



High and Sustained Participation in a Multi-year Voluntary Performance Measurement Initiative Among Primary Care Teams

Carol Mulder^{1*}, Jennifer Rayner^{2,3}

Abstract

Background: The province of Ontario, Canada has made major investments in interdisciplinary primary care teams. There is interest in both demonstrating and improving the quality of care they provide. Challenges include lack of consensus on the definition of quality and evidence that the process of measuring quality can be counter-productive to actually achieving it. This study describes how primary care teams in Ontario voluntarily measured quality at the team level.

Methods: Data for this 4-year observational study came from electronic medical records (EMRs), patient surveys and administrative reports. Descriptive statistics were calculated for individual measures (eg, access, preventive interventions) and composite indicators of quality and healthcare system costs. Repeated measures identified patient and practice characteristics related to quality and cost outcomes.

Results: Teams participated in an average of 5 of 8 possible iterations of the reporting process. There was variation between teams. For example, cervical cancer screening rates ranged from 21 to 86% of eligible patients. Rural teams had significantly better performance on some indicators (eg, continuity) and worse on others (eg, cancer screening). There were some statistical but small changes in performance over time.

Conclusion: High, sustained voluntary participation suggests that the initiative served a need for the primary care teams involved. The absence of robust data standards suggests that these standards were not crucial to achieve participation. The constant level of performance might mean that measurement has not yet led to improvement or that measures used might not accurately reflect improvement. The data reinforce the need to consider differences between rural and urban settings. They also suggest that further analysis is needed to identify characteristics that teams can change to improve the quality of care their patients experience. The study describes a practical, sustainable real-world approach to performance measurement in primary care that was attractive to interdisciplinary teams.

Keywords: Primary Care, Performance Measurement, Ontario

Copyright: © 2020 The Author(s); Published by Kerman University of Medical Sciences. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Citation: Mulder C, Rayner J. High and sustained participation in a multi-year voluntary performance measurement initiative among primary care teams. *Int J Health Policy Manag.* 2020;x(x):x–x. doi:10.34172/ijhpm.2020.186

Article History:

Received: 13 October 2019
Accepted: 23 September 2020
ePublished: 19 October 2020

*Correspondence to:

Carol Mulder
Email:
cmulder.infoaccess@gmail.com

Key Messages

Implications for policy makers

- It is possible to measure quality in a sustained, voluntary way even without a universal definition of quality or mandated set of indicators. Given the importance of voluntary engagement in improving quality, a shift away from developing mandated reporting requirements towards increased consideration and support of voluntary initiatives may be in order.
- Participation in performance measurement might be less dependent on the robustness or completeness of the indicators used than the literature and current investment of effort might suggest.
- Rurality could be more explicitly considered when examining performance on quality indicators.

Implications for the public

It has been difficult to date to decide exactly how to measure the quality of care people get from primary care teams. Part of the challenge has been in agreeing what high-quality care is and then finding data to easily track progress with quality. This study tells the story of a group of primary care teams who went ahead anyway, even though they were not required to. They did it the best way they could, knowing it might not be perfect but would at least be a start. They found that it worked. Most of the group took part not just once but in as many as 8 cycles over 4 years. This shows that it is possible to move forward with tracking progress with quality even in the face of uncertainty about the perfect way to do it.

Background

The province of Ontario, Canada has made major investments in the development of interdisciplinary primary care teams. Family Health Teams (FHTs), which are similar to the “patient medical home” concept in the United States and elsewhere¹ were introduced in 2005. The intent was to capitalize on the promise of team-based care to generate better patient perceptions on important outcomes like access² as well as clinical outcomes related to chronic disease management.³ In spite of this investment, Canada and Ontario have maintained a persistently poor showing on international comparisons of primary care quality.^{4,5} There is also an increasing sense that primary care teams are too expensive.^{6–8} This study arose out of an interest in both demonstrating and improving the quality of care provided by interdisciplinary primary care teams.

Part of the challenge in improving primary care is the difficulty in measuring quality. Firstly, there is a lack of consensus on the definition of quality. The Organisation for Economic Co-operation and Development and the Commonwealth Fund are two international organizations that routinely define and report national performance on quality of healthcare, including primary care. However, these frameworks and their associated indicators do not appear to work at the level of the individual provider or primary care team. For example, even though “access to an appointment on the same or next day” is part of the set of indicators commonly used in international comparisons of primary care quality,⁵ there is evidence that this measure is not meaningful to either patients or providers in Ontario.^{9,10} Deber and Schwartz¹¹ describe similar issues with other measures that are commonly used but do not resonate with providers. For these or other reasons, many different visions and frameworks of indicators of primary care quality have emerged for intended use at the individual provider or team level.^{12–15} This makes comparisons across settings regarding the impact of efforts to improve quality almost impossible.

Another difficulty in primary care performance measurement is the increasing evidence that the process of measuring quality can be counter-productive to actually achieving it. Berwick,¹⁶ one of America’s most fervent evangelists for measurement and quality in healthcare from 1980 onwards, now says the solution to improving quality in healthcare is to *stop* excessively measuring. There is little evidence that measurement has improved quality.^{14,17–19} The reverse is starting to surface. Among the observed unintended negative results of performance measurement is the risk of increased inequity as providers preferentially focus on healthier patients who they feel are more likely to have good outcomes.²⁰

Cutting back on the number of indicators is not a solution. Focussing on a small number (or just one) indicator has been shown to detract attention from other aspects of primary care with resulting untoward clinical consequences.²¹ Limiting measurement to just a small number of measures also makes it harder to reflect the comprehensive, relationship-based nature of primary care.^{13,14} As Starfield (and many others) have observed, the true value of primary care does not lie

in superior performance on specific clinical “body part” measures.^{22,23} Rather, it lies in the strength of the relationship between patients and their primary care providers,^{20,24,25} which in turn is based on the 4Cs: first Contact, Continuity, Comprehensiveness and Coordination of care.²⁶ Talbot²⁷ and Smith²⁸ have noted that when people are judged on measures that they do not feel truly reflect the quality of their work, they can get demoralized, which tends to further undermine quality.

Primary care teams want to demonstrate and continually improve the quality of care they provide. The Model for Improvement clearly positions measurement as a necessary element in efforts to improve.²⁹ However, measuring quality is problematic. There are many indicators and frameworks of indicators but there is no standard definition of quality nor ideal way to measure quality at the practice or team level. Herein lies a dilemma: “Measurement is only a handmaiden to improvement but improvement cannot act without it.”³⁰ This study therefore addresses the challenge of improvement by focussing on the process of measurement.

This study describes the experience of Family Health Teams who were members of the Association of Family Health Teams of Ontario (AFHTO). AFHTO is a voluntary membership-based organization which represents nearly all of Ontario’s FHTs and some Nurse Practitioner-Led Clinics (NPLCs). All teams had administrative leadership and staff and used an electronic medical record (EMR). Some teams established formal partnerships to share Quality Improvement specialist resources. Depending on the team, the clinical staff complement included physicians, nurse-practitioners, nurses, social-workers, dietitians, pharmacists, occupational therapists or other interdisciplinary professionals. The mix of clinicians and the formal nature of their relationship with the administrative structure of the team varied between teams. Together, these teams provide care for approximately 3 million people (approximately 25% of the province’s population). Primary care is defined here as a community (vs hospital)-based ambulatory service. This is similar to the concept of “community-oriented primary care” described by the Institute of Medicine which is characterized as being integrated, comprehensive and based on sustained partnership with patients.³¹ One of AFHTO’s key strategic directions concerned demonstrating and further supporting improvements in the quality of team-based primary care. Hence, the interest of AFHTO in this study.

The question this study addresses is “what does voluntary performance measurement look like in a collective of primary care teams in Ontario?” The study describes how primary care teams who were members of AFHTO voluntarily measured and reported their performance. It reports participation and performance across all teams and briefly explores differences between rural and urban teams to determine the need for stratification of performance analyses.

Methods

This longitudinal observational study took place over 4 years among primary care teams that were members of AFHTO.

Participants in the Measurement Initiative

On behalf of its members, AFHTO implemented a performance measurement initiative called “Data to Decisions” (D2D) which produced a performance measurement report approximately every 6 months from 2014 to 2018. All members of AFHTO (approximately 192 organizations, depending on the year) were invited to participate in each of the 8 iterations of D2D. Participation was voluntary. Participation was evaluated through the number of teams contributing data and the number of indicators for which data were contributed.

Data Sources for the Measurement Initiative

All data reported in D2D were submitted via a secure, web-based form to AFHTO by participating teams. All indicators were submitted at the team, vs individual patient, level. There was no patient-level data involved in D2D. The data came from three different sources: EMRs, patient surveys and administrative data reports. EMR data came from whichever EMR system was in place at the participating primary care team. This could be any of more than a dozen different systems in use in Ontario at the time of the study. Teams extracted data from their EMR according to guidance in a data dictionary publicly available via AFHTO’s website. Patient survey data came from surveys administered as part of normal operations within the team. Teams submitted data only for those survey questions that aligned with the wording in the data dictionary. Over time, some teams adjusted the wording of their patient surveys to increase their ability to submit data to D2D. The third source of data was an administrative data report³², definitions for which were published by the report’s producer, Health Quality Ontario. Because the administrative report was produced only for physicians, indicators based on data from these reports were not available to and therefore not submitted by NPLCs to D2D.

Indicators Included in the Measurement Initiative

In keeping with the voluntary, member-driven nature of the initiative, the indicators included in D2D were selected by members through a modified Delphi process.^{33,34} These indicators covered common topics in primary care performance such as cancer screening, access and patient experience. D2D also included some novel indicators which are described in more detail here. The Diabetes composite score is an example of an indicator based on EMR data. The Diabetes composite score follows the lead of Minnesota and Wisconsin³⁵ who were among the first to report a single metric reflecting several aspects of diabetes management. The version used here includes appropriate hemoglobin A1C (HbA1C) testing, appropriate HbA1C and blood pressure levels and cardiovascular protection via statin therapy, all based on clinical guidelines at the time. The score is presented as a percentage of the maximum possible total score.

Per capita healthcare system cost data is an example of an externally defined and calculated indicator sourced from administrative data reports.

Quality was represented in D2D as a composite measure composed of 14 individual performance indicators. The

composite quality score was based on the normalized performance of the individual components, each of which were weighted according to patients’ perception of the importance of the indicator in their relationship with their primary care provider.³⁶ The composite quality score intentionally combined performance data on technical measures (eg, cancer screening) with measures of patient experience (eg, perception of courtesy of office staff) in an attempt to reflect a more comprehensive view of quality. The composite score is represented as a percentage of the maximum possible total score. More details of the composite are available on the AFHTO website.

In addition to these performance indicators, teams provided data about their patient panel, self-reported status as a rural or urban team and teaching status.

Analysis of Participation and Performance Data

Descriptive statistics were calculated for participation and performance on the indicators in the D2D reports. Data from the first iteration of D2D were excluded as some of the relevant indicators were not introduced until later. Differences in the number of indicators for which teams contributed data resulted in a different number of teams with complete data for each indicator. Random effect, random null model analyses were conducted using Mixed Linear Models procedures in SPSS³⁷ to account for repeated measures in teams that contributed data for the same indicators in multiple iterations of D2D.

Linear regression scores were calculated to describe performance in relationship to rurality, a characteristic important to the teams involved in the study. Linear regression scores were also used to describe performance over time. Sample sizes for linear regressions were a function of the number of teams with data for each performance indicator for each iteration and thus are higher than the total number of teams participating in D2D.

Results

Table 1 describes the teams participating in at least one of the 8 iterations of D2D. It summarizes characteristics of the teams and their patient panels. Sample size for each characteristic varied because not all teams chose to provide data for all indicators in all iterations. The number of teams providing data for each element at least once in the eight iterations is indicated in the table.

Over the 4 years of the initiative, 174 teams contributed data to at least one iteration of D2D. On average, teams participated in 5 iterations. More than 60% of AFHTO members contributed to each of the iterations beyond the first two. Teams varied in patient panel size, setting, geographic location and teaching status.

Table 2 describes mean performance across all 8 iterations combined. It is limited to the indicators that were part of D2D in the eighth and most recent version. The descriptive statistics were calculated using repeated measures techniques to account for multiple data points of performance across multiple iterations. As with team characteristics, there was

Table 1. Characteristics of All Teams Participating in at Least One Iteration of D2D

Characteristic	Estimate	95% CI		Number of Teams	Minimum	Maximum
		Lower Bound	Upper Bound			
Mean number of D2D iterations per team	4.77	4.61	4.93	174	2	8
Rural (percent of teams)	41.82	35.25	48.38	174	n/a	n/a
Academic (percent of teams)	12.27	8.14	16.40	174	n/a	n/a
Mean patient panel size	19 772.86	16 233.15	23 312.57	150	813	197 994
Mean SAMI score	1.01	0.99	1.03	155	0.71	1.44
Hospital-EMR integration (percent of teams)	61.16	55.80	66.53	174	n/a	n/a

Abbreviations: EMR, electronic medical record; SAMI, Standardized Acute-Clinical-Groups Morbidity Index; D2D, Data to Decisions.

Table 2. Mean Performance of All Participating Teams Over All Iterations of D2D, With Indicators in Order of Appearance in D2D

Indicator (and Brief Definition – Complete Definitions in Supplementary File 1)	Data Source	Mean	SE	95% CI		No. of Teams	Minimum	Maximum
				Lower Bound	Upper Bound			
Patient-reported courtesy of office staff (% of survey respondents)	Patient survey	89.20	0.60	88.02	90.38	128	45.90	100.00
Patient-reported involvement in decision-making (% of survey respondents)	Patient survey	89.67	0.37	88.94	90.40	156	53.10	100.00
Colorectal cancer screening (% of eligible patients – indicator not available for NPLCs)	Admin. report*	67.26	0.60	66.07	68.45	161	26.10	84.00
Cervical cancer screening (% of eligible patients – indicator not available for NPLCs)	Admin. report	67.65	0.67	66.33	68.97	161	21.00	86.00
Childhood immunization (% immunized according to guidelines)	EMR	65.96	1.44	63.11	68.81	149	2.50	100.00
Diabetes management score (% of patients receiving care according to guidelines)	EMR	65.89	0.85	64.20	67.58	126	7.40	94.60
Patient reports a reasonable wait for appointment (% of survey respondents)	Patient survey	78.89	0.94	77.02	80.76	122	31.00	100.00
Patient reports ability to get appointment on same/next day (% of survey respondents)	Patient survey	56.53	2.66	51.27	61.78	155	7.14	100.00
Continuity (% of patients seeing own physician – not available for NPLCs)	Admin. report	66.27	0.90	64.50	68.04	157	13.50	90.40
Readmission within 30 days (% of patients with selected conditions – not available for NPLCs)	Admin. report	5.85	0.11	5.64	6.06	157	1.60	15.50
Total (adjusted) per capita cost (Canadian dollars per patient – not available for NPLCs)	Admin. report	2495.56	27.31	2441.45	2549.67	113	1643.07	4030.11
Overall quality score (% of maximum possible score)	Multiple	55.59	0.77	54.08	57.11	165	14.06	89.23

Abbreviations: EMR, electronic medical record; D2D, Data to Decisions; NPLCs, Nurse Practitioner-Led Clinics; SE, standard error.

*Admin. Report: MyPractice, an administrative data report produced by Health Quality Ontario.

variation between teams. For example, the percent of eligible patients screened for cervical cancer in the time period covered by each iteration ranged from 21 to 86%. The smallest range for any of the percentage-type indicators was 47% (for patient involvement in decision-making).

Univariate correlation analyses of performance of rural and urban primary care teams are presented in Table 3. The direction of the differences is inferred from the sign of the T-value where negative values signify lower performance among rural teams. Table 3 shows that rurality was significantly related to performance on all but two of the performance indicators: readmissions and childhood immunization rates. The direction of the difference varied, with rural teams showing higher performance on some indicators (eg, continuity) but lower on others (eg, cancer screening). For this reason, subsequent correlations were stratified according

to rurality.

Table 4 shows performance over time as measured through linear regression of the performance indicators against iterations of D2D. Coefficients for colorectal cancer screening and overall quality were statistically significant in both urban and rural teams, suggesting improvement over time. There were also statistically significant increases in performance for several other indicators but these differed between rural and urban teams. For example, there was a statistically significant increase in diabetes management performance in urban teams but not rural. In contrast, there was a statistically significant increase in continuity in rural teams but not urban. The practical significance of any of these observed improvements is questionable, however, since the coefficients are small and the R-squared values associated with the regression models are low, with none more than 0.06.

Table 3. Differences in Performance Between Rural and Urban Primary Care Teams, With Statistically Significant Differences ($P < .050$) Indicated in Italics and Indicators Presented in Order of Appearance in D2D^a

Indicator (and Brief Definition – Complete Definitions in Supplementary File 1)	T Value	Significance (2-tailed)	Mean Difference	95% CI	
				Lower Bound	Upper Bound
<i>Patient-reported courtesy of office staff (% of survey respondents)</i>	2.74	.01	2.11	0.59	3.62
<i>Patient-reported involvement in decision-making (% of survey respondents)</i>	-2.19	.03	-1.09	-2.07	-0.11
<i>Colorectal cancer screening (% of eligible patients – indicator not available for NPLCs)</i>	-6.13	.00	-3.68	-4.85	-2.50
<i>Cervical cancer screening (% of eligible patients – indicator not available for NPLCs)</i>	-5.23	.00	-3.37	-4.64	-2.11
Childhood immunization (% immunized according to guidelines)	0.16	.88	0.27	-3.15	3.70
<i>Diabetes management score (% of patients receiving care according to guidelines)</i>	2.73	.01	2.90	0.81	4.98
<i>Patient reports a reasonable wait for appointment (% of survey respondents)</i>	-3.53	.00	-4.15	-6.45	-1.84
<i>Patient reports ability to get appointment on same/next day (% of survey respondents)</i>	-2.37	.02	-5.97	-10.92	-1.02
<i>Continuity (% of patients seeing own physician – not available for NPLCs)</i>	5.16	.00	5.05	3.13	6.98
Readmission within 30 days (% of patients with selected conditions – not available for NPLCs)	0.79	.43	0.14	-0.20	0.48
<i>Total (adjusted) per capita cost (Canadian dollars per patient – not available for NPLCs)* (higher values less desirable)</i>	2.19	.03	59.91	6.20	113.62
<i>Overall quality score (% of maximum possible score)</i>	-3.97	.00	-3.84	-5.74	-1.94

Abbreviations: D2D, Data to Decisions; NPLCs, Nurse Practitioner-Led Clinics.

^aThe direction of the differences is inferred from the sign of the T-value where negative values signify lower performance among rural teams.

Table 4. Results of Univariate Linear Regressions of Individual D2D Indicators (Dependent Variable) on Iteration of D2D, Presented in Order of Appearance of Indicators in D2D

Dependent Variable (and Brief Definition – Complete Definitions in Supplementary File 1)	Urban				Rural			
	N	Co-efficient	P Value	R-square	N	Co-efficient	P Value	R-square
Patient-reported courtesy of office staff (% of survey respondents)	251	-0.135	.658	0.001	201	-0.614	.038	0.021
Patient-reported involvement in decision-making (% of survey respondents)	385	0.245	.184	0.005	302	0.196	.294	0.004
Colorectal cancer screening (% of eligible patients – indicator not available for NPLCs)	418	0.818	.000	0.059	338	1.038	.000	0.051
Cervical cancer screening (% of eligible patients – indicator not available for NPLCs)	401	-0.278	.146	0.005	334	0.051	.844	0.000
Childhood immunization (% immunized according to guidelines)	304	-0.718	.175	0.006	264	-1.520	.019	0.021
Diabetes management score (% of patients receiving care according to guidelines)	236	1.222	.010	0.028	218	0.283	.487	0.002
Patient reports a reasonable wait for appointment (% of survey respondents)	241	-1.083	.007	0.030	183	-0.389	.442	0.003
Patient reports ability to get appointment on same/next day (% of survey respondents)	374	-0.338	.397	0.002	284	1.023	.406	0.002
Continuity (% of patients seeing own physician – not available for NPLCs)	375	0.297	.312	0.003	313	0.992	.010	0.021
Readmission within 30 days (% of patients with selected conditions – not available for NPLCs)	405	0.141	.051	0.009	327	0.058	.164	0.006
Total (adjusted) per capita cost (Canadian dollars per patient – not available for NPLCs) (higher values less desirable)	289	4.992	.628	0.001	209	19.677	.101	0.013
Overall quality score (% of maximum possible score)	408	0.910	.009	0.017	374	0.623	.079	0.008

Abbreviations: D2D, Data to Decisions; NPLCs, Nurse Practitioner-Led Clinics.

Discussion

There was high and sustained voluntary participation in D2D, suggesting that some element of the process or content was serving a need for the primary care teams involved. At the very least, the pattern of participation illustrated that it is feasible to implement a performance measurement process

on a voluntary, from-the-ground-up basis. Participation in measurement is important because measurement is a crucial element in improvement.³⁸ Given that imposed measurement expectations tend to be poorly received,^{13,27} it is useful to see that measurement initiatives do not have to be mandated to have good uptake and thus support the improvement journey.

The constant level of performance in the face of high and sustained participation in measurement could mean that measurement has not yet led to improvement in these teams. There are many possible reasons for this. Prime among them is the belief that change takes time and that even 4 years and 8 iterations might not have been sufficiently long to show improvement. In addition, not all teams participated over the entire time period, thus shortening the potential window for observing impact. Measurement error could also be a factor, as noted in the limitations section below. On the other hand, teams may have improved in areas not reflected in the indicators used here. In his warning that “It is wrong to suppose that if you can’t measure it, you can’t manage it,” Deming³⁹ suggested that the data that can be measured and reported may reflect some *but not all* of what matters in improving quality.¹¹ For example, it may not be appropriate for a team focussing on youth mental health in response to local community issues to redirect their efforts to improving cervical cancer screening, even if their rate is lower than their peers. Participation in measurement is nonetheless valuable because it can expose assumptions,⁴⁰ disrupt the status quo and thus prompt reflection and deliberate choices regarding improvement,⁴¹ even if they are not focussed on the variation illustrated in the data initially reported.

The relationship between rurality and performance is consistent with other observations that expectations⁴² and delivery of primary care in rural areas are different from urban areas.⁴³ These data do not address whether this is as it should be. However, they do suggest that stratification on the basis of rurality might be necessary to understand differences in performance in primary care across large and heterogeneous geographic areas. Currently, performance measurement reports in primary care, at least in Ontario, are not generally stratified according to rurality. These data add to the body of evidence calling for a change in that respect.

Limitations

Quality as reported in this study is based on a limited number of indicators taken individually and collectively, in the form of a composite quality measure. The “individual indicator” approach suffers from the limitations attributed to ‘body part’ measures (described above). The composite indicator approach suffers from the limitations of most composite measures, most notably that it is hard for providers to know what to do to improve performance.⁴⁴ Therefore, despite our best efforts, this vision of performance presented is still a very limited view of what primary care teams do.

It is possible that the voluntary nature of data submission may have biased the sample towards inclusion of better performing teams. However, the eventual participation of a large proportion of the membership in the report suggests the risk of this is low. The risk that the data described here are not representative of the membership is further mitigated by the relatively high level of participation (ie, over 60% of teams in any one iteration with 70% in at least 3 iterations). Nonetheless, it is possible that the teams choosing to participate are different from their peers in important aspects

affecting performance that are not reflected in these data.

Measurement error is possible since all the data examined here were self-reported by teams. The risk of this was mitigated by intense engagement to ensure comparability of self-reported indicators between teams, something that was of primary interest to most participants. Nonetheless, teams may have diverged from the specifications in the data dictionary to better support their own quality improvement initiatives.

The data are limited to the *FHT and NPLC* models currently deployed in Ontario, which cover about 25% of the primary care sector. Therefore, even if the data are representative of these teams, they cannot be considered to be representative of the entire sector.

Conclusion

We have described a practical, sustainable real-world approach to performance measurement in primary care in which teams voluntarily chose to participate. It succeeded in generating baseline data to support improvement efforts at the local (team) and collective (association) levels. The fact that this was observed in the absence of robust data quality standards suggests that while such standards may be important in understanding measurement data they were not, in this case, important in achieving *participation in* measurement.

There is still a need to address outstanding questions about the impact of team characteristics and activities commonly understood to be enablers of quality. The observed relationship between rurality and performance suggest that the characteristics of teams can affect the quality of care their patients experience. Further analysis is needed to understand which characteristics are most tightly coupled to performance as well as how soon after implementation of performance measurement change in performance might be expected. Concrete data about the specific impact of characteristics that are within a team’s control to change can help primary care teams leverage their will and skill for performance measurement to improve quality.

Acknowledgements

We would like to thank AFHTO members for sharing your journey.

Ethical issues

This study approved by University of Toronto Research Ethics Board (protocol 31773).

Competing interests

CM received salary support from AFHTO for the duration of the initiative.

Authors’ contributions

CM: concept, design, implementation, data collection, analysis and writing. JR: design, implementation, analysis, writing.

Authors’ affiliations

¹Department of Family Medicine, Queens University, Kingston, ON, Canada. ²Alliance for Healthier Communities, North York, ON, Canada. ³Department of Medicine, Western University, London, ON, Canada.

Supplementary files

Supplementary file 1. Sample Entry From Data Dictionary for Data to Decisions.

References

1. Hutchison B, Glazier R. Ontario's primary care reforms have transformed the local care landscape, but a plan is needed for ongoing improvement. *Health Aff (Millwood)*. 2013;32(4):695-703. doi:10.1377/hlthaff.2012.1087
2. Jesmin S, Thind A, Sarma S. Does team-based primary health care improve patients' perception of outcomes? evidence from the 2007-08 Canadian Survey of Experiences with Primary Health. *Health Policy*. 2012;105(1):71-83. doi:10.1016/j.healthpol.2012.01.008
3. Kiran T, Glazier RH, Campitelli MA, Calzavara A, Stukel TA. Relation between primary care physician supply and diabetes care and outcomes: a cross-sectional study. *CMAJ Open*. 2016;4(1):E80-87. doi:10.9778/cmajo.20150065
4. Glazier RH. Reflections of the quality of primary care in Canada and Israel. *Isr J Health Policy Res*. 2018;7(1):46. doi:10.1186/s13584-018-0243-y
5. Canadian Institute for Health Information (CIHI). *How Canada Compares: Results from the Commonwealth Fund 2015 International Health Policy Survey of Primary Care Physicians*. Ottawa, Ontario: CIHI; 2016.
6. Petch J, Tepper J. The Next Challenges for Primary Care in Ontario. *Healthy Debate*; 2012. <https://healthydebate.ca/2012/10/topic/community-long-term-care/the-next-challenges-for-primary-care-in-ontario>. Accessed Aug 4, 2019.
7. Grant K. Ontario's Curious Shift Away from Family Health Teams. *The Globe and Mail*. February 15, 2015.
8. Auditor General of Ontario. Chapter 3 section 3.06: Funding Alternatives for Family Physicians. Auditor General of Ontario; 2011.
9. Kiran T, O'Brien P. Challenge of same-day access in primary care. *Can Fam Physician*. 2015;61(5):399-400.
10. Premji K, Ryan BL, Hogg WE, Wodchis WP. Patients' perceptions of access to primary care: analysis of the QUALICOPC patient experiences survey. *Can Fam Physician*. 2018;64(3):212-220.
11. Deber R, Schwartz R. What's measured is not necessarily what matters: a cautionary story from public health. *Healthc Policy*. 2016;12(2):52-64.
12. Veillard J, Denny K, Tipper B, et al. Using performance measurement and monitoring for performance improvement. In: *Managing a Canadian Healthcare Strategy*. Kingston: The Monieson Centre for Business Research in Healthcare; 2015.
13. Smith PC. Performance management in British health care: will it deliver? *Health Aff (Millwood)*. 2002;21(3):103-115. doi:10.1377/hlthaff.21.3.103
14. Rogan L, Boaden R. Understanding performance management in primary care. *Int J Health Care Qual Assur*. 2017;30(1):4-15. doi:10.1108/ijhcqa-10-2015-0128
15. Haj-Ali W, Hutchison B. Establishing a primary care performance measurement framework for Ontario. *Healthc Policy*. 2017;12(3):66-79.
16. Berwick DM. Measuring physicians' quality and performance: adrift on Lake Wobegon. *JAMA*. 2009;302(22):2485-2486. doi:10.1001/jama.2009.1801
17. Propper C, Wilson D. The use and usefulness of performance measures in the public sector. *Oxford Rev Econ Policy*. 2003;19(2):250-267. doi:10.1093/oxrep/19.2.250
18. Sheldon TA. The healthcare quality measurement industry: time to slow the juggernaut? *Qual Saf Health Care*. 2005;14(1):3-4. doi:10.1136/qshc.2004.013185
19. Adair C, Simpson L, Birdsell J, et al. *Performance Measurement Systems in Health and Mental Health Services: Models, Practices and Effectiveness. A State of the Science Review*. Edmonton, Alberta: Alberta Heritage Foundation for Medical Research; 2003.
20. Eggleton K, Liaw W, Bazemore A. Impact of gaps in merit-based incentive payment system measures on marginalized populations. *Ann Fam Med*. 2017;15(3):255-257. doi:10.1370/afm.2075
21. Goddard M, Davies HT, Dawson D, Mannion R, McInnes F. Clinical performance measurement: part 2--avoiding the pitfalls. *J R Soc Med*. 2002;95(11):549-551. doi:10.1258/jrsm.95.11.549
22. Starfield B, Powe NR, Weiner JR, et al. Costs vs quality in different types of primary care settings. *JAMA*. 1994;272(24):1903-1908. doi:10.1001/jama.1994.03520240031037
23. Stange KC, Ferrer RL. The paradox of primary care. *Ann Fam Med*. 2009;7(4):293-299. doi:10.1370/afm.1023
24. Rosen AK, Reid R, Broemeling AM, Rakovski CC. Applying a risk-adjustment framework to primary care: can we improve on existing measures? *Ann Fam Med*. 2003;1(1):44-51. doi:10.1370/afm.6
25. Starfield B. New paradigms for quality in primary care. *Br J Gen Pract*. 2001;51(465):303-309.
26. Starfield B. *Primary Care: Balancing Health Needs, Services, and Technology*. New York, NY: Oxford University Press; 1998.
27. Talbot C. Performance management. In: Ferlie E, Lynn LE Jr, Pollitt C, eds. *The Oxford Handbook of Public Management*. New York, NY: Oxford University Press; 2005:491-517.
28. Smith PC. Performance management: the clinician's tale. *Health Econ Policy Law*. 2015;10(3):357-360. doi:10.1017/s1744133114000474
29. Langley GJ, Moen R, Nolan KM, Nolan TW, Norman CL, Provos LP. *The Improvement Guide: A Practical Approach to Enhancing Organizational Performance*. 2nd ed. Hoboken, NJ: Jossey-Bass; 2009.
30. Berwick DM. A primer on leading the improvement of systems. *BMJ*. 1996;312(7031):619-622. doi:10.1136/bmj.312.7031.619
31. Institute of Medicine (IOM). *Community-Oriented Primary Care: A Practical Assessment. The Committee Report*. Washington, DC: National Academies Press (US); 1984.
32. Health Quality Ontario. MyPractice: Primary Care. <https://www.hqontario.ca/Quality-Improvement/Guides-Tools-and-Practice-Reports/primary-care>. Accessed August 12, 2019.
33. Dalkey NC. *The Delphi Method: An Experimental Study of Group Opinion*. Santa Monica: Rand Corp; 1969.
34. World Health Organization, Regional Office for South-East Asia. *Training Module on Development of Health Workforce Strategic Plans*. New Delhi, India: WHO Regional Office for South-East Asia; 2011.
35. Bloom FJ, Graf T, Anderer T, Stewart WF. Redesign of a diabetes system of care using an all-or-none diabetes bundle to build teamwork and improve intermediate outcomes. *Diabetes Spectr*. 2010;23(3):165-169. doi:10.2337/diaspect.23.3.165
36. Mulder C, Sunderji N. Balancing patient priorities for technical and interactional aspects of care in a measure of primary care quality. *Prim Health Care Res Dev*. 2019;20:e85. doi:10.1017/s1463423619000392
37. SPSS [computer program]. Version 24. Armonk, NY: IBM Corp; 2016.
38. Baker R, Grol R. Measuring the quality of primary medical care. In: Jones R, Britten N, Culppepper L, Gass D, Mant D, eds. *Oxford Textbook of Primary Medical Care*. Vol 1. New York: Oxford University Press Inc; 2005:464-469.
39. Deming WE. *The New Economics for Industry, Government, Education*. 2nd ed. Cambridge, USA: The MIT Press; 2000.
40. Cennamo C, Berrone P, Gomez-Mejia LR. Does stakeholder management have a dark side? *J Bus Ethics*. 2009;89(4):491-507. doi:10.1007/s10551-008-0012-x
41. van de Ven AH, Poole MS. Explaining development and change in organizations. *Acad Manage Rev*. 1995;20(3):510-540. doi:10.2307/258786
42. Weinhold I, Gurtner S. Rural - urban differences in determinants of patient satisfaction with primary care. *Soc Sci Med*. 2018;212:76-85. doi:10.1016/j.socscimed.2018.06.019
43. Green M, Gozdyra P, Frymire E, Glazier R. *Geographic Variation in the Supply and Distribution of Comprehensive Primary Care Physicians in Ontario, 2014/15*. Toronto: Institute for Clinical Evaluative Sciences; 2017.
44. Scholle SH, Roski J, Adams JL, et al. Benchmarking physician performance: reliability of individual and composite measures. *Am J Manag Care*. 2008;14(12):833-838.