

The Starfield Model; A Way of Measuring Performance in Primary Care George Southey FCFP

Introduction:

The following is a narrative of a small community family practice and its 12 year evolution of a system of performance measurement. The *in vivo* observations are not blinded or controlled but reflect the reality of comprehensive primary care and the goal of a meaningful and comparable measurement tool reflecting the full scope of primary care. Steps in the evolution include the use of a Composite Indicator (CI) to incorporate all indicators, the use of patient surveys to quantify patient expectations, the use of polling results to shape the CI, and the focus on the patient/provider relationship as opposed to the usual focus on services. The results suggest a meaningful, measurable, and comparable way of quantifying quality. The results also suggest that quality can be assured to the expectations of a population and total cost of care can be dramatically reduced without compromise of quality.

I will be describing my research and experimentation over the past 12 years in the area of performance measurement in primary care and the development of the model – The Starfield Model which I believe delivers on the promise of a meaningful, measurable and comparable primary care performance model.

The American researcher Barbara Starfield (of Johns Hopkins) observed a consistent association between well-functioning primary care, and system high quality, system high equity and system lower cost^{i,ii,iii,iv}.

In other words, if it is possible to measure and improve performance in Primary Care and through measurement improve performance, it might also be possible to improve system quality, improve system equity and reduce system cost. It might hold the key to our health system's sustainability.

Background:

My story actually starts about 50 years ago. Raised in the 60s I am an unapologetic Canadian patriot. It was the decade of our centennial, a new flag and Expo 67. We had national health care and my American cousins had Vietnam and riots. I loved my country and I wanted to be a doctor.

Twenty years later I was still a proud Canadian and a newly minted family doctor. I entered practice in a town where family medicine was deep and strong. I was committed to the ideals of my chosen profession: full scope primary care for my patients whenever and wherever they might need me.

I was committed to quality in family medicine as voiced to me by my colleagues: “Affability, Accessibility and Ability –in that order” (the 3 A’s).

I performed a self-assessment:

1. People seemed to be nice to me (Affable - check),
2. I worked long hours (Accessible - check) and

3. I thought I was bright (Ability-check).

I was living my dream....but, each provider only had their gut feeling about their own performance and the performance of their colleagues.

Mid-way through the eighties the world of primary care started to shift. It seemed that idealism was wearing thin.

Monique Begin brought in the Canada Health Act, inflation eroded payments, and practices started to change. It became more common to drop obstetrics, ER work and even hospital affiliation.

Storefront walk in clinics popped up and demonstrated that you could do very well without any obligations after your shift was over.

Offices became dingy as we worked to see how cheaply (“efficiently”) we could run our practices. Doctor groups existed mostly to share a roof and a haggard secretary, otherwise family doctors often worked in intellectual and professional isolation.

New Practice and the EMR

In 1989 I started a new practice in order to follow a dream of a “group practice” under capitation funding. Just as I committed to the move, the government pulled the plug on the model (HSOs were “too expensive”).

I started using my Electronic Medical Record (EMR) which I chose because of its ability to mine data. Despite being in the usual Fee for Service model I worked to build the new practice with others who wanted to be a part of a cohesive group committed to comprehensive primary care, quality, and group dynamic.

My understanding of quality was still limited to “the 3 A’s” but I started to expand my experience as I evaluated practices in a variety of circumstances (CPSO Peer Assessments and investigations, CMPA Defense, and Plaintiff expert opinions).

Good practice appeared to be quite common but quality was very complicated. No assessment tool really allowed an observer to compare practices other than at a very subjective level.

After evaluating several hundred practices in every corner of Ontario I had some insight into the variations of primary care.

In 2002 the opportunity to work in a capitated practice reappeared and our group became a Family Health Network. With the new model came a new perspective on quality – the preventive care bonus.

We could measure and be paid for flu shots, mammograms, pap smears and shots for kids. Cool, the three A’s now had numbers and indicators.

Dutifully we ran the EMR on April 1 each year to measure our “quality”. We submitted our results and got paid. We thought we were hot, but for some reason it didn’t seem to reflect the opinions of either doctors or patients.

It was impossible to consistently compare providers at the level of all preventive services. It also seemed quite shallow as a representation of overall primary care quality.

Apparently there was something more to quality than obsessive compulsiveness with prevention.

The ability to compare quality achievement is essential to allow an organization to use quality improvement to achieve improved quality, and to work toward strategic goals.

Comparison of one's own quality over time gives insight into your own progress.

Comparison with other practices allows innovation to be identified and spread to others seeking the same objective.

As an example, let's look at the preventive care bonuses.

		Practice A
Flu Shots	Shots Given	80
	Roster>65	100
	%	80.0%
	\$	\$2,200

When you look at the achievement for flu shots in a single practice, the achievement can be described 3 ways

1. The total number of shots
2. The proportion of the registry (that is the patients 65 or older) who received the shots

And,

3. The bonus paid for the achievement.

Without a rule or framework, the description of achievement is up to each observer.

When you look at 3 practices each reporting their achievement in flu shots, one realizes that comparison

		Practice A	Practice B	Practice C
Flu Shots	Shots Given	80	1	230
	Roster>65	100	1	300
	%	80.0%	100.0%	76.7%
	\$	\$2,200	\$2,200	\$1,100

will change with the way the achievement is described.

Without a framework describing a set of rules which define achievement, any of the three practices might be considered the highest achieving practice to which the other two practices should look for ideas.

		Practice A	Practice B	Practice C
Flu Shots	Shots Given	80	1	230
	Roster>65	100	1	300
	%	80.0%	100.0%	76.7%
	\$	\$2,200	\$2,200	\$1,100
Mammograms	Mams Done	80	1	230
	Women 50-70	100	1	300
	%	80.0%	100.0%	76.7%
	\$	\$2,200	\$2,200	\$2,200
Pap Smears	Paps Done	80	1	230
	Women 35-70	100	1	300
	%	80.0%	100.0%	76.7%
	\$	\$2,200	\$2,200	\$1,100
Kids Shots	Shots Given	90	1	260
	Roster age 2	100	1	300
	%	90.0%	100.0%	86.7%
	\$	\$1,100	\$2,200	\$0
Total	Done	330	4	950
	Targets	400	4	1200
	%	82.5%	100.0%	79.2%
	\$	\$7,700	\$8,800	\$4,400

When the three practices display their achievements for four preventive care measures, identifying relative achievement will vary with the observer's particular bias.

Without the direction and discipline of a rule-set or framework, the selection of highest achieving practice will be random.

Let's imagine that the strategic objective is to provide any preventive measure to as many people as possible.

In order to work toward the strategic objective (the most prevention) a framework is needed to focus achievement. In this simple example the framework would be the rule: “the total count for all preventive actions determines the highest achieving practice”.

		Practice A	Practice B	Practice C
Flu Shots	Shots Given	80	1	230
	Roster >65	100	1	300
	%	80.0%	100.0%	76.7%
	\$	\$2,200	\$2,200	\$1,100
Mammograms	Mams Done	80	1	230
	Women 50-70	100	1	300
	%	80.0%	100.0%	76.7%
	\$	\$2,200	\$2,200	\$2,200
Pap Smears	Paps Done	80	1	230
	Women 35-70	100	1	300
	%	80.0%	100.0%	76.7%
	\$	\$2,200	\$2,200	\$1,100
Kids Shots	Shots Given	90	1	260
	Roster age 2	100	1	300
	%	90.0%	100.0%	86.7%
	\$	\$1,100	\$2,200	\$1,100
Total	Done	330	4	950
	Targets	400	4	1200
	%	82.5%	100.0%	79.2%
	\$	\$7,700	\$8,800	\$4,400

With this framework, there will be consistent selection of Practice C as the highest achieving practice.

In order to be meaningful, measurable and comparable can we assume that all prevention is of equal value?

But is this really meaningful?

Ultimately, this question is a judgement by society. We can use science to inform opinion, but there is no getting off the hook. We need to engage in addressing the thorny issue of value.

If value were to be based on the Number Needed to Treat (NNT) in each of the four preventions, the list might look as follows because the NNT looks like this:

Equal Value		NNT
Flu Shots	Flu Shots	1:63
Pap Smears	--	1:1,400
Mammograms	Mammograms	1:400
FOBT	-	1:1,200

What a complicated mess, but that is our reality.

In 2004 I met Helen Lester from the UK who talked about the Quality Outcomes Framework^{v.vi} (QOF), which consists of a composite of 80 indicators resulting in a score out of 1,000 (along with lots of pay). REALLY COOL! I thought. This must be the missing link to Ontario Quality. We ran a bunch of our EMR Data using the QOF method and produced our scores.

We got scores but something wasn't right. The scores didn't reflect what our patients and the group knew to be where the best quality resided.

One of my colleagues was a really strong practitioner and by everyone's opinion, the best, but the score was mediocre.

To make matters worse, in Ontario there was, and is, a big problem with access but the QOF score didn't put much value on access.

Maybe "quality" in the UK was different from "quality" in Ontario.

I looked at all the QOF indicators and categorized the UK data according to its principal focus: Clinical, Organizational or Patient Relationship^{vii,viii}.

	QOF (UK)	Ontario
Clinical	65%	45%
Organization	25%	25%
Pt. Relationship	10%	30%

This table describes that analysis. The Ontario data I quantified later but I include it here to contrast expectations in Ontario and the focus in the UK QOF. Looking at the UK Score it became apparent that the UK emphasis on Disease Oriented indicators was heavy

but there was little importance ascribed to **access** or to the **humanistic** component to primary care.

If public opinion mattered, the scoring of the QOF would not work in Ontario.

I discussed the situation with Helen and discovered that the QOF scoring was negotiated between doctors and the government; public expectations did not play a role.

The QOF was a **Provider-centric** model of performance.

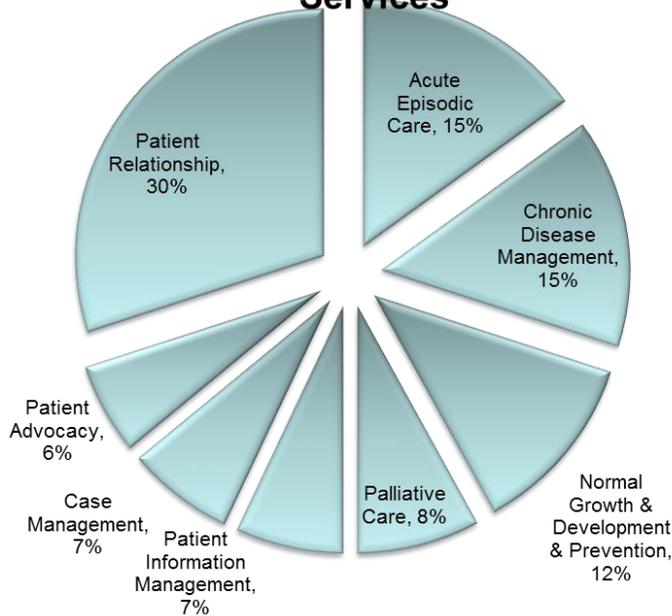
A key innovation of the QOF was the way in which all indicators were scored together allowing comparison at an overall level. Indicators were scored so that the maximum score for each indicator reflected the relative value of that indicator compared to all indicators.

Unfortunately the scoring system was not meaningful as a means of reflecting Ontario values in comprehensive primary care. There was insufficient emphasis on Access and the humanistic aspects of primary care (the Affability in the 3 As).

My curiosity turned to measuring public expectations; the art of polling. After multiple false starts, and advice from a political polling expert I started to gather results of patient expectations about their primary care.

Initially I quantified the expectations of the public for primary care **services**.

Patient Expectations of Services



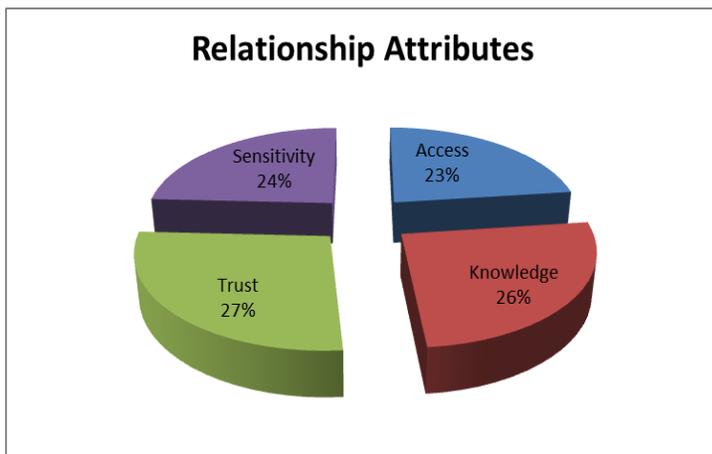
This pie-chart gives the findings from my survey and quantifies the value of the different **services in primary care**.

Near the end of her career, Starfield changed her belief regarding the way in which primary care influences the health system.

She no longer felt that the impact arose from improvement in discrete clinical services. She now felt that it was the deep committed relationship where primary care shaped the entire health system^{ix}.

In other words, she shifted focus from **What** we do, to a focus on **Why** we do primary care.

This meant that the initial polling looking at **services** needed to be reanalysed to reveal the expectations for the **relationship**



Starfield's new hypothesis was that we needed to quantify the **nature of the relationship**

After over 400 surveys, there was a consistent message that when people were asked about their future primary care, they consistently expected accessible knowledgeable and trusted service. They wanted to be known and respected. They expected their record to be current and available when and where

they needed it. They wanted their doctor to "go to bat" for them if the need arose, acting as a coordinator and as an advocate.

I realized that I was seeing a very sensible set of expectations that also reflected the goals of comprehensive primary care. From these **relationship expectations**, I weighted indicators to reflect the patient expectations.

This analysis revealed a need for indicators addressing the emotional component of the relationship, – the Affability of the 3As

To fill this need, I looked at the **Patient Survey** from CIHI^x. I found 11 survey questions addressing aspects of the emotional component between patients and their practice.

At this point, one might be concerned that tracking indicators of the relationship might only result in pleasant social interactions. Would it deal adequately with important clinical indicators such as preventive care, disease screening, and clinical parameters in chronic conditions?

The practice tracked multiple traditional indicators. These are reported on our website^{xi}. Our results suggest that providers produce quite good results when they are in an accessible, knowledgeable, trusting and sensitive relationship. We will only know for certain once data is reported from other practices for comparison.

About this time I became aware of the work of Daniel Pink and the factors influencing motivation^{xii,xiii}. Pink's three motivating factors are Purpose, Mastery and Autonomy. The health system in Ontario provides autonomy to primary care physicians. Purpose appears to be preserved within Ontario primary care and is generally described as providing the best primary care to our patients within a robust relationship. The "Mastery" motivator is fulfilled when there is a measurement process which provides feedback to providers as they progress toward their purpose.

It appeared that Pink's elements of motivation aligned with the performance model I was working on. Purpose reflected the goal of the relationship (accessible, knowledgeable, trusting and sensitive), "Mastery" was provided by the Performance Measurement system, and "Autonomy" was imbedded in Ontario's system.

It is possible that the achievement in clinical outcomes arose from assuring the relationship because it also met the criteria for motivation as described by Pink.

Now the score (of all the indicators weighted to public expectations) appeared to make sense.

- 58 Indicators
- Addressing the full patient-primary care relationship
- Weighted to reflect patient expectations
- An added benefit for practices having difficulty extracting data from their EMRs, only 8% of the indicators required EMR Data Mining. In other words, this indicator set could be used by most practices with their current capabilities.

The model achieved the three characteristics required for successful performance measurement

1. It is **Meaningful** -The measurements were meaningful to providers and patients. It reflected Pink's "Purpose".
2. It is **Measurable**
3. It is **Comparable** - At the level of comprehensive primary care, for the same practice over time and for different practices at the same time. It reflects Pink's "Mastery".

The group proceeded to track the Quality score (the composite of all indicators).

We also tracked two other key measurements.

We needed a measurement which would allow us to see how efficiently practices could provide quality care. Efficient practices manage more patients with their time.

We call this measurement Capacity. The following is a brief explanation of Capacity.

Practice A has 4,000 patients and the practice uses 100 hours of provider time each week.

Practice	Roster	Total Provider Hours	Math	Capacity
A	4,000	100	4000/100=	40
B	2,000	50	2000/50=	40
C	3,000	60	3,000/60=	50
On	13,200,000	2,220,000 *	13.2 mil/2.2 mil=	59

The Capacity is the roster divided by the hours, in this case 4000/100 or 40 patients per hour (PPH).

Practice B has 2,000 patients and operates at 50 provider hours a week. Capacity in this case is 2000/50 or 40 PPH (the same as Practice A).

* 7,400 MDs working 30 Hrs. / wk.

Practice C has 3,000 patients and operates at 60 provider hours a week. Capacity in this case is 3000/60 or 50 PPH (better efficiency than Practice A or Practice B).

All patients have different health needs. Simply counting patients is too crude; 1,000 24 year old guys are only a quarter of the needs of 1,000 geriatrics. To accurately compare different practices, adjustment to “average patients” is required. In the absence of a better method, we used the capitation table to adjust roster populations to “average patients”¹.

Ontario has 13.2 million patients, cared for by 7,400 MDs working an average of 30 hours each week. Ontario thus has an overall requirement for a capacity of 59 PPH to care for its entire population^{xiv,xv}.

A capacity of 59 PPH has been achieved Dorval Medical. It is likely that other practices have achieved similar efficiencies, but in the absence of reported data, this efficiency is hidden.

In addition to measuring Quality and Capacity, we wanted to measure cost.

Cost refers to the cost of the practice (per capita each year) and the total cost of care for the patients we served.

Small problem here...the practice had no data source to tell us about our total cost of care. Necessity being the mother of invention, I looked for a surrogate for total cost of care.

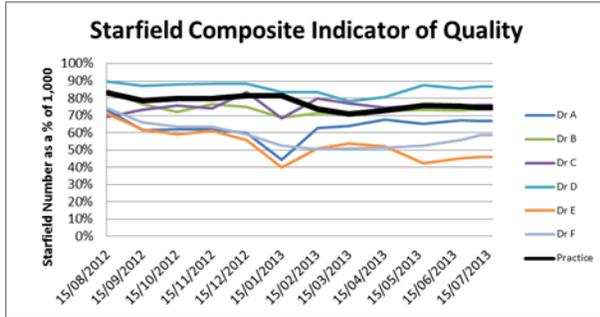
My friends in our hospital health records had an idea; why not look at the practice population’s use of the hospital. After all, hospital costs are 40% of the total cost of care^{xvi}. Not a bad idea! I tracked Length of Stay and Total Bed Days.

Now, with Quality, Capacity and Cost, we had a better picture of performance.

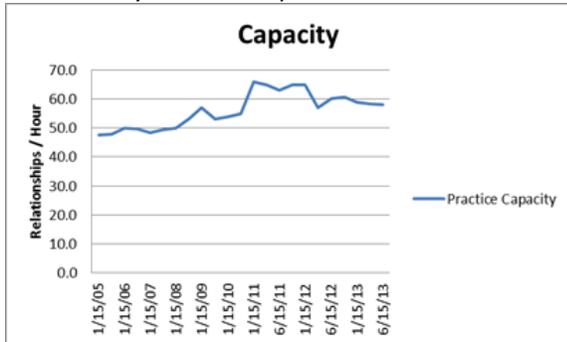
¹ Annual roster payment divided by the average roster payment gives the roster in “average patients”.

Performance is the description of three independent variables, Quality, Capacity and Cost:

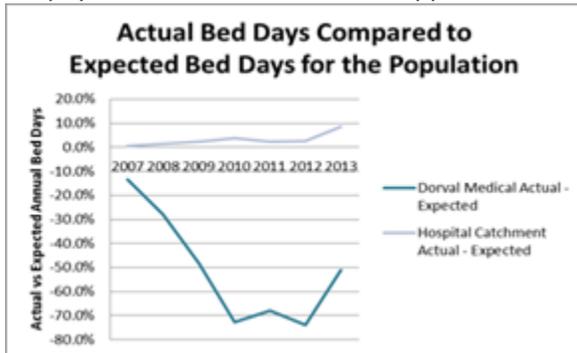
Quality is the qualitative description of what our service does. It answers the question “What do we do”



Capacity describes the amount of service we provide (also called quantity). It answers the question “How Many relationships can we sustain with a defined quality”



Cost is the cost associated with the population under our care. It describes the choice we make when we pay for the health services as opposed to using the money for other services.



Measuring these three parameters (Quality, Capacity and Cost) allows the profession and the system stewards to manage toward strategic objectives.

Let’s look at several options for strategies:

A first strategy might be called “the Silver Bullet”

Capacity is fixed at the population of Ontario 13.2 million or a capacity number of 59 PPH (see the Capacity Table).

In the “silver Bullet” strategy, cost is stabilized and the focus is on Quality.

The Quality parameter is “optimized” by a focus on seeking quality improvement without compromising Cost or Capacity.

The Silver Bullet strategy is beneficial by providing greater quality.

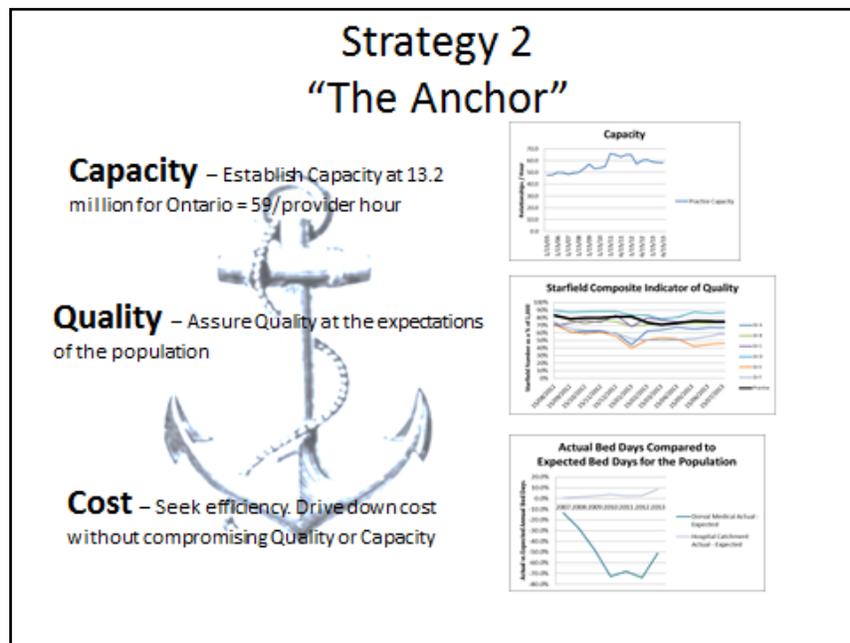
There is the hope in this strategy, that future services might be avoided, thus saving the system cost in the future.

A second Strategy example we can call “The Anchor”.

As before, Capacity is fixed at the population of Ontario 13.2 million or a capacity number of 59 PPH

In the “Anchor” strategy, Quality is the anchor for the system and is established at the expectations of the population.

The Cost parameter is “optimized” by a focus on seeking efficiency and conservation without compromising Quality or Capacity.



This was the result we appeared to achieve at Dorval Medical. Tracking performance monthly and feeding data back to my obliging and forgiving colleagues resulted in a startling observation: Quality seemed pretty good; Capacity leveled off at 59 PPH, but most interesting result was Cost. Our measurement of Cost showed a profound reduction in Total Cost of care with the practice Hospital Bed Days being substantially less than the rest of the town’s population. Apparently our patients stayed

shorter and were less likely to be admitted – almost unbelievable. I guessed that we might have a healthy population so I asked a friend who worked in ICES to run our practice data to see how it compared with the province. Our practice consists of about 10% more acute care patients than the provincial average. In other words the cost data was even more remarkable as it should have been higher than the rest of the town on the basis of our patient’s needs. This sounded like the observation Barbara Starfield observed in multiple health systems.

The Starfield Model differs from other systems of quality or performance measurement. The following table illustrates the significant differences.

Parameter	Traditional	Starfield Model
Guiding Principle	Measurement of service achievement leads to motivation for better service achievement	Assurance of the patient/provider relationship will lead to greater patient satisfaction and appropriate service achievement
Measurement Focus	Services of Primary Care	The Provider-Patient Relationship at the level of comprehensive primary care
Application of Patient Centricity	Surveys of patient experience	Patient involvement in: <ol style="list-style-type: none"> 1. Indicator selection 2. Indicator scoring 3. Indicator aggregation 4. Surveys of experience
Indicator choice	Expert consensus	Expert consensus and population polling
Indicator Subject	Expert consensus	Expert consensus and population polling
Indicator Scoring	Expert consensus	Expert consensus and population polling
Indicator selection – Dynamic?	Often slow	Intentionally fluid
Indicator Aggregation	If present - Expert consensus	Expert consensus and population polling
Aggregation Weighting method	Expert consensus	Expert consensus and population polling
Aggregation Weighting - Dynamic	Often slow	Intentionally fluid

What is next for the Starfield Model?

The Starfield Model is intended to involve patient opinion in every aspect. Current initiatives are underway to work toward this goal. Patient selection of indicators is evolving and will involve the polling of patients to determine how indicators reflect the 4 domains of the relationship (Access, Knowledge, Trust and Sensitivity). Indicator weighting is currently influenced by patient polling regarding the value attributed to the 4 domains, and will be informed by patient polling for each indicator as described.

Indicator thresholds have been arbitrarily applied using provider achievement for each indicator. The bottom 25% of providers for each indicator receives no score, and the top 25% receive the maximum

score. The middle 50% of achievement gets a portion of score based on achievement. Patient input will be incorporated by polling patients to achieve consensus on the maximum threshold.

As the practice becomes familiar with the set of indicators it should be possible to engage the patient population in selecting new indicators and altering existing indicators. Monthly polling of patients alters some parameters of indicators. The end result is an intentionally dynamic process governing indicator selection, scoring and weighting into the aggregate score for quality.

ICES has assured us that they are able to provide the Cost parameter (the total system cost for patients of a provider). This will give key feedback to practices, particularly those looking at “the Anchor” strategy”

We recognize that the adjustment to “average patient” counts is crude. ICES is developing a method to count the size of a roster which takes into account the health status of the people. This will allow most practices to be compared on the full roster parameters of Cost per patient and Capacity.

Finally, there is growing curiosity with the model. There is a group that is looking to demonstrate the Starfield Model in practices beyond Dorval Medical. Broader use of the Starfield Model will reveal important information such as normal performance in primary care. It will likely allow evolution of the Model to improve patient input and refine indicator measurement.

Conclusion:

It is possible to measure performance in comprehensive primary care by measuring the independent variables of Quality, Capacity and Cost. Quality in comprehensive primary care can be measured by a composite indicator which embraces multiple indicators, weighted to reflect the patient provider relationship. Performance measurement in this way appears to result in the assurance of Quality at the expectation of the people, establishment of Capacity at a level where all people in the province would have access to primary care, and significant reduction in total cost of care – without any compromise in Quality or Capacity. The Model appears to have characteristics which might motivate the independent Primary Care sector.

Replicating this performance measurement in other practices will provide insight into the observed response of Quality, Capacity and Cost. It might reveal a causation which would benefit the entire health system.

References

ⁱ Shi L, Starfield B, Kennedy BP, Kawachi I. Income inequality, primary care, and health indicators. *J Fam Pract.* 48 (1999), 275--84.

ⁱⁱ Starfield B. Family medicine should shape reform, not vice versa. *Fam Pract Man.* May 28, 2009; Global health, equity, and primary care. *J Am Board Fam Med.* 20(6) (2007), 511--13; Is US health really the best in the world? *JAMA.* 284(4) (2000), 483--4; Research in general practice: co-morbidity, referrals, and the roles of general practitioners and specialists. *SEMERGEN.* 29(Suppl 1) (2003), 7--16, Appendix D.

ⁱⁱⁱ Starfield B, Shi L. Policy relevant determinants of health: an international perspective. *Health Policy.* 60 (2002), 201--18.

^{iv} Starfield B, Shi L, Macinko J. Contribution of primary care to health systems and health. *Milbank Quarterly.* 83(3) (2005), 457--502.

^v <http://www.nice.org.uk/aboutnice/qof/>

^{vi} <http://www.gpcontract.co.uk/>

^{vii} <http://www.qof.ic.nhs.uk/>

^{viii} <https://catalogue.ic.nhs.uk/publications/primary-care/qof/nati-qual-outc-fram-eng-04-05/qof-eng-04-05-defi-indi-anx.pdf>

^{ix} Toward international primary care reform, Starfield, *CMAJ* • MAY 26, 2009 • 180(11)

-
- ^x http://www.cihi.ca/CIHI-ext-portal/pdf/internet/INFO_PHC_PATIENT_EN
- ^{xi} <http://www.dorvalmedical.ca/wp-content/uploads/2012/01/DMA-Performance-Indicators-July-2013.pdf>
- ^{xii} Assessing Individual Physician Performance: Does Measurement Suppress Motivation? JAMA June 27, 2012—Vol 307, No. 24
- ^{xiii} <http://www.youtube.com/watch?v=feDJ3zL23qw>
- ^{xiv} www.ophrdc.org/public/download.aspx?fileType=pdf&fileName=2010%20PIO%20Print%20Copy%20for%20Website.pdf <July 17, 2012>.
- ^{xv} www.nationalphysiciansurvey.ca/nps/2010_Survey/pdf/en/downloads/NPS2010-ON-Binder.pdf <July 17, 2012>.
- ^{xvi} CHARTING A PATH TO SUSTAINABLE HEALTH CARE IN ONTARIO, TD Economics, Special Reports May 27, 2010, Don Drummond