

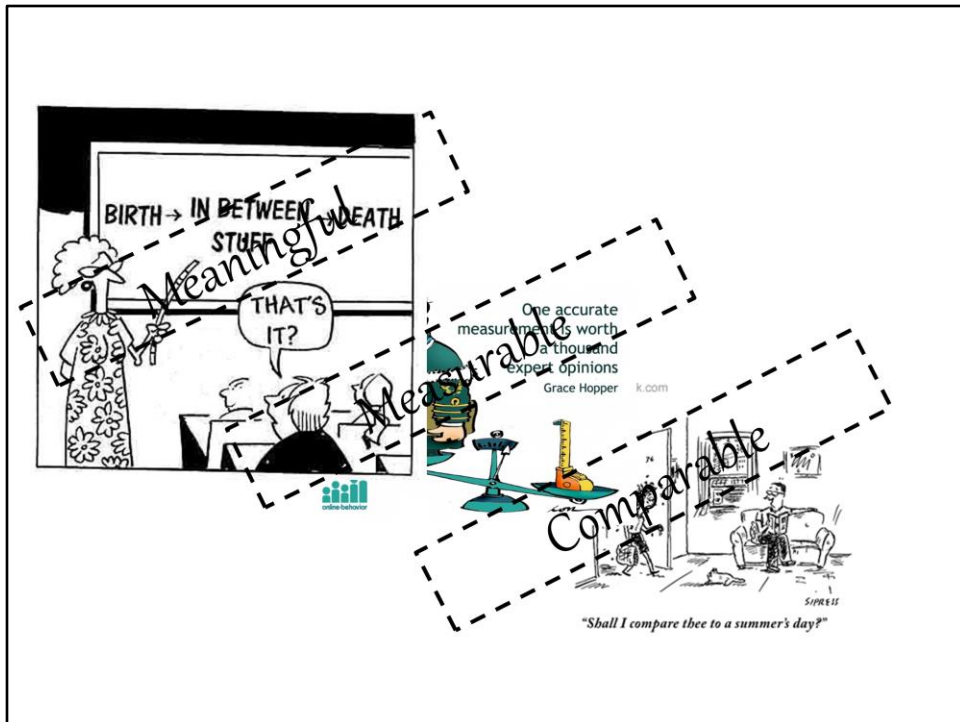


The Starfield Model

A way of Measuring Performance in Primary Care

George Southey FCFP

- I'd like to thank Danielle Martin and Nick Pimlott for inviting me to present Grand Rounds.
- 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 consistently measuring primary care performance.
- The American Epidemiologist Barbara Starfield (of Johns Hopkins) observed a consistent association between well-functioning primary care, and system high quality, high equity and lower cost.
- In other words, if it is possible to measure performance in Primary Care, it should be possible to improve system quality, improve system equity and reduce system cost. It might hold the key to our health system's sustainability.
- I will tell my story and leave plenty of time at the end for questions and discussion.



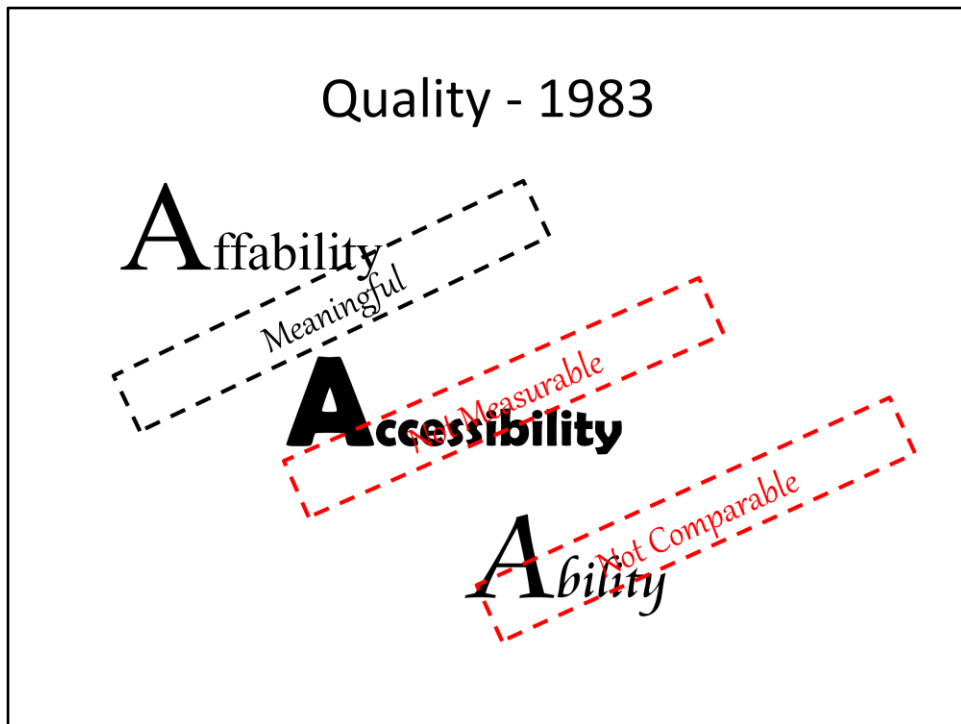
- If you only remember 3 things, they should be the three characteristics of a successful performance measurement model. It needs to be
 1. Meaningful
 2. Measurable and
 3. Comparable
- **Meaningful** to both providers and patients
- **Measurable** and
- **Comparable**- From the perspective of **comprehensive** primary care
 - Comparable for a provider **over time** and
 - Comparable **between** different providers at the same time
- Just remember the 3 words
 1. Meaningful
 2. Measurable and
 3. Comparable



- 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:

People seemed to be nice to me (Affable - check),

I worked long hours (Accessible - check) and

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.

The 80's



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

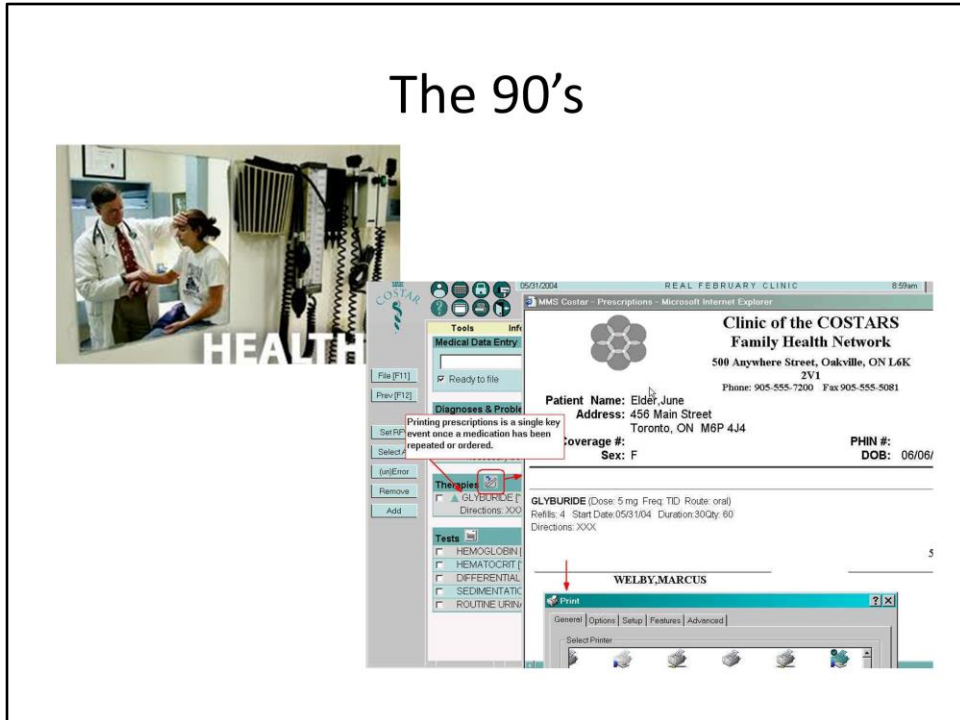
Monique Begin brought in the **Canadian Health Act**, inflation eroded payments, and practices started to change. It became more common to **drop** obstetrics, ER work and even hospital affiliation.

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 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.

The 90's



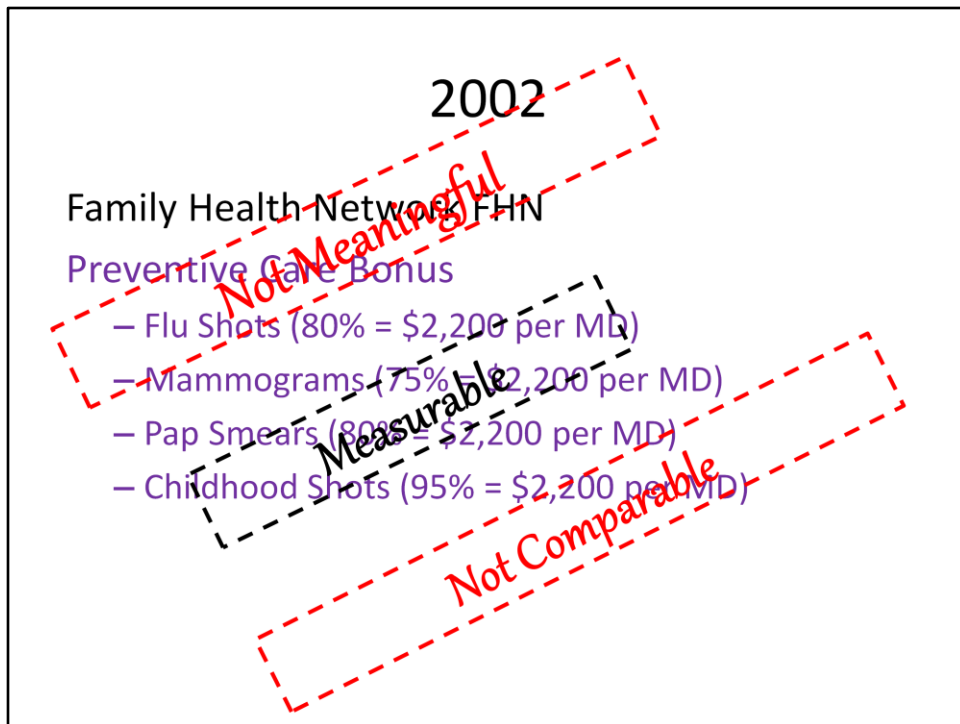
In 1989 I started a new practice in order to follow a dream of a “group practice” under capitation funding. Just as I committed to the move, the government pulled the plug on the model (**HSOs** were “too expensive”).

I started using my 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, CPSO investigation, PREP Assessments, 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 several hundred practices in every corner of the province 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 Overall Prevention

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.

Comparing Quality



The ability to **compare quality achievement** is essential to allow an organization to use **quality improvement** and with improved quality, 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.

The next few slides will demonstrate the need for a framework of rules in order to allow comparison.

One Indicator One Practice

- One Indicator (Flu Shots > 65)
- How do you describe the Quality?

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

As an example, lets look at the preventive care bonuses.

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.

One Indicator Three Practices

- One Indicator (Flu Shots > 65)

How can you compare the Quality?

		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

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

Four Indicators Three Practices

How can you compare
the Quality?

		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.

To consistently compare quality there needs to be a transparent framework

Simple Framework:

“Count the total number of people benefitting from any prevention.”

Practice C has the best Quality

		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%
	\$	\$4,180	\$8,800	\$4,400
Total	Done	330	4	950
	Targets	400	4	1200
	%	82.5%	100.0%	79.2%
	\$	\$7,700	\$8,800	\$4,400

Lets 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”.

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

2004



Quality Outcomes Framework QOF

- 80 Indicators
- Covers all services and processes
- Aggregated to a single score

In 2004 I met Helen Lester from the UK who talked about the Quality Outcomes Framework (QOF). 80 indicators gathered to result 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.

2004



Quality Outcomes Framework QOF

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

I looked at all the QOF indicators and described each according to its focus: Clinical, Organizational or Patient Relationship.

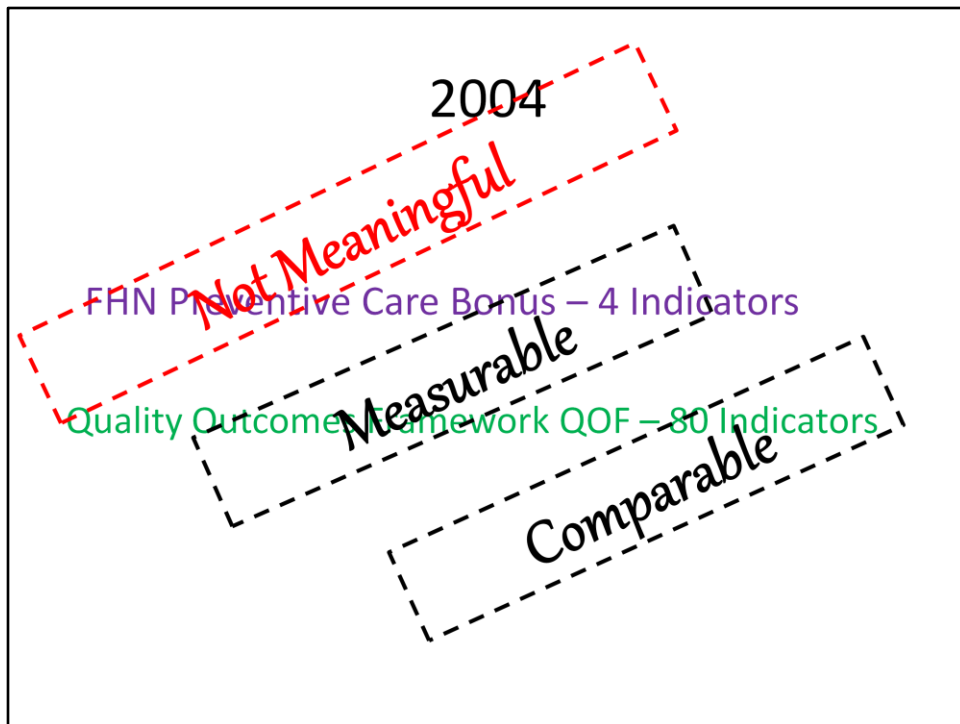
This table describes that analysis. The Ontario data I quantified later but I include it here to contrast expectations in Ontario and the focus in the QOF.

Looking at the QOF Score it became apparent that the UK emphasis on Disease Oriented indicators was heavy but there was little importance 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.

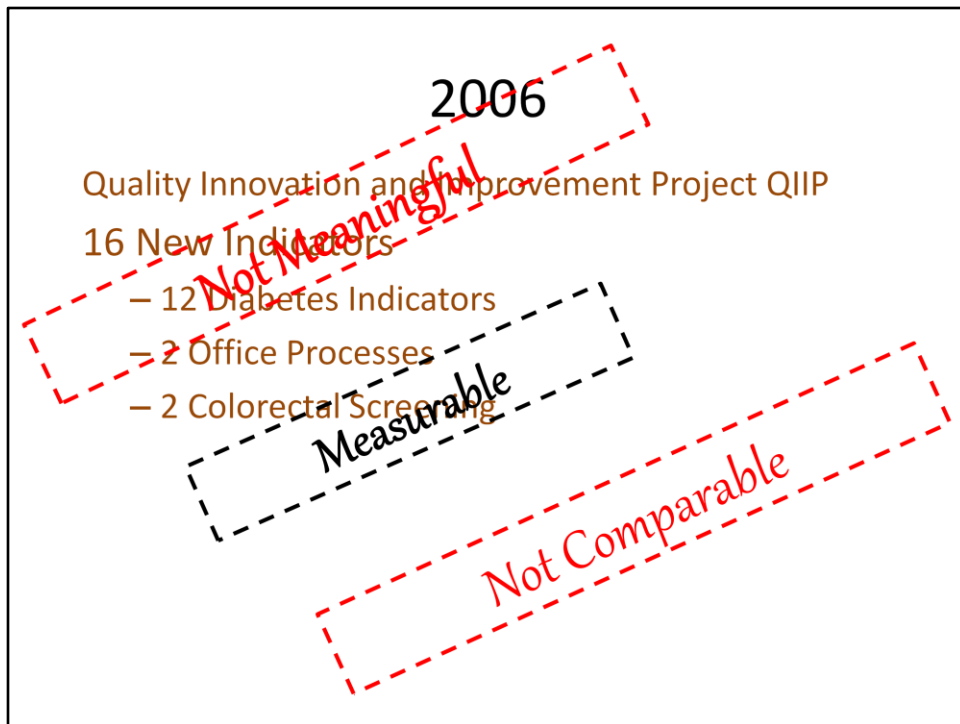
The QOF was a **Provider-centric** model of performance.



With the preventive care Indicators and the 80 QOF Indicators there appeared to be lots of things to measure.

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

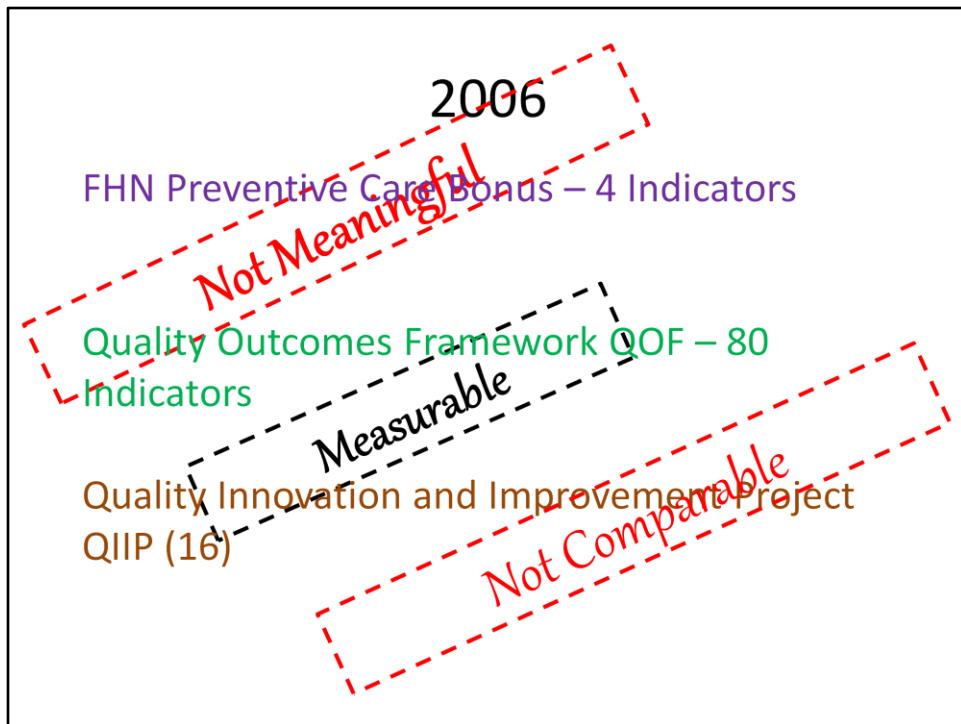
Unfortunately the scoring system was not meaningful as a means of reflecting Ontario values in comprehensive primary care. There was insufficient emphasis on Access and the humanistic aspects of primary care (the Affability in the 3 As)



In 2006, Ontario created the Quality Innovation and Improvement Partnership – QIIP

About 16 new indicators were added, providing insight into Diabetes, Colo-rectal cancer screening, and access

These new indicators did not have a method for overall scoring and as a result, practices reporting their QIIP achievements could not be compared.



Now, despite all the indicators;

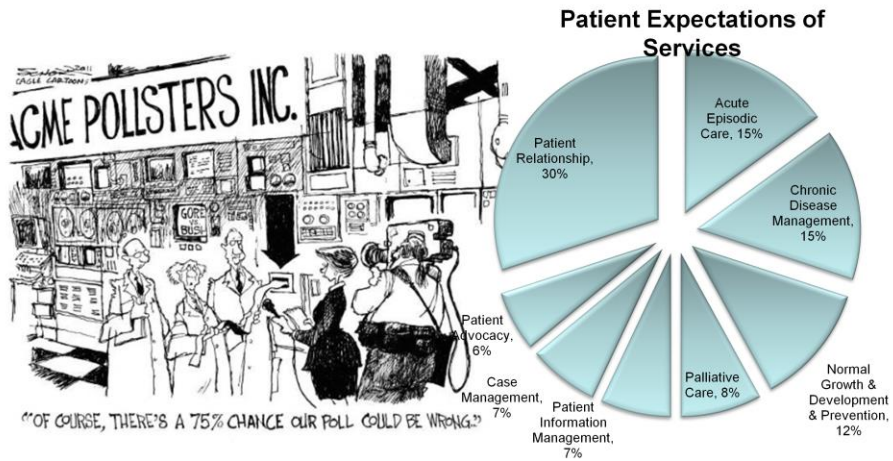
from the preventive care bonuses, to the QOF indicators and now the QIIP indicators,

we had lots of measurement, but as a group they were

not meaningful, they did not reflect Ontario values in primary care

And not comparable at the level of comprehensive primary care

2008



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**.

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

2010 - 2013



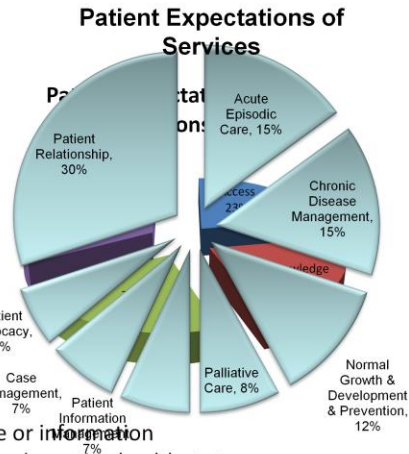
“The key components of primary care at the clinical level include access to and use of first-contact care, patient-focused (rather than disease-focused) care over time for defined populations, services that are comprehensive and timely, and coordination of care when patients need services elsewhere.”

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.

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

2010 - 2013



Access – to provider time or information
 Knowledge – of patient and practice health status
 Trust – That expected services will be provided
 Sensitivity – to emotion in the relationship

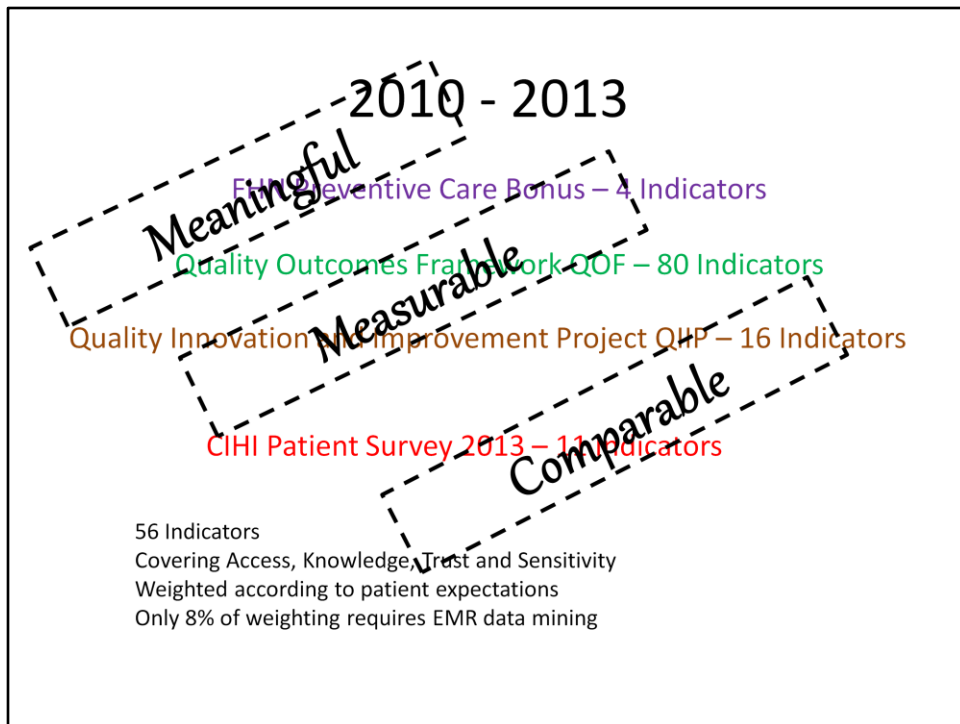
The initial survey analysis showed the relative value of the services we provide. Starfield's new hypothesis was that we needed to quantify the **nature of the relationship**

After over 400 surveys, it looked consistently that when people were asked about their future primary care, they consistently expected accessible knowledgeable and trusting 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. I found 11 survey questions addressing aspects of the emotional component between patients and their practice.



Now the score (of all the indicators weighted to public expectations) appeared to pass the “Sniff Test”.

56 Indicators

Addressing the full patient-primary care relationship

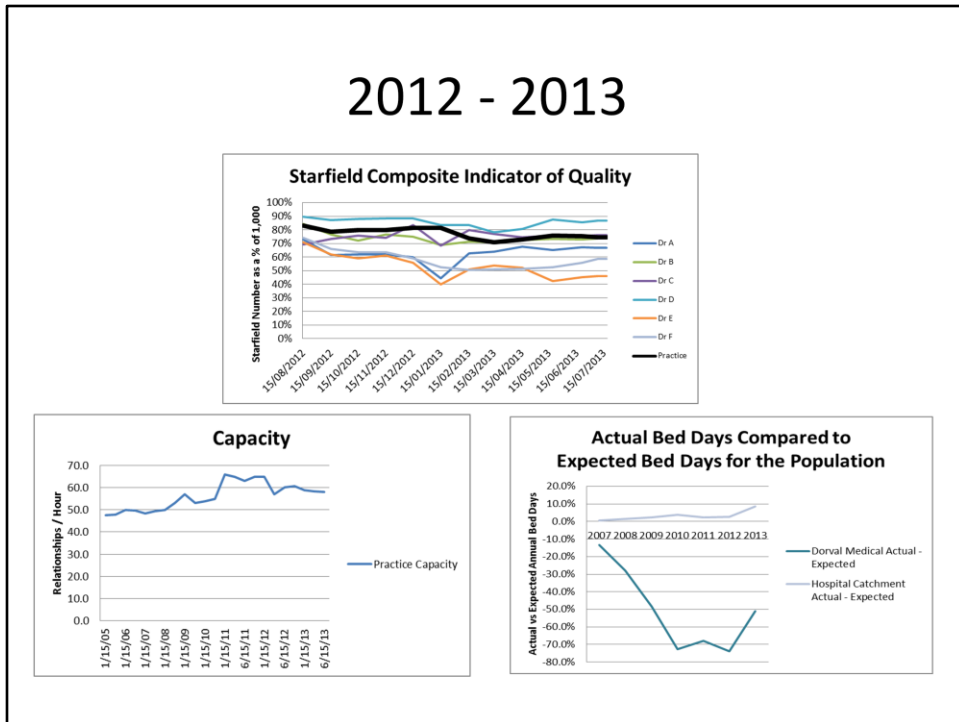
Weighted to reflect patient expectations

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 I described at the beginning:

1. It is **Meaningful** -The measurements were meaningful to providers and patients
2. It is **Measurable**
3. It is **Comparable** - A composite indicator of all indicators enabled comparison at the level of comprehensive primary care. Comparison was possible for the same practice over time, and for different practices at the same time.

2012 - 2013



The group proceeded to track the Quality score (the composite of all indicators). The top graph plots the Quality value (the Starfield Number).

We also tracked two other key measurements.

We needed to find a measurement to describe the quantity of relationships for different sized practices. Efficient practices manage more patients with the time they operate. Capacity is my measurement of how many satisfied relationships we could service given the size of the practice (described by the doctor hours each week).

8000 pts 135 hrs/wk = capacity of 59.3

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

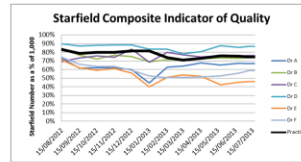
Small problem here...the practice had no data source to tell us about our total cost of care. Necessity, being the mother of invention, I looked for a surrogate for total cost of care.

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. Not a bad idea! I tracked Length of Stay and Total Bed Days.

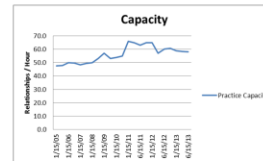
Now we had a better picture of performance

Performance

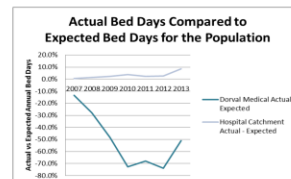
Quality – What we do in the relationship, incorporating multiple quality estimates



Capacity – How many relationships do we maintain at the described quality



Cost – What is the price for the number of relationships with their quality



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 quality”

Cost is the cost associated with the population under our care. It describes what we give up to pay for the package.

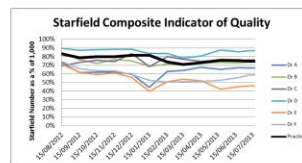
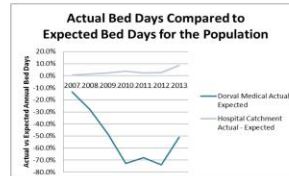
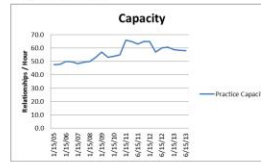
Measuring these three parameters (Quality, Capacity and Cost) allows the profession and the system stewards to manage toward strategic objectives.

Strategy 1 “The Silver Bullet”

Capacity – Establish Capacity at 13.2 million for Ontario = 59/provider hour

Cost – Stabilize cost growth

Quality – Push to improve Quality with the hope of future cost savings



A first strategy might be called “the Silver Bullet”

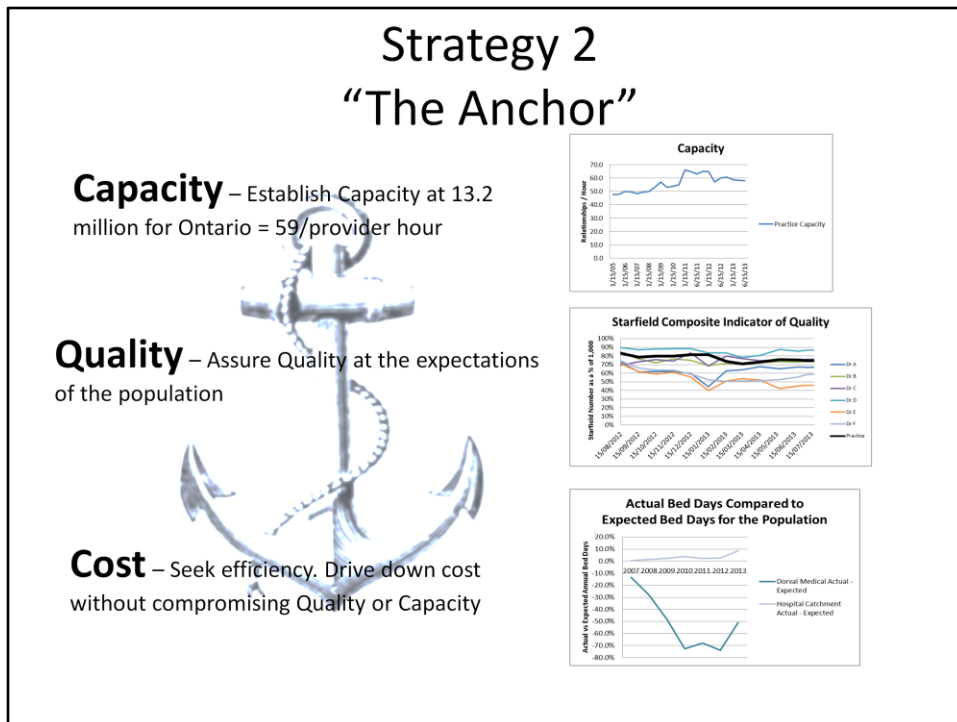
For both strategies Capacity is fixed at the population of Ontario 13.2 million or a capacity number of 59....Are people clear on the capacity measurement (back 2)

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

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, 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 was about 10% more acute 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.

What's next?

2013 -



Refine Cost – ICES will provide measurement of the total system cost, sorted by primary care provider



Refine Population Counting –

ICES is developing a method which adjusts population counts to incorporate health status



Get Involved – Any practice can participate

What is next?

Firstly, 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”

Secondly, 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 are looking to demonstrate the Starfield Model in practices beyond Dorval Medical. Pictured here is the Rat Pack who are looking at an expanded pilot: Tara Kieran and Rick Glazier of TWH, Tia Pham and Geordie Fallis of TEGH, and me.

If you are interested, read about the model and see if the rest of your group wants to play with the model.

If you want to participate with other groups, let me know.

To learn more

Google or bing



Dorval Medical

FAMILY HEALTH TEAM

gsouthey.md@dorvalmedical.ca

Thank You

The Dorval Medical web site has a page containing several documents describing the Starfield Model.

This presentation will be available on the website.

There should be a handout which shows the spreadsheet for the Model. The actual spreadsheet is available on the website and has the indicators, their current weightings, data sources and measurement rules. The spreadsheet allows a practice to calculate its overall quality score – The Starfield Number as well as the calculations for capacity and cost.

I hope you found this presentation interesting. We now have time for questions.