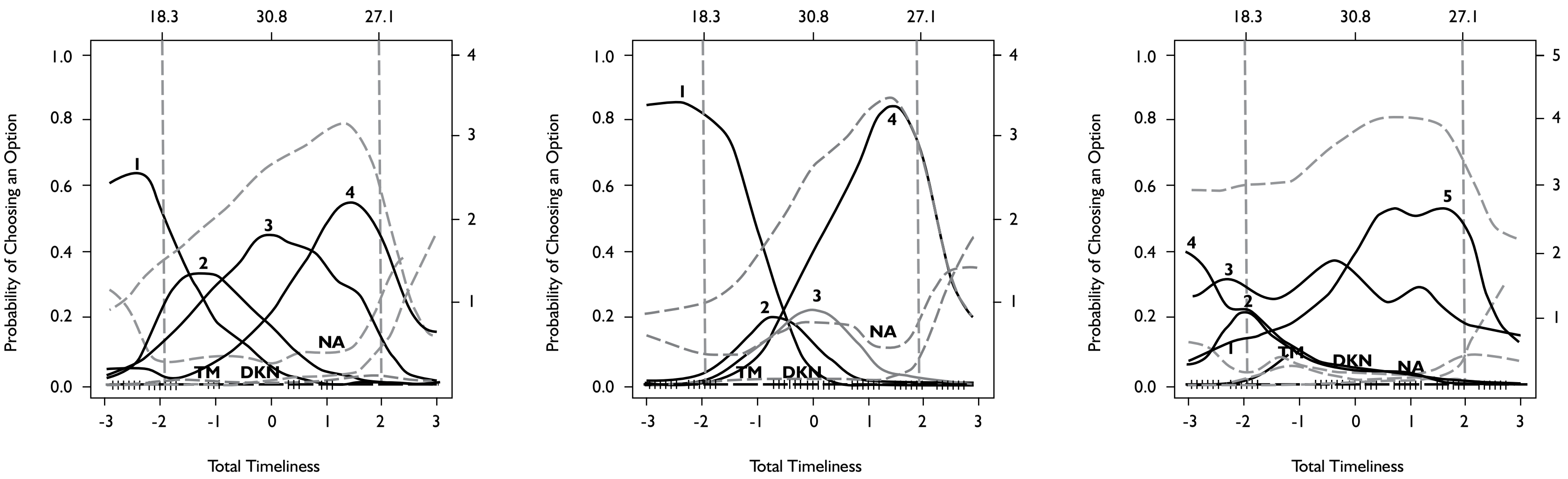
Note: Results show that the standard, second-order multidimensional model, in which individual items are located first within the scale from which they come (first-order factor) and are then linked to the general construct of accessibility (a second-order factor), also fit the data well, but not better than the unidimensional model.

1C. Second-order, multidimensional model

Chi-square = 373.90, *df* =118, *p*-value = 0.00000, RMSEA = 0.084

Note: Results show that a revised, second-order multidimensional model, in which items are first located within two different facets of accessibility, namely the timeliness of service and the extent to which patient's needs are accommodated (first-order factors) and then linked to the general construct of accessibility (a second-order factor), also fit the data well, but not better than the unidimensional model.

FIGURE 2. Response graphs for three items drawn from the PCAT-S and EUROPEP-I subscales assessing construct of timeliness within accessibility



2A.

Note: Option characteristic curves (solid lines) and expected total scores (broken lines) for item 2, "When your Primary Care Provider is open, can you get advice quickly over the phone if you need it?" from the PCAT-S are modelled as a function of total scores on these measures (top axis) and standard normal quantiles (bottom axis). Results show that this item performs well relative to other items. First, option characteristic curves cover the entire range of scores. Second, the region in which curves are most likely to be endorsed is in a distinct region that corresponds to the weight assigned to each option. The probability of true missing (TM), not applicable (NA) and don't know (DK) responses did not vary substantially as a function of total scores for the majority of participants, falling between –2 and +2 standard normal quantiles.

2B.

Note: Option characteristic curves (solid lines) and expected total scores (broken lines) for item 3, "When your Primary Care Provider is closed, is there a phone number you can call when you get sick?" from the PCAT-S are modelled as a function of total scores on these measures (top axis) and standard normal quantiles (bottom axis). Results show difficulties with some options from this item. First, only two options, namely options 1 and 4, are being endorsed with any frequency, suggesting that this item is essentially dichotomous. The probability of endorsing options 2 or 3 is relatively small compared to options 1 and 4. The probability of true missing (TM), not applicable (NA) and don't know (DK) responses did not vary substantially as a function of total scores for the majority of participants, falling between –2 and +2 standard normal quantiles.

2C.

Note: Option characteristic curves for item 2 from the EUROPEP-I, "The helpfulness of staff (other than the doctor)." Results show a number of difficulties with this item. First, the probability of endorsing option 3 is greater than probability of endorsing option 2 at all levels of total scores. Second, most option characteristic curves do not change rapidly with changes in total scores, which is captured in an essentially flat expected item score. The probability of true missing (TM), not applicable (NA) and don't know (DK) responses did not vary substantially as a function of total scores for the majority of participants.

Item Response Models

Item response analysis evaluates how well questions and individual response options perform at different levels of the underlying construct being evaluated. They provide a fine-grained level of analysis that can be used to evaluate how well individual items and options discriminate among individuals at both high and low levels of the construct and can identify items and options that should ideally be revised or, if necessary, discarded.

Defining a shared or common underlying continuum against which items can be compared is crucial, because any result will be contingent on the appropriateness of this common underlying dimension. Each item's performance was modelled two ways, as a function of (1) items drawn from the single original subscale and (2) all items from any subscale that appear to measure a common construct, for example, timeliness within accessibility. Items are likely to perform better when modelled as a function of just the instrument from which they were drawn. However, because our goal is to compare the relative performance of all items (which may come from different instruments) that are believed to assess a similar construct (i.e., accessibility), we report on items modelled on the shared or common underlying dimension (e.g., timeliness within accessibility).

We used a non-parametric item response model to examine item performance on the common underlying factor (Ramsay 2000). This is an exploratory approach (Santor and Ramsay 1998), and these techniques have been used previously to examine the psychometric properties of self-report items and to evaluate item bias (Santor et al. 1994; Santor et al. 1995; Santor and Coyne 2001). A detailed description of the algorithm used to estimate response curves has been published elsewhere (Ramsay 1991).

We supplemented our non-parametric models with parametric item response modelling to estimate the discriminatory capacity of each item within its own original subscale using Multilog (Du Toit 2003). Discriminability (the "a" parameter) indicates an item's sensitivity to detect differences among individuals ranked on the construct being measured (e.g., accessibility). It can be viewed as a slope, with a value of 1 considered the lower limit for acceptable discriminability, i.e., each unit increase in the item predicts a unit increase in the underlying construct. Items with lower discriminability in the parent construct invariably performed poorly on the common underlying dimension with non-parametric item response modelling.

Examining item performance

To illustrate how item response models can be used to evaluate item and response option performance, Figures 2A, 2B and 2C show item response graphs for three items drawn from the PCAT-S and EUROPEP-I subscales assessing construct of timeliness within accessibility. Figure 2A presents a relatively well-performing item from the PCAT-S; Figures 2B and 2C illustrate some difficulties in the other two items.

In the Figure 2 graphs, the total expected score for timeliness is presented at the top of the plot; below the horizontal axis on the bottom, it is represented as standard normal scores. Expressing scores as standard normal scores is useful because it is informative about the proportion of a population above or below integer values of standard deviations from the mean score. So in the graphs we can see that −2 SD corresponds to a total timeliness score of 18.3, the mean is 30.8, and +2 SD is 27.1. Extreme values on curves need to be interpreted with caution because, by definition, sample sizes are small in the tails of the overall distribution of scores.

The overall performance of the item is captured in the steepness of the slope of the characteristic curve (the topmost dashed lines in Figure 2A–C), which expresses item discriminability, the relationship between the cumulated item score and the total score in the construct (e.g., total timeliness). Given that we calculated items from different instruments as a function of a common continuum, slopes can be compared directly to assess performance across different subscales.

Several important features of item performance are illustrated in Figure 2A for an item from the PCAT-S. First, each of the option characteristic curves (a solid line probability curve for each response option) increases rapidly with small increases in timeliness. For example, the probability of option 1 being endorsed increases rapidly from 0.0 to 0.6 over a narrow region of timeliness, −3.0 to −1.5. Second, each option tends to be endorsed most frequently in a specific range of timeliness. For example, option 2 is more likely to be endorsed than any other option within the timeliness range of −1.0 to 0.0. Third, the regions over which each option is most likely to be endorsed are ordered, left to right, in the same way as the option scores (weights, 1 to 4). That is, the region in which option 2 is most likely to be endorsed falls between the regions in which option 1 and option 3 are most likely to be endorsed. Finally, together, the options for an item span the full continuum of accessibility, from −3 to +3. Most positive options are endorsed only at high levels of timeliness (e.g., option 5), whereas most negative options are endorsed only at low levels of timeliness (e.g., option 1).

In contrast, Figure 2B shows an item from the PCAT-S with four response options, but only options 1 and 4 are endorsed frequently. Options 2 and 3 do not provide any meaningful additional information, and the response scale functions essentially as a binary option. However, the responses cover specific and distinct areas of timeliness, making the item very discriminating, as illustrated by the steep slope of the item characteristic curve.

Figure 2C illustrates a problematic item. The response option curves do not peak rapidly nor in specific areas of timeliness, and the response options do not seem to be ordinal. The item characteristic curve is almost flat, showing little capacity for discrimination. It does not perform well to measure timeliness, which may not be surprising given that it asks about helpfulness of staff.

Conclusion

Each of the techniques described above offers a different method of examining item and subscale performance; applied together, they offer a comprehensive assessment of how the selected instruments measure performance of core primary healthcare attributes. The attribute-specific results are presented in individual papers elsewhere in this special issue.

The strength of this study was our analysis across instruments, which allowed us to identify sub-dimensions within an attribute. Sometimes a sub-dimension is unique to one subscale;

sometimes, more than one is represented. This approach will help program evaluators select the measures appropriate for their needs. Another consideration will be the missing values, and evaluators may choose not to offer "not sure" or "not applicable" options to minimize information loss. As with any study, results are sample-dependent, and items that do not function well in the present sample may still function well in a different sample of individuals or a different healthcare setting. However, the results of our study show that most of these measures can be used with confidence in the Canadian context. Ideally, any difficulties identified should be viewed as opportunities for improvement, potentially by rewriting, rewording or clarifying questions.

ACKNOWLEDGEMENTS

This research was funded by the Canadian Institutes of Health Research (CIHR).

Correspondence may be directed to: Darcy A. Santor, PhD, School of Psychology, Faculty of Social Sciences, University of Ottawa, 136 Jean-Jacques Lussier, Ottawa, ON K1N 6N5; tel.: 613-562-5799; fax: 613-562-5147; e-mail: dsantor@uottawa.ca.

REFERENCES

Borowsky, S.J., D.B. Nelson, J.C. Fortney, A.N. Hedeen, J.L. Bradley and M.K. Chapko. 2002. "VA Community-Based Outpatient Clinics: Performance Measures Based on Patient Perceptions of Care." *Medical Care* 40(7): 578-86.

Du Toit, M. 2003. *IRT from SSI: Bilog-mg, Multilog, Parscale, Testfact.* Lincolnwood, IL: Scientific Software International.

Flocke, S. 1997. "Measuring Attributes of Primary Care: Development of a New Instrument." *Journal of Family Practice* 45(1): 64-74.

Gorsuch, R.L. 1983. *Factor Analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Grol, R., M. Wensing and Task Force on Patient Evaluations of General Practice. 2000. "Patients Evaluate General/Family Practice: The EUROPEP Instrument." Nijmegen, Netherlands: Centre for Quality of Care Research, Raboud University.

Haggerty, J.L., F. Burge, M.-D. Beaulieu, R. Pineault, C. Beaulieu, J.-F. Lévesque et al. 2011. "Validation of Instruments to Evaluate Primary Healthcare from the Patient Perspective: Overview of the Method." *Healthcare Policy* 7 (Special Issue): 31–46.

Haggerty, J.L., J.-F. Lévesque, D.A. Santor, F. Burge, C. Beaulieu, F. Bouharaoui et al. 2011. "Accessibility from the Patient Perspective: Comparison of Primary Healthcare Evaluation Instruments." *Healthcare Policy* 7 (Special Issue): 94–107.

Jöreskog, K.G. and D. Sörbom. 1996. *LISREL 8: User's Reference Guide.* Chicago: Scientific Software International.

Kline, R.B. 1998. *Principles and Practice of Structural Equation Modeling.* New York: Guilford Press.

Ramsay, J.O. 1991. "Kernel Smoothing Approaches to Nonparametric Item Characteristic Curve Estimation." *Psychometrika* 56: 611–30.

Ramsay, J.O. 2000. *TESTGRAF: A Program for the Graphical Analysis of Multiple-Choice Test and Questionnaire Data (Computer Program and Manual).* Montreal: McGill University, Department of Psychology. Retrieved June 11, 2011. <ftp://ego.psych.mcgill.ca/pub/ramsay/testgraf/TestGrafDOS.dir/testgraf1.ps>.

Safran, D.G., J. Kosinski, A.R. Tarlov, W.H. Rogers, D.A. Taira, N. Lieberman and J.E. Ware. 1998. "The Primary Care Assessment Survey: Tests of Data Quality and Measurement Performance." *Medical Care* 36(5): 728–39.

Santor, D.A. and J. C. Coyne. 2001. "Examining Symptom Expression as a Function of Symptom Severity: Item

Performance on the Hamilton Rating Scale for Depression." *Psychological Assessment* 13: 127–39.

Santor, D.A. and J.O. Ramsay. 1998. "Progress in the Technology of Measurement: Applications of Item Response Models." *Psychological Assessment* 10: 345–59.

Santor, D.A., J.O. Ramsay and D.C. Zuroff. 1994. "Nonparametric Item Analyses of the Beck Depression Inventory. Examining Gender Item Bias and Response Option Weights in Clinical and Nonclinical Samples." *Psychological Assessment* 6: 255–70.

Santor, D.A., D.C. Zuroff, J.O. Ramsay, P. Cervantes and J. Palacios. 1995. "Examining Scale Discriminability in the BDI and CES-D as a Function of Depressive Severity." P*sychological Assessment* 7: 131–39.

SAS Institute. 2003. *SAS User's Guide: Statistics* (Version 9.1). Cary, NC: SAS Institute.

Shi, L., B. Starfield and J. Xu. 2001. "Validating the Adult Primary Care Assessment Tool." *Journal of Family Practice* 50(2): n161w–n171w.

Stewart, A.L., A.M. Nápoles-Springer, S.E. Gregorich and J. Santoyo-Olsson. 2007. "Interpersonal Processes of Care Survey: Patient-Reported Measures for Diverse Groups. *Health Services Research* 42(3 Pt. 1): 1235–56.

RESEARCH PAPER

Accessibility from the Patient Perspective: Comparison of Primary Healthcare Evaluation Instruments

L'accessibilité du point de vue du patient : comparaison entre instruments d'évaluation des soins de santé primaires

JEANNIE L. HAGGERTY, PHD
Department of Family Medicine, McGill University
Montreal, QC

JEAN-FRÉDÉRIC LÉVESQUE, MD, PHD
Centre de recherche du Centre hospitalier de l'Université de Montréal
Montréal, QC

DARCY A. SANTOR, PHD
School of Psychology, University of Ottawa
Ottawa, ON

FREDERICK BURGE, MD, MSC
Department of Family Medicine, Dalhousie University
Halifax, NS

CHRISTINE BEAULIEU, MSC
St. Mary's Research Centre, St. Mary's Hospital Center
Montreal, QC

FATIMA BOUHARAOUI, MSC
St. Mary's Research Centre, St. Mary's Hospital Center
Montreal, QC

Accessibility from the Patient Perspective

MARIE-DOMINIQUE BEAULIEU, MD, MSC
Chaire Dr Sadok Besrour en médecine familiale
Centre de recherche du Centre hospitalier de l'Université de Montréal
Montréal, QC

RAYNALD PINEAULT, MD, PHD
Département de Médecine sociale et préventive, Université de Montréal

DAVID GASS, MD
Department of Family Medicine, Dalhousie University
Halifax, NS

Abstract

The operational definition of first-contact accessibility is "the ease with which a person can obtain needed care (including advice and support) from the practitioner of choice within a time frame appropriate to the urgency of the problem"; accommodation is "the way healthcare resources are organized to accommodate a wide range of patients' abilities to contact healthcare providers and reach healthcare services, that is to say telephone services, flexible appointment systems, hours of operation, and walk-in periods."

Objective: To compare how well accessibility is measured in validated subscales that evaluate primary healthcare from the patient's perspective.

Method: 645 adults with at least one healthcare contact in the previous 12 months responded to six instruments that evaluate primary healthcare with four subscales that measure accessibility: the Primary Care Assessment Survey (PCAS), the Primary Care Assessment Tool – Short Form (PCAT-S, two subscales) and the first version of the EUROPEP (EUROPEP-I). Scores were normalized to a 0-to-10 scale for descriptive comparison. Exploratory and confirmatory (structural equation modelling) factor analysis examined fit to operational definition, and item response theory analysis examined item performance on common constructs.

Results: The subscales demonstrate similar psychometric measures to those reported by developers. The PCAT-S First-Contact Utilization subscale does not fit the accessibility construct. The remaining three subscales load reasonably onto a single factor, presumed to be accessibility, but the best-fitting model has two factors: "timeliness of obtaining needed care" (PCAT-S First-Contact Access, some EUROPEP-I items) and "how resources are organized to accommodate clients" (PCAS Organizational Access and most of EUROPEP-I organization of care). Items in the PCAS and PCAT-S subscales have good discriminability.

Conclusion: Only three of the four subscales measure accessibility; all are appropriate for use in Canada. The PCAT-S First-Contact Access subscale is the best measure for first-contact accessibility, and PCAS Organizational Accessibility has good metric properties and measures for accommodation.

Résumé

La définition opérationnelle de l'accessibilité de premier contact est « la facilité avec laquelle une personne peut obtenir les services requis (y compris des conseils et du soutien) du clinicien de son choix dans un délai approprié à l'urgence du problème »; l'accommodation est

« la façon dont les ressources en santé sont organisées afin de permettre à une large gamme de patients d’entrer en contact avec les fournisseurs de soins avec les services de santé, c’est-à-dire accès téléphonique, flexibilité pour la prise de rendez-vous, heures d’ouverture et périodes allouées aux visites sans rendez-vous. »

Objectif : Voir à quel point l’accessibilité est mesurée par les sous-échelles validées servant à évaluer les soins de santé primaires du point de vue du patient.

Méthode : 645 adultes ayant eu au moins un contact avec les services de santé au cours des 12 mois antérieurs ont répondu à six instruments servant à évaluer les soins de santé primaires, incluant quatre sous-échelles qui mesurent l’accessibilité : Primary Care Assessment Survey (PCAS), Primary Care Assessment Tool – version courte (PCAT-S, deux sous-echelles) et la premiere version de l’EUROPEP (EUROPEP-I). Les résultats ont été normalisés selon une échelle de 0 à 10 pour permettre des comparaisons descriptives. Les analyses factorielles exploratoires et confirmatoires (modélisation par équation structurelle) ont permis d’examiner l’adéquation à la définition opérationnelle, et l’analyse de réponse par item a permis d’en examiner la performance en fonction des construits communs.

Résultats : Les mesures psychométriques des sous-échelles sont similaires à celles indiquées par les concepteurs. La sous-échelle « utilisation de premier contact » du PCAT-S ne concorde pas avec le construit de l’accessibilité. Les trois autres sous-échelles correspondent raisonnablement à un facteur unique, qui serait probablement l’accessibilité, mais le modèle qui s’ajuste le mieux comprend deux facteurs : « rapidité d’obtention des soins nécessaires » (« accessibilité de premier contact » du PCAT-S, certains items de l’EUROPEP-I) et « comment sont organisées les ressources pour accommoder les patients » (« accès organisationnel » du PCAS et la plupart des items « organisation de soins » de l’EUROPEP-I). Les items des sous-échelles du PCAS et du PCAT-S présentent une bonne discriminabilité.

Conclusion : Seules trois des quatre sous-échelles mesurent l’accessibilité; toutes sont appropriées pour leur usage au Canada. La sous-échelle « accès de premier contact » du PCAT-S est la meilleure mesure pour l’accessibilité de premier contact et la sous-échelle « accessibilité organisationnelle » du PCAS possède de bonnes mesures et propriétés métriques pour l’accommodation.

* * *

Accessibility receives a lot of attention in research and policy debates. However, its definition and assessment pose significant challenges.

Background

Conceptualizing accessibility of healthcare services

Accessibility is a complex notion, as evidenced by the heterogeneity of definitions and conceptualizations in the literature and the almost interchangeable use of the terms “access,” “accessibility” and “utilization of healthcare services.” The *Canadian Oxford Dictionary* (1998) defines accessibility as the "condition of being readily approached." In this sense, accessibility is a characteristic of something that can readily be reached, entered or used.

Donabedian (1973) describes accessibility as characteristic of health systems that impede or promote service utilization. Thus, health services are accessible if their specific characteristics – geographic availability, organization, price, acceptability and so on – allow a broad range of persons to reach, enter and use them (Bashshur et al. 1971; Donabedian 1973; Penchansky and Thomas 1981). From this perspective, evaluation of accessibility is amenable to both objective and subjective assessment of the geographic and temporal availability of services, their organizational availability, their costs and their social and cultural acceptability (Lévesque 2006).

Evaluating the accessibility of primary healthcare services

In a consensus consultation of primary healthcare (PHC) experts across Canada to formulate operational definitions of PHC attributes to be evaluated (Haggerty et al. 2007), two distinct definitions of accessibility emerged. The first, labelled first-contact accessibility, is "The ease with which a person can obtain needed care (including advice and support) from the practitioner of choice within a time frame appropriate to the urgency of the problem." This is specific to PHC and is one of its essential functions. The second, accommodation, is applicable to all levels of healthcare: "The way healthcare resources are organized to accommodate a wide range of patients' abilities to contact healthcare providers and reach healthcare services (telephone services, flexible appointment systems, hours of operation and walk-in periods)" (Haggerty et al. 2007).

Various instruments have been developed to evaluate PHC accessibility from the user's perspective, but there is little comparative information about these to guide evaluators in their selection of tools. Our objective was to provide insight into how well validated subscales from different instruments measure accessibility. Specifically, we wanted to know how subscale scores from different instruments compare and whether the accessibility subscales measure a single construct, presumed to be accessibility. If analyses suggested more than one factor, we wanted to judge how those factors reflected our operational definitions of accessibility. Finally, we sought to examine how well individual items measured the dimensions that were common across instruments.

Method

The method of this series of studies has been described in detail elsewhere (Haggerty, Burge et al. 2011). Briefly, six validated instruments that evaluate PHC from the patient's perspective were administered to 645 healthcare users balanced approximately by English/French language, rural/urban location, high/low education, and poor, average or excellent overall PHC experience.

The analytic strategy, described in detail elsewhere (Santor et al. 2011) consisted of examining the distributional statistics and subscale correlations, followed by common factor analysis and confirmatory factor analysis (structural equation modelling) to identify dimensions common to the entire set of items. Finally, we examined the performance of individual items and response scales using parametric item response theory analysis against the original subscale and non-parametric analysis against the common factors that emerged across instruments.

Measure selection

Three of the instruments had accessibility subscales. The Primary Care Assessment Survey (PCAS) (Safran et al. 1998) six-item Organizational Accessibility subscale elicits ratings for obtaining services at the clinic using a six-point Likert response scale (1=very poor to 6=excellent). A parallel set of questions report average wait times, but these are not included in the validated subscale. The Primary Care Assessment Tool – Short Form (PCAT-S) (Shi et al. 2001) has two subscales: the three-item First-Contact Utilization (eliciting usual patterns of care-seeking), and the four-item First-Contact Access (eliciting the probability of being seen at the clinic under various scenarios with a four-point Likert scale from "1=definitely not" to "4=definitely"). However, during the attribute-mapping process, the PCAT-S First-Contact Utilization subscale was flagged as fitting poorly with the concept of accessibility, despite its label. The first version of the European general practice instrument, EUROPEP-I (Grol et al. 2000) has a seven-item Organization of Care subscale in which most items elicit ratings of accessibility using a five-point semantic differential response scale with "1=poor" and "5=excellent" as anchors of each extreme.

Results

Comparative descriptive statistics

The content of the four subscales and item distributors are summarized in Table 1; the detailed content and distributions are available online at http://www.longwoods.com/content/22635. Only two subscales have ≥5% true missing values (PCAS on being able to talk by phone to the doctor, and EUROPEP-I on the helpfulness of staff). However, both PCAT-S subscales and the EUROPEP-I offer respondents a "don't know/not sure" or "not applicable" option, which is used by a significant proportion of respondents in PCAT-S First-Contact Access items and four EUROPEP-I items. Approximately 17% were "not sure" about arrangements for getting care or advice outside office hours on the PCAT-S, or being able to contact the doctor by phone (13%) on the EUROPEP-I.

TABLE 1. Summary of accessibility subscale content and distribution of item responses. (Detailed distribution available at http://www.longwoods.com/content/22635).

Subscale and Item Description	Response Scale	Range Missing Values	Overall Modal Response	Range Item Discriminability	Comments on Distribution
PCAS Organizational Access (6 items) Rate doctor's office for: Location, hours, usual wait for an appointment; usual wait at the clinic, ability to get through to the doctor's office or to speak to the doctor by phone	Likert evaluative, 1=very poor to 6=excellent	1%–6%	4 (good)	0.83 (office location) to 2.4 (speak to doctor by phone)	Approximately 60% of responses in two most positive categories; lowest rating for "speak to doctor by phone"

TABLE 1. Continued

Subscale and Item Description	Response Scale	Range Missing Values	Overall Modal Response	Range Item Discriminability	Comments on Distribution
PCAT–S First-Contact Utilization (3 items) Likelihood of seeking care for routine or a new problem and the need for referral to see a specialist	Likert evaluative, 1=definitely not to 4=definitely	1%–2% (true missing) 0%–4% not sure	4 (definitely)	0.87 (referral) to 4.7 (first place to seek routine care)	Over 65% of responses in the most positive category
PCAT–S First-Contact Access (4 items) When clinic open: Likelihood of being seen same day; getting advice over the phone When clinic closed: Having a phone number to call; likelihood of being seen by doctor during the night	Likert evaluative, 1=definitely not to 4=definitely	2% (true missing) 6%–18% not sure	variable, according to the item	1.06 (open) to 2.99 (phone number when closed)	Item score in decreasing order; 16%–18% "not sure" when clinic closed; 40% can definitely not be seen at night
EUROPEP-I Organization of Care (7 items) Rate care by GP: Preparation for what to expect from specialty care; helpfulness of staff; getting a suitable appointment; getting through to the practice over the phone; speaking to GP on the phone; waiting time in waiting room; getting quick services for urgent problems	Semantic differential rating, 1=poor, 5=excellent	4%–5% (true missing) 1%–13% (n/a)	5 (excellent)	1.55 (preparation for other care) to 2.86 (quick services)	Lowest rating for "speak to GP on the phone." Three items with "not applicable": preparation for specialty care (12%), speak to GP on the phone (13%) and quick services (11%)

Only the PCAT-S First-Contact Utilization showed strong ceiling effects, with the majority endorsing the most positive response. The parametric estimate of discriminability within the original subscale generally showed good capacity to discriminate between different levels of the subscale score. Exceptions were the convenience of location on PCAS Organization Accessibility and need for approved referrals to specialists on PCAT-S First-Contact Utilization.

Table 2 presents the subscale scores as item means normalized to a common 0-to-10 metric to permit comparison. The normalized means and standard deviations differ substantially from one scale to another. All except the PCAT-S First-Contact Access subscale are quite skewed towards positive values, with the median higher than the mean. The PCAT-S subscales had relatively low reliability.

Table 3 presents the Pearson correlations between the accessibility subscales. With the exception of the PCAT-S First-Contact Utilization subscale, the accessibility subscales correlate strongly with one another (.45 to .68), suggesting a common construct. In contrast, the PCAT-S First-Contact Utilization subscale correlates only modestly (.24 to .29) and correlates most highly with relational continuity subscales (.28 to .37). The EUROPEP-I Organization of Care subscale correlates as strongly with subscales of other dimensions as it does with accessibility, suggesting that it is measuring an overall experience of care rather than accessibility specifically. In contrast, the PCAS Organizational Access subscale and the PCAT-S First-Contact Access subscale have much lower correlations with other attributes of care, suggesting they are more specific for accessibility.

TABLE 2. Mean and distributional scores for accessibility subscale scores normalized to 0 to 10 (n=645)*

Developer's Scale Name (# of items in scale)	Cronbach's Alpha	Mean	SD	Minimum Observed	Quartiles		
					Q1 (25%)	Q2 (50%)	Q3 (75%)
Normalized scores							
PCAS Organizational Access (6)	.83	5.9	1.8	1.3	4.7	6.0	7.3
PCAT-S First-Contact Utilization (3)	.59	8.5	1.5	0.0	7.5	9.2	9.2
PCAT-S First-Contact Access (4)	.68	5.2	3.0	0.0	3.3	5.6	6.7
EUROPEP-I Organization of Care (7)	.89	6.5	2.4	0.0	5.0	6.8	8.6

* Subscale scores calculated as mean of item values and calculated only for observations where >50% of items were complete.

TABLE 3. Mean partial correlations between accessibility.* Only correlations significantly different from zero are provided.

Questionnaire Subscale	Organizational Access PCAS	First-Contact Utilization PCAT-S	First-Contact Access PCAT-S	EUROPEP-I
PCAS: Organizational Access	1.00	0.29	0.45	0.68
PCAT-S: First-Contact Utilization	0.29	1.00	0.24	0.29
PCAT-S: First-Contact Access	0.45	0.24	1.00	0.46
EUROPEP-I: Organization of Care	0.68	0.29	0.46	1.00

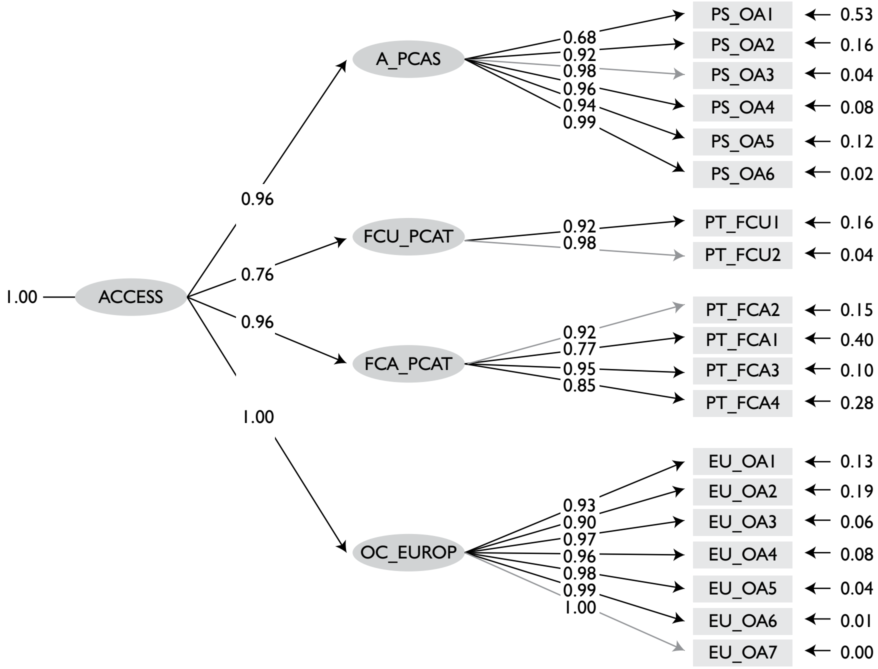
* Controlling for language, education, achievement, geographic location.

Do all items measure a single attribute?

We excluded from factor analysis all subjects who had at least one missing value on any item (listwise missing). This dramatically reduced our effective sample size from 645 to 340. Most exclusions (267/340) were for selecting the "don't know/not sure" or "not applicable" response options. Those excluded from the factor analyses were more likely than those included to be English-speaking, to report better health status and to have lower usual wait times for an appointment; they did not differ by their overall experience of care and other individual characteristics. Because this conservative approach can introduce bias, we repeated all the confirmatory factor analyses using maximum likelihood imputation of missing values (Jöreskog and Sörbom 1996) to examine the robustness of our conclusions. Running the models on a sample of 559 improved the fit statistics of confirmatory factor analysis models but did not alter any of the general conclusions.

We had little expectation that all the items would load on a single construct because we had defined two dimensions of accessibility and the PCAT-S First Utilization subscales seem to fit poorly. Nonetheless, constraining common factor analysis to one factor found loadings >.30 for all but two items. The same model with structural equation modelling generated statistics suggesting only moderate fit, with a root mean squared error of approximation (RMSEA) of p=.11, considerably higher than the .05 standard for good fit. Fit statistics improved dramatically when we removed the PCAT-S First-Contact Utilization fit, with the RMSEA=.09, a normed fit index (NFI) of .98 (well above the .90 standard) and considerably lower chi-squared and model conditional Akaike information criteria (CAIC) values (better fit with lower values). Figure 1 presents a second-order model in which items are grouped within their original subscale as one construct (first-order latent variable), which in turn emerges from a single underlying construct, presumed to be accessibility (second-order latent variable). Compared to the one-factor model, grouping items by parent subscale improves fit and shows that PCAT-S First Utilization loads much less strongly on the construct presumed to be accessibility.

FIGURE 1. Parameter estimations for a structural equation second-order model where a single underlying construct (second-order latent variable) leads to the four subscales (first-order variables) with loadings on their respective items



Chi-square = 514.03, $df = 148$, p-value = 0.00000, RMSEA = 0.090

Based on this and our a priori concerns, we decided that the PCAT-S First-Contact Utilization subscale, despite its label, does not fit the construct of accessibility. Subsequent analyses indicated a best fit with comprehensiveness of care (Haggerty, Beaulieu et al. 2011). It was removed from further analyses on accessibility.

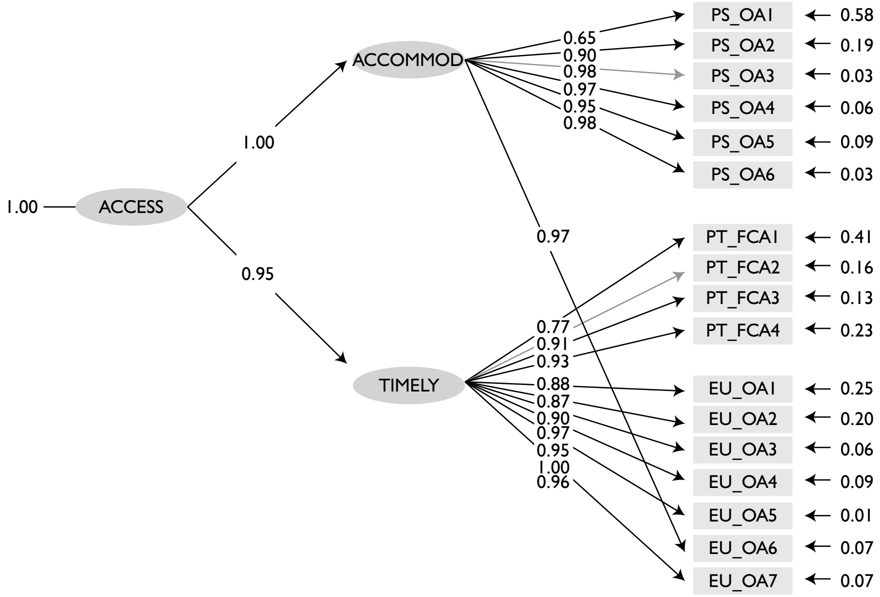
How do underlying factors fit with operational definitions?

Exploratory factor analysis suggested a two-factor model. Using our operational definitions as a guide, we judged that the first factor (eigenvalue=7.59) captured "obtaining needed care

… in a time frame appropriate to the urgency of the problem," or timeliness, and the second (eigenvalue=1.19) touched on "how resources are organized to accommodate clients," or accommodation.

Using the operational definitions and the exploratory analysis as a guide, we grouped items on timeliness and accommodation for confirmatory factor analysis, shown in Figure 2. Goodness of fit improved slightly with respect to the unidimensional model (RMSEA=.084, NFI=.98). The correlation between the dimensions of timeliness and accommodation is .95. We tried other item groupings based on our judgment, but all other configurations resulted in poorer fit statistics. The model shows that the PCAS Organizational Access subscale relates to the sub-dimension of accommodation, whereas the PCAT-S First-Contact Access subscale relates to timeliness. The EUROPEP-I Organization of Care subscale mostly measures timeliness, but one item, wait time in the waiting room, loads highly on accommodation. Figure 2 shows that some items do not have high loadings and have a high proportion of residual error (shown to the right of each item). These items relate poorly to the construct, either because they are not discriminatory or because they relate better to another construct that is not part of the latent variable.

FIGURE 2. Parameter estimations for a structural equation showing item loadings on items from different subscales on two sub-dimensions of accessibility (first-order variables), timeliness and accommodation



Chi-square = 373.90, *df* = 118, *p*-value = 0.00000, RMSEA = 0.084

A number of items were identified a priori as potentially poor indicators of accessibility. For instance, the EUROPEP-I Organization of Care subscale item, "How would you rate your general practitioner's care in preparing you for what to expect from specialist or hospital care?" does not appear to measure accessibility. However, this item correlates well with other items in the subscale (item–total correlation=.55), and removing it does not improve the model's goodness of fit. Since we had decided to respect as much as possible the developers' factor solutions, we retained this item as part of the subscale.

Individual item performance

Non-parametric item response theory graphs were modelled on dimensions of timeliness and accommodation and provide further insight into item performance. For accommodation, we found that all but one item (PCAS convenient location) in the PCAS Organizational Access subscale demonstrate excellent performance. The probability of endorsing each response option is highest in a relatively narrow and unique zone of accommodation, and is clearly ordinal, reflecting the assigned value for each option. The item score varies linearly and strongly with accommodation, demonstrating excellent discriminability. The single EUROPEP-I item that measures accommodation (EU_oa6, wait in waiting room), shows good discriminability, but the extreme response options ("1=poor," "5=excellent") are overwhelmingly endorsed, suggesting that the assigned values for middle options may not be appropriate.

For timeliness, the first two items in the PCAT-S perform relatively well (probability of being seen same day when provider is open, and getting advice over the phone). The value assigned to each response option is mostly appropriate, except the option "probably," which appears to be non-specific, as it is endorsed across the entire range of timeliness. The item score correlates highly with overall timeliness, indicating good discriminability. In contrast, the two items addressing more rare scenarios (when provider is closed) have modest discriminability, and the response options and assigned values perform less well. Only the options "1=definitely not" and "4=definitely" are endorsed with any frequency, raising the question of the appropriateness of a four-point response scale. The probability of choosing the "don't know" response option (over 15% of respondents) mimics the probability of endorsing options "2=probably not" or "3=probably," depending on the item.

All but one EUROPEP-I item performed poorly on timeliness. Only the item "getting quick services for urgent problems" demonstrated good discriminability despite problems with middle response options. The remaining items demonstrate poor discriminability, and the behaviour of the middle response options indicate that differential weights may not be warranted. Based on the content of some items (preparation to see specialist, helpfulness of staff, phone contact with the clinic, and phone contact with the general practitioner), we tested them on accommodation rather than timeliness, but item performance did not improve.

Discussion

Capacity to measure accessibility

Of the four subscales used in the concurrent validation study, three seem to evaluate clearly the attribute of accessibility, whereas PCAT-S First-Contact Utilization appears to be measuring comprehensiveness of care or possibly relational continuity (Burge, Haggerty et al. 2011). This finding implies that evaluators interested in evaluating accessibility could use any one of these three subscales. However, the PCAT-S First-Contact Access subscale measures timeliness, and the PCAS Organizational Access subscale principally measures accommodation. The EUROPEP-I Organization of Care subscale appears to capture both dimensions, but despite high internal consistency, item performance on the constructs of timeliness or accommodation are problematic. The high correlation between the EUROPEP-I Accessibility subscale and all the other scales in the evaluation study further suggests it measures a generic experience of care rather than accessibility specifically. This is not surprising, because the scale developers did not specifically intend to evaluate accessibility, but rather organization of care based on patient priorities (Grol et al. 1999). It has adequate discriminability for its own construct. Recent revisions to the instrument are minor, but may improve metric properties (Wensing 2006).

The PCAT-S First-Contact Access is the best subscale for measuring first-contact accessibility as the dimension of accessibility that is specific to PHC. The item's content well reflects policy initiatives to enhance 24/7 access to PHC, and the items are ordered by difficulty, suggesting that a summary score is informative and meaningful. However, despite good discriminability, its performance could be improved by changing the response options. First, the option "probably" is non-specific, shown both on item response analysis and discussions with respondents (Haggerty, Lawson et al. 2011). Most importantly, the "don't know" option should not be offered as it creates missing values on analysis. The item response theory analysis suggests that the developers' recommendation of imputing "not sure" as "probably not" is not appropriate. The two items referring to more rare access experiences – when the office is closed or at night – would function better with a yes/no response format but can be used to identify those with excellent access.

The PCAS Organizational Accessibility subscale demonstrates excellent performance for measuring accommodation. The good discriminability and reliability might improve further if the item on convenience of location were removed. Since accommodation is not specific to PHC, it is likely equally relevant to general healthcare accessibility. Two questionnaires that were largely derived from the PCAS (General Practice Assessment Questionnaire – GPAQ and General Practice Assessment Survey – GPAS) include a question on timely care: "If you need to see a doctor urgently, can you normally get seen on the same day?" and may be a good option for including an indicator of first-contact accessibility as well as accommodation.

Aspects not covered in the studied instruments

The subscales we studied did not address the element of "obtaining care from the provider of choice" in our operational definition, though it could be argued this is captured by the

PCAT-S First-Contact Utilization subscale that we excluded. Although studies show that patients prioritize timeliness over affiliation when faced with an acute illness (Love and Mainous 1999), on the whole, patients strongly prefer to consult their own physician for most care, as is shown on measures of relational continuity elsewhere in this special issue (Burge et al. 2011). The provider of choice – and therefore the PCAT-S First-Contact Utilization subscale – may be the link between accessibility and relational continuity, or may be particularly relevant to accessibility for non-urgent problems.

None of the identified questionnaires assess geographic accessibility, which is often the first aspect of accessibility considered (Frenk 1992). The PCAS item on "the convenience of location" is problematic owing to its skewed distribution and poor discriminability. Although long distances from care may affect a minority of Canadians, it would be important to develop a sensitive measure of geographic accessibility, given that Canadian geography makes it a major consideration in healthcare planning. We developed a measure of ease of reaching care and local availability of services that may meet this need (Haggerty, Lévesque et al. 2009).

We did not assess the subscales that evaluate economic accessibility, as these were not considered relevant for Canada because they refer to direct costs of services. However, indirect costs, such as transportation costs and pay lost when receiving medical care during working hours, can result in forgone care. We have developed a measure of economic accessibility that addresses these issues and performs very well in the Quebec context (Haggerty, Lévesque et al. 2009).

Study limits

Limiting the study to those having visited a regular provider in the previous 12 months may have selected subjects with good accessibility. However, including users and non-users in a study sample should increase the variance and improve, not compromise, measure performance. Eliminating subjects with missing values not only reduced statistical power, but may have biased the final sample, even though our sensitivity analysis using imputation of missing values did not alter our overall conclusions. The item response analysis shows that missing values tend to occur among respondents with more negative experiences of accessibility. This finding would underestimate the reported reliability and attenuate the factor analysis results, but is not expected to radically change overall differences among instruments. Finally, this study did not have the benefit of an objective assessment of aspects of accessibility at the usual source of care that could have enabled us to assess the correlation of different scales with actual measures of availability and barriers to care.

Conclusion

Despite the study's limitations, the results indicate that the PCAS Organizational Access subscale is an excellent choice for measuring accommodation. The PCAT-S First-Contact Access subscale is the best for timely first-contact care and is, therefore, highly relevant for accessibility reforms in Canada, but it could be psychometrically improved. The EUROPEP-I Accessibility subscale probably assesses a more general experience of care, including accessibility dimensions.

ACKNOWLEDGEMENTS

This research was funded by the Canadian Institutes for Health Research. During this study Jeannie L. Haggerty held a Canada Research Chair in Population Impacts of Healthcare at the Université de Sherbrooke. The authors wish to thank Beverley Lawson for conducting the survey in Nova Scotia and Christine Beaulieu in Quebec and Donna Riley for support in preparation and editing of the manuscript.

Correspondence may be directed to: Jeannie L. Haggerty, Associate Professor, Department of Family Medicine, McGill University, St. Mary's Research Centre, Hayes Pavilion – Suite 3734, 3830 Lacombe Ave., Montreal QC H3T 1M5; tel.: 514-345-3511 ext. 6332; fax: 514-734-2652; e-mail: jeannie.haggerty@mcgill.ca.

REFERENCES

Bashshur, R.L., G.W. Shannon and C.A. Metzner. 1971. "Some Ecological Differentials in the Use of Medical Services." *Health Services Research* 6: 61–75.

Burge, F., J.L. Haggerty, R. Pineault, M.-D. Beaulieu, J.-F. Lévesque, C. Beaulieu and D.A. Santor. 2011. "Relational Continuity from the Patient Perspective: Comparison of Primary Healthcare Evaluation Instruments." *Healthcare Policy* 7 (Special Issue): 124–38.

Canadian Oxford Dictionary. 1998. Toronto: Oxford University Press.

Donabedian, A. 1973. "Capacity to Produce Services in Relation to Need and Demand." In *Aspects of Medical Care Administration: Specifying Requirements for Health Care*. Cambridge, MA: Harvard University Press.

Frenk, J. 1992. "The Concept and Measurement of Accessibility." In J. Frenk, C. Ordonez, J.M. Paganini and B. Starfield, eds., *Health Services Research: An Anthology*. Washington, DC: Pan American Health Organization.

Grol, R., M. Wensing, J. Mainz, P. Ferreira, H. Hearnshaw, P. Hjortdahl et al. 1999. "Patients' Priorities with Respect to General Practice Care: An International Comparison. European Task Force on Patient Evaluations of General Practice (EUROPEP-I)." *Family Practice* 16(1): 4–11.

Grol, R., M. Wensing and Task Force on Patient Evaluations of General Practice. 2000. "Patients Evaluate General/Family Practice: The EUROPEP-I Instrument." Nijmegen, Netherlands: Centre for Quality of Care Research, Raboud University.

Haggerty, J.L. 2011. "Measurement of Primary Healthcare Attributes from the Patient Perspective." *Healthcare Policy* 7 (Special Issue): 13–20.

Haggerty, J.L., C. Beaulieu, B. Lawson, D.A. Santor, M. Fournier and F. Burge. 2011. "What Patients Tell Us about Primary Healthcare Evaluation Instruments: Response Formats, Bad Questions and Missing Pieces." *Healthcare Policy* 7 (Special Issue): 66–78.

Haggerty, J.L., M.-D. Beaulieu, R. Pineault, F. Burge, J.-F. Lévesque, D.A. Santor, F. Bouharoui and C. Beaulieu. 2011. "Comprehensiveness of Care from the Patient Perspective: Comparison of Primary Healthcare Evaluation Instruments." *Healthcare Policy* 7 (Special Issue): 154–66.

Haggerty, J., F. Burge, J.-F. Lévesque, D. Gass, R. Pineault, M.-D. Beaulieu et al. 2007. "Operational Definitions of Attributes of Primary Healthcare: Consensus Among Canadian Experts." *Annals of Family Medicine* 5: 336–44.

Haggerty, J.L., J.-F. Lévesque, D. Roberge and J. Gauthier. 2009. *Assessing Access to Primary Care: Understanding the Context and Adapting Measurement Instruments to the Specificity of Rural, Remote or Isolated Areas*. Unpublished report. Available from J. Haggerty, Department of Family Medicine, McGill University, Montreal, QC.

Jöreskog, K.G. and D. Sörbom. 1996. *LISREL 8: User's Reference Guide*. Chicago: Scientific Software International, Inc.

Lévesque, J.-F. 2006. *Inequalities in Access to Health Care in Urban South India*. Doctoral thesis dissertation. Montreal, QC: Université de Montréal.

Love, M.M. and A.G. Mainous. 1999. "Commitment to a Regular Physician: How Long Will Patients Wait to See Their Own Physician for Acute Illness?" *Journal of Family Practice* 48: 202–07.

Penchansky, R. and J.W. Thomas. 1981. "The Concept of Access: Definition and Relationship to Consumer Satisfaction." *Medical Care* 19: 127–40.

Safran, D.G., M. Kosinski, A.R. Tarlov, W.H. Rogers, D.H. Taira, N. Lieberman et al. 1998. "The Primary Care Assessment Survey: Tests of Data Quality and Measurement Performance." *Medical Care* 36: 728–39.

Santor, D.A., J. Haggerty, J.-F. Lévesque, F. Burge, M.-D. Beaulieu, D. Gass and R. Pineault. 2011. "An Overview of Confirmatory Factor Analysis and Item Response Analysis Applied to Instruments to Evaluate Primary Healthcare." *Healthcare Policy* 7 (Special Issue): 79–92.

Shi, L., B. Starfield and J. Xu. 2001. "Validating the Adult Primary Care Assessment Tool." *Journal of Family Practice* 50(2): 161–71.

Wensing, M. 2006. *EUROPEP 2006: Revised Europep Instrument and User Manual.* Nijmegen, Netherlands: Centre for Quality of Care Research, Raboud University. Retrieved July 23, 2011. <http://www.topaseurope.eu/?q=node/13>.

Accessibility from the Patient Perspective: Comparison of Primary Healthcare Evaluation Instruments

L'accessibilité du point de vue du patient : comparaison entre instruments d'évaluation des soins de santé primaires

JEANNIE L. HAGGERTY, JEAN-FRÉDÉRIC LÉVESQUE, DARCY A. SANTOR, FREDERICK BURGE, CHRISTINE BEAULIEU, FATIMA BOUHARAOUI, MARIE-DOMINIQUE BEAULIEU, RAYNALD PINEAULT AND DAVID GASS

TABLE 1. Distribution of responses to items in subscales measuring accessibility of primary healthcare services and discriminatory capacity of each item within its parent subscale. Modal response is shown in bold (n=645).

Variable Name	Instrument: Subscale	Missing Values % (n)	Per Cent (Number) by Response Option						Item Discrimination
	Primary Care Assessment Survey (PCAS): Organizational Access		**1=Very poor**	**2=Poor**	**3=Fair**	**4=Good**	**5=Very good**	**6=Excellent**	
PS_oa1	How would you rate the convenience of your regular doctor's office **location**?	1 (8)	0 (1)	1 (9)	14 (93)	27 (172)	**31** (202)	25 (160)	0.83 (0.17)
PS_oa2	How would you rate the **hours** that your doctor's office is open for medical appointments?	1 (5)	1 (6)	3 (17)	17 (111)	**38** (245)	28 (181)	12 (80)	1.84 (0.13)
PS_oa3	How would you rate the usual **wait** for an appointment when you are sick and call the doctor's office asking to be seen?	3 (22)	5 (35)	15 (95)	21 (134)	**26** (168)	19 (121)	11 (70)	2.51 (0.19)
PS_oa4	How would you rate the amount of time you **wait** at your doctor's office for your appointment to start?	2 (10)	5 (34)	12 (80)	27 (177)	**29** (190)	16 (106)	7 (48)	1.81 (0.13)
PS_oa5	Thinking about the times you have needed to **see or talk to** your doctor, how would you rate the following: ability to **get through to** the doctor's office by phone?	1 (6)	4 (27)	5 (35)	17 (110)	**31** (200)	27 (172)	15 (95)	2.10 (0.15)
PS_oa6	Thinking about the times you have needed to **see or talk to** your doctor, how would you rate the following: ability to **speak to** your doctor by phone when you have a question or need medical advice?	6 (38)	12 (79)	16 (102)	20 (132)	**25** (161)	14 (91)	7 (42)	2.40 (0.16)
	Primary Care Assessment Tool (PCAT-S): First-Contact Utilization		**1=Definitely not**	**2=Probably not**	**3= Probably**	**4= Definitely**	**Not sure / Don't remember**		
PT_fcu1	When you need a regular general check-up, do you go to your Primary Care Provider before going somewhere else?	1 (7)	2 (12)	1 (6)	10 (64)	**86** (554)	0 (2)		4.70 (0.60)
PT_fcu2	When you have a new health problem, do you go to your Primary Care Provider before going somewhere else?	1 (8)	2 (13)	2 (15)	12 (79)	**82** (528)	0 (2)		4.59 (0.54)
PT_fcu3	When you have to see a specialist, does your Primary Care Provider have to approve or give you a referral?	2 (10)	2 (16)	4 (27)	23 (151)	**65** (418)	4 (23)		0.87 (0.13)
	Primary Care Assessment Tool (PCAT-S): First-Contact Access		**1=Definitely not**	**2=Probably not**	**3= Probably**	**4= Definitely**	**Not sure / Don't remember**		
PT_fca1	When your Primary Care Provider is open and you get sick, would someone from there see you the same day?	2 (11)	6 (41)	16 (103)	43 (278)	**27 (175)**	6 (37)		1.06 (0.12)
PT_fca2	When your Primary Care Provider is open, can you get advice quickly over the phone if you need it?	2 (11)	11 (72)	17 (110)	35 (225)	**25 (161)**	10 (66)		1.06 (0.12)

TABLE 1. Continued

Variable Name	Instrument: Subscale	Missing Values % (n)	Per Cent (Number) by Response Option						Item Discrimination
PT_fca3	When your Primary Care Provider is closed, is there a phone number you can call when you get sick?	2 (13)	20 (129)	10 (63)	14 (90)	**39 (250)**	16 (100)		2.99 (0.27)
PT_fca4	When your Primary Care Provider is closed and you get sick during the night, would someone from there see you that night?	2 (12)	40 (260)	22 (140)	9 (59)	**9 (58)**	18 (116)		2.58 (0.22)
	EUROPEP-I: Organization of Care How would you rate the following care provided by your general practitioner in the last 12 months?		**1=Poor**	**2**	**3**	**4**	**5= Excellent**	**NA (does not apply)**	
EU_oa1	Preparing you for what to expect from specialist or hospital care	4 (24)	3 (20)	7 (42)	13 (87)	29 (185)	**32** (209)	12 (78)	1.55 (0.14)
EU_oa2	The helpfulness of staff (other than the doctor)	5 (29)	3 (18)	6 (39)	15 (98)	30 (192)	**37** (236)	5 (33)	1.84 (0.15)
EU_oa3	Getting an appointment to suit you	4 (24)	9 (57)	10 (67)	18 (113)	25 (158)	**34** (220)	1 (6)	2.83 (0.18)
EU_oa4	Getting through to the practice on the phone	4 (23)	5 (35)	8 (53)	19 (125)	28 (180)	**34** (217)	2 (12)	2.16 (0.15)
EU_oa5	Being able to speak to the general practitioner on the telephone	4 (27)	**19** (124)	15 (99)	**17** (107)	**18** (115)	**14** (89)	13 (84)	2.48 (0.19)
EU_oa6	Waiting time in the waiting room	4 (25)	14 (91)	12 (75)	24 (156)	**30** (192)	15 (99)	1 (7)	1.79 (0.13)
EU_oa7	Providing quick services for urgent health problems	4 (25)	8 (49)	8 (54)	17 (108)	25 (158)	**28** (179)	11 (72)	2.86 (0.21)

RESEARCH PAPER

Interpersonal Communication from the Patient Perspective: Comparison of Primary Healthcare Evaluation Instruments

La communication interpersonnelle du point de vue du patient : comparaison entre instruments d'évaluation des soins de santé primaires

MARIE-DOMINIQUE BEAULIEU, MD, MSC
Chaire Dr Sadok Besrour en médecine familiale
Centre de recherche du Centre hospitalier de l'Université de Montréal,
Montréal, QC

JEANNIE L. HAGGERTY, PHD
Department of Family Medicine, McGill University
Montreal, QC

CHRISTINE BEAULIEU, MSC
St. Mary's Research Centre, St. Mary's Hospital Center
Montreal, QC

FATIMA BOUHARAOUI, MSC
St. Mary's Research Centre, St. Mary's Hospital Center
Montreal, QC

JEAN-FRÉDÉRIC LÉVESQUE, MD, PHD
Centre de recherche du Centre hospitalier de l'Université de Montréal
Montréal, QC

RAYNALD PINEAULT, MD, PHD
Centre de recherche du Centre hospitalier de l'Université de Montréal
Montréal, QC

Interpersonal Communication from the Patient Perspective

FREDERICK BURGE, MD, MSC
Department of Family Medicine, Dalhousie University
Halifax, NS

DARCY A. SANTOR, PHD
School of Psychology, University of Ottawa
Ottawa, ON

Abstract

The operational definition of interpersonal communication is "the ability of the provider to elicit and understand patient concerns, to explain healthcare issues and to engage in shared decision-making if desired."

Objective: To examine how well interpersonal communication is captured in validated instruments that evaluate primary healthcare from the patient's perspective.

Method: 645 adults with at least one healthcare contact in the previous 12 months responded to instruments that evaluate primary healthcare. Eight subscales measure interpersonal communication: the Primary Care Assessment Survey (PCAS, two subscales); the Components of Primary Care Index (CPCI, one subscale); the first version of the EUROPEP (EUROPEP-I); and the Interpersonal Processes of Care Survey, version II (IPC-II, four subscales). Scores were normalized for descriptive comparison. Exploratory and confirmatory (structural equation) factor analysis examined fit to operational definition, and item response theory analysis examined item performance.

Results: Items not pertaining to interpersonal communication were removed from the EUROPEP-I. Most subscales are skewed positively. Normalized mean scores are similar across subscales except for IPC-II Patient-Centred Decision-Making and IPC-II Hurried Communication. All subscales load reasonably well on a single factor, presumed to be interpersonal communication. The best model has three underlying factors corresponding to eliciting (eigenvalue = 26.56), explaining (eigenvalue = 2.45) and decision-making (eigenvalue = 1.34). Both the PCAS Communication and the EUROPEP-I Clinical Behaviour subscales capture all three dimensions. Individual subscales within IPC-II measure each sub-dimension.

Conclusion: The operational definition is well reflected in the available measures, although shared decision-making is poorly represented. These subscales can be used with confidence in the Canadian context to measure this crucial aspect of patient-centred care.

Résumé

La définition opérationnelle de la communication interpersonnelle est la « capacité du clinicien à obtenir et à comprendre les inquiétudes du patient, à expliquer les problèmes concernant les soins de santé et à procéder à des prises de décisions communes s'il y a lieu. »

Objectif : Examiner à quel point les instruments validés, qui servent à évaluer les soins primaires du point de vue du patient, font état de la communication interpersonnelle.

Méthode : 645 adultes, ayant vécu au moins un contact avec les services de santé au cours des 12 mois antérieurs, ont répondu à des instruments utilisés pour l'évaluation des soins de santé primaires. Huit sous-échelles ont servi à mesurer la communication interpersonnelle :

Primary Care Assessment Survey (PCAS, deux sous-échelles); Components of Primary Care Index (CPCI, une sous-échelle); la première version de l'EUROPEP (EUROPEP-I); et Interpersonal Processes of Care Survey, version II (IPC-II, quatre sous-échelles). Les indices ont été normalisés pour procéder à des comparaisons descriptives. Les analyses factorielles exploratoires et confirmatoires (équation structurelle) ont permis d'examiner leur adéquation à la définition opérationnelle, et l'analyse de la réponse par item a permis d'en examiner le rendement.
Résultats : Les items qui n'ont pas de lien avec la communication interpersonnelle ont été retirés de l'EUROPEP-I. La plupart des sous-échelles présentent une asymétrie positive. Les indices moyens normalisés sont similaires parmi les sous-échelles, sauf pour les sous-échelles « décisions axées sur le patient » et « communication hâtive » de l'IPC-II. Toutes les sous-échelles présentent un point de saturation raisonnablement acceptable pour un facteur unique, qui est probablement la communication interpersonnelle. Le meilleur modèle comporte trois facteurs sous-jacents qui correspondent à l'« obtention » (valeur propre=26,56), à l'« explication » (valeur propre=2,45) et à la « prise de décision » (valeur propre=1,34). La sous-échelle « communication » du PCAS et la sous-échelle « attitude clinique » de l'EUROPEP-I captent toutes deux les trois dimensions. Les sous-échelles individuelles au sein de l'IPC-II mesurent chacune des sous-dimensions.
Conclusion : La définition opérationnelle est bien représentée par les mesures disponibles, quoique l'aspect ayant trait aux décisions communes y soit faiblement représenté. Ces sous-échelles peuvent être utilisées en toute confiance dans le contexte canadien pour mesurer cet aspect essentiel des soins axés sur le patient.

* * *

Effective communication between doctor and patient is a core clinical skill. In interviews with patients, doctors elicit diagnostic information and provide therapeutic advice. Effective doctor–patient communication is associated with positive health outcomes (Mead and Bower 2002; Stewart 1995) such as symptom resolution (Headache Study Group 1986; Starfield et al. 1981), improved general health (Safran, Taira et al. 1998) and physiologic measures (Stewart 1995), as well as with greater consumer satisfaction in different organizational models of primary healthcare services (Safran et al. 1994, 2000, 2002). Studies in many countries have shown that serious communication problems are common in clinical practice (Simpson et al. 1991). Indeed, most complaints by the public about medical services are not about competencies, but about communication. The majority of malpractice allegations arise from communication errors (Simpson et al. 1991).

Improving interpersonal communication between providers and patients is an important health policy issue (Simpson et al. 1991). Its importance has recently been underscored with the adoption of person-centred medicine and the "medical home" as key concepts of primary care (Stange and Acheson 2006). However, there are concerns that interprofessional teamwork, a core feature of new primary care models, may change the nature of one-on-one communication experienced by the patient (Safran 2003; Rodriguez et al. 2007). It is thus important to be able to assess interpersonal communication reliably and validly as part of monitoring the impact of new models of primary care.

Although interpersonal communication is closely related to relational continuity, trust and patient-centred care, it is considered a distinct attribute. Communication skills can be observed. Effective communication can be experienced even in a first encounter with a provider who may not be seen again. It precedes and leads to relational continuity and fosters patient-centred practice (Brown et al. 2001; Thom and Campbell 1997).

Evaluating interpersonal communication

In 2004, we conducted a consensus consultation of 20 primary healthcare (PHC) experts across Canada to formulate operational definitions of the attributes of care that should be measured in primary healthcare models (Haggerty et al. 2007). Good interpersonal communication was identified unanimously by the experts as an essential attribute, even though it is not specific to primary healthcare. The operational definition on which they agreed was "the ability of the provider to elicit and understand patient concerns, to explain healthcare issues, and to engage in shared decision-making if desired." The experts also agreed unanimously that this attribute is most validly evaluated from the patient's perspective.

Several instruments measure interpersonal communication from the patient's perspective. Some are entirely devoted to this attribute of care (Stewart 1995; Stewart et al. 1999), and others are subscales of generic instruments developed to evaluate consumers' experience with primary care (Brody et al. 1989; Flocke 1997; Safran, Kosinski et al. 1998; Stewart 1995). Each of these instruments has a slightly different approach to the definition of interpersonal communication. Roughly speaking, they all propose definitions based on the two key moments of the clinical encounter: taking the patient's history and concluding the interview. Eliciting and acknowledging patients' concerns and explaining the diagnosis and management plan are part of all definitions (Brody et al. 1989; Flocke 1997; Safran, Kosinski et al. 1998; Stewart 1995; Stewart et al. 1999). Some instruments also include consideration of patients' preferences and abilities in the decision-making process (Stewart 1995; Stewart et al. 1999).

As is the case with most instruments developed to measure primary care attributes, there is little comparative information to guide evaluators in selecting the appropriate tool for evaluating the experience of interpersonal communication. Our objective was to compare validated instruments that purport to measure interpersonal communication. We compared scores of subscales from different instruments and examined whether they measured the same construct of interpersonal communication. We expected to find some overlap between interpersonal communication and respectfulness. After discussion, we excluded trust as being a result rather than a component of interpersonal communication. Finally, we examined the psychometric performance of individual items. Our intent is not to recommend one instrument over another, but to provide insight into how well different subscales fit our operational definition of interpersonal communication.

Method

The method and analytic strategy are described in detail elsewhere in this special issue of the journal (Haggerty 2011; Santor et al. 2011). Briefly: six instruments that evaluate PHC from the patient's perspective were administered to 645 healthcare users balanced by English/French language, rural/urban location, low/high level of education and poor/average/excellent overall PHC experience. The analysis consisted of examining the distributional statistics and subscale correlations, followed by common factor and confirmatory factor analysis (structural equation modelling) to identify dimensions common to the entire set of items. Finally, we examined the performance of individual items and response scales against constructs emerging across instruments using item response theory analysis.

The confirmatory factor analysis with structural equation modelling for this analysis differed from that used for the other attributes because of the large number of items for the subjects available. We used the robust maximum likelihood (RML) method, which assumes that the variables are continuous and normally distributed, rather than the weighted least squares regression (Flora and Curran 2004). We used the Satorra-Bentler chi-square statistic, which adjusts the model chi-square for non-normality.

Measure description

Among the six validated instruments in our study, four had one or more subscales on interpersonal communication. The Primary Care Assessment Survey (PCAS) had two subscales: Interpersonal Communication, which focuses both on history taking and conclusion of the interview, and Interpersonal Treatment, which focuses more on the physician's attitude towards the patient (warmth, patience, etc.) (Safran, Kosinski et al. 1998). The Components of Primary Care Index (CPCI) had one subscale made up of six items that could be related to the dimensions of eliciting and explaining (Flocke 1997). Finally, the Interpersonal Processes of Care version II (IPC-II) had four subscales: Elicited Concerns, Explained Results, Patient-Centred Decision-Making and Hurried Communication. Permission to use the instruments was obtained from all instrument developers.

The first version of the EUROPEP (EUROPEP-I) does not include a specific subscale on communication (Grol et al. 2000; Wensing et al. 2000), but its 15-item Clinical Behaviour subscale includes questions conceptually related to the eliciting, explaining and decision-making dimensions. Although our initial intent was to respect scales as conceived and validated by the instrument developers, these items clearly did not fit the analyses of interpersonal communication so we excluded them from our analysis. With these EUROPEP-I questions included, these four instruments contained eight subscales on interpersonal communication.

We assigned items to factors or underlying subdimensions based on the exploratory factor analysis and our judgment of fit with the operational definition when items had ambiguous loadings. We used as the reference item for confirmatory factor loading the one with the highest principal components loading and apparent content fit with the latent variable.

We compared the appropriateness of a number of models in which the correlations between factors were allowed to vary or were fixed as orthogonal.

We based factor analysis only on subjects with no missing values; those with at least one missing value on any item (listwise missing) were excluded. This reduced our effective sample size. Given that this conservative approach can introduce bias, we repeated all the analyses using maximum likelihood imputation of missing values (Rubin 1987) to examine the robustness of our conclusions.

Results

Comparative descriptive statistics

The summary of the eight subscales item content and distributions are summarized in Table 1 and available in detail online at (http://www.longwoods.com/content/22636). Very few items had ≥5% missing values. Exceptions were the items in IPC-II Explained Results (5% each item) and IPC-II Patient-Centred Decision-Making (7%–9%). Additionally, many respondents declared as "not applicable" the EUROPEP-I questions on emotional problems related to health status (16%) and following advice (5%), which count as missing values in analysis. The vast majority of respondents selected the two highest response options and very few the lowest options. For subsequent analyses four items in the EUROPEP-I Clinical Behaviour subscale were removed because we judged that they do not address interpersonal communication (items EU_CB6 to EU_CB11). Confirmatory factor analysis supported this decision.

TABLE 1. Summary of Interpersonal Communication subscale content and distribution of item responses (detailed distribution available at http://www.longwoods.com/content/22636)

Subscale and Item Description	Response Scale	Range Missing Values	Overall Modal Response	Range Item Discriminability	Comments on Distribution
PCAS Communication (6 items) Rate aspect of *talking* with regular doctor: thoroughness of *questions* about symptoms; *attention* to what you have to say; *explanations* of health problems or treatments; *instructions* about further care; *advice* and help in making decisions about care Frequency of leaving doctor's office with *unanswered questions*	Likert evaluative, 1=very poor to 6= excellent	1%–2%	5 (very good)	1.8 (frequency) to 5.9 (explanation, advice)	Approximately 60% of responses in two most positive categories; ≈% in most negative
PCAS Interpersonal Treatment (5 items) Rate the personal aspects of care from regular doctor: amount of *time* spent; *patience* with questions or worries; *friendliness and warmth*; *caring and concern*; *respect* for you	Likert evaluative, 1=very poor to 6= excellent	0%–1%	6 (excellent)	3.7 (time) to 9.9 (caring and concern)	Approximately 60% of responses in two most positive categories; ≈4% in most negative

TABLE 1. Continued

Subscale and Item Description	Response Scale	Range Missing Values	Overall Modal Response	Range Item Discriminability	Comments on Distribution
CPCI Interpersonal Communication (6 items) Best description of regular doctor: Positive statements: I can easily talk about personal things; doctor always explains Negative statements: doctor does not always listen; I don't bring up things that I'm worried about; don't always feel comfortable asking questions; feel doctor ignores my concerns	Semantic differential opinion, 1=strongly disagree, 6=strongly agree	2%–3%	Most positive	1.5 (worries) to 3.8 (ignores)	Most respondents (36%–53%) select most positive opinion; 6%–8% most negative
EUROPEP Clinical Behaviour (10/15 items on communication) Rate care by GP: enough time; interest in personal situation; ease telling about problems; involvement in decisions about medical care; listening; explaining tests and treatments; telling about symptoms and/or illness; help with emotional problems; help understanding importance of advice; knowing what was done or said in previous contacts	Semantic differential rating, 1=poor, 5=excellent	2%–3% (true missing) 1%–16% (n/a)	5 (excellent)	2.9 (time) to 4.5 (listening)	Approximately 40% in most positive category; 2%–4% most negative. Two items with high proportion "not applicable": emotional problems (16%), advice (5%)
IPC-II-Elicited Concerns (3 items) Frequency of doctor's/doctors': Really finding out concerns; letting you say what was important; taking health concerns very seriously	Likert frequency, 1=never, 5=always	3%	5 (always)	3.1 (take seriously) to 4.5 (find out concerns)	Almost 80% of responses in two most frequent categories; 1% never
IPC-II-Explained Results (4 items) Frequency of doctor's/doctors' explaining: test results; results of physical exam; effect of not taking a prescribed medicine; possible side effects from a medicine	Likert frequency, 1=never, 5=always	5%	5 (always)	2.0 (side effects) to 6.3 (physical exam)	Approximately 70% of responses in two most frequent categories; 3%–7% never
IPC-II-Patient-Centred Decision-Making (4 items) Frequency of doctor's/doctors': working out a treatment plan together; asking if help needed deciding between treatment choices; asking about anticipated problems following recommendations; asking if you could do the recommended treatment	Likert frequency, 1=never, 5=always	7%–9%	4 (usually)	2.6 (plan together) to 5.5 (about problems)	Less skewed than others: ≈18% never, ≈30% usually, ≈18% always. High proportion of missing values likely due to individual case applicability, not health system context
IPC-II-Hurried Communication (5 items) Frequency of doctor's/doctors' negative behaviours: speaking too fast; using hard-to-understand words; ignoring what told; appearing distracted; seeming bothered with questions	Likert frequency, 1=never, 5=always	3%–4%	1 (never)	1.5 (hard to understand) to 3.6 (bothered)	Approximately 80% of responses in least frequent categories; 1% always

The parametric estimate of the discriminatory capacity within the original subscale is shown in the right-hand column of Table 1 (Item Discrimination), with ≥1.0 indicating that the item discriminates between different levels of the subscale score. All items demonstrate good to excellent discriminability.

Table 2 presents the descriptive statistics for the subscale scores. The normalized means and standard deviations for the subscales differ substantially from one subscale to another. All subscales except the IPC-II Patient-Centred Decision-Making are skewed towards positive values, with the median consistently higher than the mean. The internal consistency of all the scales is excellent.

TABLE 2. Mean and distributional values for interpersonal communication subscales, values normalized to a 0-to-10 scale (n=645)

				Quartiles			
Developer's Subscale Name	**Number of Items**	**Mean**	**SD**	**Q1 (25%)**	**Q2 (50%)**	**Q3 (75%)**	**Cronbach's Alpha**
PCAS Communication	6	7.33	2.11	6.00	7.67	9.00	.95
PCAS Interpersonal Treatment	5	7.44	2.17	6.00	7.80	9.60	.96
CPCI Interpersonal Communication	6	7.19	2.32	5.30	7.67	9.30	.96
EUROPEP Clinical Behaviour	10	7.70	2.30	6.50	8.25	9.50	.96
IPC-II Elicited Concerns	3	7.81	2.16	6.70	8.33	10.00	.86
IPC-II Explained Results	4	7.40	2.49	5.60	8.13	9.40	.88
IPC-II Patient-Centred Decision-Making	4	5.41	3.15	2.50	5.63	7.50	.91
IPC-II Hurried Communication	5	8.01	1.77	7.00	8.42	9.50	.85

Table 3 presents the Pearson correlations between the subscales for interpersonal communication. All the subscales correlate relatively well. Not surprisingly, the two most correlated (.82) were from the same instrument: PCAS Communication and PCAS Interpersonal Treatment. The IPC-II Patient-Centred Decision-Making showed the lowest correlation overall (.46 to .63), although the subscales correlated most highly within their own attribute family. Again not surprisingly, they also correlated highly with subscales in Relational Continuity that measure knowledge of the patient (.48 to .71) and with the IPC-II Interpersonal Style subscale, which was mapped to the attribute of respectfulness (.59–.79). The interpersonal communication subscales also correlate strongly with PCAS Trust (.49–.73), as shown in Table 3.

TABLE 3. Partial correlations§ between interpersonal communication subscales. Only correlations significantly different from zero are provided.

	PCAS* Communication	**PCAS Interpersonal Treatment**	**CPCI** Interpersonal Communication**	**EUROPEP Clinical Behaviour**	**IPC-II*** Elicited Concerns**	**IPC-II Explained Results**	**IPC-II Patient-Centred Decision-Making**	**IPC-II Hurried Communication**
PCAS Communication	1.00	0.82	0.62	0.72	0.59	0.50	0.46	0.63
PCAS Interpersonal Treatment	0.82	1.00	0.63	0.76	0.67	0.48	0.49	0.64
CPCI Interpersonal Communication	0.62	0.63	1.00	0.69	0.65	0.54	0.44	0.62

TABLE 3. Continued

	PCAS* Communication	PCAS Interpersonal Treatment	CPCI** Interpersonal Communication	EUROPEP Clinical Behaviour	IPC-II*** Elicited Concerns	IPC-II Explained Results	IPC-II Patient-Centred Decision-Making	IPC-II Hurried Communication
EUROPEP Clinical Behaviour	0.72	0.76	0.69	1.00	0.75	0.65	0.61	0.72
IPC Elicited Concerns	0.59	0.67	0.65	0.75	1.00	0.60	0.57	0.65
IPC Explained Results	0.50	0.48	0.54	0.65	0.60	1.00	0.63	0.59
IPC Patient-Centred Decision-Making	0.46	0.49	0.44	0.61	0.57	0.63	1.00	0.47
IPC Hurried Communication	0.63	0.64	0.62	0.72	0.65	0.59	0.47	1.00
PCAS Trust	0.68	0.68	0.68	0.73	0.63	0.55	0.49	0.60

§ Controlling for study design variables: province, educational achievement, geographic location.

* PCAS: Primary Care Assessment Survey

** CPCI: Components of Primary Care Instrument

*** IPC: Interpersonal Processes of Care

Do all items measure a single construct?

Missing values reduced our effective sample size for factor analysis from 645 to 427. Those patients included in the factor analysis tended to be in poorer health, to have been affiliated longer with their physician and to have consulted more often than those excluded. Imputing the missing values using a maximum likelihood method (Rubin 1987) increased our sample size and improved the fit statistics in the confirmatory factor analysis models, but it did not change the magnitude of loadings or the direction of our conclusions.

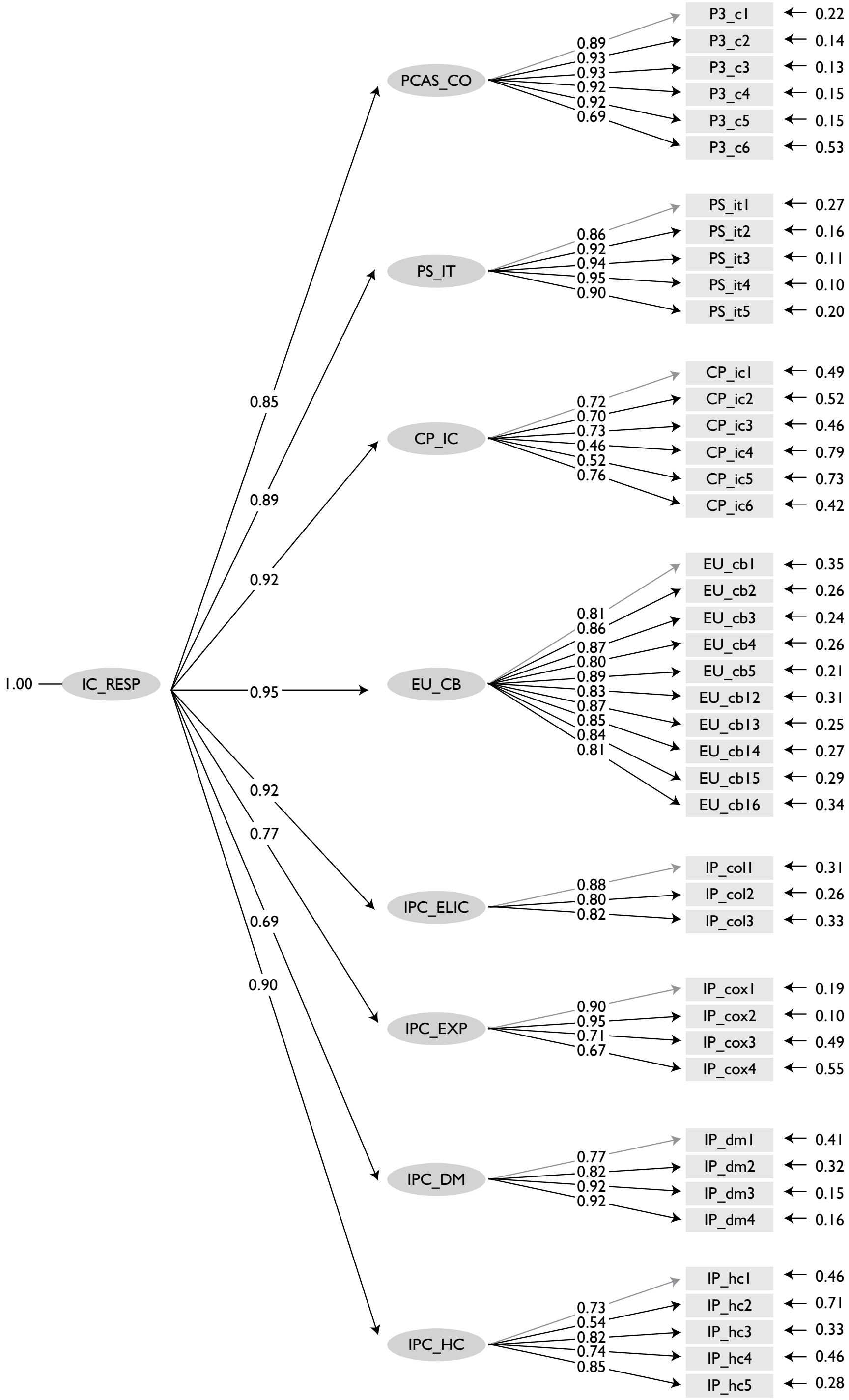
Most items load reasonably well (>.30) on a single factor with common factor analysis. However, in confirmatory factor analysis a unidimensional model did not demonstrate a good fit with a root mean square error of approximation (RMSEA) of .122, well above the .05 considered to demonstrate good fit even though the normed fit index (NFI) of .94 was above the .90 standard for good fit. Model fit improved when the items were grouped in their parent subscales and then linked to a single underlying construct presumed to be interpersonal communication, as shown in Figure 1. This model fit significantly better than the unidimensional model, as shown by the difference in chi-square values (χ^2 = 6,299 – 2,150 = 4,149, 8 *df*, p<.0001); the RMSEA of .06 and the NFI of .98 indicated good fit.

Item loadings are lowest for the CPCI Interpersonal Communication subscale, and IPC-II Explain Results and IPC-II Patient-Centred Decision-Making have weaker loadings than other subscales on the construct presumed to be interpersonal communication.

How do underlying factors fit with operational definition?

Exploratory factor analysis suggested three underlying factors in the pool of all these items. Using our operational definition as a guide, we judged that the first factor (eigenvalue = 26.56) seemed to assess the provider's ability "to elicit and understand patient concerns" (eliciting), the second (eigenvalue = 2.45) "to explain healthcare issues" (explaining) and the third

FIGURE 1. Parameter estimations for a structural equation model showing item loadings of items on parent subscales (first order) indicating the correlations between the scales



Chi-square = 2993.83, *df* = 852, *p*-value = .00000, RMSEA = .060

(eigenvalue = 1.34) "to engage in shared decision-making" (decision-making). Most items load on a first factor that relates conceptually to the eliciting dimension, including all items in PCAS Interpersonal Treatment, IPC-II Elicit and IPC-II Hurried Communication. Most CPCI Interpersonal Communication items, half of the EUROPEP-I items and the IPC-II Explain subscales load on the factor that relates conceptually to the explaining dimension, but the loadings are modest, aside from three of the four questions of the IPC-II Explain subscale. Finally, the decision-making dimension is not well represented in the scales we studied. Only six items loaded on the factor: the four questions of the IPC-II Patient-Centred Decision-Making subscale – with high loadings – and one question each on the PCAS Communication and EUROPEP-I Clinical Behaviour subscales.

A confirmatory analysis model, in which the items are grouped by eliciting, explaining and decision-making, are then associated to a single factor representing interpersonal communication (as shown in Figure 2). Compared to the unidimensional model, we see a moderate improvement in goodness of fit using the chi-square criterion (χ^2 = 6,299 – 5,039 = 1,250, 3 *df*, $p<.001$).

Figure 2 shows that the dimensions of eliciting and explaining are highly correlated (.92), but decision-making has much lower correlations with these two dimensions: .70 and –.79, respectively. Some items do not have high loadings and have a high proportion of residual error (shown to the right of each item), either because they are not discriminatory or because they relate better to another construct that is not part of the latent variable. Overall, the items in the PCAS and EUROPEP-I subscales seem to fit best with the subdimensions with which they were associated.

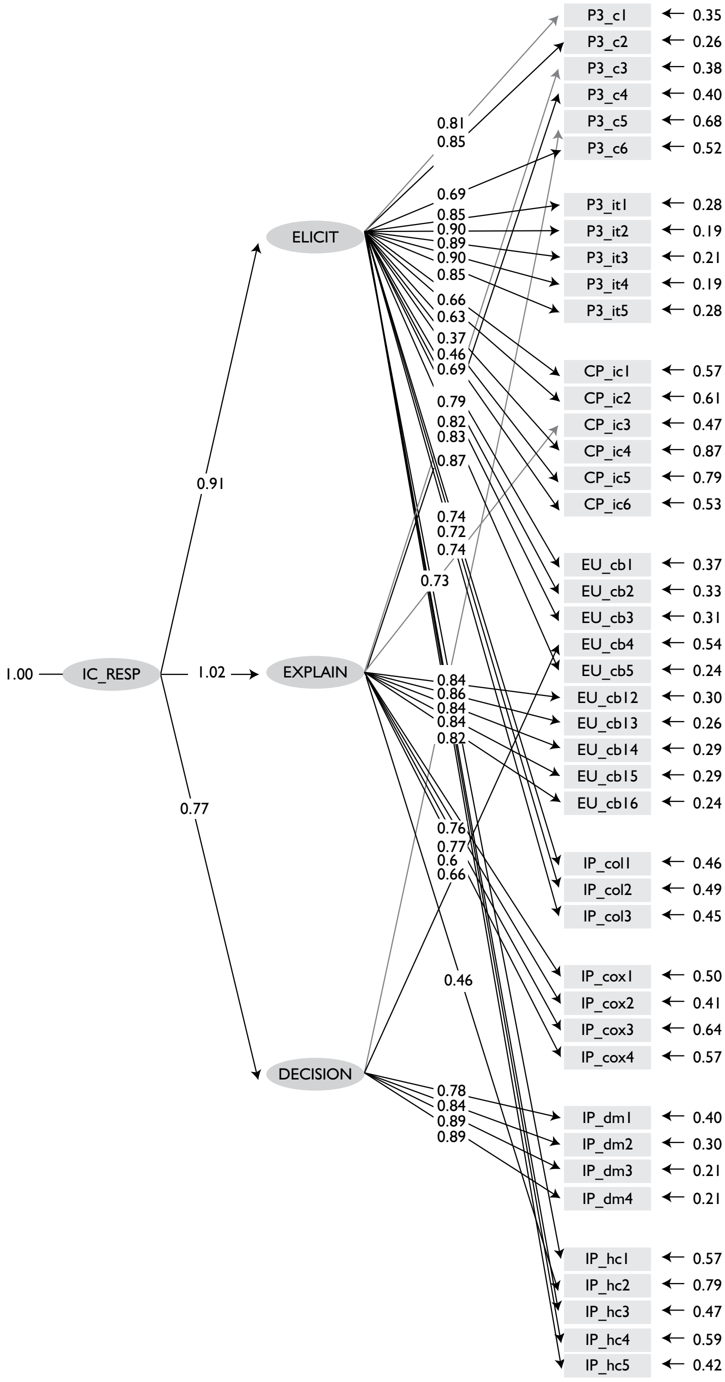
Individual item performance

We conducted item response analysis to evaluate the performance of individual items as a function of both the parent subscales and the dimensions of eliciting, explaining and decision-making.

For the eliciting dimension, the items from the PCAS and EUROPEP-I subscales demonstrated the best performance overall. The items showed relatively good discriminability, with the item score increasing with the total eliciting score. The probability of selecting low-rating options – e.g., 1 (very poor) to 3 (fair) – occurs appropriately in the negative zone of eliciting, and high ratings – 5 (good) and 6 (excellent) – in the most positive. However, the negative options are rarely endorsed, and in the PCAS, the "4=good" option is mostly endorsed when eliciting is below average; in the EUROPEP-I, "5=excellent" is endorsed at the average level of eliciting. The PCAS response scale consists essentially of responses 3 to 6, and the EUROPEP-I of 2 to 5. But peak probabilities of responding are well differentiated from one another.

The CPCI items perform the next best, showing relatively good discriminability, especially in the negative zone of eliciting. But the full response options are not fully exploited nor equally informative. The responses are either 1 (strongly agree) or 6 (strongly disagree), and the peak probabilities for the other response options are not clearly differentiated, suggesting this instrument functions as a three-point scale. This pattern is even more extreme in the

FIGURE 2. Parameter estimates for a structural equation model where items are associated with the (first-order) constructs of eliciting, explaining and decision-making, which are linked to a single underlying construct IC_RESP (second order) presumed to be interpersonal communication



Chi-square = 7651.14, *df* = 857, *p*-value = .00000, RMSEA = .060

IPC-II subscales, where any response other than "5=always" is associated with below-average eliciting and should be interpreted negatively.

These same patterns of performance are seen for each instrument in the dimensions of explaining and decision-making, although the latent variable is composed of a smaller number of items and interpretation necessarily requires prudence.

Discussion and Conclusion

Our results suggest that the validated scales that map to interpersonal communication do indeed have a single underlying construct that includes three distinct dimensions as stated in our operational definition: eliciting problems and concerns, explaining illness management and involving the patient in decision-making. This finding supports the hypothesis we formed based on our operational definition of interpersonal communication and adds to a body of knowledge along the same lines (Epstein et al. 2005; Hall et al. 1996; Mead and Bower 2002; Saba et al. 2006; Stewart 1995).

As mentioned earlier, our aim was not to evaluate the metric proprieties of entire instruments. All these instruments have been developed according to somewhat different conceptual frameworks and measurement approaches. Our aim was to compare their capacity to measure the construct of interpersonal communication according to the operational definition developed by PHC experts and providers (Haggerty et al. 2007). Unlike some others (Stewart 1995; Stewart et al. 1999), we did not explore patients' definitions of this attribute. Our aim was to ascertain whether subscales from different instruments that appear to measure this dimension actually do so.

The PCAS Communication subscale demonstrates good metric properties and discriminability and includes items that measure the eliciting and explaining dimensions of interpersonal communication plus one item on decision-making. The EUROPEP-I Clinical Behaviour subscale had only acceptable metric properties and discriminability; admittedly, however, the EUROPEP-I Clinical Behaviour component is not intended to relate specifically to interpersonal communication. The new version of the EUROPEP includes several revisions to communication items that may improve performance, but it still applies to general clinical encounters of which communication is only a part (Wensing 2006). The CPCI Interpersonal Communication subscale relates to the eliciting and explaining dimensions and, despite certain measurement problems, has good capacity to detect and discriminate problems in this dimension. The IPC-II has subscales addressing all three dimensions of interpersonal communication, but they load less on the construct and their items have poorer discriminability than subscales from other instruments. Some items map more to the respectfulness construct (PCAS Interpersonal Treatment and IPC-II Hurried Communication). The IPC-II subscales of Explained Results and Patient-Centred Decision-Making have high levels of missing values, creating problems for evaluators. The questions may not be clear or the response scale may be inappropriate for the questions, as indicated in our cognitive testing (Haggerty, Santor et al. 2011).

Patient involvement with decisions related to care is particularly poorly developed in all these instruments. In addition, much theoretical and empirical work has been done on the measurement of two important concepts related to interpersonal communication that are not considered in the instruments we studied: patient-centredness measurement tools, in which the notion of agreement on the definition of the problem and on the course of action is core (and has been associated with outcomes) (Epstein et al. 2005; Stewart 1995; Little et al. 2001), and shared decision-making measurement instruments (Elwyn et al. 2001).

As suggested earlier in this paper, very few general instruments that propose to measure communication between providers and patients have strong conceptual foundations. Those that have been associated more often with quality-of-care indicators, and less frequently with health outcomes, are the EUROPEP-I and the PCAS. For example, patients in a comparative European study based on the EUROPEP-I reported good interpersonal communication with their physicians despite important differences between countries in ratings of the organizational dimensions of care (Grol et al. 2000). Similar observations were made in the United States in studies using the PCAS (Safran et al. 1994, 2000, 2002). Although these observations may seem to suggest that patients are satisfied with their communication with their usual providers independently of the primary care model, recent comparative studies suggest that the type of organizational model makes a difference in most care experience indicators – the professional single-provider model being the best performing, and more complex organizational models performing less well (Lamarche et al. 2003). Indeed, some studies suggest that organizational interventions such as team care can disrupt the relationship between patients and their primary care providers and have a negative impact on the quality of interpersonal communication (Rodriguez et al. 2007; Safran 2003). It should be noted, however, that none of these instruments permits the evaluation of the experience of interpersonal communication from a team care perspective, an area where more research is certainly needed.

As organizational interventions increasingly change the day-to-day experience of care, they may have a growing impact on the quality of provider–patient interpersonal communication. Conversely, positive experiences with interpersonal communication and respectfulness may buffer patients from negative experiences associated with organizational changes. Indeed, we found that the Interpersonal Communication subscales provided the greatest discrimination between excellent, average and poor experience of care when our respondents were asked to rate their overall experience (Haggerty 2011). Hence the importance of being able to monitor this attribute of primary care, which is deemed essential by users, providers and decision-makers.

ACKNOWLEDGEMENTS

This research was funded by the Canadian Institutes of Health Research. Dr. Beaulieu was holder of the Sadok Basrour Research Chair in Family Medicine Research at the Université de Montréal during the conduct of this study. The authors thank Beverley Lawson for conducting the survey in Nova Scotia and Christine Beaulieu in Quebec; Fatima Bouharaoui for performing the analysis; Jocelyne Gagné and Donna Riley for support in the preparation of this manuscript.

Correspondence may be directed to: Marie-Dominique Beaulieu, Département de médecine familiale, Centre de recherche du CHUM, Hôpital Notre-Dame, Pavillon L.-C. Simard, 8ième étage, 1560, rue Sherbrooke est, Montréal, QC H2L 4M1; tel.: 514-890-8000 ext. 28046; fax: 514-412-757; e-mail: marie-dominique.beaulieu@umontreal.ca.

REFERENCES

Brody, D.S., S.M. Miller, C.E. Lerman, D.G. Smith and C. Caputo. 1989. "Patient Perception of Involvement in Medical Care: Relationship to Illness Attitudes and Outcomes." *Journal of General Internal Medicine* 4: 506–11.

Brown, J.B., M. Stewart and B.L. Ryan. 2001. *Assessing Communication between Patients and Physicians: The Measure of Patient-Centred Communication.* Working Paper Series 95-2. London, ON: Centre for Studies in Family Medicine.

Elwyn, G., A. Edwards, S. Mowle, M. Wensing, C. Wilkinson, P. Kinnersley and R. Grol. 2001. "Measuring the Involvement of Patients in Shared Decision-Making: A Systematic Review of Instruments." *Patient Education and Counseling* 43: 5–22.

Epstein, R.M., P. Franks, K. Fiscella, C.G. Shields, S.C. Meldrum, R.L. Kravitz and P.R. Duberstein. 2005. "Measuring Patient-Centered Communication in Patient–Physician Consultations: Theoretical and Practical Issues." *Social Science and Medicine* 61: 1516–28.

Flocke, S. 1997. "Measuring Attributes of Primary Care: Development of a New Instrument." *Journal of Family Practice* 45(1): 64–74.

Flora, D.B. and P.J. Curran. 2004. "An Empirical Evaluation of Alternative Methods of Estimation for Confirmatory Factor Analysis with Ordinal Data." *Psychological Methods* 9(4): 466–91.

Grol, R., M. Wensing and Task Force on Patient Evaluations of General Practice. 2000. *Patients Evaluate General/Family Practice: The EUROPEP-I Instrument.* Nijmegen, Netherlands: Centre for Quality of Care Research, Raboud University.

Haggerty, J.L. 2011. "Measurement of Primary Healthcare Attributes from the Patient Perspective." *Healthcare Policy* 7 (Special Issue): 13–20.

Haggerty, J., F. Burge, J.-F. Lévesque, D. Gass, R. Pineault, M.-D. Beaulieu and D. Santor. 2007. "Operational Definitions of Attributes of Primary Health Care: Consensus among Canadian Experts." *Annals of Family Medicine* 5: 336–44.

Haggerty, J.L., C. Beaulieu, B. Lawson, D.A. Santor, M. Fournier and F. Burge. 2011. "What Patients Tell Us about Primary Healthcare Evaluation Instruments: Response Formats, Bad Questions and Missing Pieces." *Healthcare Policy* 7 (Special Issue): 66–78.

Hall, J.A., D.L. Roter, M.A. Milburn and L.H. Daltroy. 1996. "Patients' Health as a Predictor of Physician and Patient Behavior in Medical Visits: A Synthesis of Four Studies." *Medical Care* 34: 1205–18.

Headache Study Group of the University of Western Ontario. 1986. "Predictors of Outcome in Headache Patients Presenting to Family Physicians – A One Year Prospective Study." *Headache* 6: 285–94.

Lamarche, P.-A., M.-D. Beaulieu, R. Pineault, A.-P. Contandriopoulos, J.-L. Denis and J. Haggerty. 2003. *Choices for Change: The Path for Restructuring Primary Healthcare Services in Canada.* Ottawa: Canadian Health Services Research Foundation.

Little, P., H. Everitt, I. Williamson et al. 2001. Observational Study of Effect of Patient Centredness and Positive Approach on Outcomes of General Practice Consultations. *British Medical Journal* 323(7318): 908–11.

Mead, N. and P. Bower. 2002. "Patient-Centred Consultations and Outcomes in Primary Care: A Review of the Literature." *Patient Education and Counseling* 48: 51–61.

Rodriguez, H.P., W.H. Rogers, R.E. Marshall and D.G. Safran. 2007. "Multidisciplinary Primary Care Teams: Effects on the Quality of Clinician–Patient Interactions and Organizational Features of Care." *Medical Care* 45: 19–27.

Rubin, D.B. 1987. *Multiple Imputation for Nonresponse in Surveys.* Hoboken, NJ: John Wiley.

Saba, G.W., S.T. Wong, D. Schillinger, A. Fernandez, C.P. Somkin, C.C. Wilson and K. Grumbach. 2006. "Shared Decision Making and the Experience of Partnership in Primary Care." *Annals of Family Medicine* 4: 54–62.

Safran, D.G. 2003. "Defining the Future of Primary Care: What Can We Learn from Patients?" *Annals of Internal Medicine* 138: 248–55.

Safran, D.G., J. Kosinski, A.R. Tarlov, W.H. Rogers, D.A. Taira, N. Lieberman and J.E. Ware. 1998. "The Primary Care Assessment Survey: Tests of Data Quality and Measurement Performance." *Medical Care* 36(5): 728–39.

Safran, D.G., W.H. Rogers, A.R. Tarlov, T.S. Inui, D.A. Taira and J.E. Montgomery. 2000. "Organizational and Financial Characteristics of Health Plans: Are They Related to Primary Care Performance?" *Archives of Internal Medicine* 160: 69–76.

Safran, D.G., D.A. Taira, W.H. Rogers, M. Kosinski, J.E. Ware and A.R. Tartov. 1998. "Linking Primary Care Performance to Outcomes of Care." *Journal of Family Practice* 47: 213–20.

Safran, D.G., A.R. Tarlov and W.H. Rogers. 1994. "Primary Care Performance in Fee-for-Service and Prepaid Healthcare Systems. Results from the Medical Outcomes Study." *Journal of the American Medical Association* 271: 1579–86.

Safran, D.G., I.B. Wilson, W.H. Rogers, J.E. Montgomery and H. Chang. 2002. "Primary Care Quality in the Medicare Program: Comparing the Performance of Medicare Health Maintenance Organizations and Traditional Fee-for-Service Medicare." *Archives of Internal Medicine* 162: 757–65.

Santor, D.A., J.L. Haggerty, J.-F. Lévesque, F. Burge, M.-D. Beaulieu, D. Gass and R. Pineault. 2011. "An Overview of Confirmatory Factor Analysis and Item Response Analysis Applied to Instruments to Evaluate Primary Healthcare." *Healthcare Policy* 7 (Special Issue): 79–92.

Simpson, M., R. Buckman, M. Stewart, P. Maguire, M. Lipkin, D. Novack and J. Till. 1991. "Doctor–Patient Communication: The Toronto Consensus Statement." *British Medical Journal* 303: 1385–87.

Stange, K.C. and L.S. Acheson. 2006. "Communication in the Era of 'Personalized' Medicine." *Annals of Family Medicine* 4: 194–96.

Starfield, B., C. Wray, K. Hess, R. Gross, P.S. Birk and B.C. D'Lugoff. 1981. "The Influence of Patient–Practitioner Agreement on Outcome of Care." *American Journal of Public Health* 71: 127–32.

Stewart, A.L., A. Nápoles-Springer and E.J. Pérez-Stable. 1999. "Interpersonal Processes of Care in Diverse Populations." *Milbank Quarterly* 77(3): 305–39, 274.

Stewart, M. 1995. "Effective Physician–Patient Communication and Health Outcomes: A Review." *Canadian Medical Association Journal* 152: 1423–33.

Thom, D. and B. Campbell. 1997. "Patient–Physician Trust: An Exploratory Study." *Journal of Family Practice* 44: 169–76.

Wensing, M.E. 2006. *EUROPEP 2006: Revised Europep Instrument and User Manual.* Nijmegen, Netherlands: Centre for Quality of Care Research, Raboud University. Retrieved July 8, 2011. <http://www.topaseurope.eu/?q=node/13>.

Wensing, M., J. Mainz and R. Grol. 2000. "A Standardised Instrument for Patient Evaluations of General Practice Care in Europe." *European Journal of General Practice* 6: 82–87.

Interpersonal Communication from the Patient Perspective: Comparison of Primary Healthcare Evaluation Instruments

La communication interpersonnelle du point de vue du patient : comparaison entre instruments d'évaluation des soins de santé primaires

MARIE-DOMINIQUE BEAULIEU, JEANNIE L. HAGGERTY, CHRISTINE BEAULIEU, FATIMA BOUHARAOUI, JEAN-FRÉDÉRIC LÉVESQUE,RAYNALD PINEAULT, FREDERICK BURGE AND DARCY A. SANTOR

TABLE 1. Distribution of responses for each item in subscales measuring interpersonal communication in primary healthcare services (n=645); modal response is shown in bold

Item Code	Item Statement	Missing % (n)	Per Cent (Number) by Response Option						Item Discrimination
PCAS Communication Thinking about **talking** with your regular doctor…			**1=Very poor**	**2=Poor**	**3=Fair**	**4=Good**	**5= Very good**	**6= Excellent**	
PS_c1	How would you rate the thoroughness of your doctor's **questions** about your symptoms and how you are feeling?	1 (5)	1 (6)	3 (21)	12 (76)	24 (157)	**35 (223)**	24 (157)	4.84 (.30)
PS_c2	How would you rate the **attention** your doctor gives to what you have to say?	1 (5)	1 (7)	4 (26)	11 (74)	22 (143)	29 (188)	**31 (202)**	5.30 (.35)
PS_c3	How would you rate doctor's **explanations** of your health problems or treatments that you need?	2 (10)	2 (10)	4 (23)	10 (64)	23 (147)	**32 (207)**	29 (184)	5.93 (.40)
PS_c4	How would you rate doctor's **instructions** about symptoms to report and when to seek further care?	1 (7)	2 (11)	4 (24)	11 (74)	24 (152)	**32 (205)**	27 (172)	5.72 (.37)
PS_c5	How would you rate the doctor's **advice** and help in making decisions about your care?	2 (13)	2 (11)	4 (23)	11 (73)	24 (154)	**31 (200)**	27 (171)	5.90 (.39)
PS_c6	How often do you leave your doctor's office with **unanswered questions**?	1 (4)	1 (9)	3 (19)	8 (49)	20 (132)	**41 (263)**	26 (169)	1.76 (.14)
PCAS Interpersonal Treatment Thinking about the personal aspects of the care you receive from your regular doctor…			**1=Very poor**	**2=Poor**	**3=Fair**	**4=Good**	**5=Very good**	**6=Excellent**	
PS_it1	How would you rate the amount of **time** your doctor spends with you?	1 (4)	1 (9)	5 (32)	16 (104)	27 (177)	**28 (178)**	22 (141)	3.66 (.23)
PS_it2	How would you rate doctor's **patience** with your questions or worries?	1 (5)	0 (3)	3 (19)	13 (85)	23 (150)	28 (178)	**32 (205)**	5.64 (.34)
PS_it3	How would you rate doctor's **friendliness and warmth** toward you?	0 (3)	1 (8)	3 (21)	9 (57)	24 (155)	27 (176)	**35 (225)**	7.26 (.49)
PS_it4	How would you rate doctor's **caring and concern** for you?	1 (4)	0 (3)	4 (25)	10 (67)	25 (158)	27 (174)	**33 (214)**	8.85 (.71)
PS_it5	How would you rate doctor's **respect** for you?	1 (6)	1 (8)	2 (11)	6 (41)	22 (139)	28 (183)	**40 (257)**	5.41 (.35)
CPCI Interpersonal Communication			**1=Strongly disagree**	**2**	**3**	**4**	**5**	**6=Strongly agree**	
CP_ic1	I can easily talk about personal things with this doctor.	2 (13)	7 (48)	7(48)	10 (62)	16 (102)	21(137)	**36 (235)**	2.12 (.16)
CP_ic2	Sometimes, this doctor does not listen to me.	2 (15)	**53 (339)**	17 (107)	7 (47)	9 (55)	7 (48)	5 (34)	3.20 (.22)
CP_ic3	This doctor always explains things to my satisfaction.	2 (14)	3 (22)	5 (35)	10 (66)	16 (104)	25 (163)	**37 (241)**	2.09 (.16)
CP_ic4	Sometimes, with this doctor, I don't bring up things that I'm worried about.	2 (16)	**35 (225)**	18 (119)	9 (57)	15 (96)	13 (86)	7 (46)	1.51 (.13)
CP_ic5	I don't always feel comfortable asking questions of this doctor.	3 (17)	**41 (265)**	19 (125)	8 (54)	10 (63)	11 (70)	8 (51)	1.81 (.16)
CP_ic6	Sometimes, I feel like this doctor ignores my concerns.	3 (19)	**49 (315)**	18 (119)	9 (56)	9 (55)	7 (45)	6 (36)	3.75 (.27)
EUROPEP Clinical Behaviour			**1=Poor**	**2**	**3**	**4**	**5= Excellent**	**Not applicable**	
EU_cb1	Making you feel you had time during consultations	2 (15)	3 (20)	6 (39)	15 (97)	28 (182)	**44 (282)**	2 (10)	2.91 (.21)
EU_cb2	Interest in your personal situation	2 (15)	3 (19)	8 (49)	13 (84)	31 (203)	**41 (264)**	2 (11)	3.63 (.27)
EU_cb3	Making it easy for you to tell him or her about your problems	2 (15)	3 (19)	6 (38)	16 (102)	27 (171)	**46 (295)**	1 (5)	3.80 (.28)

Item Code	Item Statement	Missing % (n)	Per Cent (Number) by Response Option						Item Discrimination
EU_cb4	Involving you in decisions about your medical care	2 (15)	2 (13)	5 (30)	12 (76)	32 (204)	**45 (292)**	2 (15)	2.99 (.20)
EU_cb5	Listening to you	2 (16)	2 (14)	4 (27)	13 (84)	28 (181)	**49 (317)**	1 (6)	4.50 (.36)
*EU_cb6	Keeping your records and data confidential	3 (19)	0	1 (5)	5 (31)	23 (147)	**66 (425)**	3 (18)	1.93 (.19)
*EU_cb7	Quick relief of your symptoms	3 (19)	2 (10)	3 (21)	16 (106)	35 (224)	**37 (236)**	5 (29)	2.41 (.18)
*EU_cb8	Helping you to feel well so that you can perform your normal daily activities	4 (23)	2 (11)	3 (19)	12 (76)	34 (220)	**42 (271)**	4 (25)	2.96 (.24)
*EU_cb9	Thoroughness	3 (17)	2 (12)	4 (23)	15 (99)	29 (185)	**47 (300)**	1 (9)	4.19 (.32)
*EU_cb10	Physical examination of you	3 (17)	3 (21)	2 (13)	12 (78)	29 (186)	**48 (312)**	3 (18)	2.95 (.22)
*EU_cb11	Offering you services for preventing diseases	3 (21)	4 (26)	5 (34)	14 (88)	29 (184)	**37 (238)**	8 (54)	2.44 (.19)
EU_cb12	Explaining the purpose of tests and treatments	3 (18)	2 (16)	5 (30)	9 (58)	30 (196)	**47 (305)**	3 (22)	3.84 (.28)
EU_cb13	Telling you what you wanted to know about your symptoms and/or illness	3 (19)	3 (17)	4 (26)	10 (67)	30 (195)	**47 (305)**	2 (16)	4.60 (.37)
EU_cb14	Help in dealing with emotional problems related to your health status	3 (18)	6 (41)	6 (39)	16 (102)	22 (143)	**31 (197)**	16 (105)	3.39 (.27)
EU_cb15	Helping you understand the importance of following his or her advice	3 (20)	2 (14)	4 (27)	17 (111)	32 (204)	**37 (238)**	5 (31)	3.66 (.26)
EU_cb16	Knowing what s/he had done or told you during previous contacts	3 (18)	4 (27)	6 (38)	15 (95)	32 (208)	**37 (238)**	3 (21)	2.98 (.20)
IPC-II-Elicited Concerns How often did the doctor(s)…			**1= Never**	**2= Rarely**	**3= Sometimes**	**4= Usually**	**5= Always**		
IP_cel1	Really find out what your concerns were?	3 (22)	2 (15)	10 (66)	13 (82)	**36 (231)**	**36 (229)**		4.46 (.31)
IP_cel2	Let you say what you thought was important?	3 (21)	1 (9)	6 (39)	11 (69)	37 (236)	**42 (271)**		3.24 (.23)
IP_cel3	Take your health concerns very seriously?	3 (22)	1 (9)	5 (31)	10 (62)	32 (207)	**49 (314)**		3.09 (.22)
IPC-II-Explained Results How often did the doctor(s)…			**1=Never**	**2=Rarely**	**3= Sometimes**	**4= Usually**	**5=Always**		
IP_cex1	Explain your test results, such as blood tests, x-rays or cancer screening tests?	5 (34)	3 (17)	8 (49)	11 (72)	26 (170)	**47 (303)**		4.76 (.34)
IP_cex2	Clearly explain the results of your physical exam?	5 (32)	3 (17)	8 (52)	10 (64)	28 (181)	**46 (299)**		6.29 (.47)
IP_cex3	Tell you what could happen if you didn't take a medicine that they prescribed for you?	5 (35)	7 (46)	10 (64)	12 (80)	29 (190)	**36 (230)**		2.41 (.18)
IP_cex4	Tell you about side effects you might get from a medicine?	5 (30)	6 (37)	12 (78)	15 (99)	27 (177)	**35 (224)**		2.00 (.16)
IPC-II-Patient-Centred Decision-Making How often did…			**1=Never**	**2=Rarely**	**3= Sometimes**	**4= Usually**	**5=Always**		
IP_dm1	You and your doctor(s) work out a treatment plan together?	7 (42)	17 (107)	14 (91)	14 (91)	**31 (198)**	18 (116)		2.58 (.17)
IP_dm2	If there were treatment choices, how often did the doctor(s) ask you if you would like to help decide the treatment?	9 (59)	18 (113)	16 (101)	12 (77)	**27 (177)**	18 (118)		3.30 (.21)
IP_dm3	The doctor(s) ask if you would have any problems following what they recommended?	7 (46)	18 (118)	13 (87)	18 (114)	**25 (162)**	18 (118)		5.53 (.32)
IP_dm4	The doctor(s) ask if you felt you could do the recommended treatment?	7 (47)	19 (125)	13 (86)	13 (83)	**26 (167)**	21 (137)		5.38 (.34)
IPC-II-Hurried Communication How often did the doctor(s)…			**1=Never**	**2=Rarely**	**3= Sometimes**	**4= Usually**	**5=Always**		
IP_hc1	Speak too fast?	3 (22)	**41 (263)**	36 (235)	15 (99)	3 (18)	1 (8)		2.28 (.18)
IP_hc2	Use words that were hard to understand?	3 (21)	35 (228)	**41 (267)**	17 (112)	2 (14)	0 (3)		1.51 (.14)
IP_hc3	Ignore what you told them?	4 (24)	**47 (300)**	32 (204)	14 (88)	3 (21)	1 (8)		3.29 (.24)
IP_hc4	Appear to be distracted when they were with you?	4 (24)	**44 (284)**	36 (230)	12 (75)	4 (23)	1 (9)		2.59 (.20)
IP_hc5	Seem bothered if you asked several questions?	4 (26)	**52 (333)**	26 (168)	12 (79)	5 (31)	1 (8)		3.56 (.26)

* Item removed from further analysis.

RESEARCH PAPER

Relational Continuity from the Patient Perspective: Comparison of Primary Healthcare Evaluation Instruments

La continuité relationnelle du point de vue du patient : comparaison entre instruments d'évaluation des soins de santé primaires

FREDERICK BURGE, MD, MSC
Department of Family Medicine, Dalhousie University
Halifax, NS

JEANNIE L. HAGGERTY, PHD
Department of Family Medicine, McGill University
Montreal, QC

RAYNALD PINEAULT, MD, PHD
Centre de recherche du Centre hospitalier de l'Université de Montréal
Montréal, QC

MARIE-DOMINIQUE BEAULIEU, MD, MSC
Chaire Dr Sadok Besrour en médecine familiale
Centre de recherche du Centre hospitalier de l'Université de Montréal
Montréal, QC

JEAN-FRÉDÉRIC LÉVESQUE, MD, PHD
Centre de recherche du Centre hospitalier de l'Université de Montréal
Montréal, QC

CHRISTINE BEAULIEU, MSC
St. Mary's Research Centre, St. Mary's Hospital Center
Montreal, QC

DARCY A. SANTOR, PHD
School of Psychology, University of Ottawa
Ottawa, ON

Abstract

The operational definition of relational continuity is "a therapeutic relationship between a patient and one or more providers that spans various healthcare events and results in accumulated knowledge of the patient and care consistent with the patient's needs."
Objective: To examine how well relational continuity is measured in validated instruments that evaluate primary healthcare from the patient's perspective.
Method: 645 adults with at least one healthcare contact in the previous 12 months responded to six instruments that evaluate primary healthcare. Five subscales map to relational continuity: the Primary Care Assessment Survey (PCAS, two subscales), the Primary Care Assessment Tool – Short Form (PCAT-S) and the Components of Primary Care Index (CPCI, two subscales). Scores were normalized for descriptive comparison. Exploratory and confirmatory (structural equation modelling) factor analysis examined fit to operational definition, and item response theory analysis examined item performance on common constructs.
Results: All subscales load reasonably well on a single factor, presumed to be relational continuity, but the best model has two underlying factors corresponding to (1) accumulated knowledge of the patient and (2) relationship that spans healthcare events. Some items were problematic even in the best model. The PCAS Contextual Knowledge subscale discriminates best between different levels of accumulated knowledge, but this dimension is also captured well by the CPCI Accumulated Knowledge subscale and most items in the PCAT-S Ongoing Care subscale. For relationship-spanning events, the items' content captures concentration of care in one doctor; this is captured best by the CPCI Preference for Regular Provider subscale and, to a lesser extent, by the PCAS Visit-Based Continuity subscale and one relevant item in the PCAT-S Ongoing Care subscale. But this dimension correlates only modestly with percentage of reported visits to the personal doctor. The items function as yes/no rather than ordinal options, and are especially informative for poor concentration of care.
Conclusion: These subscales perform well for key elements of relational continuity, but do not capture consistency of care. They are more informative for poor relational continuity.

Résumé

La définition opérationnelle de la continuité relationnelle est « une relation thérapeutique entre un patient et un ou plusieurs soignants, inscrite dans le temps et pour divers services de santé, qui donne lieu à des connaissances accumulées au sujet du patient et à des soins cohérents selon ses besoins. »
Objectif : Examiner à quel point la continuité relationnelle est mesurée par les instruments validés qui servent à évaluer les soins de santé primaires du point de vue du patient.
Méthode : Six cent quarante-cinq adultes ayant eu au moins un contact avec les services de santé au cours des 12 mois antérieurs ont répondu à six instruments qui servent à évaluer les soins de santé primaires. Parmi ces instruments, cinq sous-échelles ont trait à la continuité relationnelle : Primary Care Assessment Survey (PCAS, deux sous-échelles), Primary Care Assessment Tool – version courte (PCAT-S) et Components of Primary Care Index (CPCI, deux sous-

échelles). Les résultats ont été normalisés pour permettre des comparaisons descriptives. Les analyses factorielles exploratoires et confirmatoires (modélisation par équation structurelle) ont permis d'examiner l'adéquation à la définition opérationnelle, et l'analyse de réponse par item a permis d'en examiner le rendement en fonction de constructs communs.

Résultats : Toutes les sous-échelles présentent un point de saturation raisonnablement acceptable pour un facteur unique, qui est probablement la continuité relationnelle, mais le meilleur modèle comprend deux facteurs sous-jacents qui correspondent à (1) les connaissances accumulées au sujet du patient et (2) une relation qui se déroule sur plusieurs événements dans les services de santé. Certains items demeurent problématiques même avec le meilleur modèle. La sous-échelle « connaissances contextuelles » du PCAS est la plus discriminante entre différents niveaux de connaissances accumulées, mais cette dimension est également bien captée par la sous-échelle « connaissance accumulées » du CPCI et par la plupart des items de la sous-échelle « soins continus » du PCAT-S. Pour ce qui est de la relation continue sur plusieurs événements, le contenu des items capte la concentration des soins pour un docteur; ce qui est mieux capté par la sous-échelle « préférence pour un clinicien régulier » du CPCI et, à moindre niveau, par la sous-échelle « continuité axée sur les consultations » du PCAS, ainsi que par un item pertinent de la sous-échelle « soins continus » du PCAT-S. Mais la corrélation entre cette dimension et le pourcentage de visites au médecin personnel reste modeste. Les items fonctionnent selon une dichotomie oui/non plutôt que par choix ordinal, et ils renseignent particulièrement dans les cas de faible concentration de soins.

Conclusion : Ces sous-échelles présentent un bon rendement pour les éléments clés de la continuité relationnelle, mais elles ne permettent pas de capter la cohérence des soins. Elles renseignent davantage sur la faible continuité relationnelle.

Continuity of care is central to the delivery of primary healthcare (PHC). Indeed, in a cross-disciplinary review of the literature on continuity of care, the single largest source of research reports came from PHC (Reid et al. 2002).

Background

Conceptualizing relational continuity

The concept has evolved within some health disciplines but has remained stable in family medicine as referring to a provider–patient relationship over time and across different health events. This form of continuity, referred to as "relational continuity," is distinct from other forms that connect services received from different providers (Haggerty et al. 2003). PHC providers see relational continuity as a distinguishing characteristic of their work and a core value (McWhinney 1998). It is also prioritized by patients (Mainous et al. 2001; Baker et al. 2005; Turner et al. 2007; Cheraghi-Sohi et al. 2008). We would argue that in no other segment of the healthcare system is relational continuity more important.

Our consensus consultation of PHC experts across Canada unanimously identified relational continuity as an essential function of PHC regardless of organizational model (Lévesque et al. 2011). It is defined operationally as "a therapeutic relationship between a patient and one or more providers that spans various health care events and results in accumulated knowledge of the patient and care consistent with the patient's needs" (Haggerty et al. 2007).

Evaluating relational continuity in primary healthcare

Within PHC research, relational continuity has been inferred most commonly from the degree to which patient care is concentrated in a single physician (Steinwachs 1979; Rogers and Curtis 1980). Research evidence suggests that seeing the same provider over time for multiple health events is associated with positive outcomes, including better doctor–patient communication (Bertakis and Callahan 1992; Berry et al. 2008), greater uptake of preventive and health-promoting strategies (Ettner 1996, 1999; O'Malley and Forrest 1996; O'Malley 1997; Flocke et al. 1998), reduced diagnostic testing (Weiss and Blustein 1996), reduced emergency department utilization (Burge et al. 2003) and reduced emergency hospital admissions (Wasson et al. 1984).

Most often, relational continuity is conceived as a relationship between a patient and a single doctor. However, new models of PHC evolving in Canada and internationally are moving to interprofessional, team-based care that may be disruptive to relational continuity (Smith 1995; Rodriguez et al. 2007), making it important to measure relational continuity in the evaluation of reforms.

The objective of our study was to compare validated instruments thought to be most pertinent to the Canadian context, and in this paper we focus on how well subscales from different instruments measure the constructs of relational continuity. Specifically, we examined the equivalence of the scores of different instruments' subscales and whether all the relational continuity subscales measure a single construct or factor. If analysis suggested more than one factor, we aimed to determine how these corresponded to the operational definition. Finally, we examined how well individual items perform in measuring the common construct of relational continuity that emerged across instruments.

Method

The conduct of the study (Haggerty 2011) and the analytic approach (Santor et al. 2011) have been described in detail elsewhere. Briefly, six validated instruments that evaluate PHC from the patient's perspective were administered to 645 healthcare users balanced by English/French language, rural/urban location, low/high level of education and poor/average/excellent overall PHC experience. The analysis consisted of examining the distributional statistics and subscale correlations, followed by common factor and confirmatory factor analysis (structural equation modelling) to identify dimensions common to the entire set of items. Finally, we examined the performance of individual items and response scales using item response theory analysis.

Measure description

Three instruments in our study have five subscales that mapped to our operational definition. The Primary Care Assessment Survey (PCAS) (Safran et al. 1998) has two. The two-item Visit-Based Continuity subscale elicits how often the "regular doctor" is consulted for routine and sickness care. The Contextual Knowledge subscale contains four items asking patients to rate, on a six-point Likert response scale (1=very poor to 6=excellent), the doctor's knowledge of different dimensions of the patient and a single item on how well their doctor would know their wishes if they were in a coma; this latter item's 1-to-10 response scale is collapsed into a 1-to-6 scoring.

The Primary Care Assessment Tool – Short Form, adult (PCAT-S) (Shi et al. 2001) has a four-item Ongoing Care subscale that elicits the probability, on a four-point Likert response scale (1=definitely not to 4=definitely), that the patient always consults the "primary care provider" and that asks about the provider's knowledge of the patient.

The Components of Primary Care Index (CPCI) (Flocke 1997) has two relational continuity subscales that each use a six-point semantic difference agreement response scale (poles of 1=strongly disagree and 6=strongly agree) on various statements about the "regular doctor." The eight-item Accumulated Knowledge scale assesses the doctor's knowledge of various dimensions of the patient, including one item on the persistence of the relationship. The five-item Preference for Regular Physician subscale assesses the extent to which care is concentrated, by choice, with the regular doctor.

Results

Comparative descriptive statistics

Table 1 summarizes the item content and behaviour of the five subscales; the detailed content and distributions are available online at http://www.longwoods.com/content/22637. No items have more than 5% missing values. Most respondents select positive expressions of relational continuity, especially for items in all three instruments asking whether the same provider is consistently consulted; over 50% select the highest response option. The parametric estimates of the discriminability within the original subscale indicate that all but one item (PCAT-S, seeing same doctor or nurse) discriminate well between different levels of the subscale score ($a>1.0$). Less discriminating items are those that elicit care concentration. This finding likely reflects the different sub-dimensions revealed in factor analysis.

Table 2 presents the descriptive statistics for each subscale. The subscale scores are normalized to a 0-to-10 metric to permit comparison. The scores are skewed towards positive values, with medians higher than means, especially for the PCAS Visit-Based Continuity subscale and the CPCI Preference for Regular Physician. The normalized means for the scales differ substantially from one to another, but the standard deviations are relatively similar. The subscales have adequate reliability.

Table 3 presents the Pearson correlations between the relational continuity subscales. The PCAS Visit-Based Continuity subscale does not correlate well with other continuity subscales (r=.24, .26) and only modestly with the CPCI Preference for Regular Provider subscale

TABLE 1. Summary of relational continuity subscale content and distribution of item responses. (Detailed distribution available at http://www.longwoods.com/content/22637)

Subscale and Item Description	Response Scale	Range Missing Values	Overall Modal Response	Range Item Discriminability	Comments on Distribution
PCAS Visit-Based Continuity (2 items) Rate the frequency of seeing the regular doctor (not an assistant or partner) For a check-up or routine care; when sick	Likert frequency, 1=never to 6=always	2%	6 (always)	1.76 (sick) to 10.43 (check-up)	Over 72% of responses in two most positive categories; ≈1% in most negative
PCAS Contextual Knowledge (5 items) Rate by regular doctor: Of entire medical history; responsibilities at work or home; worries about health; values and beliefs	Likert evaluative, 1=very poor to 6=excellent	1%–2%	4–5 (good/very good)	2.74 (medical history) to 4.57 (worries)	Almost normal distribution
Rate self-knowledge by regular doctor: If unconscious, my doctor would know what I would want done for me	Likert agreement, 1 to 5 scored 1 to 6: strongly disagree to strongly agree	1%	3.5 (not sure)	1.35	
PCAT-S Ongoing Care (4 items) Likelihood of seeing same doctor or nurse each time; ability to phone the doctor or nurse who knows best; known as a person, not only as a medical problem; provider knows what problems are most important for the respondent	Likert evaluative, 1=definitely not to 4=definitely	2%–3% (true missing) 0%–7% not sure	4 (definitely)	0.74 (same provider) to 3.58 (knows most important problems)	Most respondents (33%–70%) select most positive opinion
CPCI Accumulated Knowledge (8 items) Agreement with statements about regular doctor: Positive statements: Knows a lot about my family medical history; clearly understands my health needs; been through a lot together; understands what is important for me regarding my health; takes my beliefs and wishes into account; knows whether or not I exercise, eat right, smoke or drink alcohol; knows a lot about me as a person	Semantic differential opinion, 1=strongly disagree to 6=strongly agree	2%–4%	6 (strongly agree)	1.28 (life habits) to 5.80 (understands my needs)	All items with responses in most positive category, except for the negative statement which was the opposite of that expected
CPCI Patient Preference for Regular Physician (5 items) Agreement with statements about regular doctor: Positive statements: If sick, contact a doctor in this office first; medical care improves when seeing the same doctor; important to see my regular doctor; can call this doctor if not sure to need to see a doctor Negative statement: I rarely see the same doctor when I go for medical care	Semantic differential opinion, 1=strongly disagree to 6=strongly agree	2%–4 %	6 (strongly agree)	1.08 (negative statement) to 2.55 (see my regular doctor)	All items with >31% in most positive category, except for the negative statement which was the opposite of that expected (68% in strongly disagree)

(*r*=.26 to .54). The remaining continuity subscales correlate highly with one another, but also with subscales measuring interpersonal communication, respectfulness and trust. When correlated with the reported percentage of visits to the regular physician, we found no correlation with the PCAS Visit-Based Continuity subscale or the PCAT-S Ongoing Care subscale, and only weak correlations with the PCAS Contextual Knowledge (*r*=.09), CPCI Accumulated Knowledge (*r*=.12) and CPCI Preference for Regular Provider subscales (*r*=.08).

TABLE 2. Mean and distributional values for relational continuity subscales, showing scores normalized to 0 to 10 (n=645)*

Developer's Scale Name (# of items in scale)	Scale Range	Cronbach's Alpha	Mean	SD	Quartiles Q1 (25%)	Q2 (50%)	Q3 (75%)
Normalized Scores							
PCAS Visit-Based Continuity	1 to 10	.67	8.35	2.11	7.00	9.00	10.00
PCAS Contextual Knowledge	1 to 10	.90	5.92	2.28	4.50	6.20	7.60
PCAT Ongoing Care	1 to 10	.73	7.15	2.34	5.80	7.50	9.20
CPCI Accumulated Knowledge	1 to 10	.91	6.99	2.49	5.30	7.50	9.00
CPCI Patient Preference for Regular Physician	1 to 10	.68	7.68	2.01	6.50	8.00	9.50

* Subscale scores calculated as mean of item values and were calculated only for observations where >50% of items were complete.

TABLE 3. Mean partial correlations between relational continuity subscales.* Only correlations significantly different from zero are provided.

Instrument Subscale	PCAS Visit-Based Continuity	PCAS Contextual Knowledge	PCAT Ongoing Care	CPCI Accumulated Knowledge	CPCI Patient Preference for Regular Physician
PCAS: Visit-Based Continuity	1.00		0.24		0.26
PCAS: Contextual Knowledge		1.00	0.65	0.73	0.41
PCAT: Ongoing Care	0.24	0.65	1.00	0.63	0.53
CPCI: Accumulated Knowledge		0.73	0.63	1.00	0.54
CPCI: Patient Preference for Regular Physician	0.26	0.41	0.53	0.54	1.00

* Controlling for language, educational achievement and geographic location.

We had hypothesized a priori, based on item content, that the PCAS Trust and PCAT-S First-Contact Utilization subscales might relate to relational continuity. Despite high correlations with the PCAS Trust subscale, the items did not load with other relational continuity items in common factor analysis. The First-Contact Utilization subscale, which elicits the tendency to consult the primary care provider first, neither correlates highly nor loads with continuity subscales, except for one item: "If I am sick, I would always contact a doctor at this clinic first."

Do all items measure a single attribute?

Our effective sample size for factor analysis was reduced from 645 to 495 by excluding respondents with at least one missing value on any item (listwise missing). Respondents excluded were more likely to be older and to have a chronic health problem. However, because this conservative approach to dealing with missing values can introduce bias, we repeated all the confirmatory analyses using maximum likelihood imputation of missing values (Rubin

1987) to examine the robustness of our conclusions, but the larger sample did not change the direction or essence of these.

Most of the 21 items loaded reasonably well (>.40) onto a single factor using common factor analysis; exceptions were items related to concentration of care (PCAS Visit-Based Continuity, one item of the PCAT-S and two from the CPCI Preference for Regular Provider). Confirmatory factor analysis of a one-dimensional model indicates adequate model fit, with a root mean squared error of approximation (RMSEA) of p=.086 (slightly higher than the .05 level indicating good fit, Model 1) and a normed fit index (NFI) of .98, well above the .90 level indicating good fit.

How do underlying factors fit with operational definition?

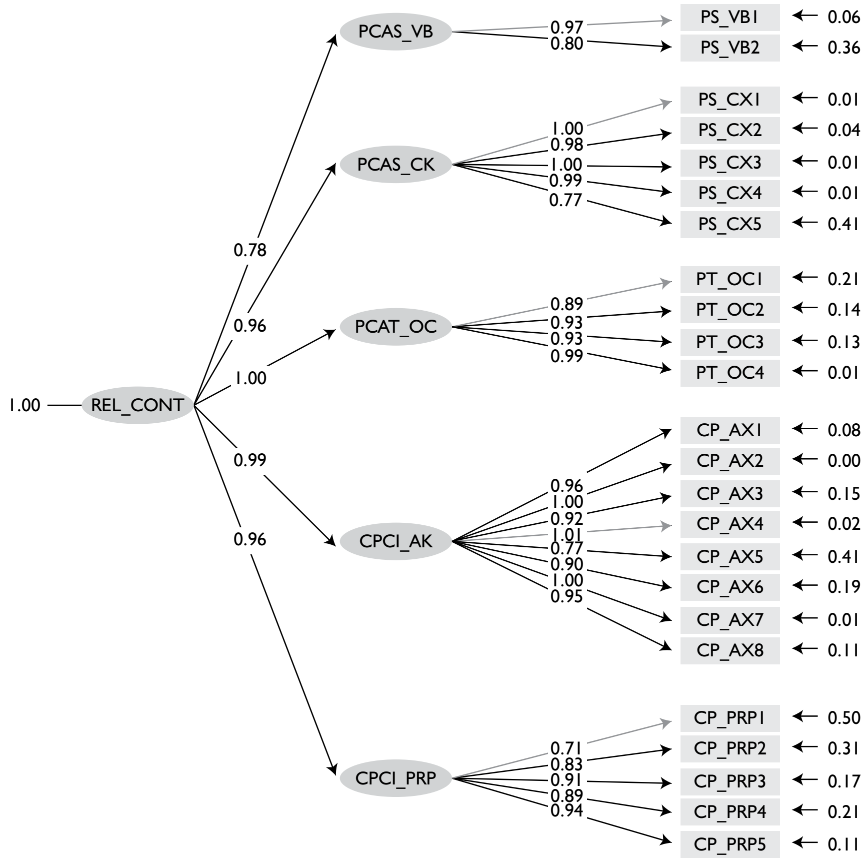
The eigenvalues of exploratory factor analysis suggest a two-factor model. Using our operational definition as a guide, we judged that the first factor (eigenvalue=10.19) captures "accumulated knowledge of the patient," while the second (eigenvalue=1.66), "relationship ... spans various health events," focuses on concentration of care rather than on duration of relationship. Both are specific to the regular provider. No items or factors captured "experienced care being consistent with individual needs."

Some individual items had problematic loadings. For instance, the PCAT-S item, "ability to talk to a known provider," loads weakly on knowledge (.46), and the content appears to relate to access rather than to relational continuity. As a whole, the CPCI Preference for Regular Physician subscale fits within concentration of care, but the first two items do not load on either factor ("contact own provider when sick," loadings .21, .16, and "care improves with concentration of care," loadings .33, .27). These were associated with concentration of care for confirmatory factor analysis. One item in the CPCI Accumulated Knowledge subscale loads very modestly on knowledge (.48), but this finding is most likely due to its reverse wording, "doctor does not know history," rather than to poor conceptual fit.

With confirmatory factor analysis, the best-fitting model items are grouped in their original subscales, which in turn are associated with a single construct, presumed to be relational continuity, as illustrated in Figure 1. The figure shows that the PCAS Visit-Based Continuity subscale has the lowest loading on the latent variable of relational continuity (.78).

The goodness-of-fit statistics for the model in which items are grouped by sub-dimensions of knowledge and concentration of care indicate an improvement over the one-dimensional model (using the difference in chi-square, change $\chi^2 = 1{,}183 - 1{,}047 = 136$, 1 *df*, $p<.001$). The model and the loadings of different items on the scales are presented in Figure 2, where we can see that some items do not have high loadings on the sub-dimension and have a high proportion of residual error (shown to the right of each item). These items may be poorly related to the construct, either because they are not discriminatory or because they relate better to another construct that is not part of the latent variable.

FIGURE 1. Parameter estimations for a structural equation model with original instrument subscales as (first-order) latent variables that relate to an underlying construct (second-order latent variable) presumed to be relational continuity



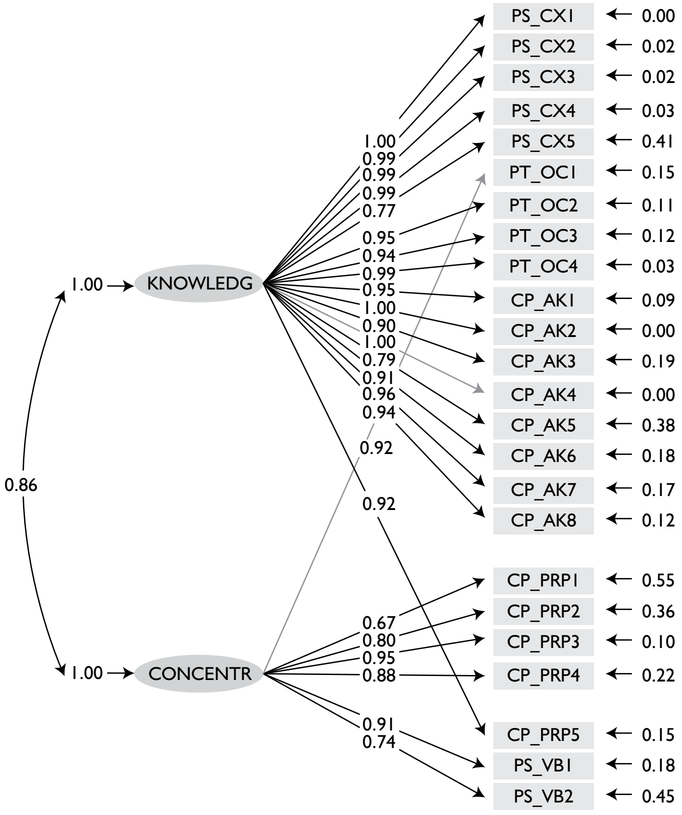
Chi-square = 966.36, *df* = 247, *p*-value = .00000, RMSEA = .077

Individual item performance

We used non-parametric item response analysis to evaluate the performance of individual items as a function of knowledge and concentration of care.

For knowledge, all but one item from the PCAS Contextual Knowledge ("knowledge of responsibilities at work or home") demonstrated good performance. The probability of endorsing each response option is highest in a relatively narrow and unique zone over the range of knowledge, and the zones are clearly ordinal, reflecting the assigned value for each option. The item score varies linearly and strongly with overall knowledge, demonstrating excellent discriminability. The exception was the agree–disagree item about "my doctor knowing what I would want done if I were unconscious or in a coma," which is transformed from a 1-to-5 to a 1-to-6 scale. The values attributed to the response options are not endorsed in an ordinal manner, and the item score does not correlate well with increasing knowledge.

FIGURE 2. Parameter estimations for a structural equation model with items loading on correlated sub-dimensions of relational continuity, accumulated knowledge and concentration of care



Chi-square = 1047.02, *df* = 251, *p*-value = .00000, RMSEA = .080

The CPCI items eliciting knowledge perform well in terms of discriminability. The best-performing items are about the doctor's understanding of health needs and taking beliefs and wishes into account. For the remaining knowledge items, the two response scale extremes (1=strongly disagree, 6=strongly agree) are most likely to be endorsed, and the mid-options are not always distinguishable from one another, suggesting that a three- or four-point scale would be most appropriate. The three knowledge items in the PCAT-S perform relatively well, though the highest response option is endorsed at average knowledge levels, so the response options principally discriminate between below-average knowledge. The value assigned to each response option is largely appropriate, except the option "probably," which appears to be non-specific as it is endorsed across the entire range of knowledge.

All the items were problematic for the dimension of concentration of care, demonstrating similar problems across instruments. The slope between the item score and the expected total concentration of care was quite steep, demonstrating good discriminability between different levels of concentration, but only in the below-average range. The most positive option is overwhelmingly endorsed starting at below-average levels of concentration. The items are discriminating because the response scales all function essentially as a binary choice of the most positive extreme versus all the other options. The result is that all these items are excellent for detecting low concentration of care, but not for distinguishing between average and excellent concentration of care.

Discussion

Our study demonstrates that these validated subscales appear to be measuring a common underlying construct of relational continuity as we conceived it. They capture two sub-dimensions: accumulated provider knowledge of the patient and concentration of care (seeing the same provider over a series of health events). Consistency of care with the individual's needs – arguably the main benefit of relational continuity – is not captured in any of the subscales we examined. It is presumed that knowledge of the patient translates into consistency, but this presumption is not directly verified.

Accumulated provider knowledge is well measured. The high correlations between PCAS Contextual Knowledge and CPCI Accumulated Knowledge, along with the similar factor loadings of the items in these subscales, lends support that each is measuring a similar sub-dimension. The PCAS Contextual Knowledge subscale shows the best capacity to discriminate between different levels of provider knowledge over the entire range of the scale, but the CPCI Accumulated Knowledge subscale and most items in the PCAT-S Ongoing Care subscale also perform well. Overall, our results show that program evaluators can count on these robust measures to monitor whether health reforms have any negative impact on provider knowledge of the patient. Being known is one of the characteristics patients value most (Turner et al. 2007; Cheraghi-Sohi et al. 2008). It is also associated with safer care (Kuzel et al. 2004), including accurate diagnosis and application of wait-and-see techniques (Hjortdahl and Borchgrevink 1991; Hjortdahl 1992). However, participants in our qualitative discussion groups of the questionnaires were divided on expectations of the depth of the physician–provider relationship; several expressed discomfort with the extent of personal knowledge suggested in the instruments, while others defended this as part of patient-centred care (Haggerty, Beaulieu et al. 2011).

Concentration of care is captured best by the CPCI Preference for Regular Provider subscale, and adequately by the PCAS Visit-Based Continuity subscale and one relevant item in the PCAT-S. These subscales are only weakly correlated, suggesting they are measuring different aspects of concentration of care. In addition, weak correlation between concentration subscales and the percentage of visits to the regular provider, as reported by respondents, raises doubts about whether relational continuity should be inferred from utilization data. Some have questioned whether concentration of care is a good proxy for relational continuity (Reid et al. 2002; Rodriguez et al. 2008). A qualitative study suggests that patients do not consider that consulting other providers necessarily diminishes their commitment to or continuity with their own physician (Roberge et al. 2001). In addition, the items function as yes/no responses rather than as the intended ordinal scales, and they are more informative and discriminatory for poor than for good concentration of care. This finding suggests they can be good indicators of discontinuity or care fragmentation, but are less useful for measuring high levels of continuity. Using any of these instruments, then, to capture improvements in the concentration of care may be useful only if the baseline experience is poor. All the analyses suggest more development is needed in both our conceptual understanding of this aspect of relational continuity and how best to measure it.

Earlier in this paper, we raised concerns about relational continuity in team-based care. Rodriguez and colleagues (2007) found that patient assessments of all aspects of care declined when the proportion of visits to their own primary care provider declined. This finding seems particularly important if the teamwork is not visible to patients (Safran 2003). The new version of the PCAS, the Ambulatory Care Evaluation Survey (Safran et al. 2006), includes a subscale on team care, but it appears to measure patients' perception of teamwork rather than the experience of relational continuity with the team per se. Starfield (1998) found that outcomes associated with "site continuity" were considerably less strong than those associated with physician continuity, unless the providers shared a common approach and philosophy of care. This finding suggests that a proxy for team relational continuity might be obtained from providers' perceptions of team cohesiveness. A recently developed measure of team relational continuity may be promising (Haggerty, Roberge et al. 2011).

Study limits

This study has several limitations. First, limiting the study to those having visited a regular provider in the previous 12 months constrains the range of relational continuity, and the resulting positive skewing of responses compromises the capacity to detect underlying factors. However, because our sampling design essentially oversampled for persons with a poor experience of care, we may have introduced greater variance than would be found in the general population. Second, eliminating subjects with missing values not only reduced statistical power but may have biased the final sample. However, sensitivity analysis using imputation of missing values did not alter our overall conclusions. Finally, our operational definition of relational continuity may be different from that of the instrument developers.

Conclusion

Overall, we found that validated subscales perform relatively well for measuring one dimension of relational continuity: accumulated knowledge. We can recommend that this dimension be used to evaluate the impact of reforms on relationships of patients with individual providers. Subscales measuring concentration of care are most useful for capturing fragmentation and discontinuity. The relational continuity dimension of making care consistent with patient needs is not captured in any instrument and may require further development, as would instruments measuring team relational continuity.

ACKNOWLEDGEMENTS

This research was funded by the Canadian Institutes of Health Research. The authors wish to thank Beverley Lawson for conducting the survey in Nova Scotia and Christine Beaulieu in Quebec; Fatima Bouharaoui for performing the analysis, and Donna Riley for support in preparation and editing of the manuscript.

Correspondence may be directed to: Frederick Burge, MD, Dalhousie University, Department of Family Medicine, AJLB 8 QEII HSC, 5909 Veteran's Memorial Lane, Halifax, NS B3H 2E2; tel.: 902-473-4742; fax: 902-473-4760; e-mail: fred.burge@dal.ca.

REFERENCES

Baker, R., G. Freeman, M. Boulton, K. Windridge, C. Tarrant, J. Low et al. 2005. *Continuity of Care: Patients' and Carers' Views and Choices in Their Use of Primary Care Services.* Report no. SDO/13b/2001. London: NHS Service and Delivery Organisation.

Berry, L.L., J.T. Parish, R. Janakiraman, L. Ogburn-Russell, G.R. Couchman, W.L. Rayburn et al. 2008. "Patients' Commitment to Their Primary Physician and Why It Matters." *Annals of Family Medicine* 6: 6–13.

Bertakis, K.D. and E.J. Callahan. 1992. "A Comparison of Initial and Established Patient Encounters Using the Davis Observation Code." *Family Medicine* 24: 307–11.

Burge, F., B. Lawson and G. Johnston. 2003. "Family Physician Continuity of Care and Emergency Department Use in End-of-Life Cancer Care." *Medical Care* 41(8): 992–1001.

Cheraghi-Sohi, S., A.R. Hole, N. Mead, R. McDonald, D. Whalley, P. Bower et al. 2008. "What Patients Want from Primary Care Consultations: A Discrete Choice Experiment to Identify Patients' Priorities." *Annals of Family Medicine* 6: 107–15.

Ettner, S.L. 1996. "The Timing of Preventive Services for Women and Children: The Effect of Having a Usual Source of Care." *American Journal of Public Health* 86: 1748–54.

Ettner, S.L. 1999. "The Relationship between Continuity of Care and the Health Behaviors of Patients: Does Having a Usual Physician Make a Difference?" *Medical Care* 37: 547–55.

Flocke, S. 1997. "Measuring Attributes of Primary Care: Development of a New Instrument." *Journal of Family Practice* 45(1): 64–74.

Flocke, S., K.C. Stange and S.J. Zyzanski. 1998. "The Association of Attributes of Primary Care with the Delivery of Clinical Preventive Services." *Medical Care* 36(8 Suppl.): AS21–AS30.

Haggerty, J.L. 2011. "Measurement of Primary Healthcare Attributes from the Patient Perspective." *Healthcare Policy* 7 (Special Issue): 13–20.

Haggerty, J.L., F. Bouharaoui and D.A. Santor. 2011. "Differential Item Functioning in Primary Healthcare Evaluation Instruments by French/English Version, Educational Level and Urban/Rural Location." *Healthcare Policy* 7 (Special Issue): 47–65.

Haggerty, J.L., F. Burge, J.-F. Lévesque, D. Gass, R. Pineault, M.-D. Beaulieu et al. 2007. "Operational Definitions of Attributes of Primary Health Care: Consensus among Canadian Experts." *Annals of Family Medicine* 5: 336–44.

Haggerty, J.L., R.J. Reid, G.K. Freeman, B.H. Starfield, C.E. Adair and R. McKendry. 2003. "Continuity of Care: A Multidisciplinary Review." *British Medical Journal* 327: 1219–21.

Haggerty, J.L., D. Roberge, G.K. Freeman et al. 2011. "When Patients Encounter Several Providers: Validation of a Generic Measure of Continuity of Care." *Annals of Family Medicine.* (Submitted.)

Haggerty, J.L., C. Beaulieu, B. Lawson, D.A. Santor, M. Fournier and F. Burge. 2011. "What Patients Tell Us about Primary Healthcare Evaluation Instruments: Response Formats, Bad Questions and Missing Pieces." *Healthcare Policy* 7 (Special Issue): 66–78.

Hjortdahl, P. 1992. "The Influence of General Practitioners' Knowledge about Their Patients on the Clinical Decision-Making Process." *Scandinavian Journal of Primary Health Care* 10: 290–94.

Hjortdahl, P. and C.F. Borchgrevink. 1991. "Continuity of Care: Influence of General Practitioners' Knowledge about Their Patients on Use of Resources in Consultations." *British Medical Journal* 303: 1181–84.

Kuzel, A.J., S.H. Woolf, V.J. Gilchrist, J.D. Engel, T.A. LaVeist, C. Vincent et al. 2004. "Patient Reports of Preventable Problems and Harms in Primary Health Care." *Annals of Family Medicine* 2: 333–40.

Lévesque, J.-F., J.L. Haggerty, F. Burge, M.-D. Beaulieu, D. Gass, R. Pineault et al. 2011. "Canadian Experts' Views on the Importance of Attributes within Professional and Community-Oriented Primary Healthcare Organizational Models." *Healthcare Policy* 7 (Special Issue): 21–30.

Mainous, A.G. 3rd, R. Baker, M.M. Love, D.P. Gray and J.M. Gill. 2001. "Continuity of Care and Trust in One's Physician: Evidence from Primary Care in the United States and the United Kingdom." *Family Medicine* 33: 22–27.

McWhinney, I.R. 1998. "Primary Care: Core Values. Core Values in a Changing World." *British Medical Journal* 316: 1807–09.

O'Malley, A.S. and C.B. Forrest. 1996. "Continuity of Care and Delivery of Ambulatory Services to Children in Community Health Clinics." *Journal of Community Health* 21(3): 159–73.

O'Malley, A.S. 1997. "Continuity of Care and the Use of Breast and Cervical Cancer Screening Services in a Multiethnic Community." *Archives of Internal Medicine* 157: 1462–70.

Reid, R., J. Haggerty and R. McKendry. 2002. *Defusing the Confusion: Concepts and Measures of Continuity of Care.* Ottawa: Canadian Health Services Research Foundation.

Roberge, D., M.-D. Beaulieu, S. Haddad, R. Lebeau and R. Pineault. 2001. "Loyalty to the Regular Care Provider: Patients' and Physicians' Views." *Family Practice* 18: 53–59.

Rodriguez, H.P., R.E. Marshall, W.H. Rogers and D.G. Safran. 2008. "Primary Care Physician Visit Continuity: A Comparison of Patient-Reported and Administratively Derived Measures." *Journal of General Internal Medicine* 23(9): 1499–1502.

Rodriguez, H.P., W.H. Rogers, R.E. Marshall and D.G. Safran. 2007. "Multidisciplinary Primary Care Teams: Effects on the Quality of Clinician–Patient Interactions and Organizational Features of Care." *Medical Care* 45: 19–27.

Rogers, J. and P. Curtis. 1980. "The Concept and Measurement of Continuity in Primary Care." *American Journal of Public Health* 70: 122–27.

Rubin D.B. 1987. "Multiple Imputation for Nonresponse in Surveys." John Wiley.

Safran, D.G. 2003. "Defining the Future of Primary Care: What Can We Learn from Patients?" *Annals of Internal Medicine* 138: 248–55.

Safran, D.G., M. Karp, K. Coltin, H. Chang, A. Li, J. Ogren et al. 2006. "Measuring Patients' Experiences with Individual Primary Care Physicians: Results of a Statewide Demonstration Project." *Journal of General Internal Medicine* 21: 13–21.

Safran, D.G., M. Kosinski, A.R. Tarlov, W.H. Rogers, D.H. Taira, N. Lieberman et al. 1998. "The Primary Care Assessment Survey: Tests of Data Quality and Measurement Performance." *Medical Care* 36: 728–39.

Santor, D.A., J.L. Haggerty, J.-F. Lévesque, F. Burge, M.-D. Beaulieu, D. Gass and R. Pineault. 2011. "An Overview of Confirmatory Factor Analysis and Item Response Analysis Applied to Instruments to Evaluate Primary Healthcare." *Healthcare Policy* 7 (Special Issue): 79–92.

Shi, L., B. Starfield and J. Xu. 2001. "Validating the Adult Primary Care Assessment Tool." *Journal of Family Practice* 50(2): n161w–n171w.

Smith, C.S. 1995. "The Impact of an Ambulatory Firm System on Quality and Continuity of Care." *Medical Care* 33: 221–26.

Starfield, B. 1998. "Patient Lists and Person-Focused Care Over Time." In B. Starfield, *Primary Care: Balancing Health Needs, Services, and Technology.* New York: Oxford University Press.

Steinwachs, D.M. 1979. "Measuring Provider Continuity in Ambulatory Care: An Assessment of Alternative Approaches." *Medical Care* 17(6): 551–65.

Turner, D., C. Tarrant, K. Windridge, S. Bryan, M. Boulton, G. Freeman et al. 2007. "Do Patients Value Continuity of Care in General Practice? An Investigation Using Stated Preference Discrete Choice Experiments." *Journal of Health Services Research and Policy* 12: 132–37.

Wasson, J.H., A.E. Sauvigne, P. Mogielnicki, W.G. Frey, C.H. Sox, C. Gaudette et al. 1984. "Continuity of Outpatient Medical Care in Elderly Men: A Randomized Trial." *Journal of the American Medical Association* 252(17): 2413–17.

Weiss, L.J. and J. Blustein. 1996. "Faithful Patients: The Effect of Long-Term Physician–Patient Relationships on the Costs and Use of Health Care by Older Americans." *American Journal of Public Health* 86: 1742–47.

Relational Continuity from the Patient Perspective: Comparison of Primary Healthcare Evaluation Instruments

La continuité relationnelle du point de vue du patient : comparaison entre instruments d'évaluation des soins de santé primaires

FREDERICK BURGE, JEANNIE L. HAGGERTY, RAYNALD PINEAULT, MARIE-DOMINIQUE BEAULIEU,
JEAN-FRÉDÉRIC LÉVESQUE, CHRISTINE BEAULIEU AND DARCY A. SANTOR

TABLE 1. Distribution of responses for each item in subscales measuring relational continuity in primary healthcare services and discriminatory capacity of each item within its parent subscale. Modal response is shown in bold (n=645).

Item Code	Instrument: Subscale Statement	Missing Values^{1} % (n)	Percentage (Number) by Response Option						Item Discrimination^{2}
	Primary Care Assessment Survey (PCAS): Visit-Based Continuity		**1=Never**	**2=Almost Never**	**3=Some of the time**	**4=A lot of the time**	**5=Almost always**	**6=Always**	
PS_vb1	When you go for a **check-up or routine care**, how often do you see your **regular doctor** (not an assistant or partner)?	2 (10)	1 (7)	2 (12)	7 (44)	5 (29)	17 (111)	**67 (432)**	10.43 (1.36)
PS_vb2	When you are **sick** and go to the doctor, how often do you see your **regular doctor** (not an assistant or partner)?	2 (13)	2 (16)	5 (33)	9 (61)	8 (53)	25 (164)	**47 (305)**	1.76 (0.15)
	Primary Care Assessment Survey (PCAS): Contextual Knowledge		**1=Very poor**	**2=Poor**	**3=Fair**	**4=Good**	**5=Very Good**	**6=Excellent**	
	Thinking about how well your doctor **knows you…**								
PS_ck1	how would you rate doctor's knowledge of your **entire medical history**?	1 (9)	2 (14)	5 (29)	16 (105)	24 (152)	**32 (207)**	20 (129)	2.74 (0.17)
PS_ck2	how would you rate doctor's knowledge of your **responsibilities at work or home**?	2 (15)	5 (31)	10 (63)	17 (108)	26 (166)	**27 (174)**	14 (88)	3.85 (0.21)
PS_ck3	how would you rate doctor's knowledge of what **worries** you most about your health?	2 (10)	3 (21)	7 (43)	21 (133)	**29 (188)**	25 (164)	13 (86)	4.57 (0.26)
PS_ck4	how would you rate doctor's knowledge of **you as a person** (your values and beliefs)?	2 (12)	7 (42)	12 (77)	19 (125)	**24 (157)**	24 (152)	12 (80)	4.35 (0.26)
			1=Strongly disagree	**2.25**	**3.5**	**4.75**	**6=Strongly agree**		
PS_ck5	If I were unconscious or in a coma, my doctor would know what I would want done for me.^{3}	1 (6)	15 (97)	16 (101)	**39 (254)**	19 (121)	10 (66)		1.35 (0.12)
	Primary Care Assessment Tool (PCAT-S): Ongoing Care		**1=Definitely not**	**2=Probably not**	**3= Probably**	**4=Definitely**	**Not sure / Don't remember**		
PT_oc1	When you go to your Primary Care Provider, are you taken care of by the **same** doctor or nurse each time?	2 (10)	4 (23)	7 (47)	17 (108)	**70 (454)**	0 (3)		0.74 (0.12)
PT_oc2	If you have a question, can you call and talk to the doctor or nurse who knows you best?	2 (13)	12 (78)	15 (94)	32 (205)	**33 (213)**	7 (42)		1.46 (0.13)
PT_oc3	Does your Primary Care Provider know you very well as **a person**, rather than as someone with a medical problem?	2 (13)	12 (76)	17 (110)	28 (182)	**38 (247)**	3 (17)		2.97 (0.22)
PT_oc4	Does your Primary Care Provider know what problems are most important to you?	3 (17)	7 (46)	14 (89)	37 (239)	**38 (244)**	2 (10)		3.58 (0.28)

TABLE 1. Continued

Item Code	Instrument: Subscale Statement	Missing Values[1] % (n)	Percentage (Number) by Response Option						Item Discrimination[2]
	Components of Primary Care Index (CPCI): Accumulated Knowledge		**1=Strongly disagree**	**2**	**3**	**4**	**5**	**6=Strongly agree**	
CP_ak1	This doctor knows a lot about my family medical history.	2 (11)	8 (52)	8 (50)	8 (53)	14 (89)	19 (120)	**42 (270)**	2.54 (0.18)
CP_ak2	This doctor clearly understands my health needs.	2 (12)	3 (21)	5 (30)	8 (53)	13 (87)	23 (151)	**45 (291)**	5.80 (0.41)
CP_ak3	This doctor and I have been through a lot together.	3 (19)	16 (105)	13 (82)	11 (69)	13 (86)	17 (111)	**27 (173)**	2.18 (0.16)
CP_ak4	This doctor understands what is important to me regarding my health.	2 (14)	4 (25)	7 (43)	11 (69)	14 (93)	25 (163)	**37 (238)**	4.95 (0.31)
CP_ak5	This doctor does not know my medical history very well. (reverse coded)	2 (16)	**46 (295)**	15 (94)	9 (58)	9 (56)	9 (59)	10 (65)	2.22 (0.17)
CP_ak6	This doctor always takes my beliefs and wishes into account in caring for me.	2 (16)	4 (23)	5 (33)	11 (68)	16 (100)	25 (159)	**38 (246)**	2.35 (0.17)
CP_ak7	This doctor knows whether or not I exercise, eat right, smoke or drink alcohol.	2 (15)	5 (30)	6 (36)	6 (38)	12 (77)	24 (155)	**46 (294)**	1.28 (0.13)
CP_ak8	This doctor knows a lot about me as a person (such as my hobbies, job, etc.).	2 (15)	14 (88)	11 (71)	13 (87)	16 (100)	19 (124)	**25 (160)**	2.80 (0.18)
	Components of Primary Care Index (CPCI): Patient Preference for Regular Physician		**1=Strongly disagree**	**2**	**3**	**4**	**5**	**6=Strongly agree**	
CP_prp1	If I am sick, I would always contact a doctor in this office first.	3 (22)	11 (68)	4 (25)	4 (26)	6 (39)	20 (128)	**52 (337)**	1.31 (0.14)
CP_prp2	My medical care improves when I see the same doctor that I have seen before.	4 (26)	5 (33)	6 (38)	7 (43)	14 (91)	20 (132)	**44 (282)**	2.11 (0.17)
CP_prp3	It is very important to me to see my regular doctor.	2 (13)	1 (9)	4 (24)	4 (24)	7 (45)	18 (114)	**65 (416)**	2.55 (0.22)
CP_prp4	I rarely see the same doctor when I go for medical care. (reverse coded)	4 (23)	**68 (438)**	11 (72)	5 (30)	3 (22)	3 (19)	6 (41)	1.08 (0.15)
CP_prp5	I can call this doctor if I have a concern and am not sure I need to see a doctor.	3 (20)	12 (79)	12 (80)	13 (81)	13 (83)	16 (104)	**31 (198)**	1.22 (0.12)

[1] Missing values: No response given to the item.

[2] Discriminatory parameter <1.0 indicates that this item does not discriminate well between individuals with low and high values of the subscale score.

[3] Item expressed as agreement on a 1-to-10 scale; distribution shows transformation to scoring suggested by the developer.

RESEARCH PAPER

Management Continuity from the Patient Perspective: Comparison of Primary Healthcare Evaluation Instruments

La continuité d'approche du point de vue du patient : comparaison entre instruments d'évaluation des soins de santé primaires

JEANNIE L. HAGGERTY, PHD
Department of Family Medicine, McGill University
Montreal, QC

FREDERICK BURGE, MD, MSC
Department of Family Medicine, Dalhousie University
Halifax, NS

RAYNALD PINEAULT, MD, PHD
Centre de recherche du Centre hospitalier de l'Université de Montréal
Montréal, QC

MARIE-DOMINIQUE BEAULIEU, MD, MSC
Chaire Dr Sadok Besrour en médecine familiale
Centre de recherche du Centre hospitalier de l'Université de Montréal
Montréal, QC

FATIMA BOUHARAOUI, MSC
St. Mary's Research Centre, St. Mary's Hospital Center
Montreal, QC

CHRISTINE BEAULIEU, MSC
St. Mary's Research Centre, St. Mary's Hospital Center
Montreal, QC

DARCY A. SANTOR, PHD
School of Psychology, University of Ottawa
Ottawa, ON

JEAN-FRÉDÉRIC LÉVESQUE, MD, PHD
Centre de recherche du Centre hospitalier de l'Université de Montréal
Montréal, QC

Abstract

Management continuity, operationally defined as "the extent to which services delivered by different providers are timely and complementary such that care is experienced as connected and coherent," is a core attribute of primary healthcare. Continuity, as experienced by the patient, is the result of good care coordination or integration.

Objective: To provide insight into how well management continuity is measured in validated coordination or integration subscales of primary healthcare instruments.

Method: Relevant subscales from the Primary Care Assessment Survey (PCAS), the Primary Care Assessment Tool – Short Form (PCAT-S), the Components of Primary Care Instrument (CPCI) and the Veterans Affairs National Outpatient Customer Satisfaction Survey (VANOCSS) were administered to 432 adult respondents who had at least one healthcare contact with a provider other than their family physician in the previous 12 months. Subscales were examined descriptively, by correlation and factor analysis and item response theory analysis. Because the VANOCSS elicits coordination problems and is scored dichotomously, we used logistic regression to examine how evaluative subscales relate to reported problems.

Results: Most responses to the PCAS, PCAT-S and CPCI subscales were positive, yet 83% of respondents reported having one or more problems on the VANOCSS Overall Coordination subscale and 41% on the VANOCSS Specialist Access subscale. Exploratory factor analysis suggests two distinct factors. The first (eigenvalue=6.98) is coordination actions by the primary care physician in transitioning patient care to other providers (PCAS Integration subscale and most of the PCAT-S Coordination subscale). The second (eigenvalue=1.20) is efforts by the primary care physician to create coherence between different visits both within and outside the regular doctor's office (CPCI Coordination subscale). The PCAS Integration subscale was most strongly associated with lower likelihood of problems reported on the VANOCSS subscales.

Conclusion: Ratings of management continuity correspond only modestly to reporting of coordination problems, possibly because they rate only the primary care physician, whereas patients experience problems across the entire system. The subscales were developed as measures of integration and provider coordination and do not capture the patient's experience of connectedness and coherence.

Résumé

La définition opérationnelle de la continuité d'approche est « l'étendue selon laquelle les divers fournisseurs offrent des services complémentaires et opportuns de telle manière qu'ils sont vécus par le patient de façon cohérente et liés entre eux ». Il s'agit d'une caractéristique fondamentale des soins de santé primaires. La continuité, telle que vécue par le patient, est le résul-

tat d'une bonne coordination ou d'une bonne intégration des soins.
Objectif : Voir à quel point la continuité d'approche est mesurée par les sous-échelles validées de coordination et d'intégration des instruments d'évaluation des soins de santé primaires.
Méthode : Les sous-échelles pertinentes du Primary Care Assessment Survey (PCAS), du Primary Care Assessment Tool – version courte (PCAT-S), du Components of Primary Care Instrument (CPCI) et du Veterans Affairs National Outpatient Customer Satisfaction Survey (VANOCSS) ont été administrées à 432 adultes qui avaient eu au moins un contact avec un fournisseur de services de santé autre que son médecin de famille au cours des 12 mois antérieurs. Les sous-échelles ont été examinées de façon descriptive, en effectuant des analyses de corrélation, des analyses factorielles et des analyses de réponse par item. Étant donné que le VANOCSS fait ressortir des problèmes de coordination par des résultats dichotomiques, nous avons employé la régression logistique pour examiner comment les sous-échelles évaluatives sont reliées aux problèmes rapportés.
Résultats : La plupart des réponses aux sous-échelles du PCAS, du PCAT-S et du CPCI sont positives, mais 83 % des répondants ont indiqué (avec la sous-échelle "coordination générale" du VANOCSS) avoir eu un ou plusieurs problèmes et 41 % avec la sous-échelle « accès au spécialiste ». L'analyse factorielle exploratoire suggère deux facteurs distincts. Le premier (valeur propre = 6,98) est la coordination des actions par le médecin de première ligne dans l'aiguillage du patient vers d'autres types de fournisseurs (sous-échelle « intégration » du PCAS et la plupart des sous-échelles « coordination » du PCAT-S). Le second (valeur propre = 1,20) sont les efforts faits par le médecin de première ligne pour assurer une cohérence entre les visites tant dans le bureau du médecin régulier qu'en dehors de celui-ci (sous-échelle « coordination » du CPCI). La sous-échelle « intégration » du PCAS est plus fortement associée à une faible probabilité de problèmes rapportés dans les sous-échelles du VANOCSS.
Conclusion : Les scores de la continuité d'approche ne correspondent que modestement aux problèmes indiqués en matière de coordination, possiblement parce qu'ils servent uniquement à évaluer le médecin de première ligne, alors que les patients connaissent des problèmes dans l'ensemble du système. Les sous-échelles ont été conçues comme mesure de l'intégration et de la coordination du fournisseur; elles ne captent pas le degré de connexion et de cohérence vécu par le patient.

* * *

CONTINUITY OF CARE, OFTEN INVOKED AS A CHARACTERISTIC OF GOOD CARE DELIVERY, has different meanings in different healthcare disciplines, but all recognize three dimensions: relational, informational and management continuity (Haggerty et al. 2003).

Background

Conceptualizing management continuity

Management continuity as a characteristic of care was proposed in 2001 in an endeavour to clarify and harmonize the different meanings of continuity of care (Reid et al. 2002). Our consensus process with primary healthcare (PHC) experts across Canada (Haggerty et al. 2007) defined management continuity as "the extent to which services delivered by different providers are timely and complementary such that care is experienced as connected and coherent." There was unanimous agreement that, while management continuity is not specific to PHC, it is an essential PHC attribute. In disease management and nursing care, continuity refers to the linking of care provided by different providers, a notion recognized in family medicine and general practice as coordination. However, coordination refers to exchanges and collaboration between providers – most of which activity is invisible to patients. We propose that patients experience coordination as management continuity (Reid et al. 2002; Shortell et al. 1996).

PHC reform has targeted increased service integration and multidisciplinary coordination. Program evaluators therefore require good information to inform their selection of tools that measure management continuity.

Evaluating management continuity

Management continuity can be evaluated from the patient or provider perspective. The reference to services being "experienced as connected and coherent" clearly pertains to the patient's perspective, while providers may be best placed to assess timeliness and complementarity of services.

Various validated instruments that evaluate PHC from the user's perspective contain subscales addressing coordination of care, which we mapped to management continuity. From the candidate instruments, we selected six that are in the public domain and that we believe to be most relevant for Canada. In this paper, we compare the equivalence of management continuity subscales from four instruments and determine whether they appear to be measuring the same construct. Where analysis suggested more than one factor or construct, we also examine how items capture different elements of the operational definition. Finally, we examine how well individual items perform in measuring the constructs that emerge across instruments. This analysis provides insight into how well different subscales fit the construct of management continuity according to our operational definition, and to provide guidance to evaluators in selecting appropriate measures.

Method

The method is described in detail elsewhere in this special issue of the journal (Haggerty, Burge et al. 2011). Briefly: six instruments that evaluate PHC from the patient's perspective were administered to 645 healthcare users balanced by English/French language, rural/urban location, low/high level of education and poor/average/excellent overall PHC experience.

Measure description

There were five relevant subscales in four instruments. They are described briefly here in the order in which they appeared in the questionnaire. Note the slight differences in reference points and eligible respondents.

The Primary Care Assessment Survey (PCAS) (Safran et al. 1998) has a six-item Integration subscale that asks patients to rate on a six-point Likert scale ("1=very poor" to "6=excellent") different aspects of "times their doctor recommended they see a different doctor for a specific health problem"; they are to be answered only by those whose "doctor *ever* recommended … a different doctor for a specific health problem" (emphasis ours).

The Primary Care Assessment Tool – Short Form, adult (PCAT-S) (Shi et al. 2001) has a four-item Coordination subscale that elicits responses on a four-point Likert scale ("definitely not" to "definitely") to questions about the primary care provider's coordination behaviours; the subscale was completed only by those who "*ever* had a visit to any kind of specialist or special service."

The Components of Primary Care Index (CPCI) (Flocke 1997) has an eight-item Coordination of Care subscale that uses a six-point semantic differential response scale (poles of "1=strongly disagree" and "6=strongly agree") to elicit agreement with various statements about the "regular doctor"; it was completed by all respondents whether or not they had seen other providers.

Finally, the Veterans Affairs National Outpatient Customer Satisfaction Survey (VANOCSS) (Borowsky et al. 2002), informed by the Picker Institute health surveys (Gerteis et al. 1993), has two relevant subscales. Both elicit the frequency of difficulties encountered using Likert-type response options, but apply a dichotomous scoring. The six-item Overall Coordination subscale pertains to "all the healthcare providers your regular doctor has recommended you see" and was answered only by respondents who had seen more than one provider *in the last 12 months*. The four-item Specialty Access subscale refers to access to and care from specialists; we indicated that it should be answered only by those who saw a specialist *in the last 12 months*.

Because the provider reference and eligible respondents were not the same for each subscale, we defined a common subgroup on which to analyze instruments: respondents who had seen more than one provider in the previous 12 months. Additionally, these instruments had two different measurement approaches. The PCAS, PCAT-S and CPCI are evaluative subscales based on the classic approach of indicator items informed by a not-directly observed (latent) variable. Operationally, the subscales are expressed as a continuous score by averaging the values of individual items so that high scores indicate the best management continuity. In contrast, the VANOCSS reports subscales that use a benchmarking approach, where the implicit performance target is to minimize the proportion of patients that experience any problems in healthcare delivery. The items are scored dichotomously, and the subscale is represented as either dichotomous (presence of any problem) or summed to the number of problems encountered (e.g., 0 to 6 problems with coordination). A score of 0 (zero) represents the best management continuity.

Analytic strategy

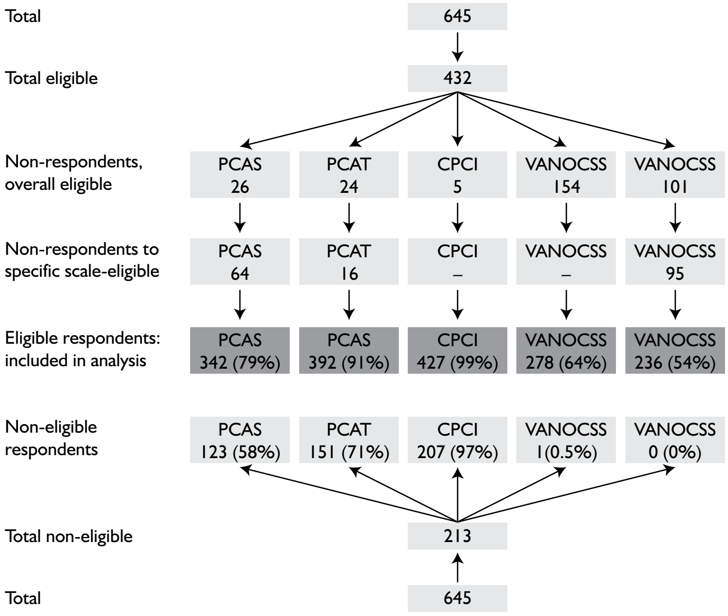
We largely applied the same analytic strategy that we used for all the PHC attributes (Santor et al. 2011), but the different scoring approach of the VANOCSS required some modification. For correlations and factor analysis, we treated VANOCSS subscales each as a single item representing the sum of reported problems. We also explored factor analysis leaving items in the original frequency options, but found that the recommended scoring performed better.

Finally, to account for the reporting versus evaluative approaches to measurement, we used binary logistic regression modelling to examine whether, in separate models, scores of the PCAS, PCAT-S and CPCI subscales predicted the presence of any reported problems with the VANOCSS Overall Coordination and Specialist Access subscales. We also explored the associations using ordinal regression models (number of problems) to determine whether there were ordinal effects.

Results

Of the 645 respondents to the overall survey, 432 reported having seen a provider other than their family physician, constituting our analytic subgroup for management continuity. As shown in Figure 1, however, some of our analytic subgroup were not eligible for a specific subscale (e.g., did not see another doctor or specialist) and, appropriately, did not respond. Others, though eligible, either did not respond (VANOCSS) or were excluded from subscale analysis because of missing values (n=179/432, 41.4%). Those excluded were more likely

FIGURE 1. Flow chart of responses to various subscales (shown in order of appearance in questionnaire)*



* Patterns of non-response in those presumed to be eligible by virtue of reporting utilization of more than one provider (n=432) and patterns of unnecessary response in those not presumed to be eligible (n=213) are demonstrated.

to have a regular clinic rather than a specific physician and appeared to be in slightly better health. The low number of responses to the VANOCSS Overall Coordination and Specialist Access subscales is likely due to respondent fatigue, as these were placed last in the questionnaire. In the analytic subgroup, 66.2% reported having a chronic health problem and 39.9% a disability limiting daily activities. Respondents had an average of 7.6 (SD=7.9) primary care visits in the last 12 months; 74% had seen a specialist.

Comparative descriptive statistics

Table 1 summarizes the item content and behaviour in the five subscales; the detailed content and distributions are available online at *http://www.longwoods.com/content/22638*. None has more than 5% missing values, though items asking about the doctor's communication with specialists (PCAS, CPCI) or involvement in care given by others (PCAS, CPCI) were at the limit of acceptable rates of missing values. Respondents endorse mostly positive assessments of management continuity, except with the VANOCSS items, where up to 59% report a problem in Overall Coordination items, and approximately 20% in Access to Specialists items. All items in the PCAS, PCAT-S and VANOCSS Overall Coordination subscales, and the majority in the other subscales, discriminate adequately between different levels of the subscale score, as indicated by discriminatory parameter $a>1$.

TABLE 1. Summary of management continuity subscale content and distribution of item responses. (Detailed distribution available at http://www.longwoods.com/content/22638).

Subscale and Item Description	Response Scale	Range Missing Values	Overall Modal Response	Range Item Discriminability	Comments on Distribution
PCAS Integration (6 items) When your doctor recommends a different doctor for a specific problem, Rate: Help deciding who to see; help in getting an appointment; involvement of doctor while being treated by a specialist; communication with specialists; help understanding what other said; quality of specialist.	Likert evaluative, 1=very poor to 6=excellent	1%–4%	4–5 (good to very good)	1.85 (quality) to 4.89 (communication)	Only 5%–11% in the two most negative categories
PCAT-S Coordination (4 items) Likelihood that primary care provider: Discussed alternatives for places to seek care; helped make the appointment; wrote information about the reason for the visit; talked about what happened at the visit.	Likert evaluative, 1=definitely not to 4=definitely	2%–3% (true missing) 3%–6% not sure	4 (definitely)	1.23 (alternatives) to 2.33 (information)	Over 52% of responses in the most positive category. Few missing values (True or Not sure)
CPCI Coordination of Care (8 items) Agreement with statements about regular doctor: Positive statements – This doctor: knows when I'm due for a check-up; coordinates all care; keeps track; follows up on a problem; follows up on visits to other providers; helps interpret tests or visits Negative statement – This doctor: does not always know about care received at other places	Semantic differential opinion, 1=strongly disagree, 6=strongly agree	1%– % (true missing) 0%–4% not sure	6 (strongly agree)	0.32 (negative statement) to 4.14 (keeping track)	Most responses (22%–62% in the most positive category; "help interpret tests" seems to have U distribution

TABLE 1. Continued

Subscale and Item Description	Response Scale	Range Missing Values	Overall Modal Response	Range Item Discriminability	Comments on Distribution
VANOCSS Overall Coordination of Care (6 items) Frequency of different providers: Being familiar with recent medical history; not knowing about tests or their results; not knowing about changes in your treatment Frequency of patient: Being confused because of different information; knowing next steps; knowing who to ask for questions about care	3–4 point frequency categories scored as problem/no problem	0%–3%	n/a	1.44 (next steps) to 2.81 (tests)	Two items with 59% presence of a problem (being familiar with recent medical history and next steps). Others had 26%–40% presence of a problem
VANOCSS Access to Specialists (4 items) Frequency of issues getting care from specialists: Access when needed; difficulty with getting an appointment; given information about who to see and why; specialists had information needed from medical record	3–4 point frequency categories scored as problem/no problem	1%–2%		0.42 (appointment) to 2.68 (information needed)	11%–20% reporting presence of a problem

Table 2 presents the descriptive statistics for each subscale, with PCAS, PCAT-S and CPCI subscale mean scores normalized to a 0-to-10 metric to permit comparison. As expected, those three subscales are skewed towards positive assessments with the PCAT-S, demonstrating the most extreme skewing. In contrast, 83% of respondents reported at least one problem on the VANOCSS Overall Coordination subscale and 41% on Specialist Access. Among the respondents in the top quartile of management continuity according to the PCAS, PCAT-S and CPCI subscales, 61%, 76% and 74%, respectively, report having one or more problems on the VANOCSS Overall Coordination subscale and 34%, 37% and 44%, respectively, on the VANOCSS Specialist Access subscale.

TABLE 2. Mean and distributional scores for management continuity subscales, showing normalized mean scores and number of problems (n=432)

Scale Range	Cronbach's Alpha	Mean	SD	Minimum Observed	Quartiles Q1 (25%)	Q2 (50%)	Q3 (75%)
Normalized Mean Scores							
PCAS Integration normalized	.90	6.99	1.97	0	5.67	7.33	8.33
PCAT-S Coordination normalized	.73	7.61	2.52	0	6.67	8.33	10.00
CPCI Coordination of Care normalized	.79	6.65	2.04	0	5.25	6.75	8.00
Number of problems							
VANOCSS Coordination of Care (Overall): Number of problems (0 to 6)	.74	2.50	1.90	0	1.00	2.00	4.00
VANOCSS Specialty Provider Access: Number of problems (0 to 4)	.54	0.60	0.90	0	0	0	1.00

Table 3 presents the Pearson correlations between the management continuity subscales. The PCAS, PCAT-S and CPCI subscales correlate highly with one another (r=.54 to .62) and negatively, but only modestly, with the number of problems reported on the VANOCSS subscales (r=−.19 to −.39). The five subscales correlate strongly with other subscales for relational continuity (mean, .40; range, .28 to .68) and interpersonal communication (mean, .42; range, .30 to .63). The VANOCSS subscales correlated weakly (and negatively) with all other subscales, underlining the difference in measurement approach.

TABLE 3. Partial Pearson correlation coefficients between subscales for management continuity and subscales evaluating other attributes of primary healthcare.* Only statistically significant correlations are shown (at p<.05).

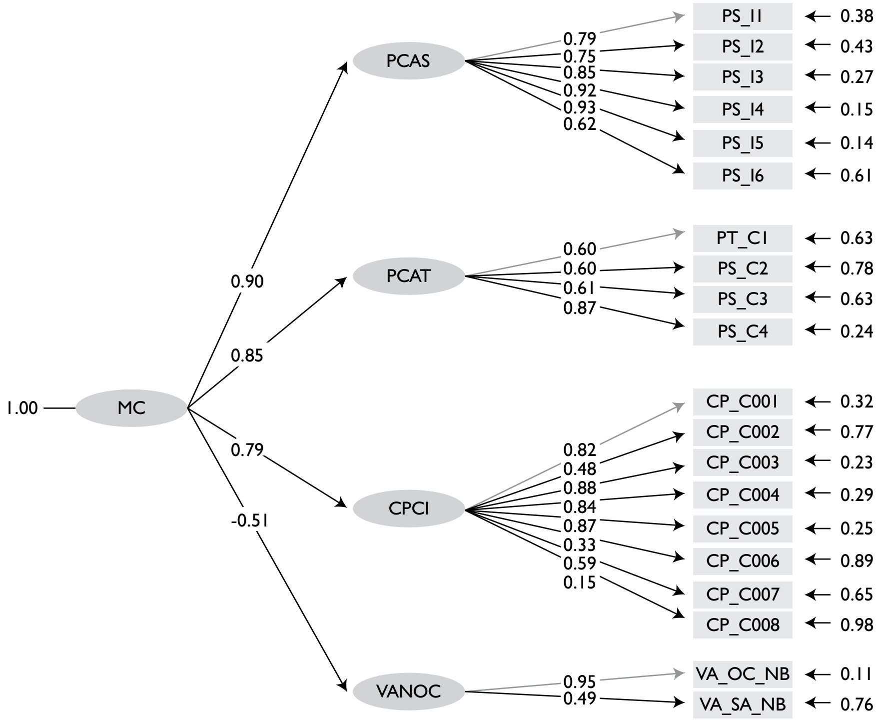
	n=246			n=132	
	PCAS: Integration	**PCAT-S: Coordination**	**CPCI: Coordination of Care**	**VANOCSS: Coordination of Care (Overall), number of problems**	**VANOCSS: Specialty Access, number of problems**
PCAS Integration	1.00	0.55	0.62	–0.39	–0.20
PCAT-S Coordination	0.55	1.00	0.54	–0.19	–0.22
CPCI Coordination of care	0.62	0.54	1.00	–0.34	–0.27
VANOCSS Coordination of Care (Overall), number of problems	–0.39	–0.39	–0.39	1.00	0.34
VANOCSS Specialty Access, number of problems	–0.20	–0.20	–0.20	0.34	1.00

* Controlling for language, education and geographic location.

Do all items measure a single attribute?

In exploratory factor analysis, most items loaded well (>.4) on a single factor. The VANOCSS subscales were added as single items summing the number of problems encountered, and had negative loadings. Items with low factor loadings were those with low discriminability values (help interpreting tests, doctor knows about other care in CPCI; difficulty making appointment with specialists in VANOCSS Specialty Access). However, in confirmatory factor analysis the goodness of fit of the one-dimensional model was barely adequate with a root mean squared error of approximation (RMSEA p=.128), considerably higher than the p=.05 criterion for good model fit, but a comparative fit index (CFI) of .92, which is above the .90 criterion. When subscale items were associated with their parent subscales and were in turn associated with a latent variable – presumed to be continuity (Figure 2) – fit statistics improved significantly over the one-dimensional model: chi-square difference = 1,042.85 – 628.16 = 414.69, *df* 4, p<.001; RMSEA=.078; CFI=.97. This finding suggests that items reflect a common underlying construct for subscales in all instruments.

FIGURE 2. Parameter estimations for structural equation model showing loadings of items on parent scales (first-order variables), which in turn load on management continuity (second-order latent variable)



Chi-square = 628.16, *df* =166, *p*-value = .00000, RMSEA = 0.078

Note: There is modest improvement compared to the unidimensional model (difference in chi-square of the models χ^2 = 1,042.85 – 628.16 = 414.69, *df* 4, p<.001) but persisting high levels of residual error (shown to the right of the items), especially for the PCAT-S and CPCI subscales. Root mean squared error of approximation (RMSEA) does not support good model fit.

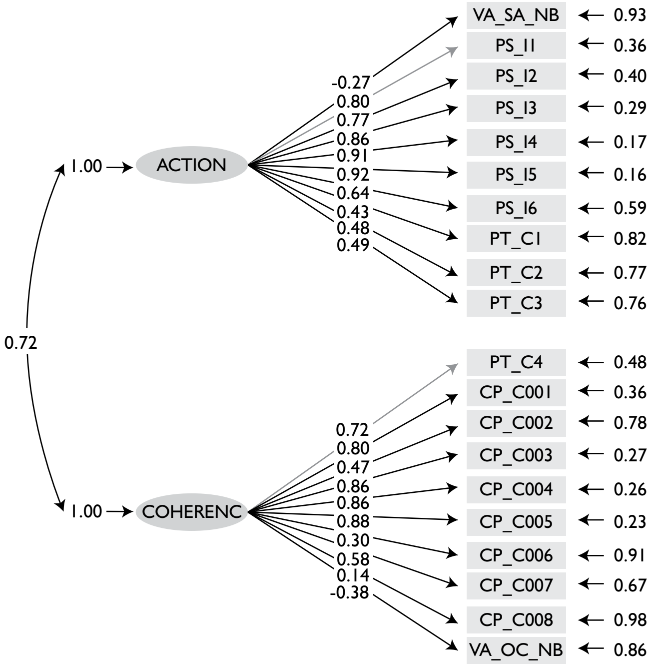
How do underlying factors fit with operational definition?

Exploratory factor analysis suggests two underlying factors (again, VANOCSS as single items). The first (eigenvalue=6.73) seems to capture observed behaviours of the primary care physician related to the transition of patient care to other providers, or coordination actions. It includes all items in the PCAS Integration and the VANOCSS Overall Coordination and VANOCSS Specialist Access, plus two items in the PCAT-S Coordination ("help with appointment" and "write information for specialist," factor loading only .35). The second factor (eigenvalue=1.58) suggests provider efforts to produce coherence between visits within and outside the regular doctor's office. All but two items on the CPCI Coordination subscales load on this factor, as well as two of the PCAT-S Coordination items ("discuss places to get help" and "what happened at visit"). Two CPCI items do not load on any factor. One may be a slightly different construct ("want one doctor to coordinate") and the other is problematic owing to reverse scoring ("doctor sometimes not aware of care").

We examined the fit of individual items in the VANOCSS subscales both in original and dichotomously coded forms. They mostly load on coordination action, except for "difficulty making appointments with specialists you wanted to see," an item that does load with any others and probably does not fit with the construct of management continuity.

The structural equation, two-factor model with items grouped by coordination actions and coherence is shown in Figure 3.

FIGURE 3. Structural equation model showing loadings of items on sub-dimensions of coordination action and experienced coherence, shown to be correlated (curved lines)



Chi-square = 671.29, *df* = 169, *p*-value = .00000, RMSEA = .084

Note: There is modest improvement compared to the unidimensional model (difference in chi-square of the models χ^2 = 1,042.85 – 671.29 = 371.56, *df* 1, p<.001) but persisting high levels of residual error (shown to the right of the items), especially for the PCAT-S and CPCI subscales. Root mean squared error of approximation (RMSEA) does not support good model fit.

Does worse management continuity predict reported problems?

Logistic regression modelling shows a statistically significant higher likelihood of reporting a problem in the VANOCSS Overall Coordination subscale for every unit of decrease in the score on the PCAS, PCAT-S and CPCI subscales (Table 4). Normalizing the scores to a 0-to-10 metric allows comparison of the magnitude of effect. Each unit decrease in the PCAS Integration subscale, the largest effect, is associated with a 2.2 higher likelihood of problems on Overall Coordination (OR=2.2); it accounts for 29% of the variance in Overall Coordination. On the non-normalized score, going from a score of 5 to 4 (very good to good) is associated with a 4.8 times higher likelihood of reporting problems. For Specialist Access, the effects are statistically significant only for the PCAS, and they are modest. The logistic model shows poor goodness of fit for the CPCI.

TABLE 4. Results of logistic regression models examining the likelihood of any problem reported on the VANOCSS Overall Coordination and Specialty Access subscales, with PCAS, PCAT-S and CPCI subscale scores. Odds ratios show the likelihood of reporting any problem associated with each unit decrease in the assessment of management continuity.*

	Odds Ratio with Normalized Score OR (95% CI)	Odds Ratio with Raw Score OR (95% CI)	Explained Variance (Nagelkerke's R2**)	Goodness of Fit (Hosmer and Lemshow p-value***)
	VANOCSS Overall Coordination			
PCAS Integration	2.2 (1.6, 3.0)	4.8 (2.6, 9.1)	.29	.11
PCAT-S Coordination	1.3 (1.0, 1.7)	2.6 (1.1, 6.3)	.06	.93
CPCI Coordination of Care	1.4 (1.1, 1.8)	2.0 (1.3, 3.2)	.08	.05
	VANOCSS Specialty Provider Access			
PCAS Integration	1.4 (1.1, 1.6)	1.8 (1.2, 2.7)	.09	.09
PCAT-S Coordination	1.1 (0.9, 1.3)	1.4 (0.8, 2.4)	.01	.15
CPCI Coordination of Care	1.1 (0.9, 1.4)	1.3 (0.9, 1.8)	.01	.03

* Each OR calculated in a separate regression model. OR-normalized refers to subscale scores normalized from 0 to 10.

** R2 is interpreted as a reflection of outcome variance explained by a variable in the model.

*** Goodness of model fit; values <.05 indicate poor fit.

Individual item performance

Parametric item response theory analysis shows that positive skewing of the PCAS, PCAT-S and CPCI subscales results in diminished capacity to discriminate between different degrees of above-average continuity but is highly discriminatory of below-average levels. The VANOCSS subscales, in contrast, are more discriminatory for above-average continuity.

Discussion and Conclusion

In this study, we found that five validated subscales measuring patient assessments of care coordination relate adequately to a common construct, which we presume to be management continuity. Exploratory factor analysis suggests that two distinct sub-dimensions underlie this pool of items: coordination actions and coherence.

Coordination actions relate to physician behaviours to facilitate transition of patient care to other providers, presumably to achieve timeliness and complementarity of services, though no subscales directly addressed these qualities. The PCAS Integration subscale covers this dimension, as do most of the items of the PCAT-S Coordination subscale. However, providers – not patients – are probably the best source for assessing timeliness and complementarity of services.

The coherence dimension is highly correlated with coordination actions but seems to address the provider's effort, directed at the patient, to link different services and avoid gaps in care. This effort includes sense-making after a series of visits for a specific health condition and planning for future care based on results of past visits. The CPCI Coordination subscale addresses this factor, but it is also captured in the PCAT-S item about talking with the patient about what happened at the specialist visit. However, none of the subscales captures connectedness and coherence as experienced by the patient, though the absence of these qualities can be inferred from the occurrence of problems reported using the VANOCSS Overall Coordination subscale. Granted, the developers were measuring coordination, not management continuity, but all assume that the patient is aware of the provider's coordination efforts. Qualitative studies suggest that patients are often unaware of critical aspects of coordination, such as agreed-upon care plans or information transfers between providers (Gallagher and Hodge 1999; Woodward et al. 2004). They presume these elements are in place, and can detect only failures or gaps. Thus, they can more validly assess discontinuity than continuity. This finding would suggest that the VANOCSS approach may be a more accurate assessment of management continuity.

In qualitative studies, patients seem to express their experience of good coordination as giving a sense of security and of being taken care of rather than as connectedness or smoothness (Burkey et al. 1997; Kai and Crosland 2001; Kroll and Neri 2003). The French term for "continuity of care," *prise-en-charge*, captures this notion but seems to have no English equivalent. The term implies the presence of a provider who takes responsibility for ensuring that required care is provided, as by case managers in mental healthcare or patient navigators in cancer care (Wells et al. 2008). Indeed, we observed strong correlations between subscales of management continuity and relational attributes of care. As measures of relational continuity or interpersonal communication increase, the number of reported coordination problems between all providers decreases.

This study has several limitations. The most striking is considering together instruments that use different reference points as well as two distinct approaches to measurement. The confirmatory factor analysis shows, not surprisingly, that the best model is the one where items are associated to their own parent instruments. The PCAS, PCAT-S and CPCI subscales focus specifically on the PHC physician and elicit predominantly positive assessments, a feature common to rating scales (Williams et al. 1998). In contrast, the two VANOCSS subscales elicit experienced difficulties across all providers, from which evaluators infer the degree of coordination or specialist access. This measurement approach, used by the Picker Institute (Gerteis et al. 1993), expressly increases sensitivity to problems in order to guide and monitor improvement efforts.

Nonetheless, the distinct approaches also create a unique opportunity to compare different formats and provide new information on how well assessments of coordination behaviours predict reported problems. Our item response analysis (not shown) suggests that subscales could be used in combination to reliably and validly identify persons with poor management continuity (low scores on the PCAS Integration and more than one problem on the VANOCSS Overall Coordination subscale) or good management continuity (high scores with no problems). We recently developed a measure of management continuity that combines both approaches; it appears to perform well and to predict continuity outcomes (Haggerty, Roberge et al. 2011).

In conclusion, these measures of coordination or management continuity seem to have a single underlying construct but capture only partially our definition of management continuity. Combining the PCAS and the VANOCSS Overall Coordination subscales is probably the most accurate way to detect both problematic continuity and good continuity. However, further development is needed of measures of how patients experience coordination as management continuity or discontinuity.

ACKNOWLEDGEMENTS

This research was funded by the Canadian Institutes of Health Research. During the conduct of the study Jeannie L. Haggerty held a Canada Research Chair in Population Impact of Healthcare at the Université de Sherbrooke. The authors wish to thank Beverley Lawson for conducting the survey in Nova Scotia and Christine Beaulieu in Quebec and Donna Riley for support in preparation and editing of the manuscript.

Correspondence may be directed to: Jeannie L. Haggerty, Associate Professor, Department of Family Medicine, McGill University, St. Mary's Research Centre, Hayes Pavilion – Suite 3734, 3830 Lacombe Ave., Montreal QC H3T 1M5; tel.: 514-345-3511 ext. 6332; fax: 514-734-2652; e-mail: jeannie.haggerty@mcgill.ca.

REFERENCES

Borowsky, S.J., D.B. Nelson, J.C. Fortney, A.N. Hedeen, J.L. Bradley and M.K. Chapko. 2002. "VA Community-Based Outpatient Clinics: Performance Measures Based on Patient Perceptions of Care." *Medical Care* 40(70): 578–86.

Burkey, Y., M. Black and H. Reeve. 1997. "Patients' Views on Their Discharge from Follow-up in Outpatient Clinics: Qualitative Study." *British Medical Journal* 315(7116): 1138–41.

Flocke, S. 1997. "Measuring Attributes of Primary Care: Development of a New Instrument." *Journal of Family Practice* 45(1): 64–74.

Gallagher, E.M. and G. Hodge. 1999. "The Forgotten Stakeholders: Seniors' Values Concerning Their Health Care." *International Journal of Health Care Quality Assurance* 12(3): 79–87.

Gerteis, M., S. Edgman-Levitan, J. Daley and T.L. Delbanco. 1993. *Through the Patient's Eyes: Understanding and Promoting Patient-Centered Care.* San Francisco: Jossey-Bass.

Haggerty, J.L., F. Burge, M.-D. Beaulieu, R. Pineault, C. Beaulieu, J.-F. Lévesque et al. 2011. "Validation of Instruments to Evaluate Primary Healthcare from the Patient Perspective: Overview of the Method." *Healthcare Policy* 7 (Special Issue): 31–46.

Haggerty, J.L., F. Burge, J.-F. Lévesque, D. Gass, R. Pineault, M.-D. Beaulieu et al. 2007. "Operational Definitions of Attributes of Primary Health Care: Consensus Among Canadian Experts." *Annals of Family Medicine* 5: 336–44.

Haggerty, J.L., R.J. Reid, G.K. Freeman, B.H. Starfield, C.E. Adair and R. McKendry. 2003. "Continuity of Care: A Multidisciplinary Review." *British Medical Journal* 327: 1219–21.

Haggerty, J.L., D. Roberge, G.K. Freeman et al. 2011. "When Patients Encounter Several Providers: Validation of a Generic Measure of Continuity of Care." *Annals of Family Medicine.* (Submitted.)

Kai, J. and A. Crosland. 2001. "Perspectives of People with Enduring Mental Ill Health from a Community-Based Qualitative Study." *British Journal of General Practice* 51: 730–36.

Kroll, T. and M.T. Neri. 2003. "Experiences with Care Co-ordination Among People with Cerebral Palsy, Multiple Sclerosis, or Spinal Cord Injury." *Disability and Rehabilitation* 25: 1106–14.

Reid, R., J. Haggerty and R. McKendry. 2002. *Defusing the Confusion: Concepts and Measures of Continuity of Care.* Ottawa: Canadian Health Services Research Foundation.

Safran, D.G., J. Kosinski, A.R. Tarlov, W.H. Rogers, D.A. Taira, N. Lieberman and J.E. Ware. 1998. "The Primary Care Assessment Survey: Tests of Data Quality and Measurement Performance." *Medical Care* 36(5): 728–39.

Santor, D.A., J.L. Haggerty, J.-F. Lévesque, F. Burge, M.-D. Beaulieu, D. Gass et al. 2011. "An Overview of Confirmatory Factor Analysis and Item Response Analysis Applied to Instruments to Evaluate Primary Healthcare." *Healthcare Policy* 7 (Special Issue): 79–92.

Shi, L., B. Starfield and J. Xu. 2001. "Validating the Adult Primary Care Assessment Tool." *Journal of Family Practice* 50(2): 161–71.

Shortell, S.M., R.R. Gillies, D.A. Anderson, K.M. Eirckson and J.B. Mitchell. 1996. *Remaking Health Care in America: Building Organized Delivery Systems.* San Fransisco: Jossey-Bass.

Wells, K.J., T.A. Battaglia, D.J. Dudley, R. Garcia, A. Greene, E. Calhoun et al. 2008. "Patient Navigation: State of the Art, or Is It Science?" *Cancer* 113: 1999–2010.

Williams, B., J. Coyle and D. Healy. 1998. "The Meaning of Patient Satisfaction: An Explanation of High Reported Levels." *Social Science and Medicine* 47(9): 1351–59.

Woodward, C.A., J. Abelson, S. Tedford and B. Hutchison. 2004. "What Is Important to Continuity in Home Care? Perspectives of Key Stakeholders." *Social Science and Medicine* 58: 177–92.

Management Continuity from the Patient Perspective: Comparison of Primary Healthcare Evaluation Instruments

La continuité d'approche du point de vue du patient : comparaison entre instruments d'évaluation des soins de santé primaires

JEANNIE L. HAGGERTY, FREDERICK BURGE, RAYNALD PINEAULT, MARIE-DOMINIQUE BEAULIEU,
FATIMA BOUHARAOUI, CHRISTINE BEAULIEU AND DARCY A. SANTOR

TABLE 1. Distribution of values on the items in subscales mapped to management continuity among the 432 respondents who saw more than one provider

Item Code	Item Statement	Missing % (n)	Percentage (Number) by Response Option						Item Discrimination[1]
Primary Care Assessment Survey (PCAS): Integration Thinking about the times your doctor has recommended you see a different doctor for a specific health problem…			**Very poor**	**Poor**	**Fair**	**Good**	**Very good**	**Excellent**	
PS_i1	How would you rate the help your regular doctor gave you in deciding **who to see** for specialty care?	1 (4)	2 (6)	3 (9)	8 (28)	26 (87)	34 (117)	26 (89)	2.30 (0.20)
PS_i2	How would you rate the help your regular doctor gave you in **getting an appointment** for specialty care you needed?	2 (6)	2 (8)	4 (13)	9 (31)	24 (82)	32 (109)	27 (91)	2.35 (0.21)
PS_i3	How would you rate regular doctor's involvement in your care when you were being treated by a **specialist or were hospitalized?**	4 (13)	1 (3)	10 (33)	12 (39)	31 (104)	26 (87)	18 (61)	3.68 (0.30)
PS_i4	How would you rate regular doctor's **communication with specialists** or other doctors who saw you?	4 (15)	2 (8)	7 (22)	14 (46)	32 (109)	25 (84)	17 (56)	4.89 (0.42)
PS_i5	How would you rate the help your regular doctor gave you in understanding what the **specialist or other doctor** said about you?	4 (12)	3 (10)	4 (13)	15 (52)	28 (95)	27 (91)	20 (67)	4.06 (0.35)
PS_i6	How would you rate the quality of **specialists or other doctors** your regular doctor sent you to?	1 (4)	2 (5)	2 (6)	8 (27)	23 (77)	38 (130)	27 (91)	1.85 (0.19)
Primary Care Assessment Tool (PCAT-S): Coordination Thinking about visit to specialist or specialized service…			**Definitely not**	**Probably not**	**Probably**	**Definitely**	**Not sure / Don't remember**		
PT_c1	Did your Primary Care Provider discuss with you different places you could have gone to get help with that problem?	2 (9)	10 (40)	9 (36)	23 (89)	52 (203)	4 (16)		1.23 (0.17)
PT_c2	Did your Primary Care Provider or someone working with your Primary Care Provider help you make the appointment for that visit?	3 (12)	10 (40)	6 (23)	16 (61)	64 (251)	2 (6)		2.07 (0.26)
PT_c3	Did your Primary Care Provider write down any information for the specialist about the reason for the visit?	3 (11)	7 (29)	8 (33)	22 (86)	54 (212)	6 (22)		2.33 (0.24)
PT_c4	After you went to the specialist or special service, did your Primary Care Provider talk with you about what happened at the visit?	2 (9)	14 (56)	8 (33)	19 (76)	53 (208)	3 (11)		2.23 (0.25)
Components of Primary Care Index (CPCI)			**Strongly disagree**	**2**	**3**	**4**	**5**	**Strongly agree**	
CP_coo1	This doctor knows when I'm due for a check-up.	2 (10)	11 (46)	11 (49)	14 (59)	12 (52)	17 (74)	33 (142)	2.35 (0.21)
CP_coo2	I want one doctor to coordinate all of the healthcare I receive.	3 (11)	3 (11)	3 (14)	4 (15)	9 (37)	18 (78)	62 (266)	1.13 (0.17)
CP_coo3	This doctor keeps track of all my healthcare.	1 (6)	4 (19)	7 (29)	7 (29)	12 (51)	22 (93)	48 (205)	4.14 (0.36)
CP_coo4	This doctor always follows up on a problem I've had, either at the next visit or by phone.	3 (13)	7 (31)	9 (38)	8 (33)	12 (51)	21 (91)	41 (175)	3.91 (0.30)

TABLE 1. Continued

Item Code	Item Statement	Missing % (n)	Percentage (Number) by Response Option						Item Discrimination[1]
CP_ coo5	This doctor always follows up on my visits to other healthcare providers.	3 (11)	6 (27)	8 (34)	12 (53)	15 (63)	20 (86)	37 (158)	3.25 (0.26)
CP_ coo6	This doctor helps me interpret my lab tests, X-rays or visits to other doctors.	2 (9)	22 (94)	15 (66)	13 (56)	11 (46)	15 (66)	22 (95)	0.43 (0.13)
CP_ coo7	This doctor communicates with the other health providers I see.	5 (21)	10 (42)	13 (55)	16 (71)	16 (71)	17 (75)	23 (97)	1.62 (0.16)
CP_ coo8	This doctor does not always know about care I have received at other places.[2]	5 (22)	11 (46)	13 (58)	17 (72)	13 (55)	20 (86)	22 (93)	0.32 (0.11)
Veterans Affairs National Outpatient Customer Satisfaction Survey (VANOCSS): Overall Coordination of Care[3]			**Problem**	**No problem**					
VA_ cco1	Were the providers who cared for you always familiar with your most recent medical history?	0 (1)	59 (165)	40 (112)					1.67 (0.26)[4]
VA_ cco2	Were there times when one of your providers did not know about tests you had or their results?	1 (3)	40 (111)	59 (164)					2.81 (0.46)
VA_ cco3	Were there times when one of your providers did not know about changes in your treatment that another provider recommended?	2 (5)	26 (73)	72 (200)					1.90 (0.35)
VA_ cco4	Were there times when you were confused because different providers told you different things?	1 (3)	28 (78)	71 (197)					1.90 (0.30)
VA_ cco5	Did you always know what the next step in your care would be?	3 (7)	59 (163)	39 (108)					1.44 (0.26)
VA_ cco6	Did you know who to ask when you had questions about your health care?	2 (5)	37 (102)	62 (171)					1.48 (0.28)
Veterans Affairs National Outpatient Customer Satisfaction Survey (VANOCSS): Access to Specialists[5]			**Problem**	**No problem**					
VA_ spa1	How often did you get to see specialists when you thought you needed to?	2 (5)	12 (29)	86 (202)					1.26 (0.36)[4]
VA_ spa2	How often did you have difficulty making appointments with specialists you wanted to see?	1 (3)	19 (44)	80 (189)					0.42 (0.23)
VA_ spa3	How often were you given enough information about why you were to see your specialists?	1 (2)	11 (25)	89 (209)					2.48 (0.62)
VA_ spa4	How often did your specialists have the information they needed from your medical records?	1 (3)	20 (48)	78 (185)					2.68 (0.56)

[1] Items were assessed against the construct of the original scale. Values > 1 are considered to be discriminating.

[2] Values were reversed for this reverse-worded item: only 11% (46) strongly agreed that the doctor did not know about care received at other places.

[3] Descriptive statistics based on 279 respondents (subscale placed second-to-last in questionnaire).

[4] Items are scored dichotomously; these are not Likert scales.

[5] Descriptive statistics based on 236 respondents (subscale placed last in questionnaire).

RESEARCH PAPER

Comprehensiveness of Care from the Patient Perspective: Comparison of Primary Healthcare Evaluation Instruments

La globalité des soins du point de vue du patient : comparaison entre instruments d'évaluation des soins de santé primaires

JEANNIE L. HAGGERTY, PHD
Department of Family Medicine, McGill University
Montreal, QC

MARIE-DOMINIQUE BEAULIEU, MD, MSC
Chaire Dr Sadok Besrour en médecine familiale
Centre de recherche du Centre hospitalier de l'Université de Montréal
Montréal, QC

RAYNALD PINEAULT, MD, PHD
Département de Médecine sociale et préventive, Université de Montréal
Montreal, QC

FREDERICK BURGE, MD, MSC
Department of Family Medicine, Dalhousie University
Halifax, NS

JEAN-FRÉDÉRIC LÉVESQUE, MD, PHD
Centre de recherche du Centre hospitalier de l'Université de Montréal
Montréal, QC

DARCY A. SANTOR, PHD
School of Psychology, University of Ottawa
Ottawa, ON

FATIMA BOUHARAOUI, MSC
St. Mary's Research Centre, St. Mary's Hospital Center
Montreal, QC

Comprehensiveness of Care from the Patient Perspective

CHRISTINE BEAULIEU, MSC
St. Mary's Research Centre, St. Mary's Hospital Center
Montreal, QC

Abstract

Comprehensiveness relates both to scope of services offered and to a whole-person clinical approach. Comprehensive services are defined as "the provision, either directly or indirectly, of a full range of services to meet most patients' healthcare needs"; whole-person care is "the extent to which a provider elicits and considers the physical, emotional and social aspects of a patient's health and considers the community context in their care." Among instruments that evaluate primary healthcare, two had subscales that mapped to comprehensive services and to the community component of whole-person care: the Primary Care Assessment Tool – Short Form (PCAT-S) and the Components of Primary Care Index (CPCI, a limited measure of whole-person care).

Objective: To examine how well comprehensiveness is captured in validated instruments that evaluate primary healthcare from the patient's perspective.

Method: 645 adults with at least one healthcare contact in the previous 12 months responded to six instruments that evaluate primary healthcare. Scores were normalized for descriptive comparison. Exploratory and confirmatory (structural equation modelling) factor analysis examined fit to operational definition, and item response theory analysis examined item performance on common constructs.

Results: Over one-quarter of respondents had missing responses on services offered or doctor's knowledge of the community. The subscales did not load on a single factor; comprehensive services and community orientation were examined separately. The community orientation subscales did not perform satisfactorily. The three comprehensive services subscales fit very modestly onto two factors: (1) most healthcare needs (from one provider) (CPCI Comprehensive Care, PCAT-S First-Contact Utilization) and (2) range of services (PCAT-S Comprehensive Services Available). Individual item performance revealed several problems.

Conclusion: Measurement of comprehensiveness is problematic, making this attribute a priority for measure development. Range of services offered is best obtained from providers. Whole-person care is not addressed as a separate construct, but some dimensions are covered by attributes such as interpersonal communication and relational continuity.

Résumé

La notion de globalité des soins s'applique tant à l'étendue des services offerts qu'à une approche holistique des soins cliniques. La globalité des services se définit comme « la prestation, directe ou indirecte, d'une gamme complète de services afin de répondre aux besoins des patients en matière de soins de santé »; les soins centrés sur le patient se définissent par « l'étendue selon dans laquelle le fournisseur de soins considère les aspects physiques, émotionnels et sociaux de la santé d'un patient et tient compte du contexte communautaire lors de la prestation de soins. » Deux des instruments d'évaluation des soins primaires comprennent des sous-échelles qui correspondent à la globalité des soins et aux composantes communautaires des soins centrés sur le patient : le Primary Care Assessment Tool – version courte (PCAT-S) et le Components of Primary Care Index (CPCI, mesure limitée des soins centrés sur le patient).

Objectif : Examiner dans quelle mesure la globalité des soins est captée par les instruments validés qui servent à évaluer les soins de santé primaires du point de vue du patient.

Méthode : Six cent quarante-cinq adultes, qui ont eu au moins un contact avec les services de santé au cours des 12 mois antérieurs, ont répondu à six instruments d'évaluation des soins primaires. Les résultats ont été normalisés pour permettre des comparaisons descriptives. Les analyses factorielles exploratoires et confirmatoires (modélisation par équation structurelle) ont été employées pour vérifier l'adéquation à la définition opérationnelle, et l'analyse de réponse par item a été utilisée pour examiner la performance en fonction de construits communs.

Résultats : Plus du quart des répondants ont laissé des questions sans réponse au sujet des services offerts ou des connaissances du médecin sur la communauté. Les sous-échelles ne correspondent pas à un seul facteur; la globalité des services et l'orientation communautaire ont été examinées séparément. Les sous-échelles « orientation communautaire » n'ont pas offert une performance satisfaisante. Les trois sous-échelles « globalité des services » se sont ajustées très modestement à deux facteurs : (1) la plupart des besoins en matière de services de santé [d'un fournisseur] (« globalité des soins » du CPCI et « utilisation de premier contact » du PCAT-S) et (2) étendue des services (« globalité des services disponibles » du PCAT-S). La performance individuelle d'items a révélé plusieurs problèmes.

Conclusion : La mesure de la globalité des soins est problématique, ce qui en fait une priorité pour le développement de mesures. L'étendue des services offerts s'obtient plus efficacement auprès des fournisseurs. Les soins centrés sur le patient ne sont pas traités comme un construit indépendant, mais certaines dimensions sont couvertes par des caractéristiques telles que la communication interpersonnelle et la continuité relationnelle.

* * *

COMPREHENSIVENESS IS ENSHRINED AS ONE OF THE FIVE PRINCIPLES OF THE *Canada Health Act* (Madore 2005) and is often applied as a qualifier of primary healthcare (PHC) (Macinko et al. 2007; Romanow 2002), distinguishing the ideal from the merely functional or, at worst, "selective" (Cueto 2004) forms of care. While this attribute evokes a sense of "good," the generality of its invocation obscures precision about its meaning, posing a major challenge for its assessment.

Background

Conceptualizing comprehensiveness

One objective of PHC reform in Canada is to expand the comprehensiveness of services, especially in health promotion and chronic disease management (Health Canada 2007).

Most PHC renewal interventions include introducing team-based care and alternative payment mechanisms. Consequently, evaluating comprehensiveness of care is vital to evaluating the renewal of PHC in Canada.

The dictionary definition of comprehensiveness – "covering completely or broadly" (Merriam-Webster 1998) – is applied in PHC to the mandate to resolve and manage the most prevalent health conditions, undifferentiated by age, sex or disease (CFPC 2006). However, comprehensiveness sometimes refers to the bio-psycho-social or whole-person approach, which sees the patient as body and soul within a specific social context (CFPC 2006). The closest French equivalent of "comprehensiveness," *globalité*, invokes an image of both scope and whole-person approach.

Evaluating the comprehensiveness of primary healthcare

A clear operational definition of comprehensiveness is a first step in any measurement. Our consensus consultation of PHC experts (Haggerty et al. 2007) unanimously identified comprehensiveness as a core attribute of PHC, but two distinct definitions emerged. The first, comprehensive services, corresponds to scope: "The provision, either directly or indirectly, of a full range of services to meet patients' healthcare needs. This includes health promotion, prevention, diagnosis and treatment of common conditions, referral to other providers, management of chronic conditions, rehabilitation, palliative care and, in some models, social services." The second definition, whole-person care, is: "The extent to which a provider elicits and considers the physical, emotional and social aspects of a patient's health and considers the community context in their care." Most experts agreed that providers and utilization data were the best data sources for comprehensive services, and the patient for whole-person care.

Our objective was to compare subscales from different validated instruments that purport to measure comprehensiveness. We compare scores and examine (a) whether comprehensiveness subscales from different instruments seem to measure the same underlying construct, or how emerging factors relate to both operational definitions and (b) how well individual items perform. Our intent is not to recommend one instrument over another, but to provide evaluators with insight into how well different subscales fit the experts' operational definitions.

Method

The method and analytic strategy are described in detail elsewhere in this special issue of the journal (Haggerty, Burge et al. 2011; Santor et al. 2011). Briefly: six instruments that evaluate PHC from the patient's perspective were administered to 645 healthcare users balanced by English/French language, rural/urban location, low/high level of education and poor/average/excellent overall PHC experience. The analysis consisted of examining the distributional statistics and subscale correlations, followed by common factor and confirmatory factor analysis (structural equation modelling) to identify dimensions common to the entire set of items. Finally, we examined the performance of individual items and response scales against constructs emerging across instruments using item response theory analysis.

Measure description

Among the six validated instruments in our study that assess PHC services from the patient's perspective, two contain subscales that mapped to our operational definitions of comprehensiveness. No subscales mapped specifically to whole-person care except for subscales for community-oriented clinical care that address one element in our definition of whole-person care: "elicits and considers the social aspects of a patient's health and considers the community context in their care."

The Components of Primary Care Index (CPCI) (Flocke 1997) has a six-item Comprehensive Care and a two-item Community Context subscale. Both elicit degree of agreement with statements about the "regular doctor" using a six-point semantic differential response scale with "1=strongly disagree" and "6=strongly agree" as anchors at each extremity.

The Primary Care Assessment Tool – Short Form, adult (PCAT-S) (Shi et al. 2001) has a four-item Services Available and a three-item Community Orientation subscale. The subscale assessing comprehensiveness with a Services Received subscale was not retained in our study because we mapped it to health promotion. The three-item subscale on First-Contact Utilization, initially mapped to accessibility, was included in the comprehensiveness attribute based on best psychometric fit (Haggerty, Lévesque et al. 2011). Questions about different aspects of care from the "primary care provider" are answered using a four-point Likert scale ("1=definitely not," "2=probably not," "3=probably," "4=definitely").

Results

Comparative descriptive statistics

Table 1 summarizes the item content and behaviour in the five subscales; the detailed content and distributions are available online at http://www.longwoods.com/content/22639. Several items have more than 5% missing values, especially items about available service by the doctor or clinic. Most problematically, a large proportion selected the "not sure" option for items in the PCAT-S Services Available and Community Orientation subscales, interpreted as missing values. Indeed, this was the model response for two items in the three-item PCAT-S Community Orientation subscale. Ceiling effects were present with a large proportion endorsing the maximum value, especially for the CPCI Comprehensive Care and the PCAT-S First-Contact Utilization items. The discriminability of items indicates good capacity to discriminate between different levels of the subscale score.

Table 2 presents the descriptive statistics by subscales. Scores are normalized to 0 to 10 to permit direct comparisons. The whole-person (community) subscales have very different means but their distribution approximates a normal curve. The comprehensive services scores are skewed towards positive values, with the PCAT-S First-Contact Utilization subscale being highly skewed; its score does not resemble that of the CPCI Comprehensive Care subscale, to which it is most similar in item content. Internal consistency (.65 to .83) is lower than for most subscales of other attributes, partly explained by the small number of items in some subscales.

Table 3 presents the Pearson correlations between the comprehensiveness subscales.

The comprehensive service subscales have low to modest correlation with one another, and whole-person (community) subscales correlate only modestly (.32). The highest correlation (*r*=.49) is between the two CPCI Comprehensive Care and Community Context subscales. The PCAT-S subscales do not correlate well with each other; the Services Available and First-Contact Utilization subscales correlate only weakly (*r*=.08), suggesting that they measure different facets of comprehensiveness. When correlated with subscales of other attributes, both CPCI subscales correlate equally or higher with measures of relational continuity (.45 to .71) than with other comprehensiveness subscales. The PCAT-S subscales correlate as well or better with measures of other attributes, though still more modestly than do the CPCI subscales; the PCAT-S Community Orientation subscale correlates most strongly (*r*=.37) with patient-centred decision-making from the Interpersonal Processes of Care instrument (Stewart et al. 2007).

TABLE 1. Summary of comprehensiveness of services subscale content and distribution of item responses. (Detailed distribution available at http://www.longwoods.com/content/22639)

Subscale and Item Description	Response Scale	Range Missing Values	Overall Modal Response	Range Item Discriminability	Comments on Distribution
Comprehensive Services					
CPCI Comprehensive Care (6 items) Agreement with statements about regular doctor: I go to this doctor for almost all my medical care; doctor handles emergencies; can take care of almost any medical problem I could go to this doctor: for help with an emotional problem; for care of an ongoing problem; for a check-up to prevent illness	Semantic differential opinion, 1=strongly disagree, 6=strongly agree	2%–6%	6 (strongly agree)	1.32 (emergencies) to 3.00 (ongoing problem)	30%–67% of respondents select most positive opinion; 3%–10% most negative
PCAT-S Comprehensiveness – Services Available (4 items) Likelihood of reporting of availability of service at the clinic… Immunizations; family planning or birth control; counselling for mental health problems; sewing up a cut that needs stitches	Likert evaluative, 1=definitely not to 4=definitely	2%–7% (true missing), 8%–24% not sure	4 (definitely)	1.84 (sewing) to 2.10 (family planning)	Approximately 15%–24% of respondents "not sure" (exception for Immunizations)
PCAT-S First-Contact Utilization (3 items) Likelihood of seeking care first from primary care provider for routine care; for a new problem; need for referral to see a specialist	Likert evaluative, 1=definitely not to 4=definitely	1%–2% (true missing), 0%–4% not sure	4 (definitely)	0.87 (referral by PCP) to 4.7 (clinic=first place for routine care)	~85% of responses in the most positive category for seeking care first with primary care provider
Community Oriented (Whole-Person Care)					
CPCI Community Context (2 items) Agreement with statements about regular doctor: This doctor knows a lot about my community; uses her/his knowledge of my community to take care of me	Semantic differential opinion, 1=strongly disagree, 6=strongly agree	4%–5%	6 (strongly agree)	Inappropriate in a two-item scale services	Responses relatively evenly distributed among options; ~32% in most positive category

TABLE 1. Continued

Subscale and Item Description	Response Scale	Range Missing Values	Overall Modal Response	Range Item Discriminability	Comments on Distribution
PCAT-S Community Orientation (3 items) Likelihood of primary care provider… Making home visit; knowing about health problems in neighbourhood; getting opinions from people to provide better healthcare	Likert evaluative, 1=definitely not to 4=definitely	2% (true missing), 27%–32% not sure	3 (probably)	1.18 (get opinions) to 2.61 (neighbourhood)	Responses relatively evenly distributed among options. High missing values due to "not sure" response.

TABLE 2. Mean and distributional scores for comprehensiveness subscales, showing normalized scores from 0 to 10 (n=645)*

Developer's Scale Name (# items in scale)	Cronbach's Alpha	Mean	SD	Quartiles Q1 (25%)	Q2 (50%)	Q3 (75%)
Comprehensiveness of Services						
CPCI Comprehensive Care (6)	.83	7.7	2.1	6.7	8.0	9.3
PCAT-S Comprehensiveness (Services Available) (4)	.72	7.7	2.5	6.7	8.3	10.0
PCAT-S First-Contact Utilization (3)	.68	9.1	1.6	8.9	10.0	10.0
Whole-Person Care						
CPCI Community Context (2)	.82	6.5	3.1	4.0	7.0	10.0
PCAT-S Community Orientation (3)	.65	4.9	2.9	3.3	5.0	6.7

* Subscale scores calculated as mean of item values and only calculated for observations where >50% of items were complete.

TABLE 3. Mean partial Pearson correlations between comprehensiveness subscales and with other subscales included in the questionnaires.* Only correlations significantly different from zero are provided.

Questionnaire Subscale	CPCI: Comprehensive Care	PCAT-S: Services Available	PCAT-S: First-Contact Utilization	CPCI: Community Context	PCAT-S: Community Orientation
Comprehensiveness of Services					
CPCI Comprehensive Care	1.00	0.28	0.34	0.49	0.22
PCAT Services Available	0.28	1.00	0.08	0.22	0.23
PCAT First-Contact Utilization	0.34	0.08	1.00	0.20	0.09
Whole-Person Care					
CPCI Community Context	0.49	0.22	0.20	1.00	0.32
PCAT Community Orientation	0.22	0.23	0.09	0.32	1.00

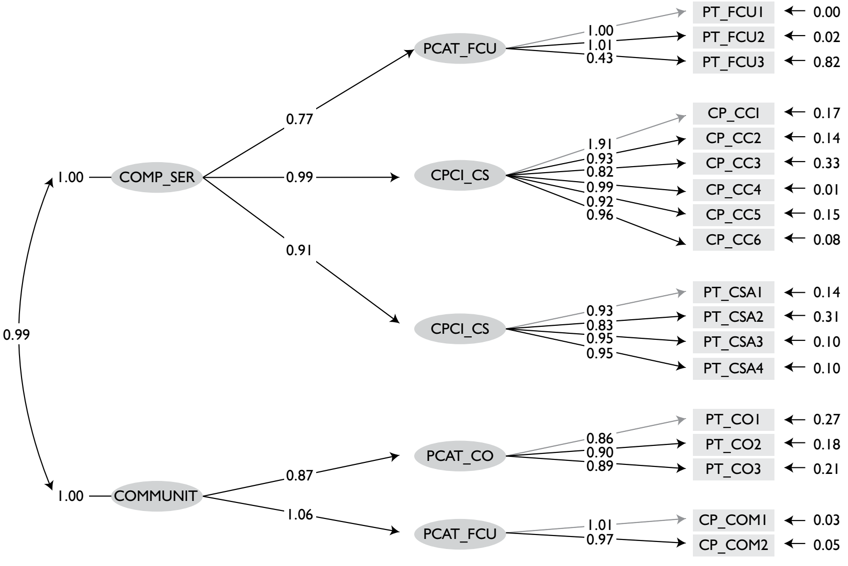
Note: Controlling for study design variables (language, educational achievement, geographic location).

Do all items measure a single attribute?

Missing values drastically reduced the effective sample size for factor analysis from 645 to 322 (listwise missing). The majority of exclusions (79% or 255/323) were due to selecting the "not sure/don't know" option. Respondents excluded were older and in worse health than those included, but were otherwise similar in all measured variables. Imputing values for missing responses increased the analytic sample to 490 and slightly improved the fit statistics but did not change the conclusions, suggesting that we principally lost statistical power and did not introduce bias by using the more conservative approach for factor analysis.

We had little expectation that all items would load on a single factor because we had two operational definitions. Indeed, a one-factor model with structural equation modelling generated fit statistics suggesting poor fit, with a root mean squared error of approximation (RMSEA) of p=.184, considerably higher than the .05 standard for good fit. We removed PCAT-S First-Contact Utilization and the RMSEA fit improved (p=.134), but at the expense of the normed fit index (NFI), which went from .96 to .91, though still higher than the .90 standard for good fit. However, removing the two whole-person (community) subscales improved model fit overall (RMSEA, p=.125; NFI=.94). Even when items are grouped within their original subscales loading on a single construct, presumed to be comprehensiveness (usually expected to improve fit), the model does not fit a single underlying factor (RMSEA, p=.165; NFI=.97). We concluded that the community subscales measure a separate construct and need to be examined in independent models. Figure 1 presents the model with items grouped within their parent instrument subscales as five first-order latent variables, which in turn emerge from two separate, though correlated, second-order constructs, community and comprehensive services (RMSEA, p=.165; NFI=.97). Remaining analyses examine community and comprehensive cervices constructs separately.

FIGURE 1. Parameter estimations for a structural equation second-order model where a single underlying construct (second-order latent variable) leads to the five subscales (first-order variables) with loadings on their respective items

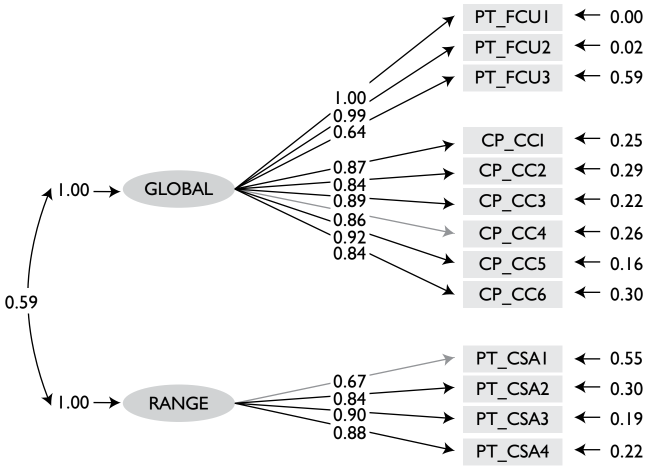


Chi-square = 810.64, df = 129, p-value = .00000, RMSEA = .161

How do underlying factors fit with operational definitions?

For *comprehensive services*, exploratory factor analysis suggested two underlying factors. Using our operational definitions as a guide to infer the factors, we judged that the first (eigenvalue=3.80) captures "meeting most of the patient's healthcare needs" or Most Needs. This item was composed of items from the CPCI Comprehensive Care and the PCAT-S First-Contact Utilization subscales. The second factor (eigenvalue=1.03) corresponds to "the provision directly of a full range of services," or Range, and consists only of items from the PCAT-S Services Available subscale. When items are grouped by Range and Most Needs, model goodness-of-fit statistics improve dramatically over the one-dimensional model (Figure 2). However, fit is only adequate (RMSEA, p=.108; NFI=.95), and Range and Most Needs correlate only modestly (.59).

FIGURE 2. Parameter estimations for a structural equation first-order model where sub-dimensions are correlated



Chi-square = 217.14, *df* = 64, *p*-value = .00000, RMSEA = .108

For whole-person care, it is clear that the CPCI Community Context and PCAT-S Community Orientation subscales respond only to the social or community elements of the operational definition. They hold together as a single factor but do not represent whole-person care completely. Exploratory factor analysis with other attributes of care found that the community items load relatively well with items from the Accumulated Knowledge subscale in relational continuity and modestly with the Elicit construct of interpersonal communication.

Individual item performance

In confirmatory factor analysis, many items have a high proportion of residual error (shown to the right of each item label in Figures 1 and 2), which may reflect either poor conceptual fit or poor item performance.

Indeed, graphs of non-parametric item response analysis suggest that item performance may be largely responsible for poor model fit. Only two items show adequate discriminability and response behaviour for Most Needs, both from the CPCI Comprehensive Care subscale: "take care of almost any medical problem" and "help with a personal or emotional problem" (CP_cc4). The discriminatory capacity of "doctor handles emergencies" (CP_cc2), while modest overall, is concentrated in the central zone of Most Needs, consequently making an important contribution to the precision of the subscale. In remaining items, the most positive option is predominantly endorsed, even at low levels of Most Needs, providing adequate differentiation between low and average levels of Most Needs but no discrimination between above-average levels. Skewed items towards positive values provide excellent discrimination and information yield in the negative range of the attribute of interest. Conversely, skewing towards negative values provides good discrimination in the positive range, as for PCAT-S Community Orientation items on home visits and knowing the important problems in the patient's neighbourhood.

Discussion

Although comprehensiveness is a core PHC attribute, we encountered difficulties in both its definition and its measurement. The definitional challenge was resolved by distinguishing between comprehensive services and whole-person care, suggesting the need for two different measures. However, we found problems with measures of both operational definitions.

Comprehensive services referring to scope of services is the most common use of comprehensiveness in the PHC and evaluation literature. However, our consensus consultation with experts indicates that providers – not patients – are the best source of information on this attribute because providers need to plan service delivery to meet the needs of a broad group of patients (Haggerty et al. 2007). This disparity may explain the poor performance items eliciting the availability of specific services. Patients can validly assess only the availability of services that they or their close associates have needed. Although the "don't know" option in the PCAT-S Services Available subscale allows patients to state their level of knowledge, this option counts as a missing value for measurement purposes, compromising the subscale's psychometric performance and resulting in loss of information. Program evaluators and policy makers interested in evaluating range of comprehensive services are advised to obtain data from the providers themselves or the records of service provision.

Given the importance of this attribute to PHC and health system performance, it would be important to define minimal standards for PHC comprehensiveness in Canada. An environmental scan of Canadian policy documents provides a composite profile of expected services: rapid management of acute and urgent health problems; timely provision of non-urgent routine care (including wellness care and chronic illness management); coverage with recommended preventive services; referral to hospitals and specialists; follow-up care after hospitalization or specialty care; primary mental healthcare; full maternity and child care;

coordinated care of the frail elderly; and end-of-life care (Haggerty and Martin 2005). These services are potentially traceable with medical service billings, but not all elements can be captured in each province or compared among provinces. Some of these elements are elicited in the National Physician Survey (CFPC et al. 2007). These periodic surveys indicate that urban physicians are increasingly developing areas of expertise and working in group settings, or with other professionals, to ensure a wide range of services, whereas rural doctors retain a comprehensive profile in their personal practice (CFPC 2008).

The dimension of whole-person care is best measured from the patient's perspective, but no subscales mapped principally to this attribute – a surprising finding, given its philosophic centrality in PHC. However, attention to both emotional and physical needs are covered in other subscales. Accumulated Knowledge subscales within relational continuity measure the primary care physician's knowledge about various dimensions of the person (medical, family, work, values) and whether the patient is known as a person and not merely as someone with a medical problem (Burge et al. 2011). Attentiveness to the patient's felt concerns are addressed in the Elicit dimension of interpersonal communication (Beaulieu et al. 2011). In contrast, the subscales in this study covered only the "social aspects of care." Whole-person care may be inferred from the use of multiple subscales, including the community scales covered in this paper.

Whole-person care relates closely to the concept of person-centred care. Two relevant instruments assessing patient-centredness of clinical encounters were not included in our study because they are visit-based (Stewart et al. 2000, 2003; Little et al. 2001). Positive assessments on both instruments have been linked to better satisfaction and symptom resolution (Little et al. 2001; Stewart et al. 2000).

The French word for "comprehensiveness," *globalité*, connotes both scope of services and a bio-psycho-social approach. The CPCI Comprehensive Care subscale has items relating both to scope (the doctor handles emergencies) and bio-psycho-social approach (can help with a personal or emotional problem). It would be the best measure of *globalité*. Item response analysis suggests that the performance of this subscale would be improved by dropping non-informative or redundant items (most medical care and checkups) and adjusting the response scale.

There are limitations to this study. Most notably, our decision to map entire subscales to our operational definitions led to our representing whole-person care only by subscales addressing community care, providing only limited coverage of whole-person care. It would be more appropriate to view our results for these subscales as assessing community-oriented care, which was the original intention. Though some measurement problems persist, "community-oriented care" is a more valid representation of the construct. Finally, the lack of another data source on the actual services offered by the provider limits our capacity to assess the validity of patients' reports of available services.

Conclusion and Lessons Learned

In sum, the CPCI Comprehensive Care subscale provides a good assessment of patients' confidence in their physician's ability to meet most of their healthcare needs, including emotional problems (*globalité*). The bio-psycho-social approach is partially covered in instruments measuring relational continuity and interpersonal communication. Demonstrated provider knowledge of the patient's community with the PCAT-S Community Orientation subscale would indicate excellent accumulated and global knowledge.

Physician report or billing data is the best source for evaluating range of services. Given the importance of PHC comprehensiveness to system performance, it is critical for policy makers to define minimal standards for service provision across provinces and to ensure that these are adequately captured in billing data or electronic medical records used to assess this critical aspect of care.

ACKNOWLEDGEMENTS

This research was funded by the Canadian Institutes for Health Research. During this study Jeannie L. Haggerty held a Canada Research Chair in Population Impacts of Healthcare at the Université de Sherbrooke. The authors wish to thank Beverley Lawson for conducting the survey in Nova Scotia and Christine Beaulieu in Quebec and Donna Riley for support in preparation and editing of the manuscript.

Correspondence may be directed to: Jeannie L. Haggerty, Associate Professor, Department of Family Medicine, McGill University, St. Mary's Research Centre, Hayes Pavilion – Suite 3734, 3830 Lacombe Ave., Montreal QC H3T 1M5; tel.: 514-345-3511 ext. 6332; fax: 514-734-2652; e-mail: jeannie.haggerty@mcgill.ca.

REFERENCES

Beaulieu, M.-D., J.L. Haggerty, C. Beaulieu, F. Bouharaoui, J.-F. Lévesque, R. Pineault et al. 2011. "Interpersonal Communication from the Patient Perspective: Comparison of Primary Healthcare Evaluation Instruments." *Healthcare Policy* 7 (Special Issue): 108–23.

Burge, F., J.L. Haggerty, R. Pineault, M.-D. Beaulieu, J.-F. Lévesque, C. Beaulieu et al. 2011. "Relational Continuity from the Patient Perspective: Comparison of Primary Healthcare Evaluation Instruments." *Healthcare Policy* 7 (Special Issue): 124–38.

College of Family Physicians of Canada (CFPC). 2006. *Standards for Accreditation of Family Medicine Residency Training Programs* ("The Red Book"). Mississauga, ON: Author.

College of Family Physicians of Canada (CFPC). 2008. *The Face of Family Medicine Today: Results of the National Physician Survey.* Mississauga, ON: Author.

College of Family Physicians of Canada (CFPC), Canadian Medical Association and Royal College of Physicians and Surgeons of Canada. 2007. *2007 National Physician Survey.* Retrieved July 28, 2011. <http://www.nationalphysiciansurvey.ca/nps/2007_Survey/2007nps-e.asp>.

Cueto, M. 2004. "The Origins of Primary Health Care and Selective Primary Health Care." *American Journal of Public Health* 94: 1864–74.

Flocke, S. 1997. "Measuring Attributes of Primary Care: Development of a New Instrument." *Journal of Family Practice* 45(1): 64–74.

Haggerty, J.L., F. Burge, M.-D. Beaulieu, R. Pineault, C. Beaulieu, J.-F. Lévesque et al. 2011. "Validation of Instruments to Evaluate Primary Healthcare from the Patient Perspective: Overview of the Method." *Healthcare Policy* 7 (Special Issue): 31–46.

Haggerty, J., F. Burge, J.-F. Lévesque, D. Gass, R. Pineault, M.-D. Beaulieu and D.A. Santor. 2007. "Operational Definitions of Attributes of Primary Health Care: Consensus among Canadian Experts." *Annals of Family Medicine* 5: 336–44.

Haggerty, J.L., J.-F. Lévesque, D.A. Santor, F. Burge, C. Beaulieu, F. Bouharaoui et al. 2011. "Accessibility from the Patient Perspective: Comparison of Primary Healthcare Evaluation Instruments." *Healthcare Policy* 7 (Special Issue): 94–107.

Haggerty, J. and C.M. Martin. 2005. *Evaluating Primary Health Care in Canada: The Right Questions to Ask.* Ottawa: Health Canada, Primary and Continuing Healthcare Division.

Health Canada. 2007 (March 1). *Primary Health Care Transition Fund.* Retrieved July 28, 2011. <http://www.hc-sc.gc.ca/hcs-sss/prim/phctf-fassp/index-eng.php>.

Little, P., H. Everitt, I. Williamson, G. Warner, M. Moore, C. Gould et al. 2001. "Observational Study of Effect of Patient Centredness and Positive Approach on Outcomes of General Practice Consultations." *British Medical Journal* 323: 908–11.

Macinko, J., H. Montenegro, C. Nebot and PAHO Working Group on Primary Health Care. 2007. *Renewing Primary Health Care in the Americas: A Position Paper of the Pan American Health Organization/World Health Organization (PAHO/WHO).* Washington, DC: Pan American Health Organization.

Madore, Odette. 2005. *The Canada Health Act: Overview and Options.* Report no. CIR94-4E. Ottawa: Library of Parliament, Parliamentary Information and Research Service.

Merriam-Webster's Collegiate Dictionary. 1998. Springfield, MA: Author.

Romanow, R.J. 2002. *Building on Values: The Future of Health Care in Canada. Final Report.* Regina: Commission on the Future of Health Care in Canada.

Santor, D.A., J.L. Haggerty, J.-F. Lévesque, F. Burge, M.-D. Beaulieu, D. Gass et al. 2011. "An Overview of Confirmatory Factor Analysis and Item Response Analysis Applied to Instruments to Evaluate Primary Healthcare." *Healthcare Policy* 7 (Special Issue): 79–92.

Stewart, A.L., A.M. Nápoles-Springer, S.E. Gregorich and J. Santoyo-Olsson. 2007. "Interpersonal Processes of Care Survey: Patient-Reported Measures for Diverse Groups." *Health Services Research* 42: 1235–56.

Stewart, M., J.B. Brown, A. Donner, I.R. McWhinney, J. Oates, W.W. Weston and J. Jordan. 2000. "The Impact of Patient-Centered Care on Outcomes." *Journal of Family Practice* 49: 796–804.

Stewart, M., J.B. Brown, W.W. Weston, I.R. McWhinney, C.L. McWilliam and T.R. Freeman. 2003. *Patient-Centered Medicine: Transforming the Clinical Method* (2nd ed.). Oxford: Radcliffe Medical Press.

Comprehensiveness of Care from the Patient Perspective: Comparison of Primary Healthcare Evaluation Instruments

La globalité des soins du point de vue du patient : comparaison entre instruments d'évaluation des soins de santé primaires

JEANNIE L. HAGGERTY, MARIE-DOMINIQUE BEAULIEU, RAYNALD PINEAULT, FREDERICK BURGE, JEAN-FRÉDÉRIC LÉVESQUE, DARCY A. SANTOR, FATIMA BOUHARAOUI AND CHRISTINE BEAULIEU

TABLE 1. Distribution of responses to items in subscales mapped to comprehensiveness of services and whole-person care, grouped within instrument subscale (n=645)

Item Code	Item Statement	Missing % (n)	Percentage (Number) by Response Option						Item Discrimination*
Comprehensiveness of Services									
CPCI Comprehensive Care									
			1=Strongly disagree	**2**	**3**	**4**	**5**	**6=Strongly agree**	
CP_cc1	I go to this doctor for almost all my medical care.	2 (10)	3 (18)	1 (8)	7 (42)	6 (41)	15 (96)	67 (430)	2.30 (0.21)
CP_cc2	This doctor handles emergencies.	6 (36)	10 (63)	8 (52)	17 (109)	12 (80)	17 (111)	30 (194)	1.32 (0.13)
CP_cc3	This doctor can take care of almost any medical problem I might have.	3 (19)	6 (40)	6 (40)	10 (64)	19 (122)	23 (148)	33 (212)	2.50 (0.18)
CP_cc4	I could go to this doctor for help with a personal or emotional problem.	3 (22)	7 (43)	7 (43)	9 (58)	13 (83)	19 (122)	42 (274)	2.51 (0.19)
CP_cc5	I could go to this doctor for care of an ongoing problem such as high blood pressure.	2 (11)	3 (17)	2 (11)	5 (33)	7 (42)	16 (102)	67 (429)	3.00 (0.26)
CP_cc6	I could go to this doctor for a check-up to prevent illness.	2 (11)	4 (24)	3 (22)	5 (30)	8 (52)	21 (134)	58 (372)	1.90 (0.16)
PCAT-S Comprehensiveness – Services Available									
Following is a list of services that you or your family might need at some time. For each one, please indicate whether it is available at your Primary Care Provider's office.			**1=Definitely not**	**2=Probably not**	**3= Probably**	**4= Definitely**	**Not sure / Don't remember**		
PT_csa1	Immunizations (shots)	2 (15)	5 (33)	3 (22)	13 (82)	68 (441)	8 (52)		2.07 (0.21)
PT_csa2	Family planning or birth control methods	7 (45)	5 (32)	2 (15)	17 (110)	55 (352)	14 (91)		2.10 (0.21)
PT_csa3	Counselling for mental health problems	6 (38)	9 (57)	12 (79)	23 (146)	29 (188)	21 (137)		2.00 (0.19)
PT_csa4	Sewing up a cut that needs stitches	5 (29)	11 (71)	12 (77)	18 (118)	30 (194)	24 (156)		1.84 (0.19)
PCAT-S First-Contact Utilization									
			1=Definitely not	**2=Probably not**	**3= Probably**	**4= Definitely**	**Not sure / Don't remember**		
PT_fcu1	When you need a regular general checkup, do you go to your Primary Care Provider before going somewhere else?	1 (7)	2 (12)	1 (6)	10 (64)	86 (554)	0 (2)		4.70 (0.60)
PT_fcu2	When you have a new health problem, do you go to your Primary Care Provider before going somewhere else?	1 (8)	2 (13)	2 (15)	12 (79)	82 (528)	0 (2)		4.59 (0.54)
PT_fcu3	When you have to see a specialist, does your Primary Care Provider have to approve or give you a referral?	2 (10)	2 (16)	4 (27)	23 (151)	65 (418)	4 (23)		0.87 (0.13)

TABLE 1. Continued

Item Code	Item Statement	Missing % (n)	Percentage (Number) by Response Option						Item Discrimination*
Whole Person Care (community aspect only)									
CPCI Community Context									
			1=Strongly disagree	**2**	**3**	**4**	**5**	**6=Strongly agree**	
CP_com1	This doctor knows a lot about my community.	4 (24)	8 (51)	9 (59)	13 (84)	15 (94)	20 (132)	31 (201)	**
CP_com2	This doctor uses her/his knowledge of my community to take care of me.	5 (30)	11 (69)	10 (67)	11 (71)	14 (92)	16 (105)	33 (211)	**
PCAT-S Community Orientation									
			1=Definitely not	**2=Probably not**	**3= Probably**	**4= Definitely**	**Not sure / Don't remember**		
PT_co1	Does anyone at your Primary Care Provider's office ever make home visits?	2 (12)	30 (193)	18 (118)	9 (61)	9 (57)	32 (204)		1.54 (0.18)
PT_co2	Does your Primary Care Provider know about the important health problems of your neighborhood?	2 (14)	20 (128)	18 (114)	23 (149)	11 (69)	27 (171)		2.61 (0.23)
PT_co3	Does your Primary Care Provider get opinions and ideas from people that will help to provide better health care?	2 (13)	4 (28)	8 (54)	38 (247)	20 (130)	27 (173)		1.18 (0.15)

- Items were assessed against the construct of the original scale. Values >1 are considered to be discriminating, indicating that each unit increase in the item score corresponds to more than one unit increase in the summed subscale score.

** Calculation of a discriminant parameter in a two-item scale is inappropriate

RESEARCH PAPER

Respectfulness from the Patient Perspective: Comparison of Primary Healthcare Evaluation Instruments

Le respect du point de vue du patient : comparaison entre instruments d'évaluation des soins de santé primaires

JEAN-FRÉDÉRIC LÉVESQUE, MD, PHD
Centre de recherche du Centre hospitalier de l'Université de Montréal
Montréal, QC

RAYNALD PINEAULT, MD, PHD
Centre de recherche du Centre hospitalier de l'Université de Montréal
Montréal, QC

JEANNIE L. HAGGERTY, PHD
Department of Family Medicine, McGill University
Montreal, QC

FREDERICK BURGE, MD, MSC
Department of Family Medicine, Dalhousie University
Halifax, NS

MARIE-DOMINIQUE BEAULIEU, MD, MSC
Chaire Dr Sadok Besrour en médecine familiale
Centre de recherche du Centre hospitalier de l'Université de Montréal
Montréal, QC

DAVID GASS, MD
Department of Family Medicine, Dalhousie University
Halifax, NS

DARCY A. SANTOR, PHD
School of Psychology, University of Ottawa
Ottawa, ON

CHRISTINE BEAULIEU, MSC
St. Mary's Research Centre, St. Mary's Hospital Center
Montreal, QC

Abstract

Respectfulness is one measurable and core element of healthcare responsiveness. The operational definition of respectfulness is "the extent to which health professionals and support staff meet users' expectations about interpersonal treatment, demonstrate respect for the dignity of patients and provide adequate privacy."

Objective: To examine how well respectfulness is captured in validated instruments that evaluate primary healthcare from the patient's perspective, whether or not their developers had envisaged these as representing respectfulness.

Method: 645 adults with at least one healthcare contact with their own regular doctor or clinic in the previous 12 months responded to six instruments, two subscales that mapped to respectfulness: the Interpersonal Processes of Care, version II (IPC-II, two subscales) and the Primary Care Assessment Survey (PCAS). Additionally, there were individual respectfulness items in subscales measuring other attributes in the Components of Primary Care Index (CPCI) and the first version of the EUROPEP (EUROPEP-I). Scores were normalized for descriptive comparison. Exploratory and confirmatory (structural equation modelling) factor analyses examined fit to operational definition.

Results: Respectfulness scales correlate highly with one another and with interpersonal communication. All items load adequately on a single factor, presumed to be respectfulness, but the best model has three underlying factors corresponding to (1) physician's interpersonal treatment (eigenvalue=13.99), (2) interpersonal treatment by office staff (eigenvalue=2.13) and (3) respect for the dignity of the person (eigenvalue=1.16). Most items capture physician's interpersonal treatment (IPC-II Compassionate, Respectful Interpersonal Style, IPC-II Hurried Communication and PCAS Interpersonal Treatment). The IPC-II Interpersonal Style (Disrespectful Office Staff) captures treatment by staff, but only three items capture dignity.

Conclusion: Various items or subscales seem to measure respectfulness among currently available validated instruments. However, many of these items related to other constructs, such as interpersonal communication. Further studies should aim at developing more refined measures – especially for privacy and dignity – and assess the relevance of the broader concept of responsiveness.

Résumé

Le respect est un élément central, et mesurable, de la réactivité des services de santé. La définition opérationnelle du respect est « la mesure dans laquelle les professionnels de la santé et le personnel de soutien satisfont aux attentes des patients en matière de traitement interpersonnel et font preuve de respect pour la dignité et l'intimité des patients. »

Objectif : Examiner dans quelle mesure le respect est capté par les instruments validés qui servent à évaluer les soins de santé primaires du point de vue du patient; et voir si les concepteurs

ont envisagé, ou non, la notion du respect dans leur conception.
Méthode : Six cent quarante-cinq adultes, qui ont vécu au moins un contact avec leur clinique ou médecin régulier au cours des 12 mois antérieurs, ont répondu à six instruments d'évaluation des soins primaires, notamment deux sous-échelles qui touchent au respect : Interpersonal Processes of Care, version II (IPC-II, deux sous-échelles) et le Primary Care Assessment Survey (PCAS). De plus, le Components of Primary Care Index (CPCI) et la première version de l'EUROPEP (EUROPEP-I) comprennent des items liés au respect dans des sous-échelles qui servent à mesurer d'autres caractéristiques. Les résultats ont été normalisés pour permettre des comparaisons descriptives. Les analyses factorielles exploratoires et confirmatoires (modélisation par équation structurelle) ont servi à examiner l'adéquation à la définition opérationnelle.
Résultats : Les échelles liées au respect sont en étroite corrélation entre elles et avec la communication interpersonnelle. Tous les items présentent un point de saturation adéquat pour un facteur unique, qui est probablement le respect, mais le modèle qui s'ajuste le mieux comprend trois facteurs sous-jacents qui correspondent à (1) le traitement interpersonnel de la part du médecin (valeur propre = 13,99), (2) le traitement interpersonnel de la part du personnel de bureau (valeur propre = 2,13) et (3) le respect de la dignité (valeur propre = 1,16). La plupart des items captent le traitement interpersonnel de la part du médecin (« style interpersonnel respectueux et avec compassion » et « communication hâtive » de l'IPC-II; « traitement interpersonnel » du PCAS). Le « style interpersonnel » de l'IPC-II (personnel de bureau irrespectueux) capte le traitement de la part du personnel, mais seulement trois items captent les aspects touchant à la dignité.
Conclusion : Dans les instruments validés disponibles, plusieurs items ou sous-échelles semblent mesurer la question du respect. Cependant, plusieurs de ces items sont reliés à d'autres construits tels que la communication interpersonnelle. Des recherches plus poussées devraient viser la conception de mesures plus précises, particulièrement pour ce qui est de l'intimité et de la dignité, et l'évaluation de la pertinence du concept général de réactivité.

* * *

In 2000 the World Health Report (WHO 2000) rated the health systems of member countries on the basis of fairness, effectiveness and responsiveness. In this important document, responsiveness was defined as "the extent to which non-medical expectations of patients are met." This report highlighted the importance of meeting patients' expectations, effectively and equitably, regarding *what* care they should receive and *when* and *how* care should be provided. More specific to primary healthcare (PHC), a recent Canadian study identified responsiveness as one of the six domains important to the public (Wong et al. 2008).

Background

Conceptualizing Respectfulness

In 2004, we conducted a consensus consultation of 20 PHC experts across Canada on attributes of primary healthcare that should be evaluated in health reforms (Haggerty et al. 2007). Reflecting the WHO report, responsiveness was identified as one of the attributes to be measured. However, when mapped to validated instruments, the WHO definition of responsiveness proved to be very broad and overlapped with many other attributes, such as interpersonal communication and relational continuity. In addition, conceptualizing responsiveness as the generic ability to respond to a broad range of needs expressed by the patient did not seem specific enough for evaluative purposes. Consequently, the PHC experts narrowed the concept of responsiveness and pinpointed "respectfulness" as fitting more specifically with how providers respond to patients' expectations, leaving aside aspects related to the range of services available and timeliness in responding to patients' needs. In purely linguistic terms, responsiveness is related to the quality of reacting or responding quickly, whereas respectfulness relates to the quality of being courteous, humble, reverent and deferential (*Canadian Oxford Dictionary* 1998).

The experts agreed to define respectfulness as "the extent to which health professionals and support staff meet users' expectations about interpersonal treatment, demonstrate respect for the dignity of patients and provide adequate privacy." They unanimously agreed that this attribute is most validly evaluated from the patient's perspective.

Currently, there is little comparative information to guide evaluators in their selection of the appropriate tool for evaluating respectfulness of care. Various instruments that evaluate primary healthcare from the user's perspective address aspects related to respectfulness, although respectfulness as a construct is addressed only by one instrument, to our knowledge.

In this paper, we present and discuss results from a concurrent validation process of existing instruments that assess primary healthcare from the patient's perspective, with regard to the evaluation of respectfulness. More specifically, our objectives were to contribute to the understanding of the concept of respectfulness, given the lack of instruments available to measure it, and to explore how various instruments' items could be linked with factors that could be mapped to our operational definition. Our aim was to discern the extent to which this dimension was captured in the various instruments, whether or not the instruments' developers had envisaged these as representing respectfulness. Such an analysis of items relating to the concept of respectfulness is crucial given the emphasis in current PHC reforms on better addressing non-clinical expectations of patients and the way care is provided.

Method

The main methodological aspects have been described in detail elsewhere (Haggerty et al. 2011). Briefly: six instruments that evaluate PHC from the patient's perspective were administered to 645 healthcare users, balanced by English/French language, rural/urban location, low/high education and poor/average/excellent overall PHC experience.

Among the six instruments, two instruments comprised three subscales that mapped to our operational definition of respectfulness. The Primary Care Assessment Survey (PCAS) (Safran et al. 1998) had one, Interpersonal Treatment. The Interpersonal Processes of Care

– Version II (IPC-II) had two: Interpersonal Style (Compassionate, Respectful) and Interpersonal Style (Disrespectful Office Staff). However, the IPC-II also had a Hurried Communication subscale that comprised three items out of five mapping with respectfulness. Additionally, there were individual items that assessed aspects of respectfulness in subscales measuring other attributes, notably in the Components of Primary Care Index (CPCI) (Flocke 1997), in the first version of the European general practice instrument (EUROPEP-I) (Grol et al. 2000) and in other subscales of the IPC-II (Stewart et al. 1999). No items or subscales directly addressed the provision of privacy, which was part of the operational definition and can be defined as the ability to seclude and reveal oneself selectively. The EUROPEP-I instrument addresses the notion of confidentiality of personal information. There was also considerable conceptual overlap with the concepts of interpersonal communication skills (ability to elicit and understand the patient's concerns) and trust. Although items from the PCAS Trust subscale were initially included in the respectfulness analyses, these items were later excluded as reflecting an outcome of care rather than an attribute of primary healthcare per se. In all, we retained 25 items (including four complete subscales) for analysis.

Analytic strategy

The analysis consisted of examining the distributional statistics and subscale correlations, followed by common factor and confirmatory factor analysis to identify dimensions common to the entire set of items, as outlined in detail elsewhere (Santor et al. 2011). Our strategy for analyzing respectfulness differed from that used for other attributes of care in this study. We departed from our overall strategy of honouring the subscales that had been validated by the instrument developers and included individual items from other subscales that mapped to our operational definition. This approach was necessary given that the creators of the measurement instruments did not identify respectfulness as a concept in their instruments but rather integrated various items that are conceptually linked with the operational definition of respectfulness that our experts identified.

Results

The item content and distribution of the responses summarized in Table 1 and the details are available online at http://www.longwoods.com/content/22640. Most of the individual items we mapped to respectfulness came from subscales related to interpersonal communication. Despite our design that specifically oversampled persons with a negative experience of care, the vast majority of respondents select the two most positive response options and very few the negative options. This is a well-known bias in the evaluation of experience of care by patients and may compromise the performance of exploratory factor analysis based on ordinary least squares regression techniques.

Table 2 presents the Pearson correlations among the three subscales that had been identified to capture respectfulness, as well as for IPC-II Hurried Communication, where three of the five items mapped to respectfulness. The correlations show that while subscales measuring

TABLE 1. Summary of respectfulness subscale content and distribution of item responses. (Detailed distribution available at http://www.longwoods.com/content/22640)

Subscale and Item Description	Response Scale	Range Missing Values	Overall Modal Response	Comments
Complete Subscales				
PCAS Interpersonal Treatment (5 items) Rate the personal aspects of care received from regular doctor: Time spent with doctor; patience with questions and worries; friendliness and warmth; caring and concern; respect	Likert evaluative, 1=very poor, 6=excellent	0%–1%	6 (excellent)	Over 32% of respondents select most positive category, except for amount of time spent (between good and very good); 3%–6% the two most negative categories
IPC-II Interpersonal Style – Compassionate, Respectful (5 items) Frequency of doctor(s) behaviours… Really respecting as a person; treating as equal; being compassionate; giving support and encouragement; being concerned about feelings	Likert frequency, 1=never to 5=always	3%–4%	5 (always)	45%–63% of responses in most positive category; should anything less be acceptable?
IPC-II Interpersonal Style – Disrespectful Office Staff (4 items) Frequency of office staff behaviours: Being rude; talking down; giving hard time; having a negative attitude	Likert frequency, 1=never to 5=always	3%	5 (never)	9%–15% of responses "always" or "sometimes"; ~65% "never"
Items from Subscales				
IPC-II Hurried Communication (3 items of 5) Frequency of doctor(s) negative behaviours… Ignoring what was told; being distracted; appearing to be bothered if several questions asked	Likert frequency, 1=never to 5=always agree	4%	5 (never)	1% of responses for most negative category; 17%–18% somewhat negative
IPC-II Communication – Elicited concerns (1 item of 3) Frequency of doctor(s)… Taking concerns very seriously	Likert frequency, 1=never to 5=always agree	3%	5 (always)	49% "always"
PCAS Organizational Access (1 item of 6) Rate waiting time before the appointment starts	Likert evaluative, 1=very poor, 6=excellent	2%	4 (good)	29% "good", 7% "excellent"
PCAS Communication (1 item of 6) Rate aspect of talking with regular doctor: Attention to what you have to say	Likert evaluative, 1=very poor, 6=excellent	1%	6 (excellent)	31% "excellent"
CPCI Interpersonal Communication (2 items of 8) Agreement with statements about regular doctor: doctor does not always listen; feel doctor ignores my concerns	Semantic differential opinion, 1=strongly disagree, 6=strongly agree	2%-3%	6 (strongly agree)	Approximately 50% strongly agree (NB: two negative statements among six other positive)
EUROPEP Clinical Behaviour (1 item of 17) Keeping records and data confidential	Semantic differential rating, 1=poor, 5=excellent	3%	5 (excellent)	66% "excellent"
EUROPEP Organisation of Care (2 items of 7) Helpfulness of staff; waiting time in waiting room	Semantic differential rating, 1=poor, 5=excellent	4%–5%	4–5	Lower rating for wait times.

physician respectfulness are highly correlated among themselves, these subscales do not correlate highly with the subscale for office staff respectfulness, suggesting that these are distinct constructs. To explore the extent to which respectfulness was distinct from or similar to other attributes, we calculated the mean of the correlations between each respectfulness subscale and the subscales from the other attributes. We see strong correlations with interpersonal communication. This could either be a true correlation between distinct constructs (two constructs that tend empirically to move in similar directions) or suggest a conceptual overlap between respectfulness and interpersonal communication (not truly distinct constructs).

TABLE 2. Partial Pearson correlations* between respectfulness subscales, and the mean correlations with subscales for other attributes

Respectfulness Subscales	**PCAS Interpersonal Treatment**	**IPC Hurried Communication (3 items of 5)**	**IPC Interpersonal Style (Compassionate, respectful)**	**IPC Interpersonal Style (Disrespectful office staff)**
PCAS Interpersonal Treatment	1.000			
IPC Hurried Communication (3 items of 5)	0.64	1.000		
IPC Interpersonal Style (Compassionate, respectful)	0.68	0.69	1.000	
IPC Interpersonal Style (Disrespectful office staff)	0.27	0.32	0.30	1.000
Questionnaire Subscale				
Accessibility (Mean)	0.35	0.36	0.32	0.27
Comprehensiveness of Services (Mean)	0.32	0.33	0.37	0.18
Relational Continuity (Mean)	0.45	0.40	0.44	0.23
Interpersonal Communication (Mean)	0.65	0.61	0.65	0.28

* Controlling for language, education, rurality and overall experience of care.

Do all items measure a single attribute?

Using the subscales and items, we performed factor analysis with items mapping conceptually to respectfulness. We excluded from factor analysis all respondents with any missing values (listwise missing). This approach reduced our effective sample size from 645 to 519. Those excluded from the factor analyses were more likely than those included to be older and to have a clinic as the regular provider. Because this conservative approach can introduce bias, we repeated all the factor analyses using maximum likelihood imputation of missing values (Rubin 1987, Jöreskog and Sörbom 1996) to examine the robustness of our conclusions. Imputation increased our sample size slightly and improved the fit statistics in the confirmatory factor analysis models, but it did not change the magnitude of loadings nor the direction of our conclusions. This finding suggests that our conservative approach resulted principally in loss of statistical power rather than bias.

Although most of the items loaded reasonably well (loading >.40; see Table 3) onto a single latent variable using common factor analysis, the scree plots suggested a three-factor model. Based on our operational definition, we judged that the first factor (eigenvalue=13.99) is a reflection of "physician's interpersonal treatment" (interpersonal treatment), the second factor (eigenvalue=2.13), "interpersonal treatment by office staff" (office staff) and the third factor (eigenvalue=1.16) is related to "respect for the dignity of the person" (dignity). The EUROPEP-I rating of confidential treatment of the medical record loaded only moderately on interpersonal treatment, and the research team decided that it related conceptually to respect for the dignity of the person.

TABLE 3. Results of exploratory analysis showing factor loadings (>.40 only) of items using principal components analysis with oblique rotation (n=519)

		Factors		
		Interpersonal Treatment	**Office Staff**	**Dignity**
		eigenvalue=13.99	**eigenvalue=2.13**	**eigenvalue=1.16**
Item Code	**Item Statement**	**loading**	**loading**	**loading**
PS_oa4	How would you rate the amount of time you **wait** at your doctor's office for your appointment to **start?**	—	—	.84
PS_c2	Thinking about **talking** with your regular doctor, how would you rate the **attention** your doctor gives to what you have to say?	.83	—	—
PS_it1	Thinking about the **personal aspects** of the care you receive from your regular doctor, how would you rate the amount of **time** your doctor spends with you?	.77	—	—
PS_it2	Thinking about the **personal aspects** of the care you receive from your regular doctor, how would you rate doctor's **patience** with your questions or worries?	.87	—	—
PS_it3	Thinking about the **personal aspects** of the care you receive from your regular doctor, how would you rate doctor's **friendliness and warmth** toward you?	.86	—	—
PS_it4	Thinking about the **personal aspects** of the care you receive from your regular doctor, how would you rate doctor's **caring and concern** for you?	.89	—	—
PS_it5	Thinking about the **personal aspects** of the care you receive from your regular doctor, how would you rate doctor's **respect** for you?	.83	—	—
CP_ic2	Sometimes, this doctor does not listen to me.	.72	—	—
CP_ic6	Sometimes, I feel like this doctor ignores my concerns.	.69	—	—
EU_oa2	The helpfulness of staff (other than the doctor)	—	.52	—
EU_oa6	Waiting time in the waiting room	—	—	.81
EU_cb6	Keeping your records and data confidential	.47	—	— †
IP_cel3	How often did the doctor(s) take your health concerns very seriously?	.70	—	—
IP_hc3	How often did the doctor(s) ignore what you told them?	.69	—	—
IP_hc4	How often did the doctor(s) appear to be distracted when they were with you?	.62	—	—

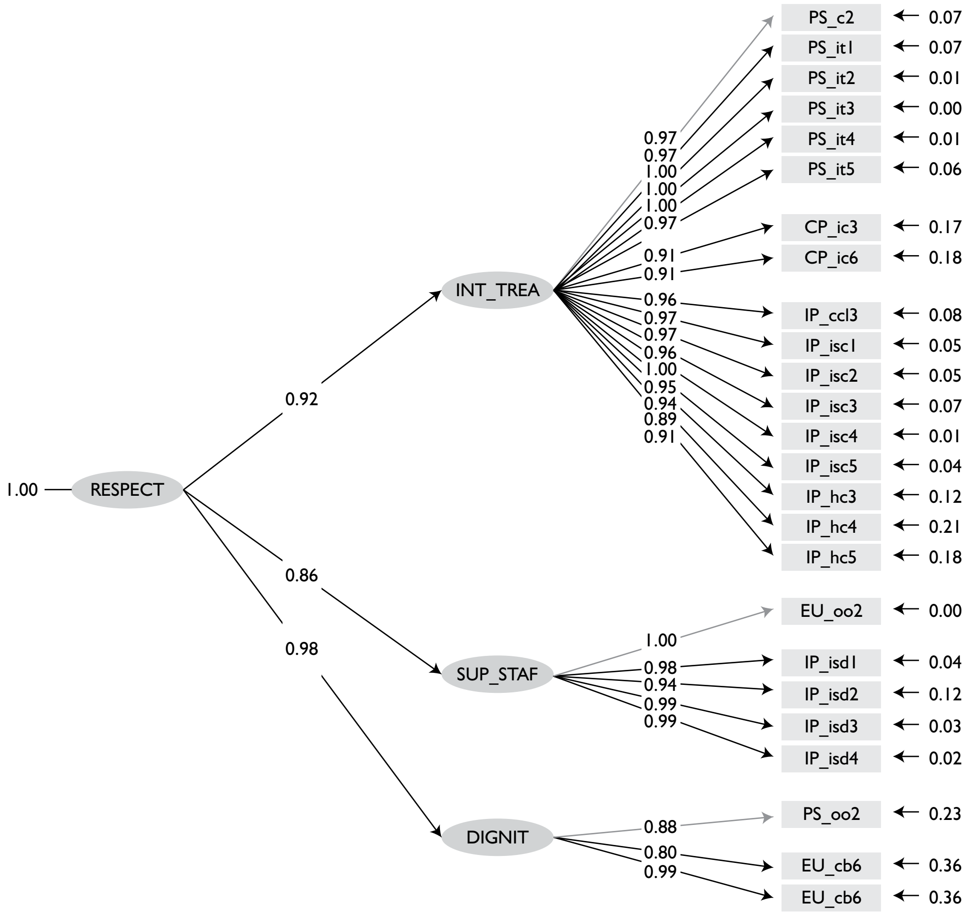
TABLE 3. Continued

Item Code	Item Statement	Factors		
		Interpersonal Treatment	Office Staff	Dignity
		eigenvalue=13.99	eigenvalue=2.13	eigenvalue=1.16
		loading	loading	loading
IP_hc5	How often did the doctor(s) seem bothered if you asked several questions?	.75	—	—
IP_isc1	How often did the doctor(s) really respect you as a person?	.81	—	—
IP_isc2	How often did the doctor(s) treat you as an equal?	.81	—	—
IP_isc3	How often was the doctor(s) compassionate?	.83	—	—
IP_isc4	How often did the doctor(s) give you support and encouragement?	.87	—	—
IP_isc5	How often were the doctor(s) concerned about your feelings?	.84	—	—
IP_isd1	How often were office staff rude to you?	—	.87	—
IP_isd2	How often did office staff talk down to you?	—	.81	—
IP_isd3	How often did office staff give you a hard time?	—	.92	—
IP_isd4	How often did office staff have a negative attitude towards you?	—	.93	—

† For confirmatory analysis, items were associated with highest-loading factors unless indicated by "†," where expert judgment prevailed.

The dimensions of interpersonal treatment and dignity are highly correlated (.91), but office staff has much lower correlations with these two dimensions (.70 and .85, respectively). Structural equation modelling confirmed the three-factor structure of the data. The model and loadings of items are presented in Figure 1.

FIGURE 1. Parameter estimations for a structural equation model showing item loadings of items on proposed sub-dimensions (first-order variables) that are subsumed on the latent factor presumed to be respectfulness



Chi-square = 1114.75, *df* = 272, *p*-value = .00000, RMSEA = .077

Discussion

We found three subscales and many items from available instruments that seemed to measure respectfulness and fit three elements in our operational definition: response to expectations in interpersonal treatment; respect for dignity; and provision of privacy by both physicians and clinic staff. Our analyses show that 18 of the 25 identified items load on one factor (interpersonal treatment). Indeed, the various experience-of-care constructs seem to correlate, and this correlation may emerge from a common latent construct. However, respectfulness by the staff is a separate construct or dimension that could come from a different reference point (e.g., the office staff rather than the physician). Finally, although our conceptual mapping placed wait time in the waiting room as a measure of respectfulness, our analysis suggests that this is probably more an indicator of accessibility, behaving very differently in factorial analyses.

Interpersonal treatment as a measure of respectfulness

The fact that a common factor seems captured by various items deserves further discussion. Many of these items, such as interpersonal communication and relational continuity, were developed for other constructs. This raises the possibility that *experience of care* in general relates to many different latent constructs, and that the same set of variables could capture different correlated constructs. This overlap could be related to the theoretical perspective adopted to subdivide the overarching concept of experience of care. In fact, the analysis of the face validity of respectfulness items suggest that healthcare providers can show their respectfulness through actions related to other attributes such as communicating well, taking time with patients, not making them wait and considering various aspects of their personality. In other words, for patients to feel respected, providers somehow have to translate their respect into actions related to communication, continuity, accessibility and comprehensiveness. Put this way, respect could be a determinant or marker of other aspects of experience of care of patients.

Interpersonal treatment and interpersonal style of communication could be part of the same construct. Given that communication and respect are closely related in the clinical encounter, communication is probably the best way to show respect, and miscommunication could generate mistrust and feelings that one is not respected or perceived as important.

Measurement implications

Our results suggest that interpersonal aspects of care have fuzzy boundaries and are probably difficult for people to distinguish. Overall, "my doctor is good and caring" might be the blanket statement that covers communication skills, relational quality and respect for the patient. The implication for measurement seems to be that dimensions related to the level of affiliation between patient and doctor are fraught with halo effects, a fact that probably impedes the clear-cut separation of concepts. Is the doctor disrespectful or a poor communicator? Is the doctor appropriate but cold? This finding might advocate for a measure of overall responsiveness, as suggested by the WHO in its framework of performance assessment, rather than the more specific concept of respectfulness. Still, such a broad concept remains difficult to measure and to link with specific improvement activities.

More recently, patient-centred care has been suggested as one essential quality of PHC and has been defined as encompassing aspects of providing care that is respectful and responsive to individual preferences, thus combining into a single concept the notion of respectfulness and responsiveness (Institute of Medicine 2001). This concept can also be useful in assessing notions of privacy and dignity, which do not seem to be captured by current instruments that assess respectfulness. This lack could be the object of further development of indices measuring patient-centred care, thus going beyond respectfulness while retaining more focus than with the broad, non-specific concept of responsiveness.

Respectfulness and quality of primary care

Our analyses suggest that respectfulness captures a wide array of the interpersonal aspects of care, and we found that among all the attributes measured, the respectfulness subscales discriminate most strongly between poor, average and excellent overall experience of care (Haggerty et al. 2011). This finding militates for a renewed interest in measuring respectfulness in healthcare evaluations.

The various Commonwealth Fund international healthcare policy surveys have highlighted that a fair proportion of the public seem to have lost confidence in the healthcare system and that what makes the difference in patients' satisfaction is more the *way* that care is provided rather than *what* care is provided (Schoen et al. 2004). Therefore, measurement of respectfulness and efforts to improve the aspects of care related to it might be levers to improve the performance of PHC care systems. This could complement the focus on medically necessary aspects of care and provide a more holistic approach to quality.

Limitations and strengths

We limited our analyses to validated instruments available in the public domain. Other instruments with various levels of validation or previous utilization exist as well and could capture some aspects related to the concept of respectfulness, such as the QUOTE questionnaire (Groenewegen et al. 2005; Kerssens et al. 2004) and the Commonwealth Fund Health Care Quality Survey (Blanchard and Lurie 2004).

We did not intend to assess respectfulness as a concept from the start. It was suggested by the operational definitions developed by the experts. These analyses represent a first assessment of the coverage and properties of items and subscales that could capture respectfulness in PHC settings. Further studies should aim at better evaluating tools and instruments for their capacity to evaluate respectfulness, develop more refined measures and assess the relevance of the broader concept of responsiveness.

Conclusion

This paper assessed how various items in current primary care experience instruments capture and relate to aspects of respectfulness. This is the first time a respectfulness lens has been applied to the most commonly used instruments. Our results show that most items related to the interpersonal aspects of treatment, while dignity and privacy are addressed poorly. However, our exploratory analysis provides guidance to evaluators to examine respectfulness in validated subscales. In the context of healthcare systems under economic stress and budgetary constraints, improvement in respectfulness could go a long way.

ACKNOWLEDGEMENTS

This research was funded by the Canadian Institutes of Health Research. The authors wish to thank Beverley Lawson for conducting the survey in Nova Scotia and Christine Beaulieu in Quebec; Fatima Bouharaoui for performing the analysis, and Donna Riley for support in preparation and editing of the manuscript.

Correspondence may be directed to: Jean-Frédéric Lévesque, Direction de santé publique de Montréal/Institut national de santé publique du Québec, 1301 Sherbrooke East, Montreal, QC H2L 1M3; tel.: 514-528-2400 ext. 3216; fax: 514-528-2470; e-mail: jflevesq@santepub-mtl.qc.ca.

REFERENCES

Blanchard, J. and N. Lurie. 2004. "R-E-S-P-E-C-T: Patient Reports of Disrespect in the Health Care Setting and its Impact on Care." *Journal of Family Practice* 53(9): 721–30.

Canadian Oxford Dictionary. 1998. Toronto: Oxford University Press.

Flocke, S. 1997. "Measuring Attributes of Primary Care: Development of a New Instrument." *Journal of Family Practice* 45(1): 64–74.

Groenewegen, P.P., J.J. Kerssens, H.J. Sixma, I. van der Eijk and W.G. Boerma. 2005. "What Is Important in Evaluating Health Care Quality? An International Comparison of User Views." *BMC Health Services Research* 5(1): 16.

Grol, R., M. Wensing and Task Force on Patient Evaluations of General Practice. 2000. *Patients Evaluate General/Family Practice: The EUROPEP instrument.* Nijmegen, Netherlands: Centre for Quality of Care Research, Raboud University.

Haggerty, J.L., F. Burge, M.-D. Beaulieu, R. Pineault, C. Beaulieu, J.-F. Lévesque et al. 2011. "Validation of Instruments to Evaluate Primary Healthcare from the Patient Perspective: Overview of the Method." *Healthcare Policy* 7 (Special Issue): 31–46.

Haggerty, J.L., F. Burge, J.-F. Lévesque, D. Gass, R. Pineault, M.-D. Beaulieu and D.A. Santor. 2007. "Operational Definitions of Attributes of Primary Health Care: Consensus among Canadian Experts." *Annals of Family Medicine* 5: 336–44.

Institute of Medicine. 2001. *Crossing the Quality Chasm: A New Health System of the 21st Century.* Washington, DC: National Academy Press.

Jöreskog, K.G. and D. Sörbom. 1996. *LISREL 8: User's Reference Guide.* Chicago: Scientific Software International.

Kerssens, J.J., P.P. Groenewegen, H.J. Sixma, W.G. Boerma and I. van der Eijk. 2004. "Comparison of Patient Evaluations of Health Care Quality in Relation to WHO Measures of Achievement in 12 European Countries." *Bulletin of the World Health Organization* 82(2): 106–14.

Rubin, D.B. 1987. *Multiple Imputation for Nonresponse in Surveys.* Hoboken, NJ: John Wiley.

Safran, D.G., J. Kosinski, A.R. Tarlov, W.H. Rogers, D.A. Taira, N. Lieberman and J.E. Ware. 1998. "The Primary Care Assessment Survey: Tests of Data Quality and Measurement Performance." *Medical Care* 36(5): 728–39.

Santor, D.A., J.L. Haggerty, J.-F. Lévesque, F. Burge, M.-D. Beaulieu, D. Gass et al. 2011. "An Overview of Confirmatory Factor Analysis and Item Response Analysis Applied to Instruments to Evaluate Primary Healthcare." *Healthcare Policy* 7 (Special Issue): 79–92.

Schoen, C., R. Osborn, P. Trang Huynh, M. Doty, K. Davis, K. Zapert and J. Peugh. 2004. "Primary Care and Health System Performance: Adults' Experiences in Five Countries: Differing Performance Levels among Countries Highlight the Potential for Improvement and Cross-National Learning." *Health Affairs* Web Exclusive: 487–503.

Stewart, A.L., A. Nápoles-Springer and E.J. Pérez-Stable. 1999. "Interpersonal Processes of Care in Diverse Populations." *Milbank Quarterly* 77(3): 305–39, 274.

World Health Organization (WHO). 2000. *The World Health Report 2000 – Health Systems: Improving Performance.* Geneva: Author.

Wong, S.T., D.E. Watson, E. Young and S. Regan. 2008. "What Do People Think Is Important about Primary Healthcare?" *Healthcare Policy* 3(3): 89–104.

Respectfulness from the Patient Perspective: Comparison of Primary Healthcare Evaluation Instruments

Le respect du point de vue du patient : comparaison entre instruments d'évaluation des soins de santé primaires

JEAN-FRÉDÉRIC LÉVESQUE, RAYNALD PINEAULT, JEANNIE L. HAGGERTY, FREDERICK BURGE,
MARIE-DOMINIQUE BEAULIEU, DAVID GASS, DARCY A. SANTOR AND CHRISTINE BEAULIEU

TABLE 1. Distribution of responses in subscales and individual items mapped to respectfulness in primary healthcare services (n=645)

Item Codes	Item Statement	Missing % (n)	Percentage (Number) by Response Option					
PCAS Interpersonal Treatment			**1=Very poor**	**2**	**3**	**4**	**5**	**6=Excellent**
PS_it1	Thinking about the **personal aspects** of the care you receive from your regular doctor, how would you rate the amount of **time** your doctor spends with you?	1 (4)	1 (9)	5 (32)	16 (104)	27 (177)	28 (178)	22 (141)
PS_it2	Thinking about the **personal aspects** of the care you receive from your regular doctor, how would you rate doctor's **patience** with your questions or worries?	1 (5)	0 (3)	3 (19)	13 (85)	23 (150)	28 (178)	32 (205)
PS_it3	Thinking about the **personal aspects** of the care you receive from your regular doctor, how would you rate doctor's **friendliness and warmth** towards you?	0 (3)	1 (8)	3 (21)	9 (57)	24 (155)	27 (176)	35 (225)
PS_it4	Thinking about the **personal aspects** of the care you receive from your regular doctor, how would you rate doctor's **caring and concern** for you?	1 (4)	0 (3)	4 (25)	10 (67)	25 (158)	27 (174)	33 (214)
PS_it5	Thinking about the **personal aspects** of the care you receive from your regular doctor, how would you rate doctor's **respect** for you?	1 (6)	1 (8)	2 (11)	6 (41)	22 (139)	28 (183)	40 (257)
IPC Interpersonal Style (Compassionate, Respectful)			**1=Never**	**2**	**3**	**4**	**5=Always**	
IP_isc1	How often did the doctor(s) really respect you as a person?	4 (23)	1 (5)	3 (18)	5 (35)	25 (159)	63 (405)	
IP_isc2	How often did the doctor(s) treat you as an equal?	4 (26)	3 (17)	4 (27)	9 (61)	29 (186)	51 (328)	
IP_isc3	How often was the doctor(s) compassionate?	4 (24)	3 (18)	7 (47)	11 (69)	27 (173)	49 (314)	
IP_isc4	How often did the doctor(s) give you support and encouragement?	4 (25)	2 (15)	9 (57)	14 (91)	23 (147)	48 (310)	
IP_isc5	How often were the doctor(s) concerned about your feelings?	3 (22)	3 (22)	10 (62)	14 (89)	25 (160)	45 (290)	
IPC Interpersonal Style (Disrespectful Office Staff)			**1=Always**	**2**	**3**	**4**	**5=Never**	
IP_isd1 Reversed	How often were office staff rude to you?	3 (20)	0 (3)	2 (12)	13 (85)	20 (131)	61 (394)	
IP_isd2 Reversed	How often did office staff talk down to you?	3 (20)	1 (5)	2 (15)	10 (62)	21 (136)	63 (407)	
IP_isd3 Reversed	How often did office staff give you a hard time?	3 (20)	0 (3)	2 (11)	7 (46)	18 (115)	70 (450)	
IP_isd4 Reversed	How often did office staff have a negative attitude towards you?	3 (21)	1 (5)	1 (8)	11 (71)	17 (110)	67 (430)	
IPC Hurried Communication (3 items of 5)			**1=Always**	**2**	**3**	**4**	**5=Never**	
IP_hc3 Reversed	How often did the doctor(s) ignore what you told them?	4 (24)	1 (8)	3 (21)	14 (88)	32 (204)	47 (300)	
IP_hc4 Reversed	How often did the doctor(s) appear to be distracted when they were with you?	4 (24)	1 (9)	4 (23)	12 (75)	36 (230)	44 (284)	
IP_hc5 Reversed	How often did the doctor(s) seem bothered if you asked several questions?	4 (26)	1 (8)	5 (31)	12 (79)	26 (168)	52 (333)	
IPC Communication (Elicited concerns, responded)			**1=Never**	**2**	**3**	**4**	**5=Always**	
IP_cel3	How often did the doctor(s) take your health concerns very seriously?	3 (22)	1 (9)	5 (31)	10 (62)	32 (207)	49 (314)	
PCAS Organizational Access (1 item of 6)			**1=Very poor**	**2**	**3**	**4**	**5**	**6=Excellent**
PS_oa4	How would you rate the amount of time you **wait** at your doctor's office for your appointment to **start**?	2 (10)	5 (34)	12 (80)	27 (177)	29 (190)	16 (106)	7 (48)

TABLE 1. Continued

Item Codes	Item Statement	Missing % (n)	Percentage (Number) by Response Option					
PCAS Communication (1 item of ...)			**1=Very poor**	**2**	**3**	**4**	**5**	**6=Excellent**
PS_c2	Thinking about **talking** with your regular doctor, how would you rate the **attention** your doctor gives to what you have to say?	1 (5)	1 (7)	4 (26)	11 (74)	22 (143)	29 (188)	31 (202)
CPCI Interpersonal Communication (2 items of ...)			**1=Very poor**	**2**	**3**	**4**	**5**	**6=Excellent**
CP_ic2 Reversed	Sometimes, this doctor does not listen to me.	2 (15)	5 (34)	7 (48)	9 (55)	7 (47)	17 (107)	53 (339)
CP_ic6 Reversed	Sometimes, I feel like this doctor ignores my concerns.	3 (19)	6 (36)	7 (45)	9 (55)	9 (56)	18 (119)	49 (315)
EUROPEP Clinical Behaviour (1 item of ...)								
EU_cb6	Keeping your records and data confidential	3 (19)		1 (5)	5 (31)	23 (147)	66 (425)	3 (18)
EUROPEP Organisation of Care (2 items of ...)								
EU_oa2	The helpfulness of staff (other than the doctor)	5 (29)	3 (18)	6 (39)	15 (98)	30 (192)	37 (236)	5 (33)
EU_oa6	Waiting time in the waiting room	4 (25)	14 (91)	12 (75)	24 (156)	30 (192)	15 (99)	1 (7)

Policy is always in the making. This journal is designed to serve readers from diverse backgrounds including health system managers, practitioners, politicians and their administrators, educators and academics. Our authors come from a broad range of disciplines including social sciences, humanities, ethics, law, management sciences and knowledge translation. They want good policy – a foundation for best practices.

www.healthcarepolicy.net

CIHR IRSC
Canadian Institutes of Health Research
Instituts de recherche en santé du Canada

UNIVERSITÉ DE SHERBROOKE

Centre de recherche
Hôpital Charles LeMoyne
Centre affilié universitaire
et régional de la Montérégie

Réseau de recherche
en santé des populations
du Québec

CIHI ICIS
Canadian Institute
for Health Information
Institut canadien
d'information sur la santé