GRAND ROUNDS®

# Grand Rounds' Approach to Quality Measurement

March 2020

**Dani G,**
Grand Rounds
Physician Assistant

# About Grand Rounds

Grand Rounds partners with over 130 plan sponsors to deliver an exceptional healthcare experience, that reduces costs and improves health outcomes. Think of us as a personal healthcare assistant who will guide your members to the right care, through our high-tech and high-touch experience. Our digital experience brings simplicity to our members' care, harnessing the latest in machine learning and clinical research to offer members one place to go for their needs. Our award-winning care team goes above and beyond to provide reassurance through advocacy services, financial guidance, benefits routing and empathetic support for all members.

Our mission is to raise the standard of healthcare for everyone, everywhere. Since 2011, we have connected nearly 5 million covered members to high-quality care, resulting in better outcomes, less waste, happier members and time back to leadership teams.

# Table of Contents

# 1 The matching problem

25% of healthcare spending is wasted on unnecessary and inappropriate care.[1] Connecting members to high-quality doctors who treat appropriately is the key to delivering real-world outcomes that reduce waste.

When it comes to finding high-quality care, we believe that there is a "matching problem" as opposed to a supply problem. When a patient searches for care, there are usually high-quality doctors who are (i) in network, (ii) accepting new patients and (iii) specialized in treating the patient's specific clinical need, but the patient still fails to "match" with the right provider. Some pockets of the country, often those in rural areas with lower population density, truly do have doctor supply issues, but in much of the country it really does come down to a matching problem.

Here is the reality: too many patients find their doctor by asking friends or running a search on their insurance carrier's provider directory. Some also search Google, Healthgrades or Yelp. None of these options are equipped to help the patient select the doctor who has the right skill to treat their specific need.


1 – Shrank, William, et al. "Waste in the US Health Care System: Estimated Costs and Potential for Savings." JAMA, 2019.

The details matter and that is where all of these tools fall short:


**Family Medicine in Seattle, WA**

- 22 years experience
- Healthgrades: 3 out of 5
- 1.5 miles away
- Next available appointment is tomorrow


**Family Medicine in Seattle, WA**

- 14 years experience
- Healthgrades: 3 out of 5
- 2.1 miles away
- Next available appointment is in two weeks

Relying on traditional doctor-finding tools, the two physicians above appear almost identical. However, these tools miss the finer details and that's the problem that patients face:


**Family Medicine in Seattle, WA**

- Frequently prescribes high-risk medications inappropriately
- Refers to low quality specialists
- Increases patients' odds of undergoing unnecessary spine surgery


**Family Medicine in Seattle, WA**

- Safe prescribing patterns
- Manages chronic diseases well
- Strong track record of retaining patients
- Refers to high-quality specialists

With proper tools, patients would connect with high-quality doctors much more often. But for most, the selection process is left to chance and it allows low-quality and dangerous doctors to continue putting patients at risk. Some of the worst horror stories make it into the news, but there are a startling number of unsafe doctors on insurance networks:

**Dr. X**
In-Network

Family Medicine in Seattle, WA
Yelp Rating: 3.8 out of 5

Prescribed controlled substances without legitimate medical purpose

**Example:** 42-yr-old man with known addiction issues died after being prescribed opioids and muscle relaxants

**Dr. Y**
In-Network

Family Medicine in Dallas, TX
Yelp Rating: 4.8 out of 5

Multiple board actions from 2008—present for failure to provide adequate care

**Example:** Caused patient's irreversible brain damage by failing to evaluate and treat prolonged low blood sodium

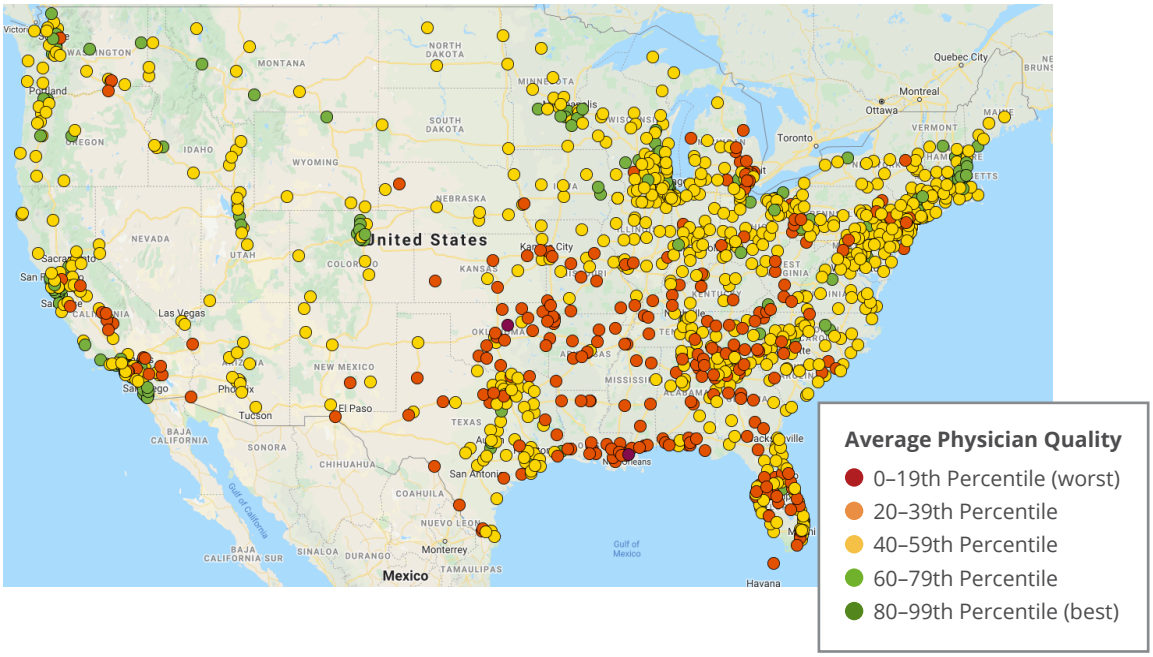
**Dr. Z**
In-Network

Family Medicine in Bloomingdale, IL
Yelp Rating: 3.9 out of 5

Sanctioned in 2012 for failing to meet surgical standards of care, putting patients at risk of serious bodily harm

**Example:** Performed surgery on the wrong finger

Left to their own devices, patients are exposed to bad doctors in every geography in the country:



And that is just the quality of care aspect of picking a doctor. There is also the cost side to consider. Patients have little visibility into whether the appointments, tests and treatments being recommended are necessary, nor do they know if they are paying a reasonable price for those services. A visit to the doctor while on a high deductible plan can cost a patient thousands of dollars that he or she does not have, and then there are all of the tests and follow-up appointments. The sad truth is that these unplanned expenses can easily bankrupt an unsuspecting patient.

Legacy tools are poorly suited to the current healthcare consumer environment. A new healthcare quality measurement approach is necessary to protect patients from dangerous doctors and financial ruin.

2.4 million

patients go to the doctor's office in the US every day.[2] The stakes are high, and time is of the essence.

## Progress, but not enough

There is growing interest in, and investment behind, measuring and comparing doctor quality. Many agree that high-quality care costs less and keeps people happier, healthier and more productive, but there is still a lot of room for improvement in the field of healthcare quality research:

### Too much focus on surgical outcomes at the expense of primary care and non-surgical quality.

**Why this matters:** Many quality measurement efforts have focused primarily on assessing surgical outcomes despite the fact that surgical encounters account for only a tiny fraction of patient's health needs and are often clinically unnecessary. For every one patient who needs a surgical procedure, there are 50 more patients whose health can most effectively be managed by a high-quality PCP or through non-surgical specialist treatment.

### Lack of adoption of the latest advances in machine learning.

**Why this matters:** Machine learning is critical to scaling the analytical horsepower of the algorithms themselves, allowing the intake of much more data and optimization across many more metrics. Without it, it is not computationally possible to (i) incorporate enough data and (ii) operate in a feedback loop of continuous improvement.

### Static doctor ratings fail to account for a patient's specific condition and personal circumstances

**Why this matters:** (i) Patients have unique circumstances and needs and (ii) physicians have unique sub-specializations and expertise. Static doctor scores, such as those in a traditional provider directory, are not sensitive to either of those elements and result in generic doctor recommendations that fail to account for whether a given provider is actually high-quality for an individual patient's specific clinical needs.

2 – Centers for Disease Control, 2019.

# 2 How Grand Rounds measures provider quality

Our investments in data and talent allow us to measure physician quality at the individual physician level across every geography and medical specialty, including primary care and all types of surgical and non-surgical specialists.

**10B+**
Clinical Data Points

**130+**
Data & Clinical Experts

We have invested over **$75M** in talent and data to enable world-class physician quality measurement

## Our team

We assembled a world-class team of 130+ data experts and clinicians to allow us to model provider quality at the cutting edge of medicine. We highly value the tight knit collaboration across these teams because it ensures that our machine learning models are trained on sophisticated and nuanced clinical insights provided by our clinicians. At Grand Rounds, we felt it was important to balance data and clinical points of view when measuring quality in order to avoid (i) over-indexing on biases that might be held by medical insiders and (ii) building and launching quality measurement models that lacked sufficient grounding in standard clinical practice. Combining data and clinical expertise offers the foundation that can transform provider quality measurement.

**World-class interdisciplinary team of data experts and clinicians**

**Data Science**
35+ data experts

Stanford, MIT, Berkeley, Georgia Tech, Amherst, UCLA, Cornell, LSE, Duke, UCSC

**Did you know?** Our Director of Data Science created the first computational simulation of a whole cell and was featured in the New York Times

**Clinical Care**
100+ clinicians

Stanford, Emory, Tufts, UCSF, University of Washington, Dartmouth, Mass General, Brigham and Women's

**Did you know?** One of our Senior Pediatric Nurse Practitioners led a team that successfully relocated 127 critically ill children between hospitals in a single day

## Data sources

We draw from a wide array of data elements to measure provider quality. This includes claims data sets, health outcomes data, as well as proprietary data that we generate through our own product offering:

| CLAIMS DATA | OUTCOMES DATA | MISC. DATA | GRAND ROUNDS DATA |
|---|---|---|---|
| Commercial claims | State registries | Medical associations | Expert medical opinions |
| Multi-carrier claims | Specialty societies | Specialty boards | After-care summaries |
| Customer claims | (soon) Member-reported outcomes | Provider offices | Member surveys |
| ... | ... | ... | ... |

The traditional commercial claims data sources provide us with over 50% of the annual billed claims in the US healthcare system. Combined with our other supplemental sources, we are able to evaluate provider quality across 96% of the practicing physicians in the country across every medical specialty recognized by the American Medical Association.

Our proprietary data sources offer us a unique edge, as we alone collect this data and incorporate it into our quality measurement approach. Expert medical opinions, for example, offer deep insights through detailed medical records and doctors' notes that claims data simply cannot provide. The same is true for the aftercare summaries generated by the quality-informed doctor referrals that we make for our members. We use the insights from these data sources to validate our analyses as well as for inspiration for new clinical metrics to evaluate.

## How our approach is different

Our approach to provider quality measurement is differentiated in several ways:

1. **Novel quality measures** developed to capture key overlooked aspects of provider quality
2. **Dynamic quality scoring** to match provider skill with member need
3. **Machine learning** is used to more accurately predict providers' performance
4. **External validation** by Veracity Healthcare Analytics

Our in-house team of data experts and clinicians develops novel quality measures that add more breadth and depth to our provider quality measurement approach. These novel metrics go beyond the industry's standard endorsed metrics, offering new levels of precision measurement across clinical specialties, including primary care.

We built a first-of-its-kind provider-patient Match Engine, which analyzes and predicts (i) physician skill and cost effectiveness and (ii) the clinical situation of the member in need, and the engine returns an optimized set of provider referrals results tailored to the member. This is a tremendous leap forward from the traditional provider directories.

For years, machine learning has been at the core of our measurement approach. By building deep learning architecture in-house, our provider quality models can ingest a virtually limitless amount of data, allowing the models to evolve and improve over time. Built in partnership by our data science and clinical teams, the models balance the optimal amount of tech-enabled sophistication and human oversight to ensure that our provider referrals improve in accuracy with each passing year.

Lastly, our provider quality measurement approach has been externally validated by experts in the provider community. Veracity Healthcare Analytics, led by a team on the faculty at Harvard Medical School, reviewed and confirmed the standards we use to measure provider skill.

## How we define quality

Our provider quality measurement methodology evaluates providers across four dimensions of performance that we consider critical to selecting a doctor:

**SAFETY**

Does the provider practice medicine that is likely to harm the member?

**EFFECTIVENESS**

Do the provider's methods resolve the member's clinical need?

**EFFICIENCY**

Does the provider avoid wasteful and unnecessary treatments?

**PRICE**

Does the provider charge a fair and reasonable fee?

Our provider quality methodology first ensures that any provider we recommend to a patient is an outstanding clinician, and then further optimizes for cost of care so that we recommend high-quality, high value providers who will deliver cost-effective care.

One of the key benefits of co-staffing our quality modeling projects with clinicians and data scientists is that we can develop novel quality metrics across our four dimensions of quality with speed and clinical precision. This enables us to go farther than traditional metrics allow.

For example, the opioid epidemic has become national news in recent years and has drawn significant attention toward harmful opioid prescribing behavior. However, benzodiazepines are misused in ways similar to the opioid crisis, but benzodiazepine prescribing has not drawn nearly the same level of scrutiny as has opioid prescribing, and a comparable breadth of prescribing metrics that have been endorsed by the medical community does not exist for benzodiazepines as it does for opioids. As a result, we developed our own benzodiazepine prescribing metrics to enhance our ability to assess prescribing risk.

| | Primary Care | Specialists |
|---|---|---|
| Traditional | • Cervical Cancer Screening<br>• Inappropriate Spine Imaging<br>• Adherence to Oral Diabetes Meds<br>• ... | • Post-Surgical Mortality<br>• Elective C-Section Rate<br>• Asthma Complication Rate<br>• ... |
| Proprietary | • High-Dose Benzodiazepine Prescribing<br>• New Patient Retention<br>• Specialist Referral Quality<br>• ... | • Conservative Treatment Prior to Spine Surgery<br>• Migraine Expertise<br>• Unnecessary Pre-Procedural Lab Testing<br>• ... |

The process for calculating provider scores is a multi-step process that incorporates our quality metrics across safety, effectiveness, efficiency and price to produce a final, comprehensive physician quality score, which is detailed at a high level below:

1. **Calculate the raw score** on each clinical metric across the four domains
2. **Calculate the member-specific weights** for each clinical metric
3. **Combine the weighted metrics** into a clinical quality score for each provider
4. **Adjust the score for efficiency and price** once clinical quality has been established

**Distilling the Calculation**

**Provider quality score** = $(\text{Metric}_1 \times \text{Weight}_1) + (M_2 \times W_2) + (M_3 \times W_3) \ldots + (M_n \times W_n)$

The final physician quality score is then combined with additional personalization factors (e.g. travel distance) to determine the provider's ranking in our provider search results. A high score will move the provider toward the top of the list, and a low score will ensure that the doctor moves toward the bottom of the results list so that more qualified doctors appear first. Each of our four quality domains plays a role in determining a physician's overall quality score. To illustrate how we measure physician performance across safety, effectiveness, efficiency and price, we will refer back to anonymized profiles of actual practicing physicians:

| |  **Primary Care Dr. A** |  **Orthopedic Surgery Dr. B** |  **Neurology Dr. C** |
|---|---|---|---|
| **Public Biographical Details:** | Solo practitioner with 42 years of experience and good online reviews | Spine surgeon with 30 years of experience and mixed online reviews | Neurologist with 42 years of experience and good online reviews |

## Safety

We optimize for the providers who practice safe medicine and consistently minimize risk. High-quality providers avoid collateral damage stemming from their treatments, such as addiction resulting from high-dose opioid prescriptions meant to treat back pain.

We use a number of metrics to measure physician effectiveness, with many endorsed by the medical community, such as unnecessary spine imaging for back pain, and some developed in-house by our R&D teams, such as high-dose benzodiazepine prescribing:

**SAMPLE SAFETY QUALITY MODELS**

### Prescribing Safety Measures

- High-Dose Opioid Prescribing*
- High-Dose Benzodiazepine Prescribing
- Barbiturate Prescribing*
- ...

### Post-Surgical Complication Measures

- Post-Surgical Mortality*
- Prolonged Length of Stay After Surgery*
- ...

### Risk of Gross Negligence

- Probability of Medical Board Sanctions
- ...

### Inappropriate Imaging Measures

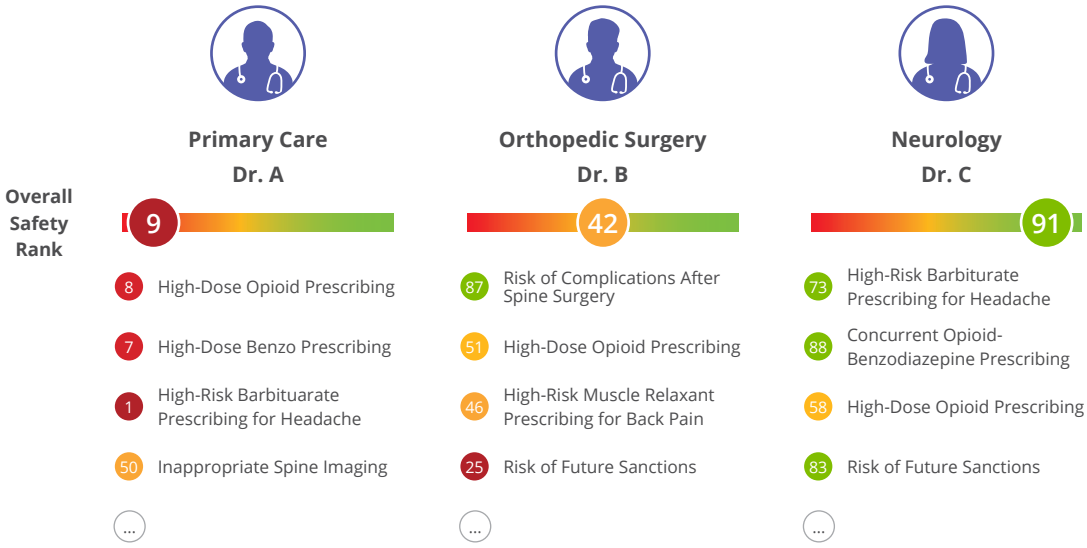
- Unnecessary Spine Imaging for Back Pain*
- Unnecessary Cardiac Stress Testing*
- ...

*Similar measures are endorsed by third-party quality organizations and/or medical societies

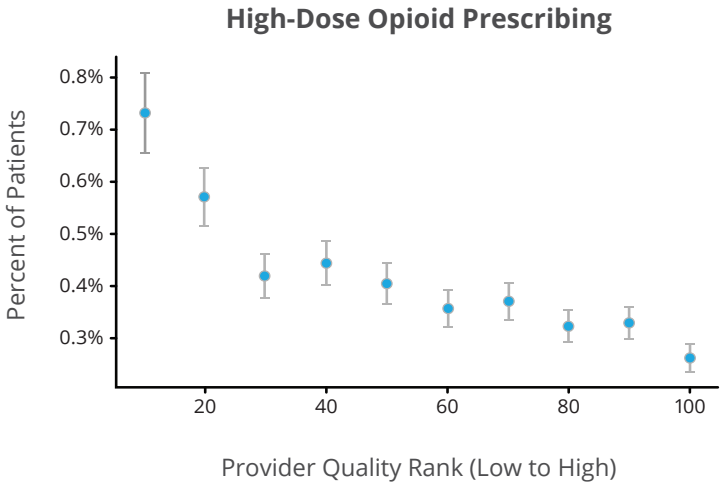
**Why this matters:** Bottom decile providers are **10x more likely to receive sanctions and 6x more likely to prescribe opioids** at unsafe doses.

Our safety models identify which providers will put their patients in harm's way so that we can help those patients avoid dangerous doctors altogether. In the example below, Dr. A receives a poor safety score among PCPs due to his prescription patterns, while Dr. B is in the middle of the pack for orthopedists, and Dr. C is rated as excellent among neurologists.

Our models are nuanced to account for the subtleties of physician practice patterns. For example, Dr. C has a moderately low score for opioid prescribing among neurologists, but the metric is not heavily weighted in his score because neurologists as a whole tend to prescribe lower rates of opioids when compared to other specialties. As a result, only scoring in the 58th percentile for opioid prescribing for a neurologist is still a decent result when compared to the broader medical community.



By incorporating a vast number of safety measures into our quality rankings, our goal is to ensure that our members avoid providers who practice unsafe medicine. To validate that this is actually the case, we measure the predictive performance of our overall quality rankings against an independent sample of claims from Grand Rounds covered members. For example, the chart below looks at the degree to which members were actually less likely to be prescribed dangerously high doses of opioids if those members had a new encounter with doctors we rated highly vs. poorly. We observed that members who visited doctors whom we rated as high-quality had far fewer high-dose opioid prescription fills compared to members who visited doctors whom we rated as low-quality. This demonstrates that our models are properly identifying the higher performing doctors.




Members are

**64%**

less likely to be treated with high-dose opioids if they see a top decile PCP rather than a bottom decile one.

Source: Grand Rounds commercial claims data from 2018.

**Department of Justice**

U.S. Attorney's Office

District of Alaska

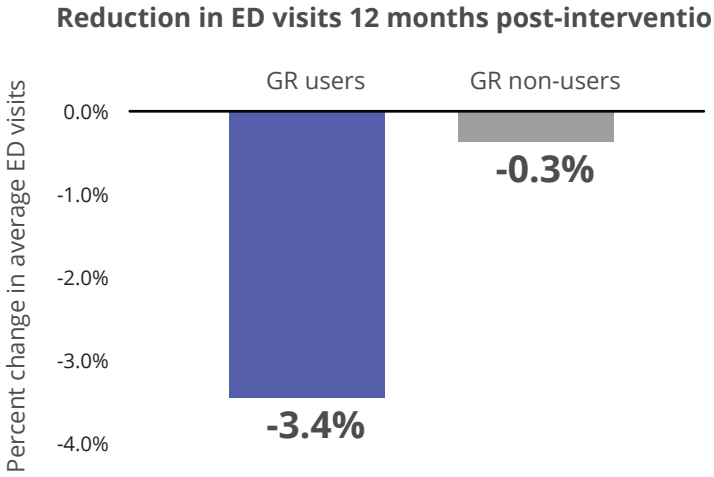
FOR IMMEDIATE RELEASE — Wednesday, October 9, 2019

## Anchorage Nurse Practitioner and Soldotna Doctor Arrested on Separate Federal Narcotics Charges

Anchorage, Alaska – U.S. Attorney Bryan Schroder announced today that Special Agents with the Drug Enforcement Administration have arrested an Anchorage nurse practitioner and a Soldotna doctor on separate federal narcotics charges, which allege that they illegally distributed large amounts of opioids and other powerful narcotics by writing prescriptions for "patients" without medical examinations and lacking medical necessity. Federal law enforcement officials executed multiple search warrants in both cases yesterday.

Source: "Anchorage Nurse Practitioner and Soldotna Doctor Arrested on Separate Federal Narcotics Charges." Department of Justice, 2019.

Most importantly, our safety measures deliver upon real world outcomes. For example, PCPs who prescribe opioids responsibly drastically lower the risk of opioid addiction and overdose, which are common causes of visits to the emergency department. Across our customer base, matched cohort studies demonstrate that users who follow through on our high-quality doctor referrals experience lower rates of emergency department utilization. We attribute the observed decrease in emergency department visits in large part to our platform routing to doctors who practice better medicine:




**Reduced Emergency Department Usage**

3.1%

Relative reduction in all-cause ED visits for GR users compared to matched control.

## Effectiveness

We optimize for providers who have a track record of practicing medicine that resolves the patient's clinical need. For example, one mark of a high-quality PCP is appropriate monitoring of diabetes progression and medication adherence. Another example is preventative cancer screening, which leads to earlier cancer detection at a point when the tumor is contained and much more treatable.

We use a number of metrics to measure physician effectiveness, with many endorsed by the medical community, such as colon cancer screening, and some developed in-house by our R&D teams, such as quality of referrals to specialists:

**SAMPLE EFFECTIVENESS QUALITY MODELS**

### Medication Adherence Measures

- Adherence to Statin Therapy*
- Adherence to Antihypertensives*
- Adherence to Oral Diabetes Meds*
- ...

### Potentially Avoidable Complication Measures

- Rate of Diabetes Complications*
- Rate of Asthma Complications*
- Rate of Cardiovascular Complications
- ...

### Preventive Screening Measures

- Colon Cancer Screening*
- Cervical Cancer Screening*
- ...

### Procedure-Related Effectiveness Measures

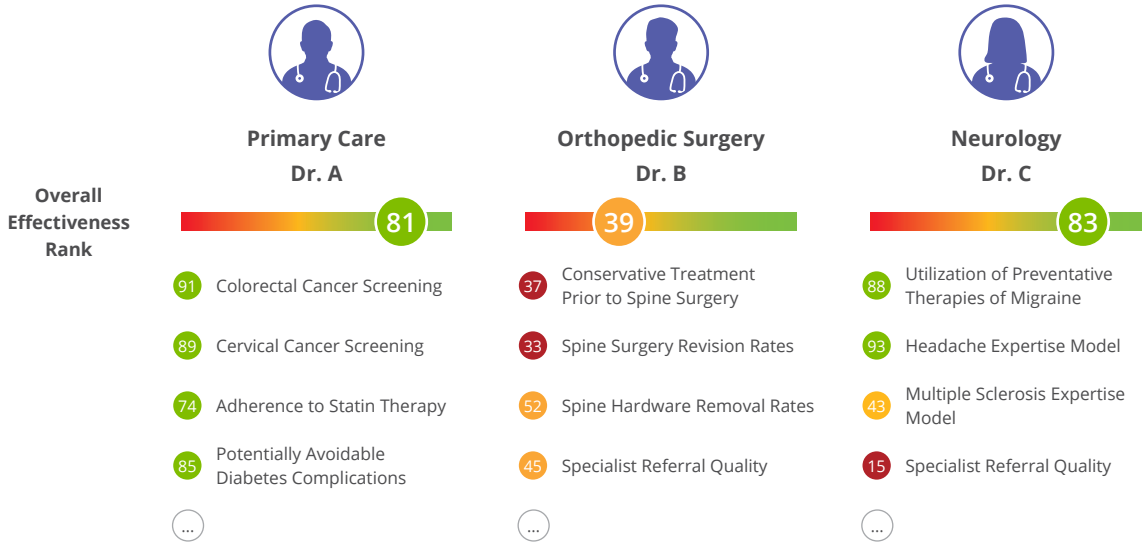
- Conservative Treatment Prior to Spine Surgery
- Revision Rates After Spine Surgery
- ...

- New Patient Retention
- Specialist Referral Quality
- Condition-Specific Expertise

**Why this matters:** Medication non-adherence alone is a **$300B** problem that leads to **hospitalizations** and mortalities, where 50% of filled prescriptions are taken incorrectly (CDC).

*Similar measures are endorsed by third-party quality organizations and/or medical societies

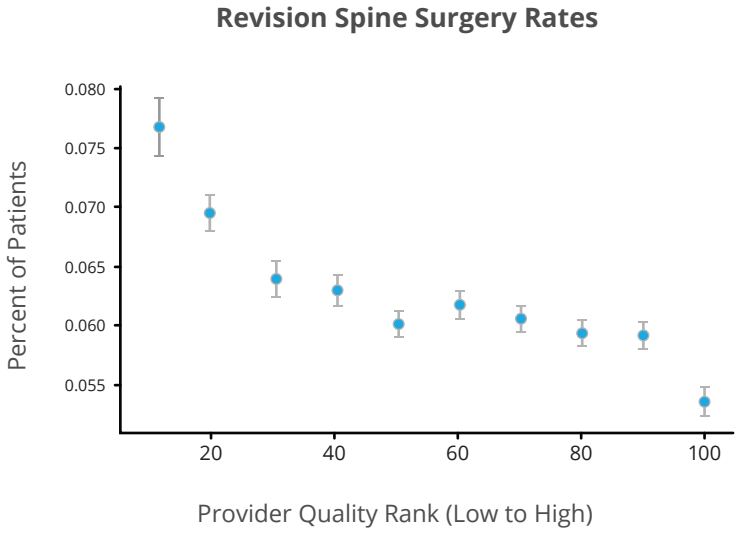
Despite scoring poorly on our safety measures, Dr. A scores very well on our effectiveness measures. His practice patterns indicate that he practices medicine known to be effective at preventing and treating complex diseases. Dr. B scores poorly on our effectiveness measures due to his pattern of circumventing conservative treatment in favor of surgery, as well as a high rate of spine surgery revisions on a risk-adjusted basis, which indicates low surgery efficacy. Dr. C scores well on account of his headache treatment patterns.




Our effectiveness measures ensure that our members avoid providers who practice ineffective medicine. For example, one element of world class medicine is exerting influence of patients to drive medication adherence. We endeavor to connect our members to physicians who are most likely to ensure that patients follow through on medications. According to the CDC, roughly 1 in 5 prescriptions is never filled, and nearly half of the prescriptions that are filled are taken incorrectly.[3]

3 – Neiman, Andrea, et al. "Improving Medication Adherence for Chronic Disease Management — Innovations and Opportunities." Centers for Disease Control, 2017.

Similar to our safety example, we validate the predictive performance of our overall quality rankings against an independent sample of claims from GR-covered members. In the chart below, we observed that members who visited doctors whom we rated as high-quality experienced substantially lower rates of revision spine surgeries compared to members who visited doctors whom we rated as low-quality. This demonstrates that our models are properly identifying higher performing doctors.




Members are

**30%**

less likely to need revision spine surgery if they see a top decile surgeon rather than a bottom decile one.

Source: Commercial claims data for the year after initial spine surgery.

# Medicare Records Reveal Troubling Trail of Surgeries

*By John Carreyrou And Tom McGinty*

Updated March 29, 2011 12:01 am ET

SHARE TEXT

PORTLAND, Ore.—Dr. Vishal James Makker had already operated on Ronald Johnson's spine six times in less than two years, but he had some grim news for the former machine-tool operator: X-rays showed Mr. Johnson needed a seventh surgery.

Source: Medicare Records Reveal Troubling Trail of Surgeries." Wall Street Journal, 2011.

## Price and efficiency

We optimize for skilled clinicians who deliver cost-effective care. We do not compromise on quality, but once we establish that doctors are of comparable quality then we want to drive volume toward the most cost effective of the high-quality options. If we simplify the total cost to the healthcare system as: Total Healthcare Cost = (# Units of Care) x (Price of Care), then we see that there are two levers to pull to reduce total costs: (i) the number of units and (ii) the price. To reduce waste and improve outcomes, we model quality such that we (i) prioritize doctors who are more resource-efficient and use the minimum number of units of care necessary, and (ii) prioritize doctors who charge lower unit prices. Optimizing for price differences is particularly critical to controlling overall healthcare spending trend, as a large portion of annual spending increases are driven by hospitals and pharmaceutical companies increasing the unit prices of services and drugs.[4]

When it comes to efficiency, this means that during the course of caring for a patient, the provider uses the minimum amount of care necessary to resolve the patient's clinical need. For example, CT scans are only prescribed when necessary and are not re-prescribed multiple times.

When it comes to price, this means that given the option of several high-quality clinicians, we will (in a vacuum) optimize for the clinician who delivers high-quality care with the lowest sticker price. If two doctors who are otherwise high-quality hip replacement surgeons perform a hip replacement surgery for $20,000 and $62,000, then priority is given to the doctor who performs the surgery for $20,000.

We deploy a number of different metrics across our efficiency and price dimensions to ensure that we connect members to high-quality providers who demonstrate good financial value for their services. Our efficiency measures include volume-based metrics that scan for wasteful and inappropriate care, including excessive rates of spine imaging, EKG testing and c-section procedures. Our price measures include cost assessments that scan for the most affordable high-quality care:

**SAMPLE EFFICIENCY & PRICE MODELS**

### Efficiency Measures

- Elective C-Section Rates*
- Unnecessary Spine Imaging for Back Pain*
- Unnecessary Cardiac Stress Testing*
- Unnecessary EKG Testing
- Unnecessary Pre-Procedural Labs
- Inappropriate Cervical Cancer Screening*
- ...

### Cost / Pricing Measures

- Total Cost of Care*
  (applied to primary care specialties)
- Condition-Based Cost of Care Episodes*
  (e.g. back pain, asthma, CHF exacerbation, diabetes exacerbation etc.)
- Procedure-Based Cost of Care Episodes*
  (e.g. bariatric surgery, knee replacement, hysterectomy, gallbladder surgery)
- Service-Specific Unit Pricing
- ...

*Similar measures are endorsed by third-party quality organizations and/or medical societies

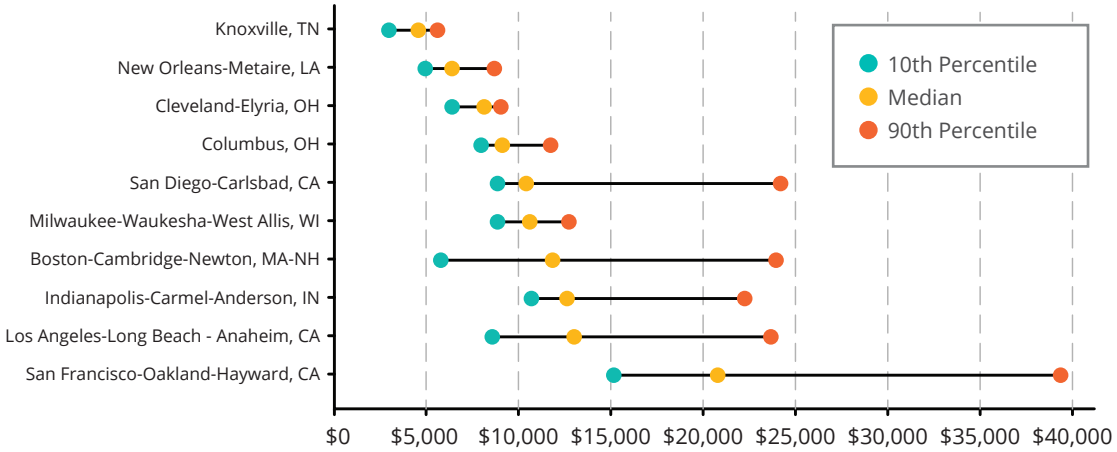
**Why this matters:** **Nearly $1 trillion, or 25%, of annual US healthcare spend is considered waste.** Routing to quality providers who are price- and resource-efficient costs less.

4 – Health Care Cost Institute, 2018.

Price is one of the most confounding aspects of healthcare. The lack of transparency around price is what prevents healthcare from tracking with normal supply and demand economics. Two doctors who (i) participate in the same insurance network and (ii) practice in the same geography may charge wildly different prices for the same procedure. In analyzing their own claims data, Blue Cross Blue Shield found that in Dallas, Texas, a hip replacement could cost between $16,772 and $61,585 depending upon the hospital.[5] The Healthcare Cost Institute found similar results across a number of procedures, including C-sections:

**C-Section Delivery**

Range of service prices for selected metro areas, 2016

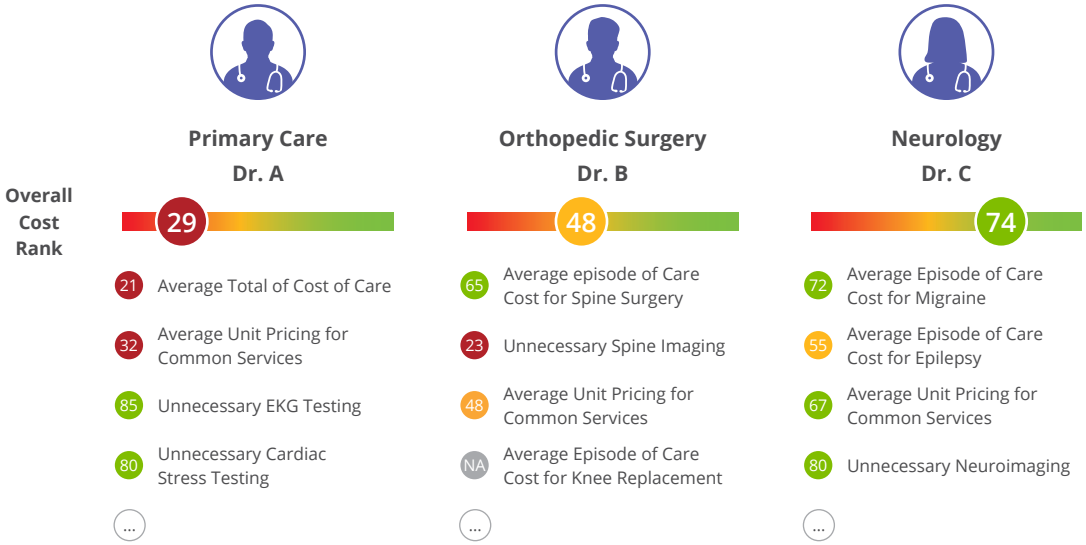



Source: Health Care Cost Institute

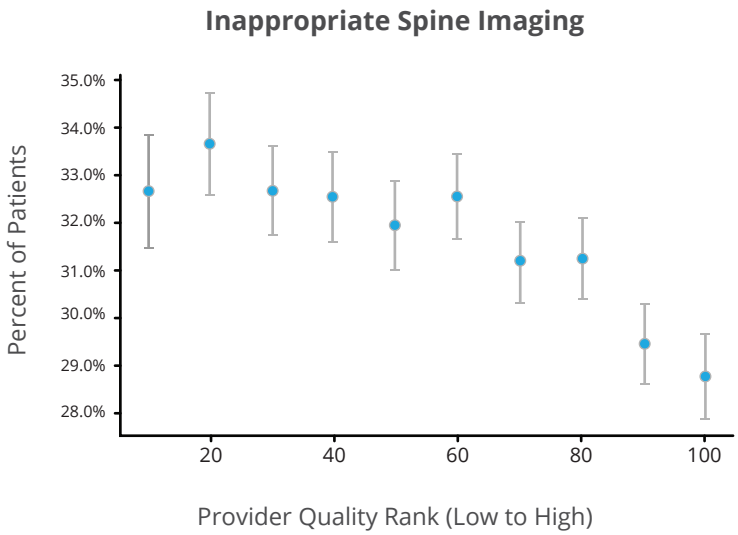
It is therefore critical to identify high-quality providers who will perform high-quality medicine at an affordable cost. And, given the lack of clear patterns in cost from one site of care to the next, this is absolutely possible.

5 – "Blue Cross Blue Shield Association Study Reveals Extreme Cost Variations for Knee and Hip Replacement Surgeries." Blue Cross Blue Shield, 2015.

In terms of overall cost rank, Dr. A ranks particularly low against other PCPs due to high unit pricing for common services, such as routine tests, and a high overall total cost of care. His good efficiency scores for lack of both unnecessary EKG testing and unnecessary cardiac stress testing were not enough to offset his score. Dr. B is middle of the pack for orthopedic surgeons for unit pricing, but is particularly inefficient when it comes to spine imaging. Dr. C has an attractive combination of competitive unit costs and a history of efficient care amongst neurologists, which ensures that he is prioritized in relevant searches for neurological care.



Similar to our safety and efficiency examples, we validate the predictive performance of our overall quality rankings against an independent sample of claims from GR-covered members. In the chart below, we observed that members who visited doctors whom we rated as high-quality experienced substantially lower rates of inappropriate spine imaging compared to members who visited doctors whom we rated as low-quality. The results demonstrate that our models can accurately predict cost efficient care, which allows us to direct our users to high-quality care that costs less.




Members are

**12%**

less likely to receive inappropriate spine imaging if they see a top decile PCP rather than a bottom decile one.

Source: Grand Rounds commercial claims data from 2018.

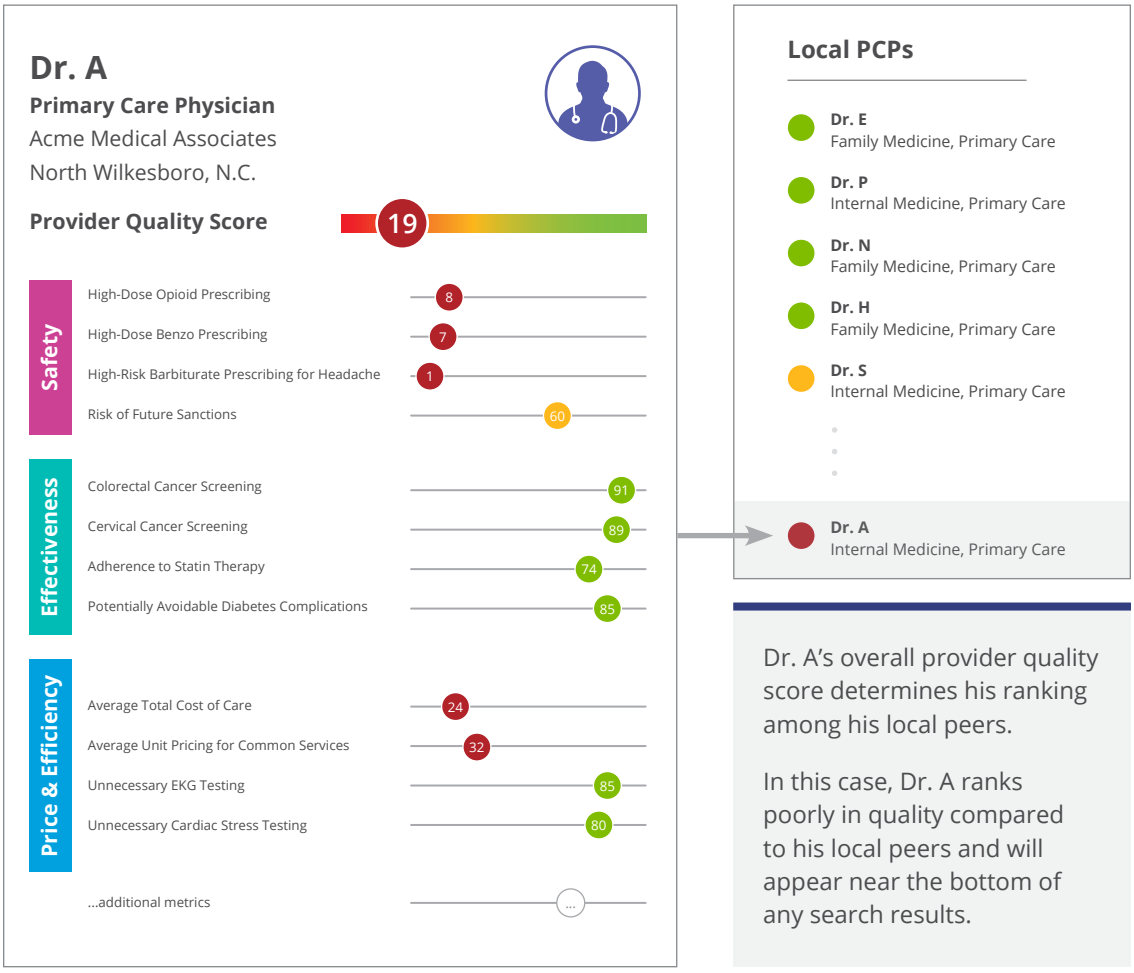
# Good or Useless, Medical Scans Cost the Same

Gail Kislevitz of Ridgewood, N.J., had an operation on her knee after an M.R.I. scan came back "uninterpretable." She learned later the cartilage had not been torn. Juan Arredondo for The New York Times

Source: Kolata, Gina. "Good or Useless, Medical Scans Cost the Same." New York Times, 2009.

# 3 Bringing it all together

Our quality measurement for each doctor comes together in the form of an aggregate provider quality score across all four of our quality dimensions. The aggregated score then determines the priority we give to that doctor in our referral rankings. The doctor with the highest quality score will appear as the first result in our referral list. Doctors with lower scores appear toward the bottom of the referral list. In this case, because Doctor A received a low overall quality score due to significant safety concerns and ranks towards the bottom of the referral list:



## Innovations in provider quality measurement

One area where we have invested heavily in the last couple of years is the concept of hyper-personalized provider matching through our Match Engine. We call this "dynamic provider scoring," which aims to redefine provider selection in terms of the nuanced clinical match between patient and provider.

A helpful analogy to provider matching is precision medicine and cancerous tumor treatment: oncologists select the cancer treatment based on the patient's genetic makeup and lifestyle preferences so that each treatment is personalized to the patient's specific need. We aim to do the same for matching every patient with any kind of high-quality provider.

An unfortunate wrinkle in healthcare today is that even if a patient connects with a reputable and generally high-quality doctor, the doctor might not be high-quality at treating the patient's actual need. Think of two patients who both need a neurologist, one for severe migraines that have not responded to standard treatments and the other for an acute exacerbation of Multiple Sclerosis (MS). The best provider for the migraine patient is likely a neurologist who is highly specialized in the treatment of migraine and only occasionally treats MS. Meanwhile, the best provider for the MS patient is likely to be a different neurologist whose expertise is in managing complex MS cases. Traditional approaches to quality would make no distinction between these two neurologists, though. Both would simply be identified as high-quality providers and recommended to both patients when, in reality, matching to the appropriate specialist isolates down to one obvious choice for each patient.

Dynamic provider scoring refers to the process through which our quality models algorithmically adjust provider quality scores in response to a patient's specific clinical need at the time of their doctor search. This means that when a patient without a history of back pain searches for back pain care, our provider quality models will react and upweight the importance of metrics that measure conservative treatment prior to surgery. The Match Engine will recommend doctors who pursue conservative first-line therapy, which will help avoid unnecessary surgeries while still delivering high-quality care.

The above examples of the provider quality scores for Providers A, B and C each assumed that a generally healthy patient without a prior history of serious condition was searching for care, and we scored the doctors as such. However, if instead the searching patients presented with preexisting conditions, our models would account for the patients' specific clinical profiles and the provider scores would adjust accordingly.

For example, imagine that these three patients each needed to find a doctor for care:

Young & Healthy

**Particular emphasis on preventive care quality**

Has diabetes and multiple cardiovascular risk factors

**Particular emphasis on ability to manage chronic diseases**

History of chronic back pain

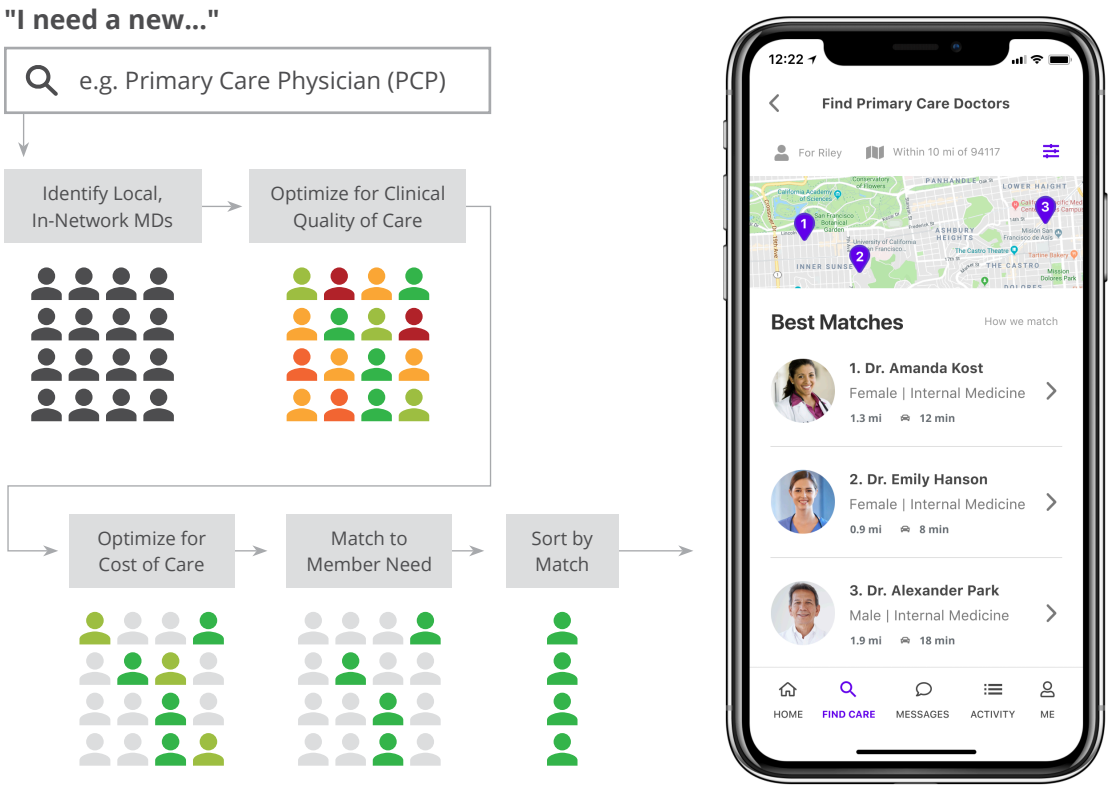
**Particular emphasis on ability to treat back pain safety and effectively**

Our models would take in known information about these patients through a combination of historical claims information and the clinical search itself, and then return a set of provider referral recommendations to each member based upon their specific needs. Of great importance is that the same provider might be the top choice for one member, but not so for the next. Something to call out is that the likelihood that a doctor goes from being rated as "outstanding" to "terrible" through the dynamic scoring process is quite low. Rather, the process will likely result in the nearest neighbors being re-ranked according to the clinical need. This can be seen below, where physician #1 is the best match for the first patient, but not so for either of the other patients, though she still ranks reasonably well consistently:

| **MEMBER PERSONALIZATION**<br>Access member's specific clinical need | **PROVIDER QUALITY MEASUREMENT**<br>Access clinical skill and financial profile | **MATCHING**<br>Rank doctors based on member-specific clinical needs |
|---|---|---|
| | |  |
|  History of chronic pain | • Opioid prescribing<br>• Muscle relaxant prescribing |  <br>Best Match |
|  Has diabetes and multiple cardiovascular risk factors | • Diabetes complication risk<br>• Diabetes med adherence |  <br>Best Match |
|  Young and healthy | • Preventive care<br>• Patient retention |  <br>Best Match |

Contrast the above with traditional provider directories: when you search for an orthopedist, every doctor tagged as an orthopedist appears in the results list regardless of whether they treat the part of the musculoskeletal system with which the patient actually needs help.

This approach results in dynamic and personalized provider referrals, and ensures every searching member is matched to a provider based upon (i) the provider's clinical skill and cost effectiveness and (ii) the member's clinical situation.



In order to build the Match Engine, we had to not only take in an unbelievable amount of data from many different sources, but we needed to build out many more clinical and financial quality metrics than the legacy measurement approaches could offer. Traditional models do not offer the amount of model diversity required to build sensitive models that can adapt to the personal profile of a given patient. It took many years of research and investment, but our Match Engine can now dynamically score physicians to the individual needs of a patient.

The biggest limitation of traditional efforts to measure provider quality has been the narrow lens through which "quality" is assessed. A surgeon might be scored based solely on their rate of complications for one or two types of procedures, with no consideration given to the necessity of those surgeries or to their other patients who instead received non-surgical care. A primary care physician might be rated based on their management of a few chronic conditions with no consideration given to the other 90% of conditions they are responsible for treating on a routine basis.

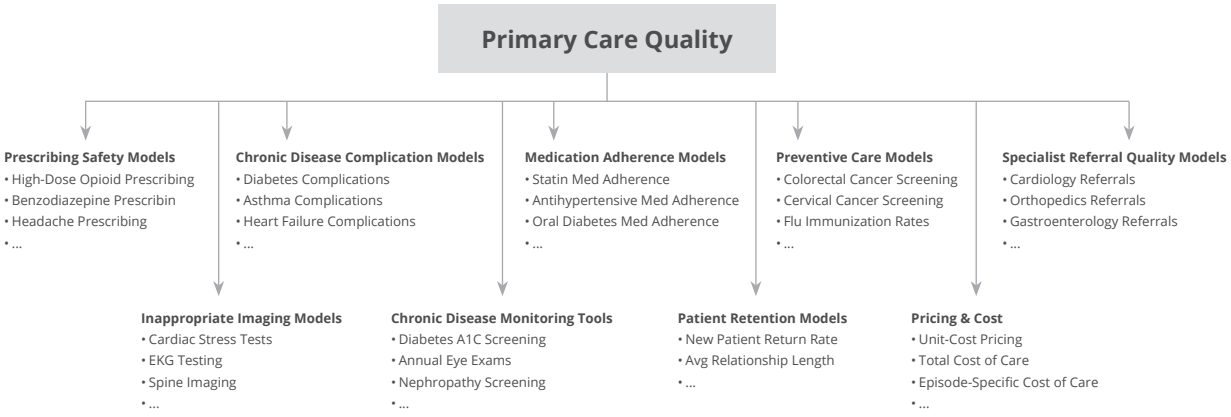
It took a long-term view and the right investments in talent to build enough data and metric coverage to unlock dynamic provider scoring. Today, we have close to 40 tenured data scientists, machine learning engineers, quantitative researchers, and data product managers who team up with over 100 on-staff clinicians to co-develop our clinical measurement algorithms. This combination of technical and clinical talent enables us to build machine learning models that model the appropriate clinical behavior in line with clinical best practice and appropriateness.

Our data scientists, data product managers and staff physicians are co-staffed on provider measurement projects in order to develop sound clinical metrics, and our lead data scientists, data product managers and staff physicians sit on model review boards to ensure the highest standards for technical and clinical measurement are met. These teams evaluate the existing universe of relevant quality metrics and then identify opportunities to develop new quality measures that capture important treatment decisions and patient outcomes that have traditionally been overlooked, particularly in the realm of primary care.

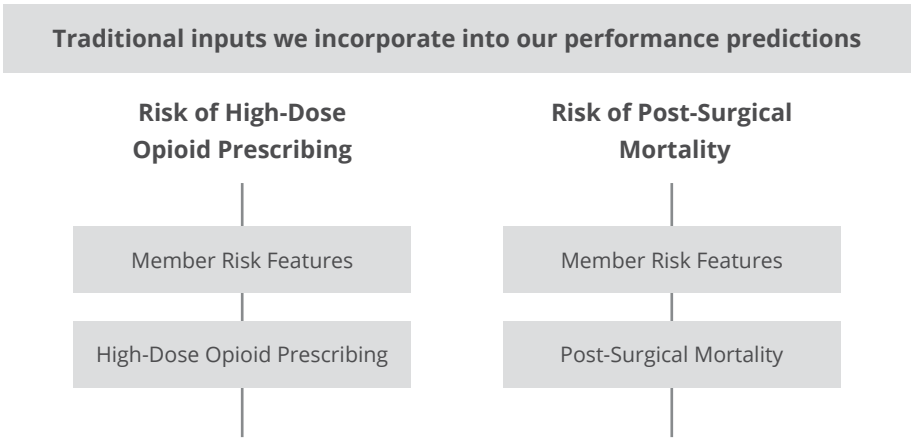
For example, four years ago, before the opioid addiction crisis rose to national prominence, we were alarmed by the opioid prescription patterns we saw in our data and launched our own opioid research initiative. Through that effort, we developed a series of inappropriate prescribing metrics to identify which primary care physicians and specialists were prescribing addictive medications at dangerously high doses or failing to follow guidelines recommending lower-risk alternative treatments.

Another set of metrics that seem simple and intuitive, but is both computationally complex and novel, are our metrics that capture the degree to which primary care physicians' build strong, enduring relationships with their patients, reduce the probability that patients receive unnecessary tests and procedures, and refer their patients to high-quality specialists.

This is particularly important because primary care needs typically account for over 50% of patients' provider search queries and often represent the best opportunity to have maximum impact early in a patient's care journey. The result is that we built the most comprehensive primary care quality measurement framework available, and that is in addition to the depth that we built across the remaining clinical specialties:

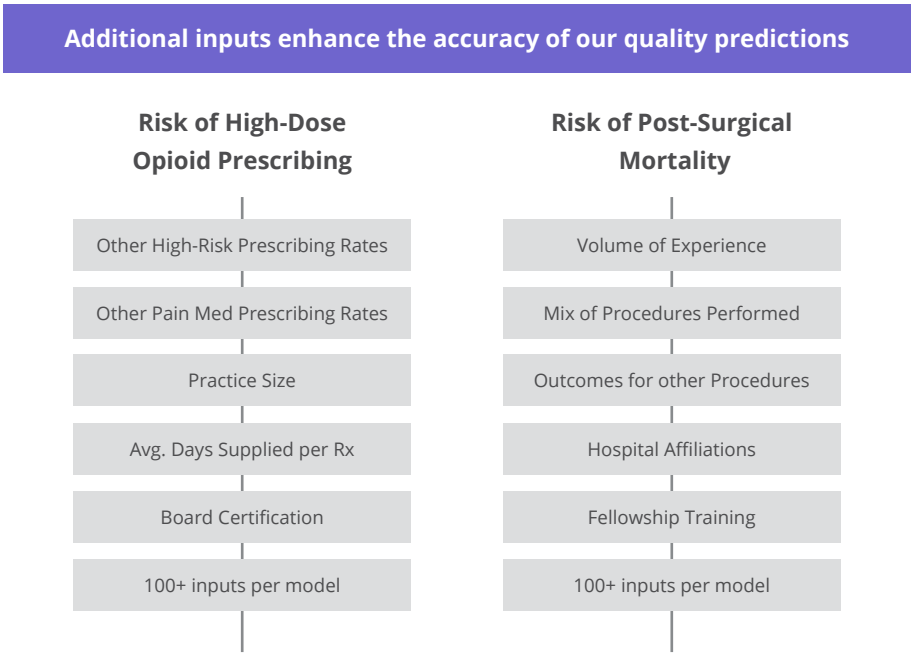



Another limitation of the traditional approaches themselves is the lack of diversity in the data that gets incorporated into the models. Too much of the available data was being ignored, which negatively impacted the predictive power of the quality models themselves:




**Why this matters:** Only looking at these inputs, alone, means ignoring 99% of the clinical and biographical information that is available on each provider.

We incorporate the traditional data inputs to harness the progress of the industry to date, but we also expanded into non-traditional data inputs to introduce new sources of potential insight that could increase the accuracy of our quality models:

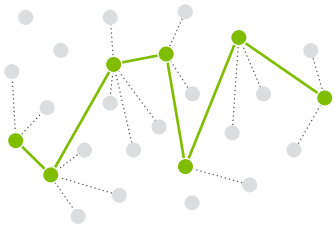



**Why this matters:** Machine learning is much more predictive when we can include additional inputs that go beyond rare events, particularly those inputs that are less obvious.

Building out many more quality metrics and introducing many more data inputs introduced a new problem: the computational complexity of what we were attempting to measure. The simplicity of the legacy quality measurement approaches makes them much easier to measure, as the quality metrics are static and the data inputs mostly consist of claims data and some outcomes measures. Our approach exponentially increased the analysis that needed to occur in order to assess the impact of each metric and data input individually, as well as evaluating all of the combinations of metrics and data inputs.

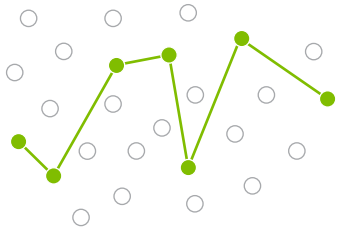
This is where machine learning (ML) plays a critical role in (i) scaling our analytics and providing the computational horsepower necessary to enable our quality measurement approach, and (ii) creating a feedback loop of continuous improvement. Machine learning is essentially a powerful optimization simulator that our ML engineers and data scientists program to seek out the combination of inputs that delivers the desired outcome. This architecture allows us to run a virtually unlimited amount of data through the algorithm and have the algorithm process the information and select the best possible combination of inputs that produces the best clinical outcome. The process also lets us easily introduce more metrics and data sources as we scale up our efforts.

We load **every available piece of data**, no matter how obscure



Thousands of computer simulations identify the "**needles in the haystack**" that predict great outcomes

The predictive inputs become part of the new model



The **ML feedback loop** ensures that models continue to improve as we add new data

**Why this matters:** Machine learning allows us to incorporate **more inputs** to build **more predictive** models at **greater speed**—but it requires significant investments in talent and data.

## Validation within the provider community

Raising the standard of care in the US cannot happen without the providers themselves, and we recognize the importance of partnering closely with the provider community in this endeavor.

Our efforts have focused most recently upon soliciting input and validation from the provider community. We shared examples of our own validations throughout this paper, but we strongly believe in the power of external validation to ensure that we hold ourselves to the standards of objective industry experts. To that end, we engaged Veracity Healthcare Analytics to review our physician quality metrics and to evaluate the clinical validity of our approach. Veracity is led by Dr. Niteesh Choudhry, a professor at Harvard Medical School and a renowned expert in the field of physician quality measurement in the research community. In short, Dr. Choudhry's validation study was overwhelmingly positive, and we will continue to work with Veracity to validate future measures as we introduce new inputs to our models:

**100%**
of metrics confirmed for clinical face validity

**92%**
of metrics supported by published studies

**95%**
of measure specifications deemed appropriate

> "In summary, virtually all of the metrics used by Grand Rounds to evaluate physician quality have clinical face validity and empirical evidence supporting their relationship with health care quality. Most have been well adapted."
>
> – Professor Niteesh K. Choudhry, Harvard Medical School

The Veracity validation process involved deep engagement from two of the foremost quality measurement experts in the country and resulted in healthy debate that ultimately resulted in our team revising four of our quality metrics according to Dr. Choudhry's recommendations.

In an effort to offer transparency to the provider community, Dr. Choudhry's validation report is now available to interested parties, and we welcome the opportunity to discuss the report and to share our latest thinking on this topic.

There is more to come on this front as we seek to further engage the broader provider community, and we look forward to sharing the results of our continued efforts in this area.

# 4 Conclusion

Our innovation in quality measurement has come a long way, but our work is not done yet. With 25% of healthcare spending wasted on unnecessary and inappropriate care, connecting people to high-quality doctors who treat appropriately is critical to delivering real-world outcomes that reduce waste.

We measure provider quality across four domains: safety, effectiveness, efficiency and price.

We invested over $75 million in data and talent to measure provider quality using both conventional claims datasets as well as proprietary data collected through our own services.

We stand apart on account of (i) dozens of proprietary metrics that we developed in-house to enhance our quality models, (ii) our use of machine learning to exponentially scale our analysis abilities and (iii) our first-of-its-kind doctor matching platform that uses our dynamic provider scoring engine to match members with doctors who possess the exact expertise to solve the member's need.

Our provider quality algorithms cover >96% of the practicing doctors in the country, effectively overlaying a virtual high-quality provider network on top of an existing insurance network.

Our provider quality measurement approach is at the forefront of innovation in the industry, and we welcome the opportunity to further discuss how optimizing for provider quality can raise the standard of care for everyone, everywhere.

GRAND ROUNDS®

# Grand Rounds' Approach to Quality Measurement

March 2020

For more information, please contact Grand Rounds:
1-866-221-3108
inquiries@grandrounds.com
grandrounds.com/employers

**Grand Rounds**
360 3rd Street, Ste. 425
San Francisco, CA 94107