6-2017

The Design of a Patient-Centric Healthcare Facility Rating Website: Consumer Choice as a Tool for Reform

Christopher August DiCapua
Union College - Schenectady, NY

Recommended Citation

DiCapua, Christopher August, "The Design of a Patient-Centric Healthcare Facility Rating Website: Consumer Choice as a Tool for Reform" (2017). *Honors Theses*. 18.
https://digitalworks.union.edu/theses/18

Running Title: A Consumer Choice Healthcare Rating Website

The Design of a Patient-Centric Healthcare Facility Rating Website: Consumer Choice as a Tool for Reform

By

Christopher August Di Capua

* * * * * * * * *

Submitted in partial fulfillment
of the requirements for
Honors in the Department of Sociology

UNION COLLEGE
March, 2017

ABSTRACT

DI CAPUA, CHRISTOPHER. The design of a patient-centric healthcare facility rating website: consumer choice as a tool for reform. Department of Sociology, March 2017.

ADVISOR: Professor Melinda Goldner

The U.S. healthcare system consistently underperforms on crucial international comparisons, thereby highlighting the need for reform. Simultaneously, there exists bipartisan and strong cultural support for patient choice; i.e. the ability of patients to assess the quality of healthcare facilities and choose amongst competing options. However, prior literature suggests that patients struggle to choose amongst competing facilities due to perceived competency barriers and insufficient information.

In this two-phased thesis project, I abstracted a model for mobilizing patient choice as a tool for healthcare reform by designing a website which presents government data on healthcare facility performance. First, three types of focus groups were conducted to: (1) establish a patient-centric definition of quality, (2) determine the appropriate level of data granularity for a facility rating website, and (3) design a user interface for online healthcare content that takes into account patient preferences.

In total, 23 subjects were recruited and split amongst the three focus groups. From the first group, a set of guidelines were extracted for a patient-centric definition of quality. Patients preferred Outcomes domains over Process or Input measures, valued Effectiveness and Safety most heavily, and had preferences that varied primarily along the lines of illness severity and length of care period. Focus Group Two illustrated the need to maintain data transparency; i.e. patients valued data on a facility's overall performance, performance in key areas (domains), and performance on individual indicators. Lastly, Group Three set guidelines on coloration and methods to efficiently disseminate data on performance.

In phase two, the focus group findings were used to guide the development of a ranking of U.S. hospitals using data included in the Centers for Medicare and Medicaid Services (CMS) database. A website design was then wire-framed using the prototyping program Axure. A post-hoc analysis revealed trends in hospital performance according to geographic location and ownership type. This line of work

exerts pressure on healthcare facilities to meet a certain standard of care. Data transparency continues to serve as a viable avenue for patient empowerment and a useful lever for healthcare reform.

Preface

Sociological, public policy, and epidemiological journals have consistently highlighted the shortcomings of the United States healthcare system; plagued by poor outcomes and rising national expenditures, current sociological research has centered on exploring alternative policy solutions to address these persistent issues. One relatively new and unexplored area of this research relates to the development of patient choice. It is a culturally evident and bipartisan belief that patients should have the right to choose amongst competing facilities when deciding where to receive their care. Classical economic theory would suggest that patients – given the opportunity to select amongst competing facilities – would choose those facilities which offer the highest quality of care at the lowest of costs.

Yet, prevailing research illustrates that this is not the case in the context of healthcare, which contains a number of market failures. For instance, evidence suggests that an asymmetry of information and power between healthcare providers and patients makes consumers hesitant to pass judgements on quality. Likewise, the presence of an insurance model – which limits the number of in-network providers a patient can choose from – dampens efforts towards exerting choice. As a result, consumer desires for a high-quality, low cost product have exerted only minimal selective pressures on the industry.

Within the context of these previous sociological studies, this research aims to improve upon the consumer's ability to select the highest-quality healthcare by constructing a patient-centric consumer choice rating website for healthcare facilities. This work had three interrelated research aims: (1) to establish a consumer-driven definition of healthcare quality, (2) to determine how much data patients need available in order to make an informed selection, and (3) to determine how complex healthcare data should be presented in order to maximize the user's understanding.

In Chapter One I explore the state of the U.S. healthcare system through landmark studies such as the 2012 Commonwealth Report. Evidence reveals the deeply entrenched issues within our current system, including poor access, rising costs, consistently low health outcomes, and a low general health status of our population. I then move on to a discussion of attempts at utilizing market forces to initiate reform; because the U.S. system is culturally capitalistic, it is argued that these market-based approaches are the most viable to implement. Following a discussion of the Affordable Care Act, I then introduce the concept of patient-choice and healthcare ratings websites as a potential lever. I then attempt to answer the question of why these ratings sites – while widely available – are consulted by only a small segment of the U.S. population. It is postulated that a healthcare ratings website must first produce a unified and transparent definition of quality. The latter sections of Chapter One are devoted to exploring commonly cited models of healthcare quality and approaches to understanding patient preferences. Chapter One concludes with a review of the mathematical underpinnings of the statistical tool, composite ratings.

Chapter Two outlines the methods employed in this study, which relied upon three sets of focus groups to address each of the three research aims. Conversely, Chapter Three is devoted to statistically and qualitatively analyzing focus group transcriptions in order to extract a series of guidelines and themes associated with each research question. The guidelines and themes gathered from this analysis are then used in Chapter Four, whereby I use data from the Centers for Medicare and Medicaid Services (CMS) to rank CMS-registered hospitals according to a novel healthcare quality framework. A design for a consumer-choice healthcare ratings website

is also constructed using a wire-framing software. Lastly, I end Chapter Four with an exploratory statistical analysis regarding the relationships between a number of hospital characteristics (geographic location, ownership type, etc.) and overall hospital quality. The implications of this work as well as potential directions for future research are outlined.

TABLE OF CONTENTS

CHAPTER 1

I. BACKGROUND AND SIGNIFICANCE

The United States healthcare system is often criticized in both academic journals and the popular media for rising individual costs, national expenditures, and significant inequities in outcomes. But while all pundits can agree that the healthcare system needs reform, we must first start by defining what it is we are trying to achieve through said reform. As such, the opening section of this chapter works to define the current state of the U.S. healthcare system – its flaws and its strengths. We then shift to a discussion of healthcare as a market. The U.S. population holds a cultural reverence for capitalistic market forces – particularly within the healthcare sect – and potential reform efforts are assessed within this ideological context.

We then focus in on the concept of consumer choice. Because consumers control demand, it is argued that an informed patient population can drive market-based correction within the healthcare industry by demanding quality care. Countless information sources are available to patients, such as television, print, radio, and online content. However, we will discuss that the majority of patients rely entirely upon anecdotal word-of-mouth from trusted family and friends to choose amongst competing healthcare facilities. The entirety of section four will be devoted to uncovering why online content has exploded in other consumer industries but lagged within healthcare. We will look critically at the websites that dominate the online health information marketplace, such as Yelp and Consumer Reports.

Lastly, we will discuss the very concept of healthcare quality. We assert that part of the reason online content has failed to gain traction is due to lacking reliable empirical data from the majority of web-based sources. But what are standardized, reliable, comprehensive data? One of the challenges of rating healthcare facilities through quantitative metrics is that it presupposes an agreed upon definition of healthcare quality. The remaining sections of chapter one focus on defining the construct of healthcare quality from both the perspective of legislatures and average consumers. We assert that current research should focus on (1) establishing a definition of quality that satisfies both individual and broader strategic policy objectives and (2) disseminating empirical data within that framework of quality in such a way that it can be understood by average consumers with minimal health literacy. These research questions will serve as the underpinning of the following thesis research.

II. The Triple Aim of Any Healthcare System

In 2008, Dr. Donald Berwick redefined the international objectives of any healthcare system through the development of the new axiom: “Cost, Access, and Quality,” otherwise termed the Triple Aim. The utility of the model is its simplicity. A system must provide care that is affordable to both the individual and society, accessible to the entirety of the population, and must provide a quality service (Berwick 2008).

Berwick (2008) states that “the components of the Triple Aim are not independent of each other” (760). Conversely, the pursuit of one goal will often spill over and affect the outcome of the other two. To provide an example, let us look at the effort to improve healthcare quality. Over the past century, the United States has pioneered pharmaceutical and medical technological innovation (Teleki et al. 2003). Many such innovations, including the development of new antibiotics and imaging technologies have drastically improved the quality of care. Yet, simultaneously, such improvements are often met with increases in cost, and as a spillover effect, decreases in access (Berwick 2008).

There is therefore a balancing act that exists in this tripartite structure of healthcare. But despite the interdependent nature of these columns, Berwick (2008) asserts that an effective healthcare system is one that uses creative solutions to simultaneously improve all three aims.

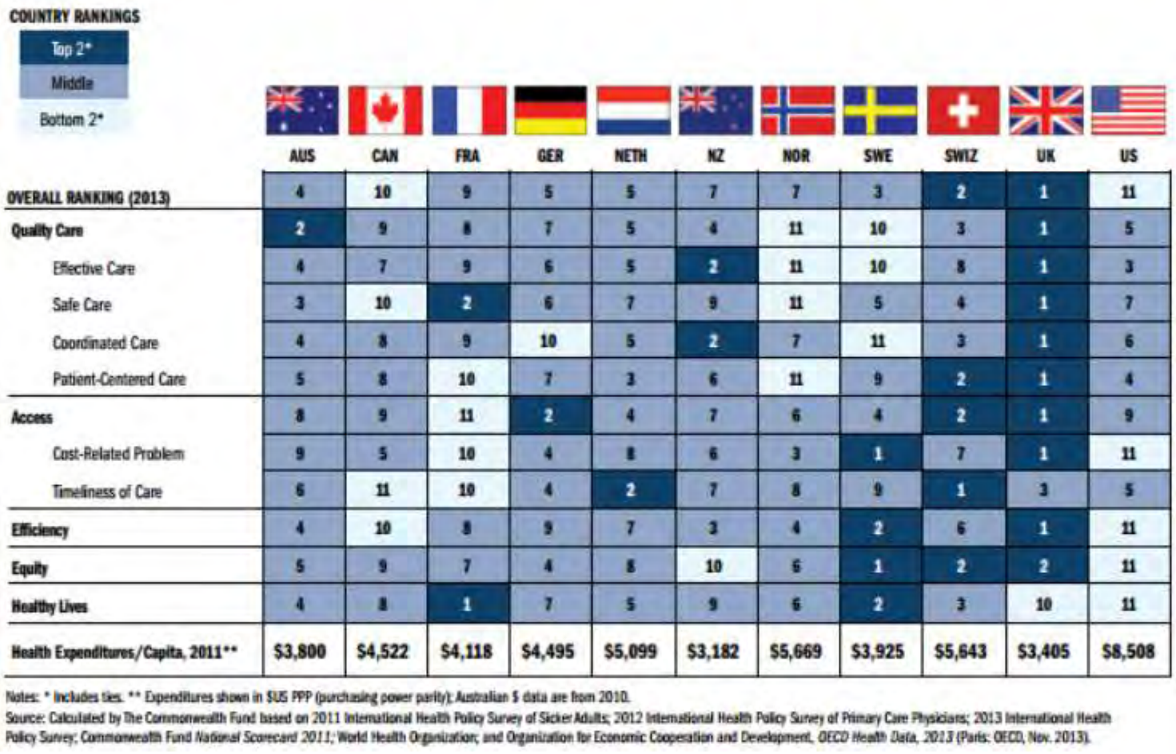
In this research study, the Triple Aim serves as the operational definition for an ideal healthcare system, both because of its comprehensiveness and its simplicity in comparison to alternative models. A system that can manage costs, access, and quality, is a system that can simultaneously assure social justice, economic integrity, and a healthy population.

If the goal of any system is to achieve Triple Aim, the obvious question to ask is how the United States fares under such a definition?

The State of United States Health Care:

Assessing the quality of the U.S. healthcare system can be a rather nebulous task. In some respects, the U.S. system is amazing; it leads the world in the development and utilization of state-of-the-art medical innovations, boasts globally recognized medical training, and possesses a number of renowned quaternary academic medical centers (Teleki et al. 2003). Yet, simultaneously, it is impossible to ignore the system's glaring shortcomings. What is more, if the U.S. is to establish new social policies to improve the healthcare system and achieve the Triple Aim, we must first identify and target specific areas for improvement.

A recent report by the Commonwealth Fund compares 11 different high-income nations along the lines of (1) quality of care, (2) access, (3) efficiency, (4) equity, (5) healthy lives, and (6) health expenditures (2014). According to these classifications, the United States performs last overall and in the individual categories of efficiency, equity, healthy lives, and total expenditures (2014: Figure 1). Let us pick apart these findings one-by-one.



COUNTRY RANKINGS

Top 2*

Middle

Bottom 2*

	AUS	CAN	FRA	GER	NETH	NZ	NOR	SWE	SWIZ	UK	US
OVERALL RANKING (2013)	4	10	9	5	5	7	7	3	2	1	11
Quality Care	2	9	8	7	5	4	11	10	3	1	5
Effective Care	4	7	9	6	5	2	11	10	8	1	3
Safe Care	3	10	2	6	7	9	11	5	4	1	7
Coordinated Care	4	8	9	10	5	2	7	11	3	1	6
Patient-Centered Care	5	8	10	7	3	6	11	9	2	1	4
Access	8	9	11	2	4	7	6	4	2	1	9
Cost-Related Problem	9	5	10	4	8	6	3	1	7	1	11
Timeliness of Care	6	11	10	4	2	7	8	9	1	3	5
Efficiency	4	10	8	9	7	3	4	2	6	1	11
Equity	5	9	7	4	8	10	6	1	2	2	11
Healthy Lives	4	8	1	7	5	9	6	2	3	10	11
Health Expenditures/Capita, 2011**	$3,800	$4,522	$4,118	$4,495	$5,099	$3,182	$5,669	$3,925	$5,643	$3,405	$8,508

Notes: * Includes ties. ** Expenditures shown in $US PPP (purchasing power parity); Australian $ data are from 2010.
Source: Calculated by The Commonwealth Fund based on 2011 International Health Policy Survey of Sicker Adults; 2012 International Health Policy Survey of Primary Care Physicians; 2013 International Health Policy Survey; Commonwealth Fund *National Scorecard 2011*; World Health Organization; and Organization for Economic Cooperation and Development, *OECD Health Data, 2013* (Paris: OECD, Nov. 2013).

Figure 1: Commonwealth Fund Healthcare System Country Rankings

Quality of care: In the Commonwealth Fund's (2014) report, the term quality is subdivided into three equal-weighted categories, including effectiveness, safety, coordination, and patient-centeredness. Before assessing the quantitative results for each division, operational definitions must be established. Effectiveness represents "the degree to which patients receive services that are effective and appropriate for preventing or treating a given condition and controlling chronic illness" (Radley et al. 2011: online). Within this framework, the United States performs relatively well, ranking 3rd in the Commonwealth Fund's assessment (2014: See Appendix 1 for effectiveness measures).

The term safe care, on the other hand, is defined by the Institute of Medicine as "avoiding injuries to patients from care that is intended to help them" (IOM 2001; Commonwealth Fund 2014). Appendix 2 lists the metrics used to rank safety, but importantly the U.S. ranks 7th overall in this category, which, although rather low, represents an improvement from the 2010

Commonwealth Fund Report. The U.S. now leads all nations in controlling rates of hospital-acquired infections.

> Coordinated care
>
> throughout the course of treatment and across various sites of care helps to ensure appropriate follow-up treatment, minimize the risk of error, and prevent complications. Failure to properly coordinate care raises the cost of treatment, undermines the delivery of appropriate, effective care, and puts patients' safety at risk (Commonwealth Fund 2014).

The importance of effective coordination is only magnified when care is placed in the context of changing patient needs in the 21st century. As the Baby Boomer generation continues to age, and the healthcare system experiences the shift from an acute- to chronic-disease burden, successful coordination of providers will be essential to delivering high quality care (Berwick 2008). In this measure, the U.S. performs average, ranking 6th (Appendix 3).

The shift towards patient-centered care has been relatively recent in comparison to the discussion of the other dimensions of quality; nonetheless, the Commonwealth Fund's *National Scorecard* defines it as "care delivered with the patient's needs and preferences in mind" (Why not the Best 2011). In the United States, where patient choice, preference, and autonomy are touted as ideological axioms, one would expect that patient-centered care would be a priority. And, the results do, to some extent, support this claim, with the U.S. receiving a relatively strong 4th place rank in this category (Appendix 4).

Access: Care is accessible if it is both affordable to the individual and received within a timely manner (Commonwealth Fund 2014). Looking first at cost, the U.S. possesses the highest proportion of citizens who are unable to receive care due to price when compared to any of the 10 other studied nations (Commonwealth Fund 2014) (Appendix 5). More specifically, 37% of the population reported that they went without recommended care, necessary prescriptions, or doctor visits due to lacking financial means (Commonwealth Fund 2014).

Timeliness of care is a slightly more complex matter. While the U.S. performed a moderate 5th overall with respect to timeliness, drastic variation existed when bifurcating the results for specialist and primary care services (Commonwealth Fund 2014) (Appendix 5). Only 6% of U.S. patients wait 2 months or more to see a specialist while that number is 29% in Canada and 26% in Norway. Conversely, patients in the U.S. reported significantly lower access to primary care services compared to emergency services (Commonwealth Fund 2014).

Efficiency: An efficient healthcare system is one that maximizes clinical outcomes with minimal resource input (Commonwealth Fund 2014). More frequently, efficiency is defined as Value, which is the ratio of outcomes and costs. It is in this category, perhaps more than anywhere else, that the United States system underperforms (Appendix 6).

The U.S. healthcare system spends 17.7% of its Gross Domestic Product (GDP) on healthcare services, nearly 6% more than 2nd most costly system – the Netherlands (11.9%) (Commonwealth Fund 2014). Simultaneously, the U.S. devotes disproportionately large funding to administrative expenses; while administrative expenses total only 0.6% in Norway, the U.S. devotes 7.1% of all healthcare dollars to maintaining healthcare administration and insurance (Commonwealth Fund 2014).

It is important to recognize that these increased expenditures have spillover effects into care quality and access; providers in the United States were the most likely to report that insurance coverage restrictions limit their ability to provide medically necessitated tests or treatments (Commonwealth Fund 2014). Issues of cost, in the U.S., therefore impact the availability of services.

Equity: In the Institute of Medicine's "Crossing the Quality Chasm" 2001 report, equity is highlighted as a principle objective for an effective healthcare system. Care should "not vary in quality because of personal characteristics such as gender, ethnicity, geographic location, and socioeconomic status" (3). In this regard, the U.S. performed last overall and in nearly all individual equity metrics (Appendix 7). For example, the U.S. had the highest proportion of patients below the national median income that went without medical treatment because of cost, waited 2 months or longer to see a specialist, waited 2 hours or more in the ER, rated the quality of their care as poor, and rated their doctor as poor (Commonwealth Fund 2014).

Healthy Lives: Unsurprisingly, issues of equity, access, efficiency, and quality, result in lower healthcare outcomes. The U.S. has the highest mortality rate due to healthcare services (96 in 100,000), infant mortality rate (6.1 per 1,000), and the 2nd lowest life expectancy at age 60 (17.5), compared to each of the countries studied (Appendix 8). These same findings were mirrored by the Institute of Medicine's 2013 Report, "Shorter Lives, Poorer Health."

Discussion of Commonwealth Fund Findings: Now, there are some important caveats that should be recognized when interpreting the results of the 2014 Commonwealth Report. For one, effectiveness is measured exclusively in terms of preventative services and successful management of *chronic* illness. This metric therefore favors systems which have a stronger foundation of primary care services as opposed to specialist physicians that focus on acute and rare disease. Yet, despite the bias in this measurement, it is nonetheless valid; unlike in the 19th and 20th centuries when the principle causes of death were due to bacterial infection, diarrhea, and other acute illnesses, 21st century patients now need effective chronic care services (Mascie-Taylor and Karim 2003). Effectiveness in the 21st century is therefore far different than effectiveness in centuries prior.

Moreover, safety measurements are based upon patient-response data in the Commonwealth Report (2014). This leaves such results vulnerable to bias due to cultural differences between populations. However, while this is true, countless other studies report poor safety in the American healthcare system. For example, the Institute of Medicine's 1999 Report, "To Err is Human: Building a Safer Health System," cites that 44,000-98,000 people die in American hospitals annually due to preventable medical errors.

Summary - the current state of U.S. healthcare: The U.S. system is failing its patients. Rising national expenditures threaten long-term sustainability, high costs for individuals threaten access, lacking primary and preventative care threaten effectiveness, high rates of hospital-acquired conditions threaten safety, and variability in insurance coverage threaten equity. All of these combined effects have produced a system with significantly lower health outcomes compared to comparable middle- and high-income westernized nations.

It is difficult to blame just the healthcare system. According to the U.S. Office of Disease Prevention and Health Promotion, the health of a population is affected by 5 interconnected social determinants: (1) economic stability, (2) education, (3) social and community context, (4) health and healthcare, and (5) the neighborhood and built environment (HealthyPeople 2020 2015).

Figure 2: Social Determinants of Health

In other words, actual medical services make up only one component of a system. Yet, nonetheless, while an effective healthcare system cannot necessarily guarantee a healthy population, it certainly possesses considerable influence on health outcomes within a society.

III. MARKET-BASED CORRECTION IN THE HEALTHCARE INDUSTRY

If the Commonwealth Report (2014), IoM (1999 and 2013), and IHI (2008) teach us anything, it is that the U.S. healthcare system needs *targeted* reform. There are a number of specific issues – increasing global and personal costs, poor chronic disease management, lacking preventative services, and insufficient equity, that must be addressed if the system is to be sustainable in the long-term.

To improve healthcare delivery, a number of agreed upon essential changes come to mind. For example, the 2001 IoM “Crossing the Quality Chasm” report cites the need for greater integration and coordination of healthcare. With today’s population often suffering from multiple simultaneous chronic conditions, it is now more important that care be *organized* around collaboration between departments (IoM 2001). This means shifting from an individual to team-based model of care delivery (IoM 2001). Interdisciplinary departments must coordinate the multiple care needs of an individual.

Time also becomes a far more important factor; while acute, episodic care can be managed by an individual, chronic illnesses must be treated over a number of years (IoM 2001). Because such care requires a greater time investment, it is essential that certain responsibilities be delegated from specialists to effective primary care teams (and even the individual patient). Shifts towards self-management, prevention, and management through integrated primary care, are known to improve patient outcomes while decreasing total expenditures (IoM 2001).

Yet, while the “what” of reform is often straightforward, “how” to implement such change becomes far more complex from a policy perspective. In the United States, where the system is a dispersed and piece-wise mixing of public and private interests, enacting wide-sweeping simultaneous change has been extremely challenging (IoM 2001; Brown 2003; Henwood et al. 2003; Holahan and Peters 2014).

To date, a number of suggested approaches have been tested. For example, in 2004, Dr. Donald Berwick and the Institute for Healthcare Improvement launched the “100k Lives Campaign,” a program to prevent 100,000 deaths from hospital acquired conditions in U.S. hospitals. The program was founded on a simple premise: if healthcare organizations shared strategies and evidence from quality improvement programs, the entire network could benefit and improve patient care (Berwick 2008). Successful hospitals volunteered to serve as “mentors” to new organizations entering the program while “nodes” functioned as regional campaign offices. Through information collaboration and a team-based approach to care delivery, the 100k Lives Campaign was a resounding success (Gosfield and Reinertsen 2005).

Unfortunately, such collaborations are few and far between. In healthcare, specialization breeds separation. Private organizations simply lack the incentive to communicate and collaborate with competitor institutions in the region. To improve the care of another organization would be to risk personal market share. In other words, a constant tension exists between the need for greater continuity, integration, and collaboration, and the current market-based drive towards separation (Brown 2003).

Healthcare is riddled with these market failures that undercut current efforts towards quality improvement. Take, for example, the fee-for-service payment system. Under this model, healthcare organizations are compensated for each individual test or procedure administered (Schroeder and First 2013). This payment model inevitably incentivizes greater volume; organizations that provide more care receive greater reimbursement (Haas-Wilson 2001;

Schroeder and First 2013). Yet, more is not always better in the context of healthcare. Not only does such an incentive cause rising national expenditures, but it also threatens patient safety (Quanstrum and Hayward 2010; Schroeder and First 2013; IoM 1999; Enthoven 1988). All tests and procedures carry with them a degree of risk, and providing them when not medically necessitated places patients at greater chance for iatrogenic healthcare effects (Quanstrum and Hayward 2010; Schroeder and First 2013).

The Affordable Care Act

If unaligned market forces are damaging our system, then maybe properly organized incentive systems could rectify it (Haas-Wilson 2001; Schroeder and First 2013; VanLare and Conway 2012; Chien and Rosenthal 2016). This is the philosophical premise of the Affordable Care Act (ACA) (2010) (VanLare and Conway 2012; Chien and Rosenthal 2016; Enthoven 1988).

The ACA (2010) is a dauntingly large document, spanning more than 3,000 pages. Nonetheless, its market-based correction strategies can be summarized through three central points: (1) the development of health insurance exchanges (HIX), (2) the shift towards value-based payment modifiers, (3) and incentivizing advanced care models. Let us discuss each in turn.

Health insurance exchanges: The health insurance market has traditionally allowed companies a great degree of freedom regarding plan coverage options (Oberlander 2014; Enthoven 2004). In other words, an insurance company would charge different amounts for different plans each possessing drastically different coverage options (Enthoven 2004; Holahan and Peters 2014). Because of the complexity of plans, the insurance market has been largely price inelastic (Enthoven 2004). From an economic perspective, this suggests that demand for insurance has failed to adjust following increases in price or decreases in coverage options (Blumenthal and Collins 2014; Ringel, Hosek, Vollard, and Mahnovski 2010).

To re-establish the potential of market forces, the Affordable Care Act (2010) creates health insurance exchanges – i.e. online websites where insurance companies register and present different healthcare plans to the public (Blumenthal and Collins 2014). Importantly, the plans are tightly regulated – with certain coverage options mandated. As such, with relative consistency in quality, consumers can thereby judge plans based off of cost disparities between companies (Holahan and Peters 2014).

Value-based payment: Another strategy to utilize market forces has been through the changing of payment models by the Centers for Medicare and Medicaid Services (CMS). Starting in 2005, hospitals were mandated to annually submit both outcome (e.g. standardized mortality ratios) and process measures (e.g. use of electronic health records) to CMS, which were used to assess hospital quality (Jerrard 2008). At the time, the results of this quality report did not affect hospitals financially – i.e. data submission was the only requirement for completion. However, following the implementation of the Affordable Care Act (2010), the Pay for Reporting program has evolved into the Pay for Performance Program (otherwise known as Value-Based Purchasing) (VanLare and Conway 2012). CMS also implemented the “Present on Admission”

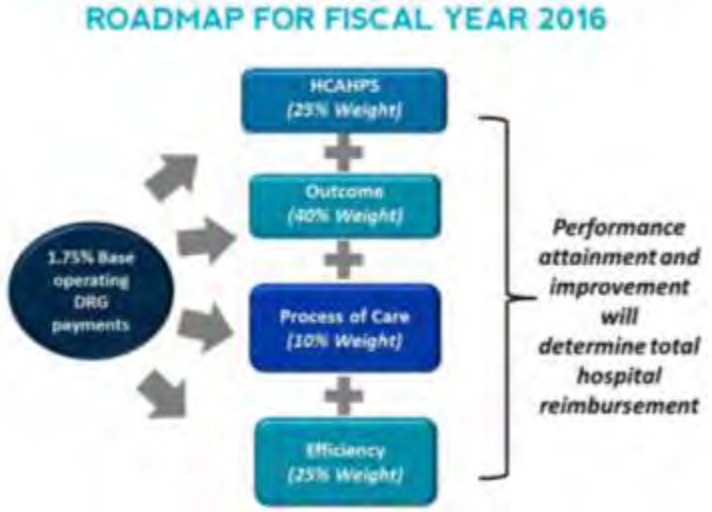


Figure 3: Relative weights of domains in CMS VBP program in 2016

and “Never Events” rules in 2007 and 2008, respectively, whereby the government refuses to pay for hospital-acquired infections or damages from negligent care (e.g. bed sores or pressure ulcers).

All of these policy adjustments together create a system where reimbursement is *adjusted* based upon the quality of healthcare facilities (Figures 3). In the VBP program, payment adjustments of up to 2.0% (by 2017) are made based upon (1) Patient Experience, (2) Safety, (3) Efficiency and Cost Reduction, (4) Process Measures, and (5) Clinical Care Outcomes (VanLare and Conway 2012).

Incentivization of advanced care models: The ACA also includes provisions to incentivize greater coordination, integration, IT utilization, primary care, and primary prevention through the development of Accountable Care Organizations (ACOs) and Patient-Centered Medical Homes (PCMHs).

The National Center for Quality Assurance (NCQA), an independent non-profit organization, currently sets the standards for PCMH and ACO recognition. The PCMH model is based off of a systems approach, whereby care and reforms are considered in terms of inputs, processes, and outcomes, with a continual evidence-based feedback loop (Appendix 10). Founded on the principles of effective HIM, multidisciplinary care (i.e., the utilization of mental health specialists), team-based care delivery, and patient empowerment, the PCMH is designed to address issues of lacking integration and continuity (AAFP 2015; Barr, 2016).

Similarly, ACOs are provider-led organizations that agree to take on the responsibility of maintaining the health and wellness of a defined population (AHRQ 2015; Barr, 2016). By taking responsibility for the entire *population* and not simply the patients that walk through the hospital doors, ACOs stress shifts towards hospital-primary care-community services integration.

While ACOs and PCMHs have been met with controversy and varying success, what is important to recognize is that these programs have been *incentivized* by the U.S. government. Currently, the Comprehensive Primary Care (CPC) Initiative and Federally Qualified Health Center (FQHC) Advanced Primary Care (APC) Program provide financial bonus payments to physicians that shift towards the PCMH model. Similarly, in the case of Accountable Care Organizations, the Medicare Shared Savings Program – established by section 3022 of the ACA – provides financial incentives for achieving ACO accreditation (CMS Shared Savings Program).

Room for Further Marketization

In the American cultural context, market forces represent the philosophical underpinning of our healthcare system (Oberlander 2014). It is believed that healthcare represents a market good, not a social one, and that the forces of supply and demand hold the power to maximize efficiency and quality (Oberlander 2014; Barr, 2016). Unfortunately, a number of market failures have undercut the effectiveness of this model; marketization through the ACA represents a step towards re-aligning market forces with desired clinical and economic outcomes. Moreover, I attest that market-based solutions represent the most viable opportunity for American reform because of its compatibility with American cultural and political ideology (Oberlander 2014).

Nonetheless, I recognize one area of missed opportunity: the patient and provider interaction. In the ACA, VBP incentivizes the provider to improve the outcomes of the patient, HIXs incentivize the insurer to improve options for the patient, and PCMH/ACOs incentivize providers to improve the outcomes of the patient. But no incentive exists that enables *patient behavior to directly influence the care of the providers.* The power of current market based reforms exist between the government and the physician or the government and the insurer, but not between the consumer and the physician.

Empowering the patient is crucial to effective market based reform. Patient choice, freedom, and autonomy, are held up as universal axioms in the U.S. system (Oberlander 2014; Andreassen and Trondsen 2010). Because healthcare is philosophically considered a market good, then its dissemination should be governed by laws of supply and demand. That is, an individual patient assesses the quality of care at varying institutions, compares that to varying cost, and makes an informed decision regarding where to receive care. In practice, this is not how the dynamic currently plays out. I assert that what exists instead is the illusion of choice.

Issues of health literacy, price transparency in a retrospective payment system, a provider-patient power and knowledge asymmetry, and lacking quality of care data availability, make it nearly impossible for the average consumer to assess the value of care at different institutions. Without such transparency, it is no wonder that American healthcare expenditures have risen to 17.6% of total GDP with appallingly unsatisfactory global health outcomes. Patients simply lack power within the system to influence how care is delivered.

I postulate that if such empowerment were achieved, it would initiate a ground-up incentive for quality improvement and cost containment. Issues identified in the Commonwealth Report (2014) – lacking integration, coordination, and preventative services – would have a driving force for rectification.

IV. CONSUMER CHOICE AND MARKET-BASED REFORM – HOW FAR ARE WE?

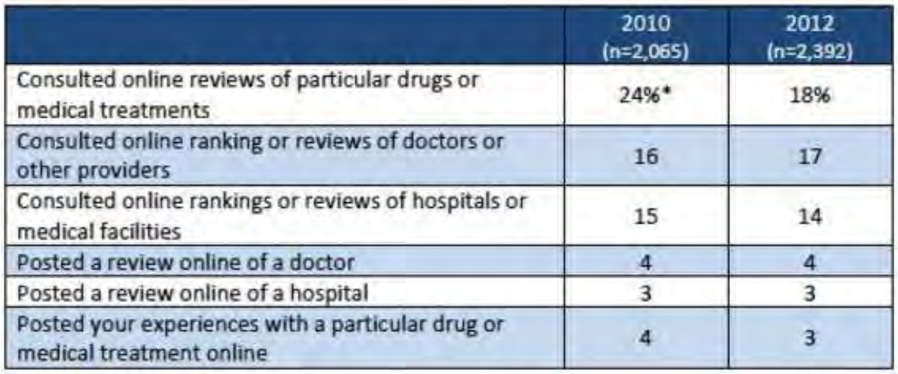
The average consumer has a number of different media forms at her disposal when seeking healthcare information. Much of this has changed drastically in the internet age.

Traditionally, healthcare information was gathered through TV, radio, newspaper, magazines, billboard advertisements, and through family/friends/coworkers/referrals from other providers (Cutilli 2010; Berkowitz, 2010).

Following internet expansion, consumers now possess a far wider number of resources at their disposal: (1) social media (Facebook, Twitter), (2) healthcare facility websites, (3) medical advice websites (e.g. WebMD), (4) new venues for advertisements, and (5) hospital rating resources (e.g. Yelp, Hospital Compare, Vitals, etc.) (Richards et al. 1998; Eysenbach and Diepgen 2001; Henwood et al. 2003; Rothenfluh et al. 2016).

Thus, the internet age has drastically changed information access – but not without a qualifier. A 2013 study by the Pew Research Center found that while 93% of patients use word-of-mouth from family and close friends to find services, only 17% supplemented this with evidence from alternative internet sources (Table 1). In other words, while the internet has made vast quantities of information available, a large segment of the patient population has been hesitant to rely on web-based sources for health information (Pennbridge et al. 1999).

Table 1: Percent of citizens using the internet to find doctors or medical facilities.



	2010 (n=2,065)	2012 (n=2,392)
Consulted online reviews of particular drugs or medical treatments	24%*	18%
Consulted online ranking or reviews of doctors or other providers	16	17
Consulted online rankings or reviews of hospitals or medical facilities	15	14
Posted a review online of a doctor	4	4
Posted a review online of a hospital	3	3
Posted your experiences with a particular drug or medical treatment online	4	3

Thomas Friedman (2005) calls the internet “the great global flattener;” a globalizing equalizer, which has enabled billions to access unlimited quantities of information instantaneously. With the explosion of the internet within every other industry, lacking healthcare penetration could only be perceived as perplexing. Why has internet usage expanded so drastically in some areas of healthcare – for example, diagnostic advice from WebMD – but so slowly in others – for example, social media usage amongst hospitals?

Limitations of the Internet as a Tool for Consumer Empowerment:

Rothenfluh *et al.* (2016) point to the personal nature of choosing a physician. In one interview, the subject reports that

> "such a physician rating website would probably be useful to make a first contact, but after that it is obviously very much about the feeling you get, the appearance and impression once you get there. With a hotel, you book and then you say afterwards 'okay, that was great' and you may go again some other time. There it is about the best offer at that moment. It isn't really that relevant" (Rothenfluh et al. 2016: 7).

In their research, Rothenfluh *et al.* (2016) refer to this phenomena as the theme of "trust." Because seeking medical treatment is such a personal experience – i.e. the patient is placed into an increasingly vulnerable position as illness severity increases – these less tangible "feelings" play a far more important role when settling on a provider. As such, an individual is far less likely to trust crowd-ratings over traditional resources such as family, friends, and coworkers (Rothenfluh et al. 2016, Pew Research Center 2013).

Another theme that Rothenfluh *et al.* (2016) uncovered relates to perceived competency. In their study, participants were asked to evaluate the quality of hotels and the quality of providers through similar ratings websites; the subjects displayed confidence when rating hotels but *perceived* themselves as less able to assess physician quality:

> "I don't understand any of it (the diagnosis and treatment prescription), so I trust in that, what he tells me and then I just take that (the medication) and do what he tells me. Obviously there are differences among physicians, it always depends upon what your problem is. But in the end… Yeah well, you also don't know which one is better than the other ones. You never know!" (7).

It is important to remember that the doctor-patient interaction is defined by its power and information asymmetries (Haas-Wilson 2001; Henwood et al. 2003; Rothenfluh et al. 2016).

This dynamic, Rothenfluh *et al.* (2016) suggest, makes patients less confident in their ability to judge more qualified and educated providers.

This theme of lacking competency is supported by alternative psychological studies. For example, the Milgram experiments found that average men would perform acts against their own conscience when instructed to do so by a formal authority figure - in this case, a physician wearing a standard white laboratory coat (Blass 2004). In other words, confronting those with perceived authority is challenging because patients are socialized not to in American society. Already vulnerable patients simply feel underqualified to trust anonymous crowd-ratings over the perceived expertise of a trained clinician.

As such, with lacking standard evidence, individuals typically fall back on trusted family and friends; while they may not be clinicians themselves, they still possess a degree of familial credibility that anonymous sources lack (Verhoef et al. 2014; Stvilia, Mon, and Jeong 2009).

Current Forms and their Limitations:

In the United States, a number of web-based consumer choice healthcare websites exist. This section will serve as an encompassing review of each site within the context of prevailing research regarding patient "trust" and "perceived competency." To date, 8 websites dominate the online healthcare ratings market: (1) Consumer Reports, (2) Hospital Compare (and other CMS versions), (3) Health Grades, (4) Leapfrog, (5) U.S. News and World Report, (6) Vitals, (7) ZocDoc, and (8) Yelp.

Within this list, options can be categorically bifurcated as either (1) anonymous crowd-rating sites or (2) standardized evidence-based reviews. Let's discuss each type in turn.

Anonymous Crowd-Rating Sites: ZocDoc, Yelp, Vitals, and HealthGrades produce physician-level scores based exclusively on anonymous star ratings. On the Vitals website, users search for physicians by specialty, location, and insurance type (Vitals.com 2016). Physicians are then listed according to star rating rank, with clickable names which direct the user to a "more information" page specific to that provider. Certain standard identifying information are provided (specialty, address, phone number, directions, insurance acceptance) as well as a list of each individual rating and comments from previous patients (Vitals.com 2016).

ZocDoc constructs a similar interface; i.e. an initial search page broken down by specialty and location, a second page with lists of providers ranked by star rating, and a subsequent page for each physician (including each individual rating and identifying information) (ZocDoc.com 2016). ZocDoc differs from Vitals in that it offers a "Book an Appointment" calendar for each physician practice. Physicians can register their practices with ZocDoc and link their scheduling calendars, thereby enabling consumers to make appointments and confirm insurance type directly on the page (ZocDoc.com 2016).

Yelp is undoubtedly the most common of all crowd-rating websites, for it offers reviews in all industries – not just medicine. Unlike Vitals and ZocDoc, Yelp provides ratings for *both* physicians and healthcare facilities. However, the rating methodology is the same as other competitors. Each facility or doctor gets its own page, 1 to 5 star rating, and list of individual reviews alongside useful identifying information (Yelp.com 2016).

Of all of these, HealthGrades provides the most comprehensive reviews. Similar to its competitors, its initial search page is broken down by specialty service (HealthGrades.com 2016). Following pages provide lists of doctors with corresponding anonymous star ratings and identifying information. HealthGrades' innovation is that it offers a list of hospital affiliations when you visit a doctor's specific page. Users can click on each hospital and get re-routed to a new window, which offers further information on that specific *facility.* Facility-level data are evidence-based, which is a strength, but there are some notable caveats.

For one, HealthGrades does not provide global or departmental ratings for individual facilities. Instead, data are limited to individual variables, which users are left to make sense of alone. For example, HealthGrades lists out the "percent of patients that would definitely recommend this hospital" – a common metric used in national HCAHPS surveys (HealthGrades.com 2016). However, we are left wondering what the national average and variance are for the "percent who would recommend" variable score. Without a point for comparison, or a standardized score for all HCAHPS survey questions, it becomes challenging to assess patient satisfaction at the facility level. Moreover, while patient satisfaction is undoubtedly an important metric, HealthGrades emphasizes these data and seems to de-emphasize data which are harder to obtain, such as clinical outcomes and cost efficiency. This kind of convenience sampling is inconsistent with broader definitions of healthcare quality, which will be discussed later.

Evidence-Based Review Sites: Sites like Hospital Compare, Leapfrog, Consumer Reports, and U.S. News and World Report, each provide standardized reviews on healthcare providers and/or facilities. For example, Leapfrog produces its own standard annual survey, which it disseminates to hospitals throughout the nation (LeapFrog.com 2016). However, it is important to recognize first that Leapfrog surveys are limited to hospitals. No other facility types are included, which leaves a large gap in coverage of patient transparency needs.

Second, participation in the Leapfrog survey is entirely voluntary, and it shows; 96 of 153 hospitals in Pennsylvania declined to report in 2016's annual survey (LeapFrog.com 2016).

Likewise, Hospitals can pick-and-choose which categories they provide data for and which they don't. While 1,260 hospitals reported on C-Sections, only 984 reported on mortality rates for surgeries (LeapFrog.com 2016). Thus, hospitals are enabled to highlight categories in which they do particularly well and hide unflattering data. Most frustratingly, for hospitals that do report the entirety of the dataset, no global scores are provided. Instead, scores for each individual variable are provided, thereby complicating interpretation for the average consumer who must sift through approximately 100 individual metrics to estimate the quality of care at a particular facility.

Consumer Reports suffers from similar shortcomings. For one, it only rates hospitals. Second, similar to Leapfrog, we are faced with data overload. Consumer Reports provides data within safety score components, outcomes components, and experience components, but no overall scores for each component or total global score (ConsumerReports.com 2016). Likewise, data are listed at the hospital level; no ratings are provided by service type or department. Individuals searching for heart care are thereby given the same data as individuals seeking dermatology services. Without providing consumers with a method to navigate this complex data, Consumer Reports and Yelp undermine the value of the information.

CMS provides its own ratings service through the Hospital Compare, Nursing Home Compare, and Dialysis Compare websites. While these websites are by far the most comprehensive, they too present a number of shortcomings. For one, each facility type gets a separate ratings website. Thus, patients may know about hospital compare, but not know about Nursing Home Compare, etc. Second, hospitals receive global scores, which are useful and unique, but similar to Consumer Reports, Hospital Compare does not break down scores by service type. Instead, one total global score is offered, which may mean different things to patients with varying needs.

Likewise, data are broken down into separate categories (domains of quality): patient experience, timely and effective care, complications, readmissions and deaths, use of medical imaging, and payment and value of care. In the ratings process, these different domain scores are aggregated through a weighted average to produce the overall score. However, Hospital Compare does not provide a rating for each domain – nor does it offer the user an explanation of each domain’s relative weights. Thus, we have no idea how CMS arrived at the overall score. Nor are consumers able to assess the quality of specific individual domains which he or she finds particularly important. Instead, within each domain, data percentage scores are provided for each individual variable. In total, this aggregates to 76 total percentages that consumers are left to sift through and value independently (Hospital Compare, 2016). In summary, we are presented with the same data overload problem seen in Leapfrog and Consumer Reports.

The Health section of *U.S. News and World Report* offers ratings for Hospitals and Nursing Homes. Like the other sites evaluated, this leaves unresolved the issue of lacking facility coverage. Likewise, Top 100 Rankings are offered for a variety of specialties and procedures. However, the list of ranked hospitals is therefore only 100 long (health.USNews.com 2016). Those outside those bounds are not given global ranks or overall scores, making comparison between facilities in a local area particularly challenging. In other words, *U.S. News and World Report* highlights each year’s top performers in particular specialties, but is not designed to serve as a comprehensive search engine of all American healthcare facilities.

Lasting Impressions:

The U.S. healthcare market is saturated with a variety of healthcare rating websites. Yet despite their immense prevalence, the Pew Research Center's 2014 report indicates that internet utilization in healthcare lags far behind other consumer industries (Pew Research Center 2014). An in-depth review of current online resources brings insights as to why: the majority of sites are limited to anonymous crowd-ratings, which fail to overcome barriers of "trust" and "perceived competency" as outlined by Rothenfluh et al. (2016). And, for those sites which do utilize standardized datasets, gaps in coverage, inconsistent rating methodologies, and data overload, may undermine the dissemination of content (Verhoef et al. 2014; Stvilia, Mon, and Jeong 2009; Malat 2001). Future sites should therefore focus on providing standardized, reliable, comprehensive data, in a format easily internalized by the lay public.

V. WHAT IS QUALITY?

But what are standardized, reliable, comprehensive data? One of the challenges of rating healthcare facilities through quantitative metrics is that it presupposes a theoretical framework of healthcare quality. Unless we know what an excellent facility *looks like,* it will be impossible to provide comprehensive ratings.

Donabedian's Model:

The history of quality evaluation in medicine begins with Dr. Avedis Donabedian, a former Public Health Professor at the University of Michigan, who wrote the landmark article "Evaluating the Quality of Medical Care" in 1965. In his paper, Donabedian (1965) argued that healthcare can only be assessed once a fundamental operational definition of quality is established. Moreover, he recognized that the agreed upon framework will in turn have "profound influence on the approaches and methods" of healthcare delivery (Donabedian 1965: 692).

In Donabedian's (1965) framework, he first defines medicine through a systems model, where all aspects of the healthcare system are categorized as inputs, processes, and outputs. Under this model, inputs are "not the process of care itself, but the settings in which it takes place" (Donabedian 1965: 693). Broader structural variables include financial inputs, facilities and equipment, the qualifications of medical staff, and the administrative structure which serves as support for medical operations. Moreover, processes represent the total number of acts that comprise medical services delivered, such as physical examinations, diagnostic tests, treatment procedures, etc. (Donabedian 1965: 693). Processes are not concerned with clinical outcomes (the end result of care), but are instead measures of whether "good" care has been applied. This includes measurements of continuity, coordination of care, acceptability of care to the patient, justifications of diagnostic tests and procedures, etc. Lastly, outputs represent clinical outcomes (Donabedian 1965: 693). How the patient fares following medical treatment is thereby a measurement of the quality of treatment itself.

Importantly, Donabedian (1965) stresses the importance of each component of the systems model. As he rightly notes, traditional healthcare quality assessment has focused heavily on outcomes, thereby ignoring processes and inputs. But outcomes are not always valuable metrics. For example, Donabedian (1965) affirms that the individual outcome measure must always be questioned. While mortality ratios may be useful for assessing the quality of life-threatening ailments, such a measurement may have no value when placed in the context of routine procedures. Likewise, measurements of short-term outcomes have little value when assessing care for chronic illnesses, where outcomes take years (if not decades) to materialize. Thus, while outcomes appear the most concrete of all systems measurements, other complimentary data are essential for adequately assessing quality.

The same is undeniably true of inputs and processes. Different illnesses require different kinds of (and numbers of) specialists, varying degrees of continuity (acute v. chronic illness), different justifications for tests and procedures, and so on. Input and Process measures, like outcomes, must have *“specified relevant dimensions, values and standards to be used in assessment”* [italics added for emphasis] (Donabedian 1965: 694).

Donabedian’s (1965) initiated the movement towards quality assessment in healthcare, but his framework is not without criticism. For example, Mitchell, Ferketich, and Jennings (1998) argue that his model is too linear. While Donabedian (1965) separates quality into inputs, process, and outputs, his framework fails to show how these categories interact and affect one another reciprocally (Mitchel et al. 1998). Moreover the Donabedian (1965) model presents limitations with respect to lay public dissemination. Meaning, the model was never intended for use by consumers. As such, these categories may appear abstract to average consumers in comparison to alternative models of healthcare assessment.

Numerous other organizations have spearheaded movements towards quality assessment in healthcare following the research pioneered by Donabedian (1965) (Besiki et al. 2009; Martin et al. 2015; Schang et al. 2016; Carayon et al. 2014; Backman et al. 2016; Mosadeghrad 2012; NCQA Guide to Healthcare Quality 2011; NQF 2002; OECD Kelly and Hurst 2006; WHO Quality of Care 2006). The most commonly cited frameworks include: (1) the World Health Organization, (2) Organization for Economic Cooperation and Development, (3) the Institute of Medicine, (4) and the National Quality Forum. Let’s discuss each in turn.

World Health Organization (WHO) Quality Framework:

In 2006, the World Health Organization produced a report titled “Quality of Care: A Process for Making Strategic Choices in Health Systems,” with the aim of defining the domains of healthcare quality to guide future national quality improvement efforts. In its opening paragraphs, the writers mirror Donabedian’s (1965) tone and claim that “every initiative taken to improve quality and outcomes in health systems has as its starting point some understanding of what is meant by ‘quality’”(WHO Quality of Care 2006: 9).

Importantly, the WHO definition of quality focuses on the health system as a whole (WHO Quality of Care 2006). With a more system-wide perspective, this framework encompasses six separate dimensions: (1) effectiveness, (2) efficiency, (3) accessibility, (4) acceptable/patient-centeredness, (5) equitability, and (6) safety (WHO Quality of Care 2006).

Effective care is adherent to an evidence-base and results in improved outcomes for both individuals and communities. Interestingly, this dimension has roots in Donabedian’s (1965) two dimensions of Processes and Outputs. That is, effective care represents evidence-based care (the right care/the right process) and also yields beneficial outcomes (has desirable outputs).

Efficient care can be said to maximize resource use while simultaneously avoid unnecessary waste (WHO Quality of Care 2006). In the context of Donabedian’s (1965) original model, this domain can be encompassed by both inputs and outputs. Resources are inputs, so maximizing resource use represents maximizing structural inputs (capital, staff, education, technology) into the system. Simultaneously, waste represents an output – a consequence of efficient or inefficient processes.

Accessible care is defined by the WHO as timely, geographically reasonable, and provided in a setting where skills and resources are matched to medical need. Likewise, this domain can be encapsulated by Donabedian's (1965) framework. What care is delivered, where it is delivered, and how it is delivered and obtained, represents the Process of care delivery.

Acceptable/patient-centered care is care that accounts for the preferences (cultural or philosophical) of a particular patient or community. This serves as a measure of process. The services that are delivered should be limited by the parameters of patient consent and preference.

Equitable care is meant as consistent care. It does not vary in quality due to demographic or personal patient characteristics. This is also a measure of process. It presupposes that care delivered should be the same for all patients that interact with the system.

Lastly, safety comes from care processes that minimize the risk of harm and the rate of harm to patients. In Donabedian's (1965) model this could be broken into the categories of both process and outcomes. On one end, correctly administered care minimizes the risk of harm by using evidence-based guidelines and agreed upon processes. Likewise, rate of harm represents a safety outcome domain.

Interestingly, when each of these six domains are placed within Donabedian's model, 5 possess roots in process, 3 in outcomes, and only 1 in inputs. An important conclusion can be drawn from this exercise. While outcomes are the most obvious measures of quality (Donabedian 1965), process measures are of great (almost deterministic) significance. Processes – i.e. the 'who, what, when, where, how, and why' of care delivery – define a large segment of what the healthcare system *is.* While outcomes are undeniably important, it is hard to view them as anything but a consequence of inputs and processes.

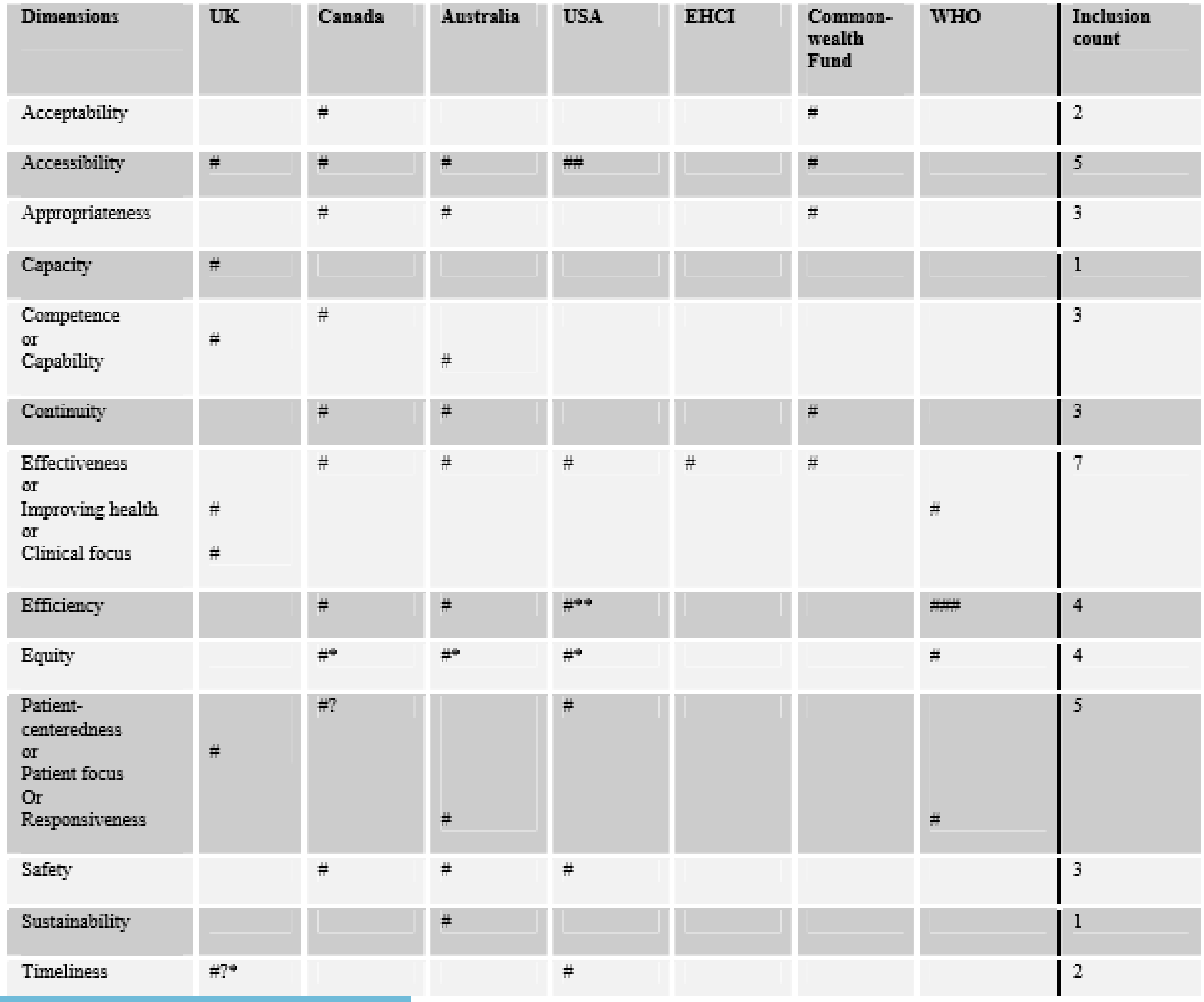
Thus, there is an important strategic approach to the WHO model of care quality; in the forward of the document, the writers state that the report is designed to serve as a framework to select "*new interventions and strategies for quality improvement*" (WHO Quality of Care 2006: 9). If the framework included largely outcomes domains, then it would not provide directive suggestions for *how* to improve. Conversely, by framing domains largely in the form of processes, the WHO offers concrete locations and interventions for targeted improvements. In other words, outcomes can only be improved by modifying inputs and processes. Even if inputs, processes, and outputs are equally important, emphasizing changes in processes might allow for more strategic leverage within the context of broader policy objectives.

The OECD Quality Framework:

Like the WHO report, the OECD produced its own report on quality indicators in 2006, which it titled "The Health Care Quality Indicators Project" (Kelly and Hurst 2006). The project was guided by a panel of experts sourced from 23 countries to answer (1) "**what** concepts, or dimensions of quality of health care should be measured and" (2) "**how**, in principle, should they be measured" (Kelly and Hurst 2006: 3). The Health Care Quality Indicators Project (HCQI) was started in 2001 as a method for defining quality that fit within the needs-based context of broader societal population health and healthcare needs. Likewise, it aimed to synthesize the multiple approaches of quality measurement in each of the 23 contributing nations to come up with a universally applicable definition, framework, and set of measurements for quality. Table 2 illustrates a full list of dimensions from six member countries as well as which domains each nation includes.

Interestingly, the most common domain from the nations was “Effectiveness or Improving Health or Clinical Focus,” which represents a measure of output. The second most common measure was “Patient Centeredness or Patient Focus or Responsiveness,” which is a measure of process. Thus, most nations seem to emphasize the importance of outcomes. Outcomes are tangible and emotionally charged; a healthcare system is measured by how its patients fare. Measures of process – how a nation achieves its outcomes – are commonly secondary.

Table 2: Dimensions of Quality by Nation



Dimensions	UK	Canada	Australia	USA	EHCI	Common-wealth Fund	WHO	Inclusion count
Acceptability		#				#		2
Accessibility	#	#	#	##		#		5
Appropriateness		#	#			#		3
Capacity	#							1
Competence or Capability	#	#	#					3
Continuity		#	#			#		3
Effectiveness or Improving health or Clinical focus	# \#	#	#	#	#	#	#	7
Efficiency		#	#	#**			####	4
Equity		#*	#*	#*			#	4
Patient-centeredness or Patient focus Or Responsiveness	#	#?	#	#			#	5
Safety		#	#	#				3
Sustainability			#					1
Timeliness	#?*			#				2

The HCQI model for quality was thereby constructed by synthesizing the most common dimensions from host nations with broader models outlined by the Institute of Medicine, Canadian Health Indicator Framework, and Australian ECHI project (Kelly and Hurst 2006) (Figure 4). While the framework includes effectiveness, safety, responsiveness, accessibility, and cost, the variables accessibility and cost are removed to focus the HCQI project upon the four other domains (which have been given precedent weight) (Kelly and Hurst 2006).

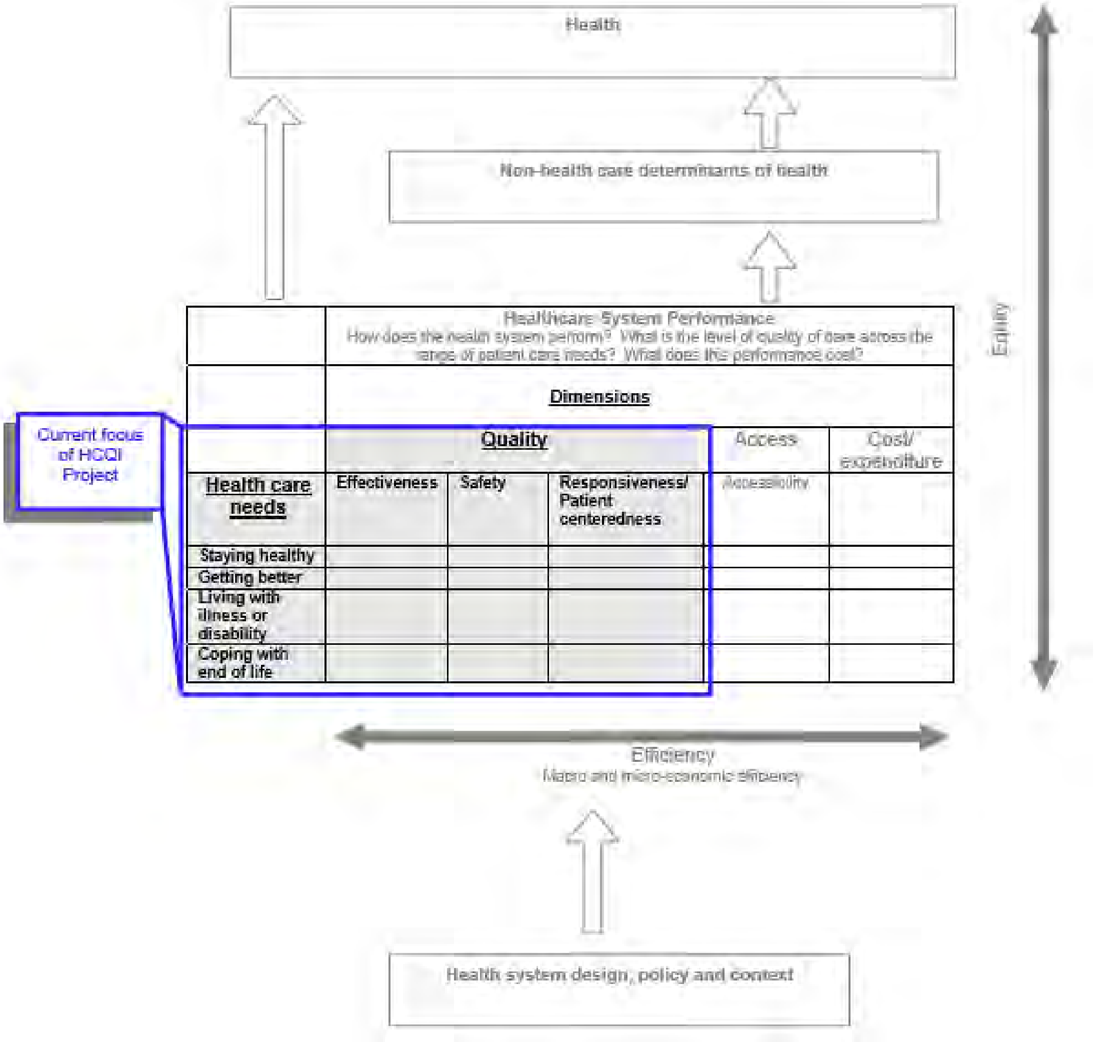


Figure 4: Dimensions of Quality in OECD HCQI Project

Figure 4 is particularly interesting because it aims to contextualize quality domains within other determinants of health, such as policy climates, non-health care determinants, issues of societal equity and societal efficiency (Kelly and Hurst 2006).

Perhaps most important in the OECD model is that the writers firmly state that indicators should be primarily based off of processes and outcomes (Kelly and Hurst 2006). According to Kelly and Hurst (2006), structure indicators “may represent necessary conditions for the delivery of a given quality of health care but they are not sufficient. Their presence does not ensure that appropriate processes are carried out or that satisfactory outcomes are achieved by the health system” (Kelly and Hurst 2006: 16). Conversely, process indicators are argued to represent “the closest approximation of health care offered and are the most clinically specific of the three types of indicators” (Kelly and Hurst 2006: 16). Likewise, outcome indicators represent “measures of health improvement (or deterioration) attributable to medical care” (Kelly and Hurst 2006: 16).

With this ‘system definition,’ the OECD takes a firm stance that processes and outputs are the most essential indicators to include in quality domains. This philosophy is mirrored by the 5 domains included in the HCQI model: effectiveness, safety, responsiveness, accessibility, and cost. Each of these domains relate most strongly to processes and outcomes.

Institute of Medicine Quality Framework:

The Institute of Medicine’s (2001) report, “Crossing the Quality Chasm,” represents another attempt at defining the quality of a healthcare system for targeted improvement. The IoM’s (2001) framework contains six domains, such as safe, effective, efficient, personalized, timely, and equitable. In many ways, these domains largely mirror the frameworks proposed by the OECD and WHO. However, unlike other organizations, the IoM (2001) offered “Ten Rules for Redesign,” which are listed below:

1. *“Care is based on continuous healing relationships.* Patients should receive care whenever they need it and in many forms, not just face-to-face visits. This implies that the health care system must be responsive at all times, and access to care should be

provided over the Internet, by telephone, and by other means in addition to in-person visits.

2. *Care is customized according to patient needs and values*. The system should be designed to meet the most common types of needs, but should have the capability to respond to individual patient choices and preferences.
3. *The patient is the source of control*. Patients should be given the necessary information and opportunity to exercise the degree of control they choose over health care decisions that affect them. The system should be able to accommodate differences in patient preferences and encourage shared decision making.
4. *Knowledge is shared and information flows freely*. Patients should have unfettered access to their own medical information and to clinical knowledge. Clinicians and patients should communicate effectively and share information.
5. *Decision making is evidence-based*. Patients should receive care based on the best available scientific knowledge. Care should not vary illogically from clinician to clinician or from place to place.
6. *Safety is a system property*. Patients should be safe from injury caused by the care system. Reducing risk and ensuring safety require greater attention to systems that help prevent and mitigate errors.
7. *Transparency is necessary*. The system should make available to patients and their families information that enables them to make informed decisions when selecting a health plan, hospital, or clinical practice, or when choosing among alternative treatments. This should include information describing the system's performance on safety, evidence-based practice, and patient satisfaction.
8. *Needs are anticipated*. The system should anticipate patient needs, rather than simply react to events.
9. *Waste is continuously decreased*. The system should not waste resources or patient time.
10. *Cooperation among clinicians is a priority*. Clinicians and institutions should actively collaborate and communicate to ensure an appropriate exchange of information and coordination of care" (IoM Crossing the Quality Chasm 2001: 3-4).

Importantly, these ten rules for redesign are simply suggestions for changes in processes. For example, claiming that "the patient is the source of control," suggests that care models should be adjusted to deliver services in accordance with patient wishes (IoM Crossing the Quality Chasm 2001: 3). Likewise, using "evidenced-based decision making" represents a shift towards standardized processes. Even safety, which is traditionally defined as an outcomes measure, is shifted to a process measure in this rule's list; the IoM claims that systems should be designed to prevent and mitigate errors. In other words, the IoM is relating processes and outcomes in an almost causal fashion. While recognizing that not all outcomes are linked directly to processes, the Ten Rules for Redesign implies that modifications to care methodologies can improve care quality and corresponding outcomes.

National Quality Forum Quality Framework:

The National Quality Forum (NQF) was originally established as an independent third party non-profit to improve healthcare quality by standardizing its measurement (NQF Quality Framework 2002). In its 2002 report, the NQF convened 9 "highly respected" quality improvement content experts with the aim of establishing a new framework for care quality.

Following its completion, the NQF (2002) put forward 6 standard principles of quality: safe, beneficial, timely, patient-centered, efficient, and equitable. This model is particularly similar to the framework posited by the Institute of Medicine.

Conclusions from Various Approaches to Defining Quality:

While the search for a universal quality definition has been around since Dr. Donabedian's landmark 1965 article, variation still persists amongst prominent healthcare organizations such as the OECD, WHO, IoM, NQF, and Commonwealth Fund. Moreover, each have interpreted Donabedian's work differently, placing varying emphasis on inputs, processes, or outputs for one strategic reason or another (Mosadeghrad 2012).

However, before we proceed, it is important to recognize two emerging tensions within these definitions. First, there is a tension between macro- and micro-level definitions of quality. While Donabedian's model may work best for defining the quality of care at an institution, the WHO plainly states that its model exists to define the quality of care system-wide. Although first-principles suggest that quality care at the individual should aggregate to quality care at a societal level, this may not be true in practice. For example, system-wide measurements of quality, such as the WHO and OECD's, emphasize the importance of accessibility and equity. These domains relate to broader principles of social justice and ethics. In other words, the strategic political objectives of more macro-level definitions of quality may not always align with definitions of quality that focus upon healthcare's relationship to the individual.

Moreover, there is tension between these frameworks of quality and the Triple Aim introduced in the opening paragraphs of this document. As Dr. Berwick (2008) explains, the universal objective of any healthcare system is to maintain cost, access, and quality. Under this framework, cost efficiency and accessibility are entirely separate from quality care. Conversely, cost and access are included in the OECD, WHO, Commonwealth Fund, and NQF definitions of quality.

These inconsistencies make the development of a quality framework particularly challenging. Domains emphasized in one model are de-emphasized or deliberately omitted in another for philosophical or strategic policy reasons. In his in-depth interviews, Mosadeghrad (2012) cites nine separate constituent groups that possess alternative values with respect to quality: patients, patients' relatives, providers, managers, policy makers, suppliers, payers, accreditors, and quality managers.

Lastly, measuring quality is further complicated by the interplay between inputs, processes, and outputs. While Donabedian (1965) emphasized each as essential, the WHO seemed to emphasize inputs and processes, while the OECD highlighted processes and outcome measures. Each model presupposes that inputs and processes possess a causal relationship to outcomes. If this were so, then choosing to rate facilities based on either inputs/processes or outputs would yield the same results. Unfortunately, no such research confirms this association between input/process indicators and clinical outcomes. While it seems intuitive that utilizing evidence-based guidelines would improve care quality, it is easy to be skeptical of the validity of these guidelines or of how they are executed in practice.

VI. WHAT PATIENTS WANT TO KNOW

Patient Satisfaction and Patient Experience Studies:

Countless studies have been conducted to assess how patients rate the quality of healthcare institutions following their episode of treatment (Attree 2001; Chang et al. 2006; Nelson et al. 2010; Geun-wan et al. 2015; and Anderson et al. 2007). For example, Anderson et al. (2007) conducted a qualitative analysis of 2917 online patient surveys and structured comment fields to identify patterns in how physicians were rated. Their analysis yielded 24 thematic nodes – i.e. the common criterial foundation for how patients defined a poor or excellent care experience (Table 3) (Anderson et al. 2007).

Table 3: Patterns of Patient Experience Definitions of Quality Care

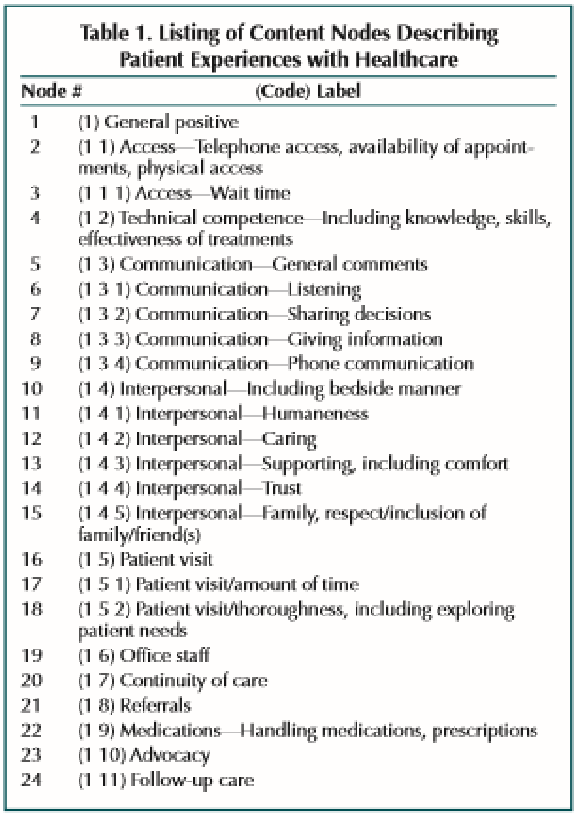


Table 1. Listing of Content Nodes Describing Patient Experiences with Healthcare

Node #	(Code) Label
1	(1) General positive
2	(1 1) Access—Telephone access, availability of appointments, physical access
3	(1 1 1) Access—Wait time
4	(1 2) Technical competence—Including knowledge, skills, effectiveness of treatments
5	(1 3) Communication—General comments
6	(1 3 1) Communication—Listening
7	(1 3 2) Communication—Sharing decisions
8	(1 3 3) Communication—Giving information
9	(1 3 4) Communication—Phone communication
10	(1 4) Interpersonal—Including bedside manner
11	(1 4 1) Interpersonal—Humaneness
12	(1 4 2) Interpersonal—Caring
13	(1 4 3) Interpersonal—Supporting, including comfort
14	(1 4 4) Interpersonal—Trust
15	(1 4 5) Interpersonal—Family, respect/inclusion of family/friend(s)
16	(1 5) Patient visit
17	(1 5 1) Patient visit/amount of time
18	(1 5 2) Patient visit/thoroughness, including exploring patient needs
19	(1 6) Office staff
20	(1 7) Continuity of care
21	(1 8) Referrals
22	(1 9) Medications—Handling medications, prescriptions
23	(1 10) Advocacy
24	(1 11) Follow-up care

Care was reviewed as excellent if patients received immediate appointments, had providers that listened, communicated information effectively, were supportive and understanding, friendly, trustworthy, didn't rush visits, provided regular follow-ups and referrals, and had friendly/effective office staff (Anderson et al. 2007). Conversely, poor care was defined by poor communication, not listening, poor follow-up, low interpersonal skills, rushed visits, and excessive wait times (Anderson et al. 2007).

Other studies mirrored these findings. Attree (2001) conducted 37 qualitative interviews with acute medical patients to uncover consistent themes in consumer-based evaluations of healthcare quality. Importantly, Attree's (2001) findings were "in opposition with the received view that patients place greater emphasis on the technical aspects of care tasks" (Attree 2006: 1365). Instead, quality care was characterized as patient focused, humanistic (demonstrated by caring staff who showed commitment and concern), and individualized (Attree 2006). Poor care was therefore both impersonal and accompanied by uncaring staff (Attree 2006).

In summation, patient ratings of healthcare are often grounded in a consumer's subjective experiences with the healthcare system (Park et al. 2016). A high focus on interpersonal relations (e.g. friendliness, compassion, concern, etc.) seems to exclude evaluations of clinical effectiveness as determined by outcomes (Park et al. 2016).

However, it is important to distinguish the traits patients use to evaluate medical care from what it is they *want to know* about medical care. In other words, when patients rate the experience of their care, they are limited by what they themselves can discern from the interaction. Immediately visible characteristics of their experience include the interpersonal skills of providers and staff as well as appointment wait times and the length of their visit. On the other hand, it is far more difficult for a patient to determine whether the care they are receiving is medically necessitated, in accordance with evidence-based guidelines, or medically effective.

As previously outlined in our discussion of Rothenfluh et al's (2016) study of patient-based provider ratings, consumers struggle with lacking "perceived competency." An underlying power asymmetry permeates the doctor-patient interaction; in result, patients are hesitant to question the technical aspects of a provider's care (Rothenfluh et al. 2016; Heritage 2013). Patient satisfaction studies are thus often limited to more subjective, interpersonal, and non-medical traits.

Unresolved is whether patient-satisfaction themes are correlated with the actual effectiveness of medical services (Verhoef et al. 2014). Simultaneously, another question is raised: would patients define quality care differently if they could be provided with any information of their choosing?

Patient-Experience Data vs. Desired Global Quality Data:

In 2005, Sofaer et al. lead 16 focus groups with participants from four major American cities – Baltimore, Los Angeles, Phoenix, and Orlando – with the aim of defining domains of quality to guide the construction of today's government HCAHPS Surveys. These surveys now serve as the standard assessment of patient satisfaction for the federal government and represent 25% of the risk-potential in the value-based payment program (VanLare and Conway 2012).

Their research uncovered a disconnect between what is measurable through patient experience surveys and what patients define as quality medical care within hospitals. For example, participants stressed the importance of structural features – such as available technologies and staffing ratios – while others highlighted particular outcomes variables – including mortality rates, 30-day hospital readmission rates, rates of hospital infections, etc (Sofaer et al. 2005). However, Sofaer et al. (2005) rightly noted that such variables "are not best derived from a patient experience survey" (2018). When limited to the list of questions that were under consideration for the HCAHPS, respondents were then most concerned with variables such as doctor communication, staff communication, responsiveness, and cleanliness in wards (Sofaer et al. 2005).

In other words, while patient satisfaction is often determined by interpersonal/non-medical factors, patients nonetheless are deeply concerned with the effectiveness of the medical care they receive. Unfortunately, little research has been conducted regarding the relative weights patients ascribe to different domains of quality. Therefore, it is unclear what patients value most within the context of Donabedian’s (1965) model of Structures, Processes, and Outcomes. Further research should aim to uncover patient-definitions of quality: what information they want before choosing healthcare facilities. These patient-level definitions of quality should be compared to broader definitions of quality (WHO, OECD, IoM, etc.) to arrive at a pluralistic theoretical framework that encapsulates both individual and policy objectives within the healthcare industry (Mosadeghrad 2012).

VII. MATHEMATICAL UDERPINNINGS OF COMPOSITE RANKINGS

Thus far, we have established that quality transparency could be utilized as a mechanism to enable patient choice, ameliorate information-based market failures, and incentivize improvement within the American healthcare industry. Moreover, we have discussed various system-level frameworks for quality and have identified a need to strengthen our understanding of patient-desired quality data. However, once an agreed upon theoretical framework is constructed, a systematic approach must be utilized to evaluate a healthcare facility’s performance according to that new definition. Moreover, for patient choice to be enabled, data on facility quality must be disseminated in a way that patients can decode and transform into actionable decision-making.

Health Literacy:

Medical information is inherently complex (Atree 2006). Approximately ninety-million Americans struggle to understand and act upon health information (Nielsen-Bohlman et al. 2004). In practical terms, this means that half of the American population is unable to interpret a research consent form, a privacy notice, the content of their own health record, the details of their current treatment, or data relating to a healthcare provider or institution (Nielsen-Bohlman 2004).

The issue of literacy is particularly pertinent when placed in the context of the doctor-patient interaction. In Emanuel's (1992) discussion of the physician-patient relationship, he outlines four models that operate along a spectrum of power: (1) the paternalistic, (2) the informative, (3) the interpretive, and (4) the deliberative models.

In model one, the provider has full autonomy in medical decision-making. He or she decides the intervention and delivers corresponding treatment unless the patient overtly objects (Emanuel 1992). In the second model, the clinician's job is to inform the patient of all relevant treatment options and to enable the consumer's self-determination (Emanuel 1992). The interpretive model, however, aims to elucidate the patient's underlying values and find the treatment option that best aligns with those philosophical underpinnings (Emanuel 1992). And lastly, the deliberative model exists as one of balance. The provider lists out a patient's options, makes suggestions, and discusses the patient's values and objectives from care (Emanuel 1992).

Undoubtedly, there are certain instances where one model may fit better than others (Emanuel 1992). For routine, clear-cut procedures, a paternalistic approach may be ideal, while in medically and philosophically ambiguous circumstances, an interpretive, deliberative, or informative model fit best. What's important to understand is that every model other than paternalism necessitates a foundational level of health literacy. For consumers to exercise a degree of autonomy with regards to their medical decision making, they must first have an understanding of their condition, their options (benefits and harms), and their personal values.

Strong health literacy is therefore a *tool* which enables patient choice and empowerment (Koh et al. 2012; Bastian 2008). Data transparency on healthcare facility quality is therefore fundamentally an intervention designed to improve upon the health literacy of patient consumers. However, for the benefits of transparency to be realized, information must be disseminated in a format that the general public can internalize. In other words, it must be disseminated in language congruent with the *current* level of health literacy for the American population.

This is no easy task. As Koh and Berwick (2012) rightly note, lacking health literacy is the largest barrier to making the public a major player in improving the health care for all Americans (Bastian 2008). Traditional public health models of information dissemination often falter because the message itself is not received or understood.

The Rise of Composite Indicators:

In recent decades, composite indicators have proliferated as tools designed to condense complex information into easily digestible, immediately recognizable forms. The Organization for Economic Cooperation and Development (OECD) was one of the leaders in popularizing composite indices: in its case, composite indices were utilized to compare partnered nations across a variety of different industries for the purpose of

> "helping governments respond to new developments and concerns, such as corporate governance, the information economy, and the challenges of an ageing population. The Organization provides a setting where governments can compare policy experiences, seek

> answers to common problems, identify good practice and work to coordinate domestic and international policies" (Hoffman et al. 2008: 4).

With its extensive expertise, the OECD (2008) – through a joint partnership with the Econometrics and Applied Statistics Unit for the Joint Research Center of the European Commission – published a comprehensive two-hundred page handbook on the construction of composite indicators intended for use by policy-makers, academics, and the media (Hoffman et al. 2008).

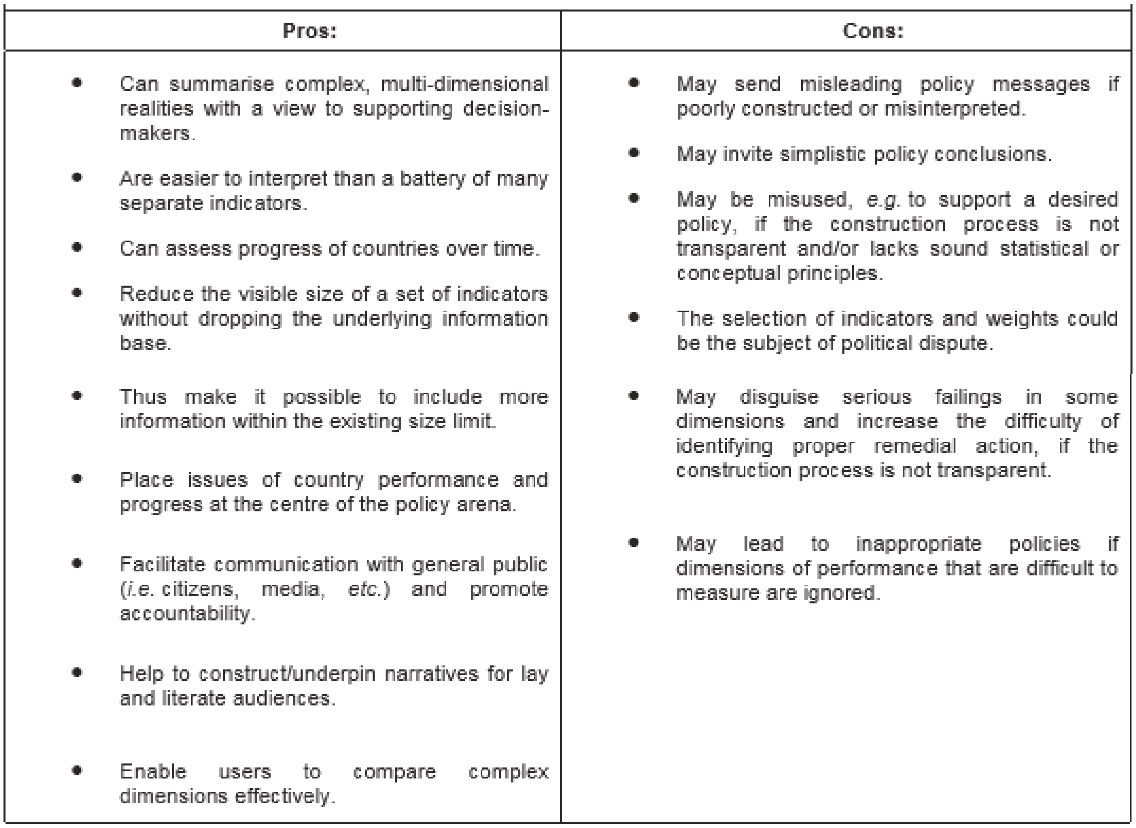
Benefits and Cons of Composite Indicators:

Composite indicators are frequently used because they "often seem easier for the general public to interpret than to identify common trends across many separate indicators" (Saltelli 2007). In other words, aggregate numerical indicators are easily disseminated and carry with them 'big picture' conclusions regarding multidimensional concepts that cannot be assessed through a single variable (e.g. the environment, economy, technological development, or even healthcare) (Hoffman 2008).

However, while composite indicators can summarize large quantities of complex information, they are only as useful as the underlying quality of their construction (Hoffman et al. 2008) (Table 4). At their best, composite indicators clarify complex problems, but at their worst, scores can send misleading policy messages, promote overly simplistic conclusions, use indicators and weights that lack sound justification, or may disguise serious failings in one or more dimensions through the process of data aggregation (Hoffman et al. 2008) (Table 4).

Thus, while composite indicators are defined as mathematical models, successful construction owes "more to the craftsmanship of the modeler than to universally accepted scientific rules for coding" (Hoffman et al. 2008: 16). This implies that the development of a

Table 4: Strengths and Weaknesses of Composite Indicators



Pros:	Cons:
• Can summarise complex, multi-dimensional realities with a view to supporting decision-makers. • Are easier to interpret than a battery of many separate indicators. • Can assess progress of countries over time. • Reduce the visible size of a set of indicators without dropping the underlying information base. • Thus make it possible to include more information within the existing size limit. • Place issues of country performance and progress at the centre of the policy arena. • Facilitate communication with general public (*i.e.* citizens, media, *etc.*) and promote accountability. • Help to construct/underpin narratives for lay and literate audiences. • Enable users to compare complex dimensions effectively.	• May send misleading policy messages if poorly constructed or misinterpreted. • May invite simplistic policy conclusions. • May be misused, *e.g.* to support a desired policy, if the construction process is not transparent and/or lacks sound statistical or conceptual principles. • The selection of indicators and weights could be the subject of political dispute. • May disguise serious failings in some dimensions and increase the difficulty of identifying proper remedial action, if the construction process is not transparent. • May lead to inappropriate policies if dimensions of performance that are difficult to measure are ignored.

useful composite indicator is not a straightforward, formulaic process. This ambiguity has made composite indicators a highly disputed technique in statistical and political literature. Nonetheless, their use has grown drastically in recent years; Bandura (2006) estimates that more than 160 indicators are currently used for information dissemination and policy purposes.

Steps to Composite Indicator Construction:

In their technical users guide, the OECD (2008) outlines Ten Steps to composite indicator development:

1. **Theoretical Framework:** A framework should provide the basis for the selection and combination of indicators into a meaningful composite based upon a ‘fitness-for-purpose’ principle (Hoffman et al., 2008).
2. **Data Selection:** Indicators should be selected based upon their “analytical soundness, measurability, coverage, relevance to the phenomenon being measured, and relationship to each other” (Hoffman et al., 2008: 17).

3. **Imputation of Missing Data:** Different approaches for imputing missing values exist. Developers should select an approach that best fulfills the computational needs of the indicator (Hoffman et al., 2008).
4. **Multivariate analysis.** Exploratory analysis should be used to investigate “the overall structure of the indicators, assess the suitability of the data set and explain methodological choices, e.g. weighting, aggregation” (Hoffman et al., 2008: 17).
5. **Normalization:** Indicators must be normalized so that they can be compared. Extreme values and skewed data must be accounted for (Hoffman et al., 2008).
6. **Weighting and Aggregation:** The chosen method of weighting and aggregation should be based upon the underlying theoretical framework. Correlation (and duplication) issues between individual indicators must be identified and corrected during the aggregation process (Hoffman et al., 2008).
7. **Robustness and Sensitivity:** Analysis should assess the robustness of the indicator by including/excluding indicators, modifying the normalization scheme, changing imputation methodologies, and altering weighting/aggregation methods (Hoffman et al., 2008).
8. **Back to the Real Data:** Indicators should be transparent to potential users and replicable – i.e. can be decomposed into original indicators or values (Hoffman et al., 2008).
9. **Links to other Variables:** The composite indicator should be compared to other published indicators for the chosen phenomenon to assess correlation and other linkages (Hoffman et al., 2008).
10. **Presentation and Visualization:** Indicators can be presented in many ways. Disseminating the results of the composite indicator in an interpretable fashion is essential (Hoffman et al., 2008).

The importance of each of these ten steps is summarized in Appendix 11 (Hoffman et al. 2008). Let us now delve into a discussion of each step in the process of constructing appropriate composite indicators.

Theoretical Framework: In Hoffman et al’s (2008) words, “What is badly defined is likely to be badly measured” (22). Thus, before a composite indicator is constructed, the developer should first have a sound definition of the phenomenon he or she is trying to measure. In the case of healthcare, CheckUp aims to assess the phenomenon of healthcare quality. Yet, what is quality? Numerous definitions, from Donabedian’s (1965) model to the WHO’s (2002) framework are often cited and have strategic provisions for influencing future policy. Conversely, no framework of quality exists from the perspective of patient consumers. Thus, the first step to developing a composite indicator of healthcare quality is establishing a justifiable framework to define the phenomenon. As Hoffman et al. (2008) explain, both the overall concept and determined sub-groups must be defined, justified, and clearly outlined to the reader.

Selection Criteria: The underlying theoretical framework will in turn affect the individual variables selected for analysis. Because a composite indicator is merely the aggregation of individual variables, selection should be based upon relevance, analytical soundness, and availability (Hoffman et al. 2008).

Often, the process of data selection is rather subjective and limited by the availability of relevant data relating to the particular phenomenon. The developer is inevitably constrained by what data are available and must work within these parameters (or determine that not enough sound data exist to make a reliable composite – thereby abandoning the project). Composites should be constructed with the intent to evolve as new data become available. This is particularly pertinent in the case of healthcare as new data are often made available through updated government or insurer mandates.

Before moving on to data analysis, the developer should create a table summarizing the characteristics of chosen variables. For example, source, type (hard, soft, input, process, output, etc.), availability, and scale (Hoffman et al. 2008). The issue of scale is particularly important. Data can be categorical (i.e. nominal scaling), where observations are grouped into qualitative classes such as marital status or gender (Hoffman et al. 2008). Data can also be quantitative, such as in ordinal scales where options are ranked, interval scales (e.g. temperature), or ratio scales (e.g. age, height, rates of disease, etc.) (Hoffman et al. 2008). Each variable's 'type' must be identified for it will affect how it is treated in later analysis.

Imputation: As just discussed, composites are limited by the available data. However, few data sets are likely to be entirely complete. As such, developers must choose how to handle missing data for individual variables before beginning the aggregation process.

As Hoffman et al. (2008) explain, data can be missing in either a random or non-random fashion. Data that are missing completely at random (MCAR) do not depend on the variable being analyzed or any other variables in the dataset. However, data missing at random (MAR) do not depend on the selected variable but may be related to the status of other variables in the data set. For example, the availability of income data would be "MAR if the probability of missing data on income depends on marital status but, within each category of marital status, the probability of missing income is unrelated to the value of income" (Hoffman et al. 24). Lastly, data are not missing at random (NMAR) if the probability of missing a value depends on the values themselves. For example, if low income households were less likely to report their income (Hoffman et al. 2008). In reality, while all variables are MCAR, MAR, or NMAR, it is often impossible to determine which form is at play (Hoffman et al. 2008).

When dealing with missing data, developers must choose between three options: (1) case deletion, (2) single imputation, or (3) multiple imputation. In the first approach, missing data are simply left omitted and aggregation weights are adjusted proportionally for each case. However, this assumes that data are MCAR and often produces biased estimates (Hoffman et al. 2008). Thus, methods of imputation are more frequently employed.

In single imputation, all missing data are substituted with a single value for all cases in the data set. This value can be a statistical point of center (i.e. mean/median/mode) or from regression imputation, hot-and-cold-deck imputation, or expectation-maximization imputation (Hoffman et al. 2008). In hot-deck imputation, blank values are replaced with available data of a similar type. For example, an individual's income could be replaced with the income of someone else with similar demographic characteristics (Hoffman et al. 2008).

In regression imputation, however, a more statistical approach is taken. Because individual indicators in a composite are often correlated, missing values can be estimated based upon the status of other highly correlated variables from the dataset (Hoffman et al. 2008). It is important to note that different indicators have varying degrees of intercorrelation. Therefore, the accuracy of a regression imputation method can be estimated by the R^2 value, residual mean square, the value of Mallows' C_k, and stepwise regression (Hoffman et al. 2008).

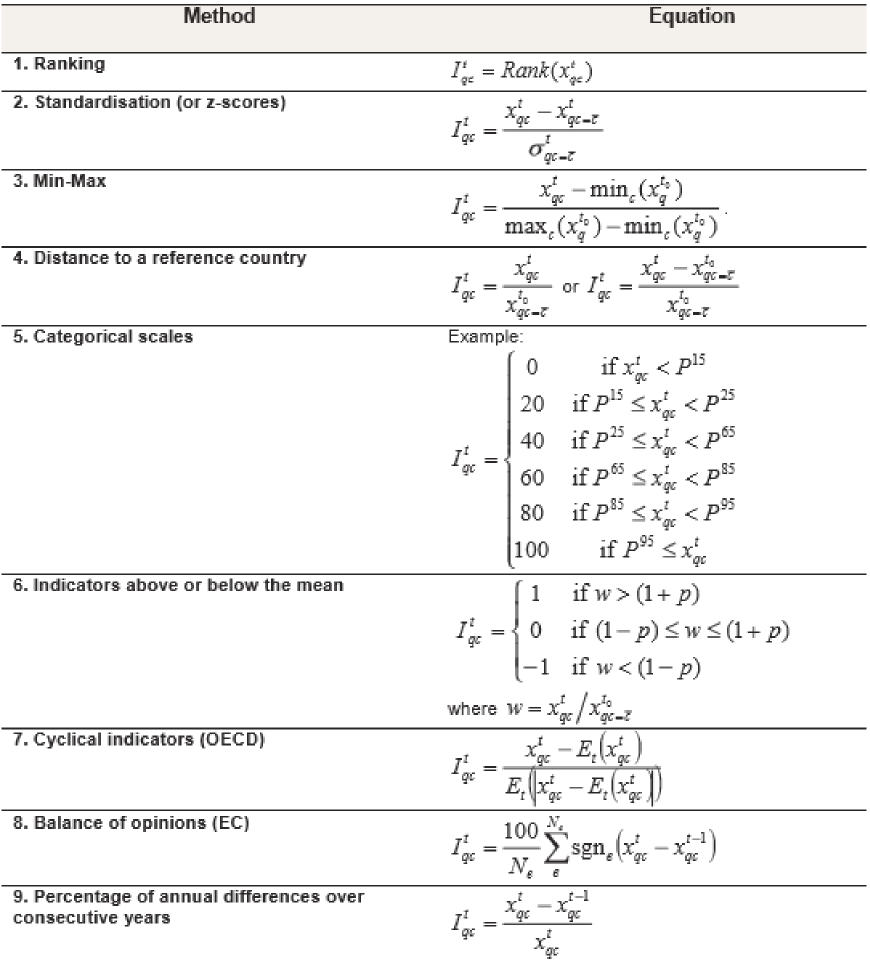
Likewise, in multiple imputation approaches, each missing case is given a unique value based upon a probabilistic estimate of the actual value (e.g. Markov Chain). Each of these imputation methodologies carries with them certain assumptions regarding the data.

Multivariate Analysis: Hoffman et al. (2008) define this portion of development as an "art" (26). Individual indicators should be analyzed to uncover their relationships to each other. This is a particularly useful preliminary step to aggregation because it can uncover strong intercorrelations between variables that would result in 'double counting' if left unchecked.

One method to analyze multiple variables simultaneously is principle components analysis (PCA), which is designed to reveal how different variables change in relation to each other. During a PCA, correlated variables are transformed into a *new* set of uncorrelated variables of a reduced number by using a correlation matrix (Hoffman et al. 2008). Likewise, Cronbach coefficient alpha (c-alpha) can also be used to estimate internal consistency within a model.

Normalization of Data: Before data can be aggregated, they must first be placed on an equivalent scale. Different measurements exist in different units and scales, and must all be converted to a common denominator before comparisons can be drawn. Different methods of normalization include ranking (all variables are converted to relative numerical positions between cases), standardization/z-scoring (coverts indicators to a scale with a mean of zero and a standard deviation of one), min-max normalization (limits all indicators to the same range [0, 1]), distance to reference (converts all indicators to positive or negative distances from an agreed upon benchmark or standard), and categorical scales (e.g. star ratings or points systems) (Hoffman et al., 2008). Table 5 outlines the mathematical formulae for each normalization approach.

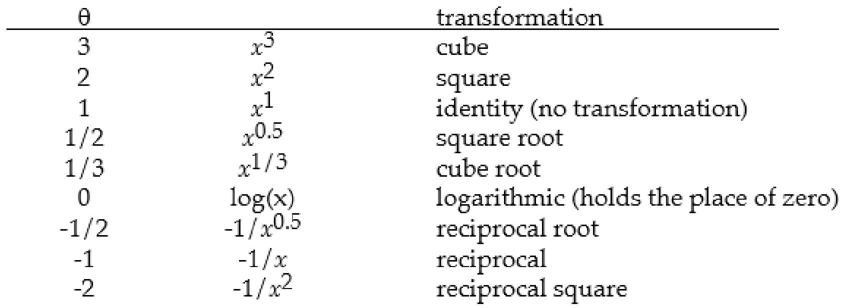
Table 5: Different methods of data normalization



Method	Equation
1. Ranking	$I_{qc}^{t} = Rank(x_{qc}^{t})$
2. Standardisation (or z-scores)	$I_{qc}^{t} = \frac{x_{qc}^{t} - x_{qc=\overline{c}}^{t}}{\sigma_{qc=\overline{c}}^{t}}$
3. Min-Max	$I_{qc}^{t} = \frac{x_{qc}^{t} - \min_c(x_q^{t_0})}{\max_c(x_q^{t_0}) - \min_c(x_q^{t_0})}$.
4. Distance to a reference country	$I_{qc}^{t} = \frac{x_{qc}^{t}}{x_{qc=\overline{c}}^{t_0}}$ or $I_{qc}^{t} = \frac{x_{qc}^{t} - x_{qc=\overline{c}}^{t_0}}{x_{qc=\overline{c}}^{t_0}}$
5. Categorical scales	Example: $I_{qc}^{t} = \begin{cases} 0 & \text{if } x_{qc}^{t} < P^{15} \\ 20 & \text{if } P^{15} \le x_{qc}^{t} < P^{25} \\ 40 & \text{if } P^{25} \le x_{qc}^{t} < P^{65} \\ 60 & \text{if } P^{65} \le x_{qc}^{t} < P^{85} \\ 80 & \text{if } P^{85} \le x_{qc}^{t} < P^{95} \\ 100 & \text{if } P^{95} \le x_{qc}^{t} \end{cases}$
6. Indicators above or below the mean	$I_{qc}^{t} = \begin{cases} 1 & \text{if } w > (1+p) \\ 0 & \text{if } (1-p) \le w \le (1+p) \\ -1 & \text{if } w < (1-p) \end{cases}$ where $w = x_{qc}^{t} / x_{qc=\overline{c}}^{t_0}$
7. Cyclical indicators (OECD)	$I_{qc}^{t} = \frac{x_{qc}^{t} - E_t(x_{qc}^{t})}{E_t(\lvert x_{qc}^{t} - E_t(x_{qc}^{t}) \rvert)}$
8. Balance of opinions (EC)	$I_{qc}^{t} = \frac{100}{N_e} \sum_{e}^{N_e} \operatorname{sgn}_e(x_{qc}^{t} - x_{qc}^{t-1})$
9. Percentage of annual differences over consecutive years	$I_{qc}^{t} = \frac{x_{qc}^{t} - x_{qc}^{t-1}}{x_{qc}^{t}}$

Moreover, before data can be aggregated, datasets for skewed indicators must be transformed to render comparable distributions for later comparison. According to the commonly cited "Ladder of Powers," data can be transformed to achieve semi-normal distributions by applying a particular exponent to each case within the indicator dataset (Table 6).

Table 6: Ladder of Powers for Transforming Data to Near-Normal Distributions



θ		transformation
3	x^3	cube
2	x^2	square
1	x^1	identity (no transformation)
1/2	$x^{0.5}$	square root
1/3	$x^{1/3}$	cube root
0	log(x)	logarithmic (holds the place of zero)
-1/2	$-1/x^{0.5}$	reciprocal root
-1	$-1/x$	reciprocal
-2	$-1/x^2$	reciprocal square

Weighting and Aggregation: Eventually, individually selected variables must be aggregated into the final composite score. However, special consideration should be placed on choosing a weighting methodology that reflects the desired purpose of the composite indicator.

Some weighting methods are statistical in nature (Hoffman et al. 2008). For example, principle component analysis, data envelopment analysis, and the unobserved components model (UCM) each estimate the relative impact of each variable on the overall phenomenon being tested (Hoffman et al. 2008). These methods are particularly useful for accounting for instances of high correlation (and thus, double-counting). For example,

> "in the composite indicator of e-Business Readiness, the indicator I_1 'Percentage of firms using Internet' and indicator I_2 'Percentage of enterprises that have a website' display a correlation of 0.88 in 2003: given the high correlation, is it permissible to give less weight to the pair (I_1, I_2) or should the two indicators be considered to measure different aspects of innovation and adoption of communication technologies and therefore bear equal weight in the construction of the composite?" (Hoffman et al. 2008: 32).

Statistical methods such as principle component analysis are designed to transform original variables into a new set of reduced-number uncorrelated variables, thereby eliminating the effects intercorrelations between the indicators that make up the composite (Hoffman et al. 2008).

However, other weighting methods are participatory. For example, the Budget Allocation Process (BAP) asks subjects to assign a relative number of ‘dollars’ to each variable in the data set while the Analytic Hierarchy Process (AHP) and Conjoint Analysis (CA) have subjects rank the ‘worth’ of different variables used in the composite (Hoffman et al. 2008). Moreover, if a participatory weighting methodology is to be used, special consideration should be made regarding subject choice. Subjects can be either experts in the field, lay persons, or strategic constituents, and selection is reflective of the underlying objectives of composite developers (Hoffman et al. 2008).

Weights can also be selected based upon the quality or availability of the data for certain indicators. Under this approach, higher weights are assigned to variables with broader coverage. However, this convenience-approach is often criticized because it implies that available data are more important than scarce data, which is not often the case (Hoffman et al. 2008).

Also, another option is to apply an equal weighting approach – i.e. to assume that all variables are of the same importance (Hoffman et al. 2008). Hoffman et al. (2008) note that this is by far the most common approach, but that alternative methodologies allow for more purposeful composites.

Lastly, developers must choose between both linear and geometric aggregation methodologies. While linear aggregation is based upon the relative weights of different variables, geometric approaches are designed to reduce compensation between variables – i.e. when poor performance in one dimension is compensated by high performance in another. Geometric aggregations increase the relative weight of low values. In one respect, this is beneficial; it incentivizes organizations to improve on particularly low-performance dimensions. However, simultaneously, geometric aggregation could be said to penalize organizations with more erratic displays and fail to accurately characterize their performance as a whole.

Robustness and Sensitivity Analysis: Uncertainty and sensitivity analysis should then be performed to determine how changes in the data or development methodology affect the outcome of eventual composite scores (Hoffman et al. 2008). This can be done by (1) including or excluding individual indicators, using alternative imputation schemes, alternative normalization methods, weighting approaches, and aggregation choices. Different combinations of approaches should be executed and compared to determine the degree to which data/methodological variation modifies the final composite indicator (Hoffman et al. 2008).

Data Presentation: Finally, after a composite indicator is constructed, it is only of value if the information contained therein can be disseminated efficiently and clearly. Hoffman et al. (2008) cite multiple formats for displaying the final indicator scores, such as tabular data, bar charts, and line charts.

VIII. UNANSWERED QUESTIONS

The OECD's (2008) handbook on the construction of composite indicators reveals the vast number of choices a developer must make before arriving at a useful final model. However, beyond that, we are left with a number of unresolved questions with regards to how to develop and disseminate a useful healthcare quality composite.

First, a justified theoretical framework should be constructed to define the phenomenon of healthcare quality and its respective domains. This framework will guide the selection of individual indicators and the eventual aggregation methodology.

Second, a methodological decision must be made regarding the weighting of individual indicators. While statistical approaches allow for a more nuanced correction of intercorrelations between individual indicators, participatory models may construct a model that best-reflects the values of the individuals this website is intended for.

Third, issues of data presentation should be addressed. Unlike the OECD, which uses composites to characterize phenomena for a small subset of partnered countries, CheckUp will eventually rank tens of thousands of healthcare facilities. Traditional tabular, bar graph, or line graph methods of data presentation are not possible alone. Alternative methods of data narrowing through filtered search functions will be essential to ensuring website usability.

IX. FURTHER QUESTIONS AND CONCLUSIONS

A gap exists within the current healthcare information marketplace. Existing online patient resources are limited in scope and accuracy. However, while it is easy to criticize flaws in current models, it is far more challenging to identify the optimal content and format for future web-based healthcare information.

Future research must be conducted on patient definitions and weightings of healthcare quality domains in order to bring insights into the *types* of information that should be included in future consumer choice health websites. Moreover, research should be conducted on how to effectively disseminate this information to consumers with minimal health literacy. Lastly, user interfaces considerations are crucial to ensuring that consumer websites can be navigated by potential consumers searching for key information. These three questions will be the primary focus of this research.

CHAPTER TWO: METHODOLOGY

I. RESEARCH PURPOSE AND QUESTIONS

The aims of this research were to study patient-definitions of healthcare quality in comparison to prevailing theoretical models, determine how much/what kinds of granular data patients want available when assessing quality, and lastly, to gauge consumer preferences regarding the display of information in an online website. The ultimate goal of this work was to establish a set of guidelines for how healthcare content should be disseminated to the public in an online format. To study these three interrelated components, subjects were grouped into three corresponding types of focus groups, each of which focused on addressing one of the following three research questions:

1. Do consumers share a unified conceptual definition of healthcare quality?
2. What types of granular information do patients want access to when deciding between a list of healthcare facility options?
3. In what format do consumers want to interact with these data on a web-based platform?

The study and recruitment of participants was approved by the Union College Institutional Review Board.

II. SUBJECT RECRUITMENT

Building on previous research, focus groups were designed to explore the personal experiences and preferences of stakeholder groups in order to understand how a consumer-centric healthcare rating website would affect their interactions with the healthcare system. In order to understand the informational preferences of active healthcare consumers, subject recruitment had to be limited to individuals of independent healthcare decision-making and purchasing age; i.e. of the age where they choose where they or their dependents receive healthcare services. As such, the minimum age for participation was set at 26 years old to account for the Affordable Care Act's (2010) clause, which states that dependents can remain on their parents' insurance until the age of 26.

Sampling was achieved using both convenience and snowball techniques. Personal contacts of the researcher were contacted via email notifications. Once an initial list of subjects was secured, snowballing techniques were employed via email-forwarding to contact other potential participants based on the recommendations of these personal contacts. Google Calendars was used to schedule willing parties into designated focus group time slots based upon the scheduling availability of the participants.

In total, 23 subjects were recruited for participation in this research study, with 7 participating in Focus Group Type I, 6 in Focus Group Type II, and 10 in Focus Group Type III. Tables 7 and 8 illustrate the demographic characteristics of the subject participants.

Table 7: Demographic Characteristics of Study Participants

Descriptive	Birth	Children < 5	Children 5-13	Children 14-18	Children >19	Total Children
μ	44.1	0.1	0.3	0.1	1.0	1.5
σ	12.0	0.4	0.5	0.5	1.1	1.1
Minimum	26	0	0	0	0	0
Maximum	58	2	2	2	3	4

Participants aged from 26 to 58, with an average age of 44.1 years. The majority of the participants were female (52.2%). Likewise, the most common level of education was the completion of a four-year bachelor's degree (73.9%). The majority of subjects were employed; however, 26.0% were unemployed or full-time homemakers.

Table 8: Nominal Demographic Characteristics of Study Participants

Demographic Variable	Category	Frequency	Percent
Race	Caucasian	18	78.3
	Black	1	4.3
	Hispanic/Latino	4	17.4
Marital Status	Married	16	69.6
	Never Married	5	21.7
	Widowed	1	4.3
	Separated	1	4.3
Level of Education	Graduate Degree	3	13.0
	Completed College	17	73.9
	Some College	1	4.3
	High School	1	4.3
	Some High School	1	4.3
Employment	Employed	15	65.2
	Unemployed	1	4.3
	Full Time Homemaker	5	21.7
	Retired	1	4.3
	Student	1	4.3
Sex	Male	11	47.8
	Female	12	52.2

III. SUBJECT PROTECTIONS

Before obtaining consent, all subjects were informed that their identities would be kept confidential. During data collection, subjects were recorded during focus groups if prior permission was granted from the entire group. However, all identifying characteristics were omitted from written reports. Pseudonyms were used to protect individual identities. At no time were subject names recorded on tape. Instead, subjects were given number identifiers. A list matching actual names to numbers was kept at a separate location on a piece of paper.

Focus group discussions were transcribed using the original recordings and the corresponding tapes were then stored in a locked location. Once the research concluded, the tapes were destroyed. Transcribed focus group conversations were stored on USB drives and stored in the locked cabinet along with the corresponding tapes. No data were stored on the hard drive of my personal computer. The only other person to have access to these transcriptions was my thesis advisor, Professor Melinda Goldner.

Both verbally and in the informed consent document, it was made clear to subjects that participation was completely voluntary. Subjects were told that they could leave the focus group at any time and refuse to answer any questions. Attempts at randomization in focus group assignment were made using a random number generator; however, in some instances randomization was impossible due to the scheduling restraints of some participants recruited.

IV. PROCEDURES

Upon arrival for focus group participation, all subjects were first asked to sign an informed consent document and fill out a demographic survey (Appendix 11). The research design involved gathering qualitative in-person focus group data, using semi-structured and probative open-ended questions. Focus groups took place in a private setting, such as a conference room in a privately owned apartment building.

Moreover, the research required that subjects interact with prototype models of consumer-choice healthcare rating websites. To construct these various models, Axure RP prototyping software was used to create dynamic, web-based content that subjects could interact with during focus group discussions.

Focus Group One Procedure:

Focus Group Type I, which centered on exploring patient-centric definitions of healthcare quality, consisted of four task and discussion components. In Task One, subjects were first asked to individually brainstorm factors they believed described a quality healthcare facility (Appendix 12). After the expiration of the 5-minute period, answers were shared (if willing), and a series of probing questions were asked to justify the responses.

In Task Two, subjects were exposed to a list of quality domains documented in the literature (with corresponding definitions) (Appendix 13). Participants were tasked with identifying any domains of interest, and probing questions were used to spark discussion about the merit of domains included.

Task Three was an opportunity for synthesis; subjects were asked to individually choose the 5 domains they believed were the most important for defining healthcare quality (Appendix 14). Then, the Budget Allocation technique was utilized to force subjects to assign relative weights of importance to each of the chosen domains.

Lastly, Task Four explored the degree of rigidity of subject domain choices by positing two distinct healthcare scenarios; (1) seeking out services for a hip replacement, and (2) seeking out primary care services for the flu (Appendix 15). Subjects were given the choice to retain or modify the domains and their relative weights, thereby determining whether patients view quality domains as universal or scenario-specific.

Focus Group Two Procedure:

While Focus Group One explored consumer-based definitions of healthcare quality, Focus Group Two instead studied the desired degree of granularity of the health data presented. To re-cap, domains are really umbrella terms designed to describe an aspect of healthcare quality that should apply to multiple types of healthcare facilities. Within each domain, a variety of indicators can be used and aggregated to create a de facto measurement score for that domain. For example, let's assume we are discussing the WHO's "safety" domain for nursing home services. While the domain may be safety, two of the *indicators* included in the domain may be (1) rates of resident falls and (2) the percentage of patients that develop pressure ulcers (i.e. bed soars). By aggregating measurements from a variety of indicators, domain 'scores' were established.

In Focus Group Two, the aim was to study how much information should be provided on the individual indicators that make up the domain- and composite-level scores for different healthcare facilities. Here, there appeared to be three options: (1) to provide a list of the indicators included in each domain for each facility type (without providing any facility-level data on individual indicators), (2) to provide the raw scores for each indicator for each facility, and (3) to provide the raw and standardized score for each indicator for each facility.

In the single Task for Focus Group Type Two, subjects were shown three versions of a prototype consumer-based health care quality rating website, each of which varied by the level of granular information provided at the indicator level for the facilities rated (Appendix 16). A series of probing questions were asked on subjects such as the potentially misleading nature of raw scores, calls for information transparency, and risks of data overload.

Focus Group Three Procedure:

Focus Group Three aimed to study patient-preferences regarding the design of the user interface for healthcare quality rating websites. To explore the different aspects of a web-based user interface, the process was divided into five separate tasks, which explored the following axes of design change: (1) color schemes, (2) facility rating scales, (3) facility search methods, (4) the visual presentation of facility distances, and (5) options for user-directed domain weighting. Discussions regarding each of these user interface components were initiated by walking subjects through a variety of Axure website prototype models that differed according to these design components (Appendix 17). Probing questions were asked to spark debates over the merits of different user interface approaches.

V. DATA ANALYSIS

The analytical approach for each of these focus groups was largely qualitative in nature. For each of the focus groups, discussions were transcribed and first separated by task. Then – within each task – focus group transcriptions were coded into sets and subsets of themes. These themes were then discussed using supporting quotes. The coding, analysis, and development of themes from focus group data identified a broad spectrum of perceptions and preferences regarding each of the three explored research areas.

Importantly, while most of the analysis was qualitative, Focus Group One's study design allowed for some quantitative analysis by using the written responses from Tasks 1-4. For example, Task Two asked subjects to (1) circle, (2) cross-out, (3) leave blank, or place a (4) "?" adjacent to any domains that were (1) important, (2) not important, (3) neutral, or (4) confusing/unclear as potential healthcare quality domains. Because all subjects filled out this written task, subject responses for each variable were codified as nominal. Because the data were nominal, statistical tests were limited to descriptive statistics such as frequency and percentage distributions, but nonetheless offered insights into the varying opinions of subjects.

CHAPTER THREE: RESULTS AND DISCUSSION

By combining the findings from each of the focus groups, I was able to construct a set of guidelines for the construction of a consumer-centric healthcare website, which I then used to rank and present data on all of America's CMS registered nursing homes.

Some of the findings that guided the eventual ranking of nursing homes were controversial. For example, the diversity of opinions in Focus Group One made it impossible to render a singular unified definition of healthcare quality for all healthcare scenarios. However, nonetheless, consistent patterns and patient-preferences did emerge in both written reports and verbal discussions. These patterns – for example, the importance of Effectiveness and Safety as quality domains – were then integrated into a set of guidelines that were followed during the construction of the theoretical framework for nursing home quality. The results from Focus Group Two were similar in that the opinions of participants seemed to occasionally contradict other findings, making a straightforward conclusion difficult to grasp. For example, all participants favored models of indicator presentation that included all available granular data; but this contradicted the reality that the majority of subjects struggled to interpret this information when it was made available. Balancing the desire for more information with the challenge of making that information comprehendible is an issue discussed in the conclusion of this paper. Unlike the first two groups, Focus Group Three provided easily interpretable data that were used to formulate a list of guidelines regarding the visual presentation of healthcare quality data in an online format. These guidelines were referenced when wire framing the website design for nursing home data.

VI. FOCUS GROUP ONE: PATIENT-CENTRIC DEFINITIONS OF HEALTHCARE QUALITY

Demographics for Focus Group One:

Because each of these focus groups were used to analyze three separate research aims, subjects were sourced independently for all three and will therefore be analyzed distinctly. Table 9 illustrates the demographic characteristics of subjects included in Focus Group Type 1; of the 7 participants, 5 were female (71%), 100% were Caucasian, and 100% were currently married. On average, the participants had 2.3 children ($\sigma = 0.76$). While no participants had any children under 5, three had children between 5-13, one had a child between 14-18 years, and 5 of 7 had children older than 19 years (Table 9). All of the participants completed college, with two completing graduate degrees. Lastly, 5 were currently employed while two were full time homemakers (Table 10).

Table 9: Demographic Characteristics for Focus Group One

Descriptive	*Birth*	*Children < 5*	*Children 5-13*	*Children 14-18*	*Children >19*	*Total Children*
μ	51.3	0	0.6	0.1	1.6	2.3
σ	3.7	0	0.8	0.4	1.1	0.8
Minimum	44	0	0	0	0	2
Maximum	55	0	2	1	3	4

In total, this represented a relatively homogeneous group of participants. This presents both a benefit and a limitation. In one respect, the prevailing literature on focus group research suggests that a homogeneous group minimizes power asymmetries and is ideal for increasing the likelihood of equal participation (Morgan 1996). However, simultaneously, collecting data on such a homogeneous group limits the variety of potential experiences and opinions of participants in comparison to a more diverse subject pool.

Table 10: Nominal Demographic Characteristics for Focus Group One

Demographic Variable	*Category*	*Frequency*	*Percent*
Race	Caucasian	7	100
Marital Status	Married	7	100
Level of Education	Graduate Degree	2	29.6
	Completed College	5	70.4
Employment	Employed	5	70.4
	Full Time Homemaker	2	29.6
Sex	Male	2	29.6
	Female	5	70.4

Focus Group One, Task One:

Four separate interrelated tasks were completed in Focus Group One. In task one, subjects were asked to brainstorm the aspects they believed best defined a quality healthcare facility. In this analysis, overlapping written and verbal responses were organized into overarching themes. Table 11 lists these themes (with descriptions) and their relative written frequency.

Table 11: Themes identified in written and verbal responses from Focus Group One, Task One

Theme	*Frequency*	*Percent*
Access - the ease with which an individual can obtain needed services	6	85.7
Recommendations - word-of-mouth recommendations by friends/family	5	71.4
Experience - Years of physician experience and where they received training	4	57.1
Timeliness - Reduction of wait times and appointment flexibility	4	57.1
Clinical Outcomes - Desirable and undesirable clinical consequences of care	3	42.9
Affordability - The cost felt by the consumer when obtaining services	3	42.9
Time Spent - How long provider spends with patient during consultation	3	42.9
Physical Status (Non-Medical) - Describing the physical facility; i.e. cleanliness, décor, parking availability, quietness, etc.	3	42.9
Coordination - Organization of patient care activities between multiple participants involved in care	3	42.9
Technology - Availability of advanced on-site medical technologies	2	28.6
Interpersonal (Medical Staff) - The social characteristics of providers (friendliness, courtesy, respect, etc.)	2	28.6
Empowerment - The opportunity for self-determination in healthcare	2	28.6
Interpersonal (Ancillary Staff) - The social characteristics of ancillary staff (friendliness, courtesy, respect, etc.)	1	14.3

Of all of the included responses, 6 of 7 subjects mentioned Access as an important component of healthcare quality (85.7%). This was the most commonly noted theme; however, other responses demonstrated similar levels of overlap. For example, 5 of 7 subjects mentioned provider or facility Recommendations by trusted family or friends as useful indicators of healthcare quality, while 4 subjects mentioned provider Experience and the Timeliness of Care as useful aspects of quality (Table 11).

Some brainstormed aspects of quality were less common, but nonetheless mirrored the sentiment of domains often found in the prevailing literature: Outcomes (n = 3 of 7), Affordability (n = 3 of 7), Time Spent (n = 3 of 7), the Non-Medical Physical Status (n = 3 of 7), Coordination (n = 3 of 7), Technology available on sight (n = 2 of 7), Interpersonal qualities of medical staff (n = 2 of 7), opportunities for Patient Empowerment (n = 2 of 7), and lastly, the Interpersonal qualities of ancillary staff (n = 1 of 7). Interestingly, the brainstormed contributions largely overlay previous research regarding patient satisfaction. For example, Anderson et al. (2007) listed 24 domains for patient satisfaction aggregated through an online survey. All but four of the domains cited by participants in Focus Group One were included in some form in Anderson et al’s (2007) aggregation of patient satisfaction response data. The only domains that were not included were (1) Experience, (2) Outcomes, (3) Physical Status (non-medical), and (4) Technology on site. However, the HCAHPs survey, which is used to measure patient satisfaction scores of American hospitals by the Centers for Medicare and Medicaid Services (CMS) does include questions related to the physical status (non-medical) of hospital facilities.

Thus, when consumers think of healthcare quality, initial responses often lean towards domains that they can physically sense and measure through their own experience. For example, patients can gather information on provider or facility recommendations, can assess the site cleanliness, technology available, affordability, and interpersonal qualities of medical and ancillary staff. These domains are what patients often must use to make assessments of healthcare quality, and therefore it is unsurprising that these same domains are what first come to mind.

The justifications that subjects made for why they selected different domains reveal the subjective nature of how healthcare quality is often assessed by consumers. Subject Two stated that Timeliness was an important indicator of quality: “When I get through the door, it has to be on time. I don’t want to be waiting all day because that shows bad management and makes me nervous.” Alternatively, Subject Six then rebutted "When I can get an appointment immediately for a specialist, I assume they don't have a lot of patients." Thus, the same experience can mean two different things for patients approaching the interaction through separate lenses.

A similar kind of subjective justification was used for experience. Subject Five stated “I also look at the age. I don't want you to be 65 and I don't want you to be 25. In that middle ground, you're still learning new things but you have the experience." To her, age was indicative of quality by weighing years of experience against up-to-date training.

Perhaps the most interesting domain was Recommendations. Five subjects mentioned Recommendations as an important indicator of quality, and through discussion, one underlying sub-theme became universally apparent: trust. Subject one stated, “I like referrals and recommendations from people I trust. That matters most for me.” Another subjected reiterated, “Yes, there are different hospitals I feel are geared towards different things. Let's say [a specific hospital] is more for Nephrology – that’s what my family and friends tell me – so I'd go there for that." Thus, for many patients, quality is assessed through subjective experience. In the absence of standardized data, subjects define quality through what they and trusted friends and family can interpret from their personal experience.

Evidence becomes frequently anecdotal. Assessments become increasingly equivocal. In Chapter One I discussed various healthcare rating websites that rely exclusively on crowd-ratings. What became clear through Focus Group One, Task One is that such an approach is challenging to justify. The same experience is often interpreted very differently by individual patients depending upon how they choose to subjectively assess quality.

However, a more interesting question is whether a patient's desired domains *change* once they are exposed to previously inaccessible domains often measured by organizations such as CMS or the WHO.

Focus Group One, Task Two

In Task Two, subjects were provided a list of domains cited in the literature and were given the opportunity to select those which they found the most important and reject those they thought were of little-to-no value. Table 12 lists the frequency of responses for each domain case; while only three subjects thought of Outcomes in Task One, all 7 selected Effectiveness as important in Task Two. These domains are *not* identical. Effectiveness represents the frequency of positive clinical outcomes while Outcomes includes the frequency of both positive and negative outcomes. However, there is considerable overlap in their conceptual definitions. Likewise, no subjects thought of Safety in Task One, but all 7 circled it as an important aspect of quality in Task Two. Moreover, certain domains that were commonly noted in Task One were considered to be of less relative importance when overlapped with Task Two. For example, while Affordability was mentioned by 3 of 7 subjects initially, no subjects labeled it as important in Task Two. Instead, four felt neutral about its value, and 3 considered it to be of no value.

Table 12: Written responses regarding accepted healthcare quality domains from the literature

Domain	*Frequency Important*	*Frequency Unimportant*	*Frequency Neutral*	*Frequency Confusing/Unclear*
Effectiveness	7	0	0	0
Safety	7	0	0	0
Coordination	5	1	1	0
Centeredness	5	1	1	0
Timeliness	3	2	2	0
Affordability	0	3	4	0
Efficiency	1	2	1	3
Equity	2	2	3	0
Expenditures	1	1	5	0
Satisfaction	3	3	1	0
Guidelines	3	1	2	1

Lastly, only 3 subjects listed Satisfaction as an important quality domain in Task Two, with 3 subjects considering it to be of no value; however, 9 of the 13 themes brainstormed in Task One were cited as measures of Patient Satisfaction by Anderson et al. (2007). In other words, while patients tend to think of quality in terms of Patient Satisfaction, less than half still consider it of value once exposed to other, previously inaccessible domains.

In Task Two, I observed a shift in perspective. Once subjects were provided with domains that assessed quantitative differences between facilities – e.g., Effectiveness or Safety – traditional subjective methods of assessing quality were often dismissed. For example, Subject Two had previously stated “If you walk in for a mammogram. You are terrorized by the front office staff. You are treated like you're not even a human being. So you walk in nervous and concerned. In my mind that's not good patient care. In terms of the patient experience, then the way the office staff interact with you is essential." But then, when discussing domains in Task Two, she changed her perspective: “I'd rather have someone be an ass to me and be the best damn doctor than have someone who's nice and never gets to the right answer." Originally, Subject Two felt that the interpersonal skills of office and medical staff was important because it might be indicative of higher quality care. However, once domains on outcomes were introduced as a possibility, the relative importance of interpersonal skills seemed to dissipate; i.e. as more concrete evidence of performance became available, subjective assessments of quality decreased in perceived value.

Focus Group One, Task Three

In Task Three, subjects were told to select the 5 most important domains from Tasks One and Two and to weight them accordingly in terms of level of importance by distributing a total of $100 (Table 13). Unsurprisingly, Effectiveness – which was circled by all subjects in Task Two – was included by 100% of subjects in their Task Three list. Likewise, Safety was also deemed important by the majority of subjects (n = 5). While those two domains were included consistently, the other domain selection preferences varied drastically.

Table 13: Frequency of Responses from Top 5 Most Important Domains Lists

Theme	*Frequency*	*Percent*
Recommended	1	14.3
Effectiveness	7	100.0
Safety	5	71.4
Patient-Centeredness	2	28.6
Timeliness	2	28.6
Access	1	14.3
Coordination	2	28.6
Health Expenditures	1	14.3
Experience	1	14.3
Technology	1	14.3
Time Spent	1	14.3

One subject selected only Effectiveness and Safety, with $80 and $20 budget allocations, respectively, while another put all of his money on Effectiveness, justifying it with "I put all of my money in effectiveness. I care only about effectiveness. I assume that if the care is effective, then implicitly, the care must have been safe. If I have timeliness and coordination, but if it wasn't effective, then who cares?"

Other subjects seemed to disagree. Subject four used the maximum number of 5 domains, selecting Timeliness, Effectiveness, Access, Coordination, and Health Expenditures, and rebutted “I don't view everything as looped in with effectiveness. I can be cured by one doctor and have it take a year because of a bunch of trials and tests, and another doctor can cure me in a month, both are technically effective. But those are two different things.” In other words, there was internal debate within the focus group regarding the definitions of each of these different domains. This is understandable. The challenge of working with domains is that they are defined conceptually to participants but not operationally. As such, it is easy to debate the level of overlap and areas of distinction between different options.

Regardless, there is a visible evolution between the domains listed in Tasks One and Three. While subjects initially regarded measures of patient satisfaction to be the most useful measures of quality, by Task Three most subjects had shifted towards weighting less subjective outcomes metrics more heavily. Subject Two, for instance, listed Recommendations, Experience, Access, and Time Spent as the most important domains in Task One, but then shifted to include only Effectiveness and Safety in Task Three. Her list completely changed, rejecting anecdotal evidence and including only standardized performance data. In fact, only one subject included Recommendations or Experience in his Task Three combined list.

Focus Group One, Task Four

In Task Four, subjects were posited with two vastly different healthcare scenarios (1. seeking out services for a hip replacement v. 2. seeking out primary care services for the flu) in order to test the universality of the quality definitions constructed in Task Three. The responses here were very mixed. Three of the 7 subjects chose to conserve 100% of the domains chosen in Task Three, regardless of scenario (Table 14). Three other subjects conserved some – but not all – of the domains between the two scenarios, and one subject constructed two lists with no overlap whatsoever (Table 14). For example, one subject explained, "Patient centeredness, I want that for primary care. This is more of a relationship. I will be around them for many years. For knee replacements, I just want to get in and get out." In this case, the type of treatment affected the patient's preferences regarding care. Likewise, another subject explained that the severity of illness also changes her preferences: "If you look at this from the perspective of the illness, the severity of your illness changes my answers, too."

While the overarching objective of current healthcare quality research is to construct a single universal definition of quality that applies to all patient scenarios, this approach might only satisfy the preferences of a certain subsect of the population. For other patients, preferences regarding how care is delivered change drastically depending upon the severity of illness or type of treatment necessitated.

Table 14: Conserved vs. Un-conserved Top-5 Domains for Varying Healthcare Scenarios

Subject	% Domains Conserved	If 100% Conserved, Are Budgets Conserved? (Y/N)
1	100	Y
2	50	N/A
3	100	N
4	0	N/A
5	100	N
6	20	N/A

7	33	N/A

Focus Group Type One Conclusions

The diversity in Focus Group Type One's findings illustrate the challenges that exist in establishing a unified patient-centric definition of healthcare quality. However, a number of useful conclusions can be drawn.

1. When patients initially think of quality, they usually think of patient experience.
2. Current patient methods for assessing quality are very subjective, very abstract, and often contradictory.
3. Patient definitions of quality tend to evolve when consumers are offered concrete data on outcomes; measures of satisfaction and experience are often regarded as less important than bottom-line standardized performance.
4. Effectiveness and Safety are essential domains in a patient-centric definition of quality.
5. There is often confusion and debate regarding what distinguishes and connects these domains. Conceptual definitions must be clearly operationalized for any ranking methodology to have clarity.
6. Results regarding the universality of quality definitions are mixed but stances are firm; some subjects firmly believe that their domain list is universal while others argue that their preferences adapt depending upon circumstance.

In summary, currently employed patient definitions of healthcare quality are largely subjective, anecdotal, and lead individuals to gather vastly different conclusions from similar clinical scenarios. This presents patients with consistent challenges as they attempt to navigate the immense choice offered in the healthcare marketplace. These findings illustrate the need for more standardized methods of approaching the assessment of healthcare quality and also add validity to the statement that crowd-rated online healthcare ranking sites are deeply flawed. Moreover, there is also evidence to suggest that patient-definitions of healthcare quality *evolve;* i.e. when presented with standardized data, patients often prefer such information over subjective experience-based evidence. With this evolution came near-universal support for the inclusion of both Safety and Effectiveness into the patient-centric definition of quality. However, despite this progress, discussions of healthcare quality were often stymied by confusion over the conceptual definitions of different domains. Future research should attempt to operationalize such concepts before subjects construct their preferred quality definitions.

Also, subjects vary widely in their support for a universal definition of quality. Due to the diversity of opinions, it appears unlikely that a unified definition can be constructed that both fulfills all patient preferences and applies to all healthcare scenarios/severities. Such a finding is reminiscent of earlier theoretical work in the realm of sociology. Sociologists of the classical cannon, such as Marx, Durkheim, and Weber all aimed to establish Grand Theories, which could describe human behavior in its totality. Likewise, current research in healthcare management aims to establish all-encompassing definitions of healthcare quality. However, maybe such aims are too broad. While some subjects demonstrated consistency in their definitions between scenarios, many had completely varying preferences.

One of the original aims of Focus Group Type One was to compare patient quality definitions to existing frameworks. At the most basic level, all themes could be classified according to Donabedian’s Model of Inputs, Processes, and Outputs. The two most frequently cited quality definitions – the WHO and OECD frameworks – both place the most emphasis on outputs, followed by processes, and minimize the importance of inputs. This approach was supported by Focus Group One's Finding. As Table 15 shows, subjects placed an average of $80 on themes classifiable as outputs in Task 3, an average of $17.10 on processes, and an average of only 2.90$ on inputs.

Table 15: Budget allocations in Task Three Classified as Inputs, Processes, or Outputs

	Budget Allocation at Each Level in Task 3 Domain List ($)		
Subject	Inputs	Processes	Outputs
1	0	10	90
2	0	0	100
3	0	50	50
4	0	15	85
5	0	35	65
6	20	10	70
7	0	0	100
Average Allocation	2.9	17.1	80.0

Likewise, the two most commonly cited domains were Effectiveness and Safety, both of which are included in the WHO, OECD, and National Quality Forum definitions of quality. In these respects, patient-definitions of healthcare quality are consistent with the predominant definitions in the field.

However, there are two points of distinction. First, WHO, OECD, and NQF definitions all emphasize Access and Equity; these definitions are appealing to broader strategic policy objectives and calls for social justice. In Task Three, only one subject listed Access as a domain and no subjects included Equity. Patients – at least when deciding where to receive their own services – understandably ascribe less weight to these broader ethical and social aims. Second, it appears that while Output-related themes are given the highest budget allocation on average overall, they were given a higher allocation for knee replacements ($85.70) compared to primary care ($63.30) (Table 16). Likewise, Processes were weighted more heavily in primary care services ($36.70) compared to knee replacements ($11.40) (Table 16). As subjects noted, as the perceived risk potential increases, the weight assigned to Outcomes-related domains increases.

Conversely, as the length of the care period increases, the weight assigned to Processes increases.

Table 16: Budget allocations for Inputs, Processes, and Outputs Under Two Healthcare Scenarios

Budget Allocation of Each Level in Task 3 Domain List ($)						
	Primary Care			Hip Replacement		
Subject	Inputs	Processes	Outputs	Inputs	Processes	Outputs
1	0	10	90	0	10	90
2	0	5	95	0	0	100
3	0	50	50	0	40	60
4	0	100	50	20	0	80
5	0	30	70	0	10	90
6	0	30	70	0	0	100
7	0	50	50	0	20	80
Average Allocation	0.0	36.7	63.3	2.9	11.4	85.7

With all of these considerations in mind, we can now construct a series of guidelines regarding the construction of patient-centered quality domains.

1. Include Effectiveness and Safety – these domains are almost universally supported as essential components of healthcare quality by consumer participants.
2. Operationalize all quality domains in rankings so that patients understand the distinctions between similar-sounding domains.
3. The majority of domain weights should relate to Outcomes, not Inputs or Processes.
4. In scenarios where illness severity and treatment risk are especially high, ascribe more weight to Outcomes.
5. In scenarios where length of care period is long-term, ascribe slightly higher weights to Processes (although Outcomes should still be in majority).

VII. FOCUS GROUP TWO: APPROPRIATE GRANULARITY OF INDICATOR DATA

Demographic Data for Focus Group Two

Six subjects were acquired for Focus Group Two, with 4 participating in session one and two participating in session two. As shown in Table 17 and 18, the average age of subjects in Focus Group Type Two was 43.5 years (σ = 11.9 years) with subjects ranging from 30-59 years of age. Moreover, the majority of subjects were male (n = 4), married (n=5), Caucasian (n =5) (however, there was one Hispanic/Latino participant), and held four-year college degrees (n=5). Five subjects were currently employed while one was a fulltime homemaker. Lastly, half of the participants had children, all of which were over the age of 18.

Table 17: Demographic characteristics for Focus Group Type Two

Descriptive	*Birth*	*Children < 5*	*Children 5-13*	*Children 14-18*	*Children >19*	*Total Children*
μ	43.5	0	0	0	1.2	1.2
σ	11.9	0	0	0	1.3	1.3
Minimum	30	0	0	0	0	0
Maximum	59	0	0	0	3	3

Table 18: Nominal Demographic Characteristics for Focus Group Two

Demographic Variable	*Category*	*Frequency*	*Percent*
Race	Caucasian	5	83.3
	Hispanic/Latino	1	16.7
Marital Status	Married	5	83.3
	Never Married	1	16.7
Level of Education	Graduate Degree	1	16.7
	Completed College	5	83.3
Employment	Employed	5	83.3
	Full Time Homemaker	2	16.7
Sex	Male	4	66.7
	Female	2	33.3

While Focus Group Type One was completed through a single 7-subject session, Focus Group Type Two was split into two separate sessions. It was noted that subjects in these latter two sessions were less vocal; moreover, because there were fewer participants per group, there appeared to be less variation in opinions within each session, thereby limiting the frequency of back-and-forth debates between participants.

Nonetheless, the findings in Focus Group Type Two were far more straight-forward than in Type One and could be summarized by two interconnected themes: (1) consumers believe that more information is better and (2) subjects struggled to understand the mathematical difference between raw scores and standard scores.

Theme One: ***The belief that more information is always better***

Participants were posited three alternative approaches for presenting data on indicators: (1) a list of all indicators used to construct the facility score, (2) a list of all indicators paired with their corresponding raw scores, and (3) and a list of all indicators paired with raw and standard scores (Appendix 16).

When asked which method was preferred, subjects believed that more information would be useful for making a decision. As one woman explained, “I would prefer more information than less. Seeing what makes up effectiveness or standards of care. Looking at that further, it's nice to see how that facility is rated with the raw value and then compare it with the national." This outlook was mirrored by a male subject in his mid-50s, “Without knowing the underlying indicators of effectiveness, I think I would want to know what makes that up. Effectiveness could be very broad. Something might stand out to me that's more important than looking at just effectiveness." In other words, this subject felt that the domain-level score is not enough on its own. Because effectiveness could be measured in many ways, he felt that it would be useful to look at the individual indicator-level scores; i.e. to pick and choose the indicators within each domain that he felt were the most important.

This is an interesting finding. It suggests that subjects are not always completely satisfied with domain-level scores. Because consumers take the process of purchasing healthcare very seriously, many want the ability to dig through the individual indicator-level scores before making a decision. Moreover, some might use indicator scores as an opportunity to pick-and choose the individual indicators that they find most important. These subjects would thereby potentially disregard (or place less weight on) domain-level and overall scores and would instead compare facilities according to performance on a handful of selected indicator scores. As one woman explained, “I think it depends on the person. I'm a little neurotic when it comes to my healthcare. But I would spend the time to figure it all out. But not everybody would do that. My dad would just look at the first page you showed us and be very happy with that. That would be enough for him. It wouldn’t be enough for me. If I could have more I would want more - just keep going and going and going. But that's my personality. I could get obsessed."

This finding is somewhat unsettling. If subjects disregard domain scores and instead focus on indicators, then decisions regarding where to receive care would be based upon only a handful of individual data points. It is important to recognize that these individual indicators do not exist in a vacuum; often, it is essential to see them in combination in order to gain an understanding of a facility’s performance on a particular domain. For example, when measuring the effectiveness of pneumonia care, one subject might only care about the indicator titled “Pneumonia 30-day mortality rate.” However, to truly understand the facility’s effectiveness in treating this disease, it would also be important to look at “Pneumonia 30-day readmission rates.” These two indicators are linked. Thirty-day readmission rates would obviously be low if the 30-day mortality rate was at 100%! In other words, the tendency for subjects to want to pick and choose the individual indicators that they find most important undermines the value added through aggregative statistics.

Theme Two: ***The mathematical distinction between raw and standard scores is not easy to explain or understand.***

While all subjects appeared to understand the distinction between these types of scores by the end of both sessions, it took approximately 10 minutes of questions and answers to get all subjects comfortable with interpreting these mathematical outputs. One subject struggled to understand that the raw score was not at a national level; i.e. that it pertained to an individual facility and was not the average value nationwide. Another subject struggled to understand why a high raw score didn't necessarily equate to a high standard score. In the end, all subjects agreed that a written description of each type of score would need to appear on the Indicators page for the data to be of potential utility. But that solution, while valid in theory, will be challenging to execute in practice. To explain the distinction between each type of score for participants with little-to-no background in statistics (keeping in mind that the participants in Focus Group Type Two all possessed college degrees) would require paragraphs of text. Perhaps more alarming, if the explanation is not effective and subjects misinterpret the data, these scores could create completely faulty assessments of healthcare facility performance. This would undermine the very objective of such a site.

Focus Group Two Conclusions

The findings from Theme One and Theme Two are at odds. In one respect, subjects displayed a strong desire to have as much information available as possible. This is in line with Berkowitz's (2014) book on healthcare marketing, which states that buyer behavior is dependent upon two factors (1) involvement and (2) risk (Figure X). When a decision is of high importance and risk, consumers invest more time and energy into the decision-making process. As such, it is unsurprising that subjects conveyed an interest in indicator-level information; when assessing healthcare services, consumers will likely be willing to do extensive, time-consuming research.

However, this preference is at odds with the objective of a healthcare rating website that uses aggregative statistics. If subjects make decisions off of handpicked individual indicators of self-determined importance, then the domain and overall scores are rendered useless. Consumers would arrive at drastically different assessments of healthcare quality based upon the individual indicators they decide to select during decision-making. Moreover, if consumers are unable to understand the difference between raw and standard scores, then providing this information could leave users at a detriment.

One way to balance these themes would be to provide standard scores but not raw scores. If all indicator scores were presented on the same scale as domain scores, then much of the confusion surrounding interpretation would be eliminated. That is to say, a consistent and unified rating scale at the overall-, domain-, and indicator-levels allows for the easiest possible interpretation of the data.

While this makes the interpretation of indicators less confusing, it still leaves unresolved the issue of subjects using indicator scores in lieu of domain or overall measures. This is far trickier to address. The only way to eliminate the risk of subjects making decisions based upon indicator scores would be to eliminate those scores altogether. In this scenario, the Indicators page would simply list out the indicators measured in each domain category but would include no information on facility performance for each of the measurements. This scenario would violate theme one, which states that subjects want complete data transparency. Likewise, if indicator standard scores are included, use of domain- and overall-level scores (which are the foundation of the website) are put at risk. Either way, a core tenant or principle is violated.

To resolve this, I fall back on the ideological position that healthcare rating websites should be based on *what consumers want to know.* A healthcare rating website only provides value for consumers if it presents the information that consumers want to know. With that in mind, this paper will argue that indicator-level standard scores should be included in consumer-centric healthcare rating websites.

VIII. FOCUS GROUP THREE: THE UI OF A CONSUMER-CENTRIC HEALTHCARE RATING WEBSITE

Demographic Data for Focus Group Three

In total, 10 subjects were acquired for participation in Focus Group Type Three (Table 19 and 20). Due to scheduling, 3 subjects were allocated to time slot one, and 7 participated in time slot number two. The size difference between the groups caused a visible difference in the discussion dynamics. In group one (n=3), verbal commentary was noted as being more sporadic, discussions were less openly debated, and one subject appeared to dominate much of the conversation. However, in group two, no individual subject was able to dominate in the discussion. Similarly, discussion appeared more impassioned as subjects disagreed over certain arguments. These results suggest that larger focus groups allow for a more rich and lively discussion.

Table 19: Demographic characteristics of subjects in Focus Group Three

Descriptive	*Birth*	*Children < 5*	*Children 5-13*	*Children 14-18*	*Children >19*	*Total Children*
μ	39.4	0.2	0.2	0.2	0.6	1.2
σ	14.1	0.6	0.4	0.6	1.0	1.0
Minimum	26	0	0	0	0	0
Maximum	587	2	1	2	3	3

Fifty percent of participants in Focus Group Type three were female, and the average age was 39.4 (Tables 19 and 20). However, the variation in ages was far greater than in the other focus group types (σ = 14.1 years), with the minimum age as low as 26 and a maximum age of 58. The majority of participants were Caucasian (n=6, 60%), with one Black participant and three Hispanics/Latinos (Tables 19 and 20). Four participants were married, four had never been married, one was separated, and one was widowed. Only one participant had children under the age of 5 years, while two had children between 5-13 years, one had a child between 14-18, and the plurality (n=4) had children over the age of 19. On average, the participants had approximately 1.2 children. Level of education also varied amongst the focus group participants. While the majority completed college (n=7), one attended some college, one completed high school, and one completed some high school. In terms of employment status, 50% of participants were employed, while 20% were full time homemakers, one was a current student, and finally one subject was retired (Tables 19 and 20).

Table 20: Nominal Demographic Characteristics of Focus Group Three Subjects

Demographic Variable	*Category*	*Frequency*	*Percent*
Race	Caucasian	6	60.0
	Hispanic/Latino	3	30.0
	Black	1	10.0
Marital Status	Married	4	40.0
	Never Married	4	40.0
	Separated	1	10.0

	Widowed	1	10.0
Level of Education	Graduate Degree	0	0.0
	Completed College	7	70.0
	Some College	1	10.0
	Completed High School	1	10.0
	Some High School	1	10.0
Employment	Employed	5	50.0
	Full Time Homemaker	2	20.0
	Unemployed	1	10.0
	Student	1	10.0
	Retired	1	10.0
Sex	Male	5	50
	Female	5	50

Color

Theme One: ***Bright pictures and colors delegitimize content – similar to an advertisement***

Respondents in task one felt that a consumer-directed healthcare rating website must first legitimize itself as a valid and trustworthy source. As such, there was broad consensus that overtly bright colors and pictures felt analogous to an advertisement. To quote participant number 3, “I just feel like I see all of this shit on the subway. Flash and bright colorful advertisements for healthcare.” Another subject than reiterated, “Bright colors don’t seem like a website that is a legitimate source. Not something for healthcare.” And lastly: “The most objective source possible seems like it would take away all of the flash.” In other words, while the subjects agreed that the website design with a picture of a patient (Appendix 17) was the most engaging, they retorted that it doesn’t appear trustworthy – as if the picture implies that the website is trying to sell or convince you of the validity of a product. Objectivity and the academic nature of the website’s intention were touted as characteristics to convey through the color scheme.

Theme Two. ***A uniform color scheme looks professional – consider "medical" colors (but not red)***

When asked about possible color combinations, subjects proposed that to make the site look professional, a uniform color scheme should be employed. Subject one brought up Consumer Reports, claiming that "Consumer Reports has a very uniform color scheme. It's red. And their logo is black with gold letters. It's like incredibly to the point. Obviously don't take red because this is healthcare, but maybe blue." All subjects in groups one and two agreed that red, blue and white represented standard healthcare related colors. However, all subjects also agreed that blue and white appeared the "neatest" and most "professional": "Certain colors make me think of healthcare. Like ambulances are certain kinds of colors… But red is too bold. A blue might be a bit more…" Once again, the argument from subjects was that the color scheme should be professional, identifiable as healthcare related, and trustworthy.

Thus, although subjects supported the inclusion of traditional healthcare related colors (blue and white), there was surprising universal rejection for the color red: "I would prefer blue as a healthcare color. I would stay away from red…" (all subjects nod in agreement) "…Yes that's three no's on red." In response, I probed, asking why. Subject respondents were unable to provide a clear answer. Instead, they all simply agreed that the color red was too "harsh." Theme Two thereby appears to relate to Theme One – that bright colors and pictures seem to delegitimize content. While blue and white are more neutral, soothing, and identifiable as healthcare-related, red is bold and "hot," according to one participant, and for some reason undesirable.

It is interesting that trustworthiness was mentioned consistently by participants. As previously argued by Rothenfluh et al. (2016) in their interview-based research, crowd-rating sites often falter because of barriers related to trust. Because the healthcare experience is intimate, close family and friends are often called upon first when choosing a provider. The subjects in Focus Group Type Three mirrored this sentiment, stating that the color scheme should convey a trustworthy tone, not that of an advertisement. To accomplish this aim, plainer, medical-related colors were preferred over “hot” or “bright” color schemes with large pictures.

Theme Three: ***If the graphic is colorful, avoid “hot background colors,” visa versa***

The subjects in Focus Group Type Three group two suggested that the website could have a “hot” colored background or a “hot” colored graphic on any page, but not both. In other words, if the Search Page (Appendix 17) has a bright picture on it, then the corresponding background should be less bright. Likewise, if the background is somewhat bright, then the graphic selected for any given page should be calmer. Subject 6 summarized the point best, explaining “"It depends on the graphic too. If the graphic is colorful then you don't want a hot color. If the graphic is like a grey coverage map with red dots then maybe it’s okay to have a colorful side. But if it’s colorful then maybe use a grey panel.” In an effort to provide balance – without making the website seem harshly bright – subjects agreed that a strategy should be employed to provide such balance.

Check Boxes v. Dropdown Menus

Theme One. ***Checkboxes take up excess space on the page and add too much text***

The results regarding checkboxes vs. dropdown menus with respect to the Search Page (Appendix 17) were unanimous; checkboxes create “clutter” on the page and can make the visual appearance overwhelming. In Appendix 17, 11 different potential facility types are listed as check boxes. However, a future healthcare rating site would need to compile information on all types of healthcare facilities; as such, the checkbox approach would quickly become unsustainable. Even as it currently stands in the Axure prototype models, subjects still felt that the design should be neat to maintain a professional demeanor.

Rating Scales

Theme One. ***More granular scales imply more available information***

Opinions regarding rating scale methodologies varied drastically. But the disagreements appeared to center around one underlying question: “I think the question we need to know to pick the best one is what kind of variation exists between hospitals. Are there like hospitals with rats in it? If there is that much variation, my first instinct is to eliminate the more simplistic rating systems. ABCD ratings don’t cover it.” The amount of data coupled with the quality/availability of the data used to make facility assessments determined whether or not subjects preferred less granular or more granular rating scales.

For example, a 0-100 scale can be called very granular, for it attempts to separate hospitals with 96 points from hospitals with 97 points. A scale with this much granularity inherently implies that a vast quantity of data were used to compile these scores; otherwise it would be impossible to separate facilities so close in performance. Conversely, if it were assumed that less data were available to assess facilities, then subjects tended to prefer more simplistic scales, such as 5-star ratings or letter grades (i.e., A-F).

Theme Two. ***Numerical scales are the most accessible in the context of healthcare***

With that in mind, I then proposed that subjects assume that the amount of data available to make these determinations is infinite. With the quality/amount of data not in question, I then asked which rating scale they would prefer. Responses gravitated toward the numerical options (0-10 or 0-100). It did not appear that the other scales were difficult to understand. As one subject explained, "I think one star, two star, three star is really easy to understand. Like restaurant rating scales." However, even though subjects felt these scales were easy to understand, many believed they did not fit within the context of healthcare. As one mother explained, "I don't like the A-F scale. It doesn't tell me anything. I'd rather see healthcare in numbers as opposed to letters or stars." Another subject reiterated, "I'm comfortable with the 1 to 10 because of familiarity. The stars only relate to hotels for me. I like the simpler approach."

Theme Three. ***More information in the scale limits the number of necessary clicks***

Star scales might be the easiest to conceptualize, but they also provide less information due to their lower granularity. Subjects asserted that one of the principle objectives of any website is to limit the number of clicks necessary to obtain the maximum quantity of information. For example, initially one subject argued that "If you're making tiers. Each star being a tier. And each rating has details about a hospital. Even if all have three stars than you can still differentiate by digging deeper." That is to say, a simpler rating scale might convey less information, but a subject can still acquire that information if he/she is willing to click onto the facility's individual page to find the raw scores for the individual countless indicators that comprise the domain and overall scores.

Yet, while this is theoretically possible, it requires the user to more actively seek out the information. Another subject then responded, “Does that give the user too much credit? I don't think people are going to do that... The whole idea of website development and design is to eliminate the total number of clicks." With that in mind, a more detailed rating scale – which provides more information on individual facilities – should theoretically reduce the number of eventual clicks that the user must perform to successfully perform facility comparisons.

Taking this philosophy of click-minimization to its logical extreme, the objective should be to include as much comparative information on facilities into one page as possible (while still maintaining an interface that is neither cluttered nor unclear). This leads us into our next section’s discussion on mapping.

Mapping

One Axure prototype model (Appendix 17) included a map on the List Page that pin-dropped each facility’s location. The purpose being to visualize where different facilities are located in relation to the user’s original location. However, reviews regarding the usefulness of this visualization were mixed amongst participants.

Theme One. ***Maps provide more information than just listing out distance***

The first criticism that subjects mentioned was that the map provides information using a large section of the page that can be conveyed simply by listing out the distance from the user in miles. However, the ensuing discussion explained that this is often not the case: “"The map is important. The miles does not dictate the amount of time. Having it on the map shows you how long it will take if you are using public transport." Raw distance might be enough for some users who commute by walking, driving, or cycling, but for those who rely upon public transport, the physical location (as much as the total distance) matters greatly when determining which facility is the most geographically accessible.

Theme Two: ***A smaller map would leave more room for information on ranking-related content***

While subjects in both groups eventually agreed that the map was useful, some mentioned that it did not need to be as large as prototyped in Appendix 17. As stated in Theme 3 in our section on Rating Scales, subjects felt that the objective should be to provide as much information as possible on ranking-related criteria on the List Page, with the aim of reducing the eventual number of required clicks. To achieve this, subjects in group two suggested including domain-level scores (traditionally reserved for the Facility Page) on the List Page. One subject in her mid-twenties suggested "What if you only put the address information when you hover over the map, and then put the criteria where the address is, so that it's just three bullets and if you want to see where it is you hover it." In this model, identifying information such as phone number, address, distance, bed count, etc., would appear if the user hovered the cursor over the pin-drop located on the map. This would free up space on the List Page to include not only each facility’s overall score, but also the domain-level scores. Other subjects seemed to support this approach, stating “I like that. That makes it based more on criteria. The criteria come first and foremost and the map comes up on the side. The map and identifiers become a smaller thing. And the criteria becomes more important. You get more information on the first page."

Theme Three: ***Order facilities by rating, not distance***

One-hundred percent of participants in both groups agreed that facilities should be list-ordered by rank on the List Page, not by distance. To use one subject's wording, “"You already set your travel radius to what you're willing to do. So you assume everything within it is reasonable. Order by rank."

User-Dependent Domain Weighting

In one prototype (Appendix 17), the design enabled users to determine how the domains should be weighted prior to aggregation for the overall score. The discussion of this prototype relates to the findings of Focus Group Type One. In Focus Group Type One, it was found that subjects had varying preferences regarding the types and weights of domains included when defining healthcare quality for distinct healthcare scenarios. Because subjects have varying preferences, an intuitive solution would be to enable subjects to decide how domains are weighted in the aggregation process; that way, subjects with varying ideological preferences could place greater weights on the domains they found most important. However, this idea was quickly struck down in Focus Group Type Three's discussion.

Theme One. ***Divesting control over weighting de-legitimizes the website's proposed service***

While subjects understood the sentiment that subjects might have varying preferences regarding domain weighting, the prevailing sentiment was that a healthcare rating website should have a firm and defined method for ranking facilities: “That's the whole point of consulting a separate source. That's your service. Letting them know how much to weight each domain." One subject felt that patient preferences could often become misguided, thereby undermining the validity of rankings. He explained, “Even if they do, they are wrong. Safety matters. You can’t just eliminate an entire domain." That is to say, while users might have varying preferences, it should not be possible for a user to lower the weight or completely eliminate one or multiple domains from the ranking system. Philosophically, this subject was arguing that the user should have less healthcare related knowledge than the website creators; therefore, the website should take a firm and justifiable stance regarding the domain weightings chosen.

Theme Two. ***Subjects found the concept of domain weighting confusing***

Perhaps more important, it was apparent that a number of subjects struggled to understand what was meant by “domain weighting.” When one subject was offered the ability to modify the relative weights ascribed to each domain, she responded “If everything is weighted at 25. Then am I looking for something that is 25%. Is that the optimum?" Another subject in group one seemed to agree, reiterating “I think it is a good idea if you could manipulate the numbers. But I don't know who it is geared for. It may be very useful for people that are familiar with medical things, but it might be very confusing for people that are first starting out… If it's targeted for the regular guy, then I think it would be too confusing to manipulate the numbers. I would take the top three then dig deeper into the individual domains." These subjects agreed that the sentiment of divesting control to the user was valid; nonetheless, they found the feature of user-inputted domain-weighting challenging to use in practice.

Conclusions

Here I explored five "axes" that could be manipulated when designing the user interface of a consumer-centric healthcare rating website: (1) color, (2) search method (check boxes v. dropdown menus), (3) rating scale, (4) method for visualizing facility locations, and (5) opportunities for user-control over ranking the methodology. The theme findings for each of these axes can now be combined in order to establish a set of guidelines for how a consumer-healthcare rating website should look.

1. Avoid large pictures and use "medical" colors, such as white and blue (but avoid red).
2. Use dropdown menus or user text input fields to organize the Search Page.
3. If the data quality permits, use numerical, points-based rating scales to allow for more granularity.
4. Maps are essential for understanding facility distance, but should be left small to make room for total and domain scores on the List Page.
5. Facilities should be ordered by rating, not by distance.
6. A feature to enable users to select their preferred domain weighting method should not be offered.

With a set of evidence-based guidelines established, it is now possible to rate and present data on healthcare facilities in such a way that balances varying patient preferences.

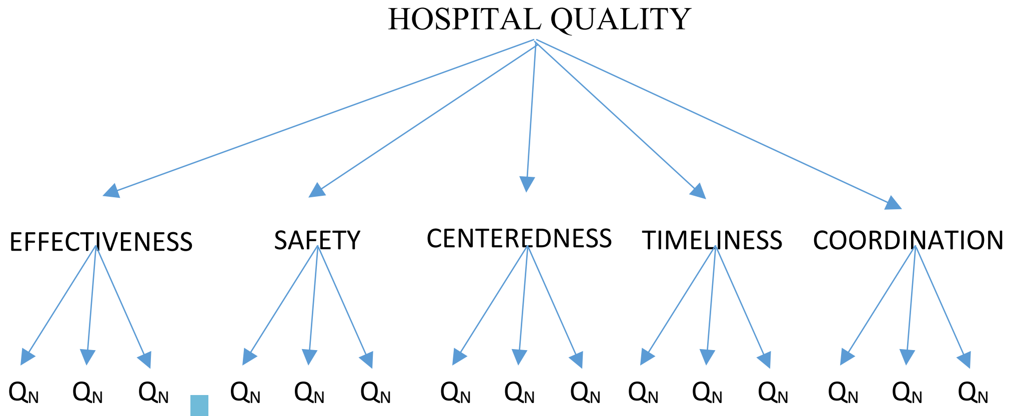
CHAPTER FOUR: ATTEMPT AT WIRE-FRAMING AND RATING U.S. HOSPITALS

In the final chapter of this thesis, I attempted to rate all of the CMS-registered hospitals in the United States using the guidelines extracted from Focus Groups One, Two, and Three. The mathematical steps administered to rate these hospitals were outlined in Chapter Two: (1) theoretical framework, (2) data selection, (3) imputation of missing data, (4) normalization, (5) weighting and aggregation, (6) robustness and sensitivity, (7) back to the real data, (8) links to other variables, and (9) presentation and visualization. Let's discuss each of these in turn.

Theoretical Framework:

All of Focus Group One was devoted towards establishing a patient-centric quality framework. While the variation in responses was too vast to create a singular unified theory, certain guidelines were extracted and used to rank hospitals. For example, both Effectiveness and Safety are included in this model because of their near-universal frequency of inclusion in participants' top-5 most important domains tables. Patient-centeredness, Timeliness, and Coordination were also included, because they were the only other domains mentioned more than once by participants (Table 13). The theoretical structure of quality domains for hospital services can therefore be modeled by Figure X and follows the First-Order/Second-Order model put forward by Huang (2002).

Figure 5: Theoretical Framework for Hospital Quality



In terms of the traditional Donabedian Model, Effectiveness, Safety, and Patient Centeredness could be termed as Outputs, while Timeliness and Coordination represent Process domains (Donabedian, 1965). The latent construct of hospital quality is therefore measured through the combination of the 5 underlying and related domains (Effectiveness, Safety, Centeredness, Timeliness, and Coordination), each of which is estimated through Q_N indicators.

Data Selection: Indicators should be selected based upon their “analytical soundness, measurability, coverage, relevance to the phenomenon being measured, and relationship to each other” (Hoffman et al., 2008: 17).

The indicators included in this analysis were collected from the CMS website and were submitted to CMS by all CMS-registered U.S. hospitals. In total, 86 relevant indicators were available from the CMS database at the time this analysis was conducted (Tables 21, 22, 23, 24, 25). Tables 21-25 show these extracted indicators organized into the domains of the theoretical framework outlined in the previous section.

Table 21: Effectiveness Measures from CMS Database

Effectiveness Measure Description	**Label**
Acute Myocardial Infarction (AMI) 30-Day Mortality Rate	MORT_30_AMI
Heart failure (HF) 30-Day Mortality Rate	MORT_30_HF
Pneumonia (PN) 30-Day Mortality Rate	MORT_30_PN
Acute Myocardial Infarction (AMI) 30-Day Readmission Rate	READM_30_AMI
Heart failure (HF) 30-Day Readmission Rate	READM_30_HF
Pneumonia (PN) 30-Day Readmission Rate	READM_30_PN
Rate of readmission after hip/knee surgery	READM_30_HIP_KNEE
Rate of readmission after discharge from hospital (hospital-wide)	READM_30_HOSP_WIDE
Rate of unplanned readmission for chronic obstructive pulmonary disease (COPD) patients	READM_30_COPD
Rate of unplanned readmission for stroke patients	READM_30_STK
Death rate for chronic obstructive pulmonary disease (COPD) patients	MORT_30_COPD

Death rate for stroke patients	MORT_30_STK
Rate of unplanned readmission for Coronary Artery Bypass Grafting (CABG)	READM_30_CABG
Death rate for CABG	MORT_30_CABG

Table 22: Safety Measures from CMS Database

Safety Measure Description	**label**
Deaths among Patients with Serious Treatable Complications after Surgery	PSI_4_SURG_COMP
Collapsed lung due to medical treatment	PSI_6_IAT_PTX
Serious blood clots after surgery	PSI_12_POSTOP_PULMEMB_DVT
A wound that splits open after surgery on the abdomen or pelvis	PSI_14_POSTOP_DEHIS
Accidental cuts and tears from medical treatment	PSI_15_ACC_LAC
Serious complications	PSI_90_SAFETY
Rate of complications for hip/knee replacement patients	COMP_HIP_KNEE
Central line-associated bloodstream infections (CLABSI) in ICUs and select wards	HAI_1_SIR
Central line-associated bloodstream infections (CLABSI) in ICUs only	HAI_1a_SIR
Catheter-associated urinary tract infections (CAUTI) in ICUs and select wards	HAI_2_SIR
Catheter-Associated Urinary Tract Infections (CAUTI) in ICUs only	HAI_2a_SIR
Surgical Site Infection from colon surgery (SSI: Colon)	HAI_3_SIR
Surgical Site Infection from abdominal hysterectomy (SSI: Hysterectomy)	HAI_4_SIR
Methicillin-resistant Staphylococcus Aureus (MRSA) Blood Laboratory-identified Events (Bloodstream infections)	HAI_5_SIR
Clostridium difficile (C.diff.) Laboratory-identified Events (Intestinal infections)	HAI_6_SIR

Table 23: Centeredness Indicators from CMS Database

Centeredness Measure Description
Cleanliness star rating
Nurse communication star rating
Doctor communication star rating
Staff responsiveness star rating
Pain management star rating
Communication about medicine star rating
Discharge information star rating
Care transition star rating
Overall rating of hospital star rating
Quietness star rating
Overall likelihood to recommend hospital star rating

Table 24: Coordination Measures from CMS Database

Coordination Measure Description	**Label**
Percent of mothers whose deliveries were scheduled too early (1-2 weeks early), when a scheduled delivery was not medically necessary	PC_01
Pneumonia patients given the most appropriate initial antibiotic(s) Higher percentages are better	PN_6
Surgery patients who were taking heart drugs called beta blockers before coming to the hospital, who were kept on the beta blockers during the period just before and after their surgery Higher percentages are better	SCIP_CARD_2
Surgery patients who were given the right kind of antibiotic to help prevent infection Higher percentages are better	SCIP_INF_2
Ischemic or hemorrhagic stroke patients who were evaluated for rehabilitation services Higher percentages are better	STK_10
Ischemic stroke patients who received a prescription for medicine known to prevent complications caused by blood clots at discharge Higher percentages are better	STK_2
Ischemic stroke patients with a type of irregular heartbeat who were given a prescription for a blood thinner at discharge Higher percentages are better	STK_3
Ischemic stroke patients needing medicine to lower bad cholesterol, who were given a prescription for this medicine at discharge Higher percentages are better	STK_6
Ischemic or hemorrhagic stroke patients or caregivers who received written educational materials about stroke care and prevention during the hospital stay Higher percentages are better	STK_8
Patients who got treatment to prevent blood clots on the day of or day after hospital admission or surgery Higher percentages are better	VTE_1
Patients who got treatment to prevent blood clots on the day of or day after being admitted to the intensive care unit (ICU) Higher percentages are better	VTE_2
Patients with blood clots who got the recommended treatment, which includes using two different blood thinner medicines at the same time Higher percentages are better	VTE_3
Patients with blood clots who were treated with an intravenous blood thinner, and then were checked to determine if the blood thinner caused unplanned complications Higher percentages are better	VTE_4
Patients with blood clots who were discharged on a blood thinner medicine and received written instructions about that medicine Higher percentages are better	VTE_5
Patients who developed a blood clot while in the hospital who did not get treatment that could have prevented it Lower percentages are better	VTE_6
Children and their caregivers who received a home management plan of care document while hospitalized for asthma	CAC_3
Heart failure patients given an evaluation of left ventricular systolic (LVS) function Higher percentages are better	HF_2

Patients assessed and given influenza vaccination Higher percentages are better	IMM_2
Healthcare workers given influenza vaccination Higher percentages are better	IMM_3_OP_27_FAC_ADHPC
Abdomen CT Use of Contrast Material	OP_10
Thorax CT Use of Contrast Material	OP_11
Outpatients who got cardiac imaging stress tests before low-risk outpatient surgery	OP_13
Outpatients with brain CT scans who got a sinus CT scan at the same time	OP_14
MRI Lumbar Spine for Low Back Pain	OP_8
Mammography Follow-up Rates	OP_9

Table 25: Timeliness Measures from CMS Database

Timeliness Measure Description	**Label**
Median Time to Fibrinolysis	OP_1
Heart attack patients who got drugs to break up blood clots within 30 minutes of arrival Higher percentages are better	AMI_7a
Heart attack patients given a procedure to open blocked blood vessels within 90 minutes of arrival Higher percentages are better	AMI_8a
Average (median) time patients spent in the emergency department, before they were admitted to the hospital as an inpatient	ED_1b
Average (median) time patients spent in the emergency department, after the doctor decided to admit them as an inpatient before leaving the emergency department for their inpatient room	ED-2b
Average (median) time patients spent in the emergency department before leaving from the visit A lower number of minutes is better	OP_18b
Outpatients with chest pain or possible heart attack who got drugs to break up blood clots within 30 minutes of arrival Higher percentages are better	OP_2
Average (median) time patients spent in the emergency department before they were seen by a healthcare professional A lower number of minutes is better	OP_20
Average (median) time patients who came to the emergency department with broken bones had to wait before getting pain medication A lower number of minutes is better	OP_21
Percentage of patients who left the emergency department before being seen Lower percentages are better	OP_22

Percentage of patients who came to the emergency department with stroke symptoms who received brain scan results within 45 minutes of arrival Higher percentages are better	OP_23
Average (median) number of minutes before outpatients with chest pain or possible heart attack who needed specialized care were transferred to another hospital A lower number of minutes is better	OP_3b
Outpatients with chest pain or possible heart attack who received aspirin within 24 hours of arrival or before transferring from the emergency department Higher percentages are better	OP_4
Average (median) number of minutes before outpatients with chest pain or possible heart attack got an ECG A lower number of minutes is better	OP_5
Surgery patients who were given an antibiotic at the right time (within one hour before surgery) to help prevent infection Higher percentages are better	SCIP_INF_1
Surgery patients whose preventive antibiotics were stopped at the right time (within 24 hours after surgery) Higher percentages are better	SCIP_INF_3
Surgery patients whose urinary catheters were removed on the first or second day after surgery Higher percentages are better	SCIP_INF_9
Patients who got treatment at the right time (within 24 hours before or after their surgery) to help prevent blood clots after certain types of surgery Higher percentages are better	SCIP_VTE_2
Ischemic or hemorrhagic stroke patients who received treatment to keep blood clots from forming anywhere in the body within 2 days of hospital admission	STK-1
Ischemic stroke patients who got medicine to break up a blood clot within 3 hours after symptoms started Higher percentages are better	STK_4
Ischemic stroke patients who received medicine known to prevent complications caused by blood clots within 2 days of hospital admission Higher percentages are better	STK_5

From this list, I chose to include only those indicators which had greater than 40% of the data available. While CMS requests data on a vast number of indicators, many facilities fail to report for a variety of reasons. Also, variables were excluded if little-to-no variation existed between cases. For example, CMS collects data on the percent of providers vaccinated; however, nearly all facilities sustain rates of vaccination coverage between 98 and 100%. Thus, the indicator is of no comparative value if compliance is near-identical at all sites. The highlighted rows in Tables 21-25 represent the indicators excluded from this analysis.

Imputation of Missing Data: *Different approaches for imputing missing values exist. Developers should select an approach that best fulfills the computational needs of the indicator*

The issue of imputation became particularly important when working with CMS data, because a large portion of cases are often left unreported by facilities. A number of different methods were considered – hot and cold, regression, expectation maximization, and multiple imputation methods. Hot deck imputation involves imputing a missing value with a case's previous and most recent available value for the given indicator. Cold imputation, conversely, uses values from another similar dataset (i.e. data on a similar variable is used if more readily available). Expectation maximization is an iterative process that calculates the probabilities for each possible value of the imputation to complete the case. The imputed value that has the highest probabilistic likelihood represents the maximum, and is imputed for that case. However, I arrived at unconditional mean imputation due to its relative simplicity and widespread use in the literature. Meaning, if a facility failed to report on a certain indicator, its score was determined by providing the national average for that indicator.

Normalization: *Indicators must be normalized so that they can be compared. Extreme values and skewed data must be accounted for*

According to the OECD composite indicators guide, indicator-level data must first be converted to the same scale before any first-order or second-order aggregation method can be used. Options included using z-scores, the distance to a reference, min-max normalization, indicators above or below the mean, or the cyclical indicators method. Z-scoring is by far the most popular approach when working with ratio variables, so was selected as the normalization technique for this analysis.

Moreover, Hoffman et al. (2008) note that indicator data must first be transformed to account for skew and kurtosis prior to aggregation; i.e. aggregation methods assume that all indicators are normally distributed. Thus, prior to normalization, all data were graphed on a frequency distribution to visualize the distribution. Skew and kurtosis were calculated in JMP using the Shapiro-Wilk and Kolmogorov-Smirnov test; however, Kim (2013) explains that such tests – while useful for small to medium samples (e.g., n < 300) – are unreliable for larger samples. Because the CMS data on hospitals contain thousands of cases, visual inspection became the primary method of assessing skew and kurtosis. Moreover, Kim (2013) argues that another method rule-of-thumb for analyzing skew and kurtosis for large samples is to assume that an absolute skew > 2 and an absolute kurtosis > 7 represents substantial evidence for non-normality.

With these two tools in mind, all indicators were initially plotted on a histogram and both skew and kurtosis were calculated. In instances where skew/kurtosis were above 2 and 7 respectively and when the histogram was visibly non-normal, transformations were applied using the Power of Ladders approach until the corresponding distribution was within normal bounds. After normality was achieved, distributions were then normalized using z-scores. Lastly, to limit the effect of outlier indicators on domain and overall scores, all indicators were Winserized at z-scores of -3 and +3.

Weighting and Aggregation: <u>The chosen method of weighting and aggregation should be based upon the underlying theoretical framework. Correlation (and duplication) issues between individual indicators must be identified and corrected during the aggregation process</u>

Weighting and aggregation occurred through a two-phased process in the construction of hospital composite scores. First-order weighting and aggregation corresponded to the weighting and aggregation of indicators into domain scores, while second order aggregation was completed while weighting and aggregating domains into the overall measurement of the latent variable quality. Options for weighting included conducting a Principal Component Analysis (PCA), Factor Analysis, data envelopment analysis, benefit of the doubt approach, using the analytical hierarchy process, or a conjoint analysis, each of which is discussed in the review of literature.

However, in or design, Domains were weighted using the Budget Allocation Process (BAP) completed in tasks 3 and 4 of Focus Group One. Because this is a website designed *for* consumers, our definition of healthcare quality is insistent upon adherence with patient preferences. Focus Group One allowed us to estimate preferred patient domain-level weightings. While based off of a small sample size and limited healthcare scenarios, I believe these are still useful results for guiding the development of domain weighting choices.

It was estimated that \$63.30 should be allocated to outputs during primary care/long-term services while \$85.70 should be allocated to outputs during severe/acute services (Table 16). Likewise, \$39.30 should be allocated to inputs/processes during PC/long-term services and \$14.30 should be allocated to inputs/processes during PC/long-term services (Table 16).

The challenge here is that hospitals provide both short-term PC services and long-term treatments for severe illnesses. Because ranking hospitals *overall* is broader than ranking a specific service type at hospitals, output and input/process weights were averaged between the two scenarios. In total, domains classified as outputs were allocated $74.50 while domains classified as inputs/processes were allocated $25.50.

1. Effectiveness, Safety, and Patient-Centeredness allocated $74.50 in total
2. Coordination and Timeliness allocated $25.5 in total

Within the category of outputs, the domains Effectiveness, Safety, and Patient-Centeredness were allocated varying dollar amounts based upon their relative frequency of inclusion in Focus Group One subjects' top-five most important domains lists. Effectiveness was referenced 7 times, Safety 5 times, and Patient-Centeredness 2 times, totaling 14 mentions in total. Using their relative proportions of the total, Effectiveness was allocated $37.25, Safety $26.60, and Patient-Centeredness $10.65. The same process was completed for the input/process domains, yielding the following allocations: Timeliness = $12.75 and Coordination = $12.75. In total, the 5 domain weights add to $100.

Aggregation amongst domains was completed using a weighted arithmetic mean. Hoffman et al. (2008) also discussed the use of geometric aggregation, which is designed to reward facilities with minimal variation between domain-level scores. However, I opted for an arithmetic approach, which I believe more transparently reflected the intent of subjects when asked to weight their chosen domains in Focus Group One.

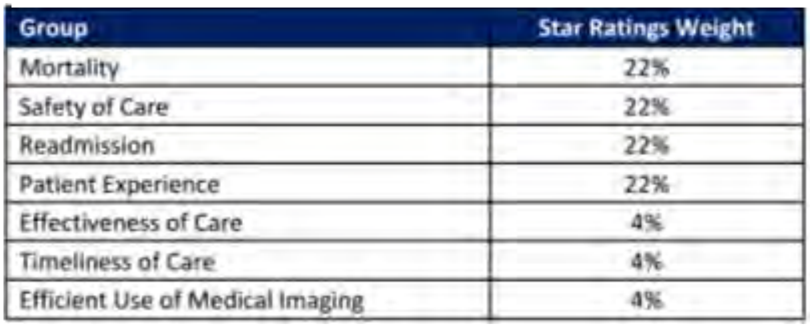
Weighting and aggregation at the indicator-level was a more challenging issue to address. Once indicators were populated into their respective domains, had mean values imputed, were transformed, and normalized, a correlation matrix was constructed to ensure that no indicators were duplicative. Meaning, if two indicators in the Effectiveness domain were correlated at a level above 0.9, it could be said that these two indicators are measuring the same construct. In other words, are vaccination rates amongst providers that different a measure than vaccination rates amongst patients? Or, should the two be combined into one measure: i.e. vaccination rates. Because answering this question through logical arguments appears overtly subjective, the approach taken in our composite ranking method was statistical in nature. The threshold for indicator duplication was set at 0.8. A correlation matrix was constructed for all indicators within each domain, and in instances that indicators were found to be duplicative, their values were averaged prior to aggregation (thereby giving each indicator ½ weight) (Appendix 18 shows the correlation matrix for the indicators included in the Effectiveness Domain).

In our model, all indicators were equally weighted. While other options include using statistical methods such as PCA or Factor Analysis, who is truly to say that Staff Responsiveness is more important than Communication about Medications within the Patient-Centeredness domain. All indicators were deemed to be of equal importance in this analysis, and thus were weighted equally after accounting for duplicates. Aggregation of indicators was performed through the calculation of an unweighted arithmetic mean (Hoffman et al. 2008).

One of the challenges of working with missing data is establishing a threshold for minimum required data availability. Meaning, while my ranking system imputes the indicator average for missing cases, if all indicators are missing for a facility, how could I justify providing domain or overall scores? This would be impossible. If a facility reports no data, the facility score would end up equaling the national average (5.0 on a 10.0 point scale), because the national average would have been imputed for all indicator scores. This would make little sense; there would be no evidence to support the claim that the facility performed at the level of the national average. Thus, a threshold must be set prior to the aggregation of indicators for domain and total scores.

Deciding what this threshold should be appears to be somewhat arbitrary. There are no set rules of thumb or statistical guidelines that can be followed to answer this question. However, it can be stated that the higher the threshold, the more rigorous the approach, the more reliable and accurate the facility scores. CMS sets a relatively low threshold; i.e. of the 57 indicators included in Hospital Compare, all facilities received scores if they reported on at least 9. Moreover, hospitals also receive scores if one or more domains are completely missing. The relative weights of the other domains are simply increased proportionally to account for this missing data. For example, Table 26 shows the domains included in Hospital Compare and their relative weights. If a facility's data are missing for Mortality, Safety of Care, Readmissions and Patient Experience, that facility still receives a score in Hospital Compare, simply by adjusting the weights of Effectiveness, Timeliness, and Efficient Use of Medical Imaging proportionally. This seems unwise, given that these three domains should theoretically only account for 12% of the total star rating if all data are available.

Table 26: CMS Hospital Compare Domain Weights



Group	Star Ratings Weight
Mortality	22%
Safety of Care	22%
Readmission	22%
Patient Experience	22%
Effectiveness of Care	4%
Timeliness of Care	4%
Efficient Use of Medical Imaging	4%

I decided to construct our healthcare ratings under much stricter data availability parameters. While this decreased the number of facilities which received scores, it also increased the reliability of those scores. First, I set the minimum reporting threshold for a domain score at 50%. That is, 50% of all indicators within a domain must be reported on for that domain to receive a score. Second, all domains must have scores for an overall quality score to be calculated. Meaning, if Patient-Centeredness is the only domain that has enough data to receive a score, then no overall quality score will be estimated in our model, because not enough data are available to accurately estimate the latent construct of healthcare quality.

Lastly, it is important to note that the domain-level scores are standardized and Winserized again following the unweighted averaging of indicators. This ensures that all domain-level scores are z-scored and follow a normal distribution with 0 as the national average and a standard deviation of 1. This step is essential to ensuring that all domains are on the same scale prior to the aggregation of domains into the overall quality score. Also, because z-scores are challenging to interpret, all domain scores are converted onto a 0.0-10.0 point scale by using the following formulae:

1. Data are converted onto a -1 to +1 scale; Score = $(x - (\text{Max} + \text{Min})/2)/[(\text{Max} - \text{Min})/2]$
2. Data are converted from -1 to +1 scale to a 0-10 scale; Score = $(x*0.5 +0.5)*10$

This same process is repeated following the aggregation of domain scores into the total score; meaning, the total score is standardized, Winserized, and converted to a 0-10 scale to ensure that 5 is the national average for total score performance.

Back to the Real Data: *Indicators should be transparent to potential users and replicable – i.e. can be decomposed into original indicators or values*

After all domain and overall scores have been calculated, the principal objective is to make sure that the data are accessible for the lay consumer. Much of this process was guided by the data obtained during Focus Groups Two and Three.

First, for each facility, the following information are included on indicators: (1) the name/description of each indicator included in each domain and (2) the standard score for each indicator included in each domain (Axure prototype shown in Appendix 19). The standard score (on a 0-10 scale) is used in order to maintain internal consistency regarding how all ratings are constructed. While I considered including the raw score, Focus Group Two illustrated that including both the raw and standardized score is confusing to participants, who struggle to interpret the raw value out of context.

Also, a technical user's guide PDF will be made available on the About Us page, outlining the theoretical framework and mathematical techniques employed in the construction of these composite scores. On the List and Facility Pages, the relative weights of each domain with respect to the total score are shown (Appendix 19). The objective here is complete transparency regarding what indicators are included in each domain, their corresponding scores, and how domains are weighted to arrive at the total facility score.

Presentation and Visualization: *<u>Indicators can be presented in many ways. Disseminating the results of the composite indicator in an interpretable fashion is essential</u>*

The physical layout of the website is an essential component of maintaining both usability and transparency for data dissemination. The structure and design of the website was based off of the results gathered during Focus Group Type Three (Appendix 19). A blue and white color scheme is used to convey a professional image consistent with medical themes, no pictures are displayed to avoid appearing like an advertisement, dropdown menus are employed on the Search Page as opposed to check boxes, a 0-10 point rating scale is used for indicators, domains, and total scores, maps are used to visualize the distances of facility options, facilities are ordered on the List Page by Total Score (not distance), and the List Page includes not only overall scores but also domain-level scores (and their respective weightings) (Appendix 19). It is our hope that these design features improve the usability of the website and the transparency of the data displayed therein.

I. TRENDS IN HOSPITAL RANKINGS DATA

The value in ranking U.S. hospitals extends beyond the utility it presents to consumers who may consult consumer-choice healthcare websites. To date, no other database attempts to rate the quality of all U.S. hospitals through a patient-influenced theoretical framework and strict minimum reporting standards for inclusion. With a rigorous metric of overall quality established, it is now possible to perform exploratory analyses of hospital quality across multiple independent variables; state (or other form of geographic region), bed count, ownership type, payer mix, religious affiliations, etc. Some preliminary findings in this area are reported here.

Let’s start our discussion with geography. Figure 6 illustrates the variation in average hospital quality when broken down by state (a data table is also presented in Appendices 20 and

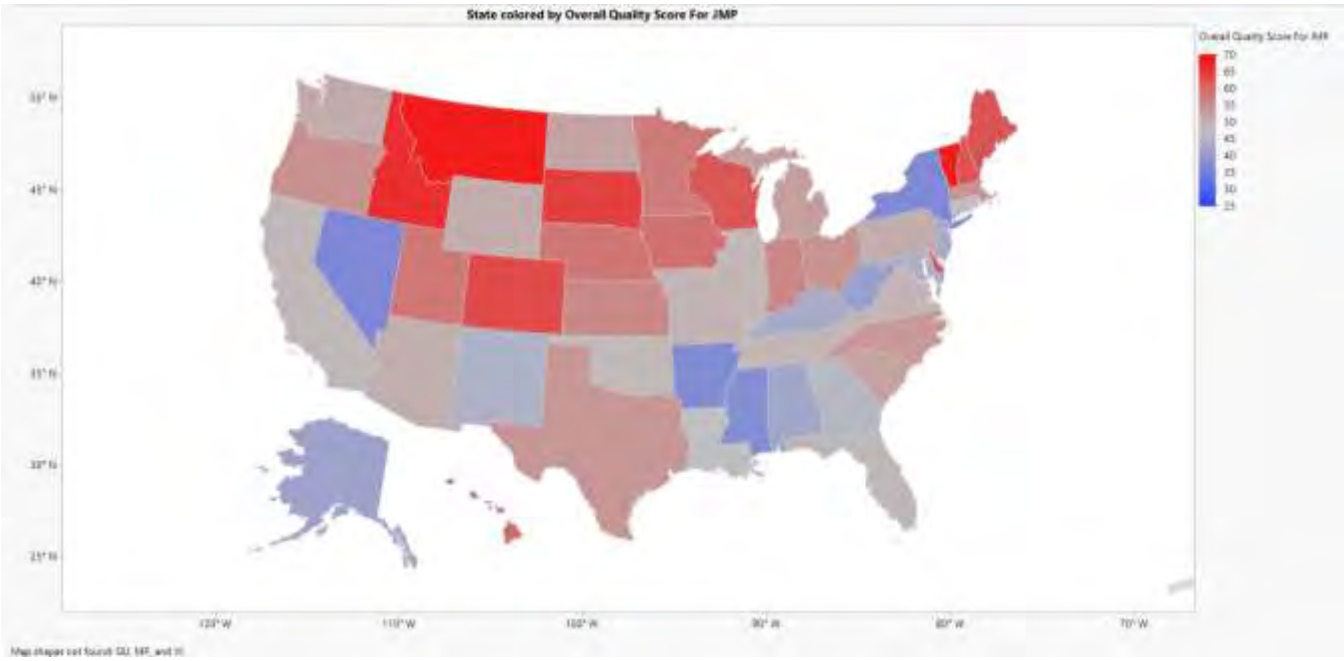


Figure 6: Average state hospital performance color-mapped. Red states perform highest compared to blue states.

21). Visually, it appears that south-eastern states perform relatively poorly compared to the rest of the country, while New England and Mid-Western states perform well on average. There are plenty of potential explanations for this trend, many of which are beyond the scope of this text. However, some possible contributors might be systemic differences in hospital funding (due to government subsidies) or disparities in average uninsured percentages. While this trend is visualized in Figure 6, an Analysis of Variance was conducted and uncovered that the pattern is significant at the 0.0001 level.

Visual inspection also reveals a pattern when comparing states along population size. Some of the highest performing states are those with smaller populations, such as Montana, Idaho, and Vermont. Unfortunately, hospital site population data are not included in the CMS database, but could likely be sourced from other databases for further studies.

I also conducted an ANOVA of hospital performance across multiple categories of hospital ownership type: Voluntary non-profit – Church, Government – Federal, Voluntary non-

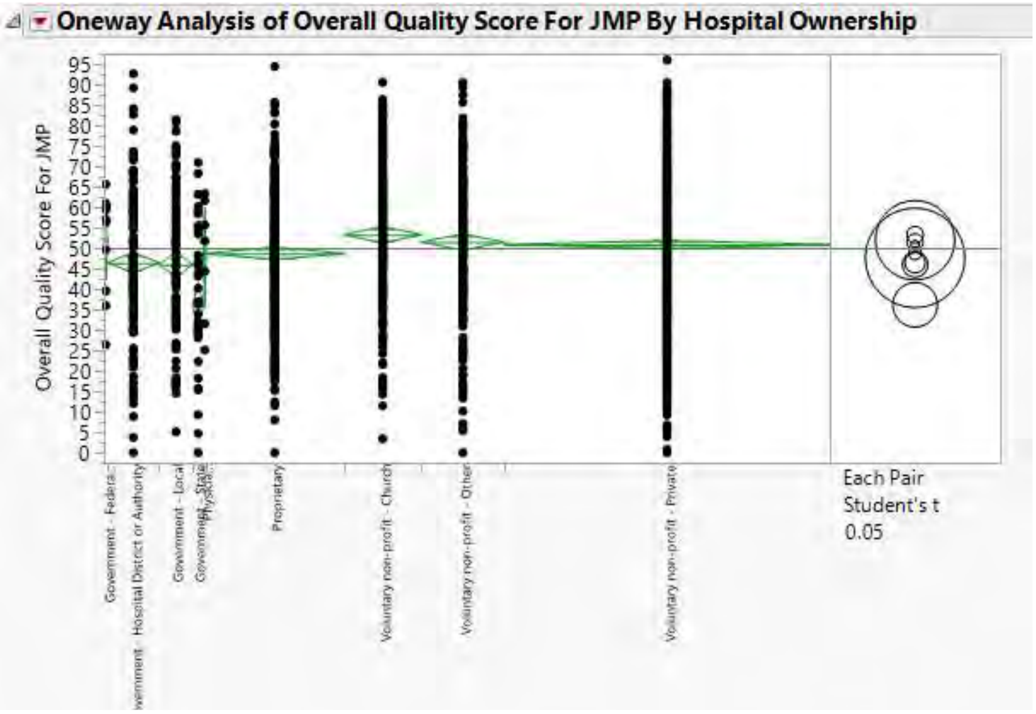
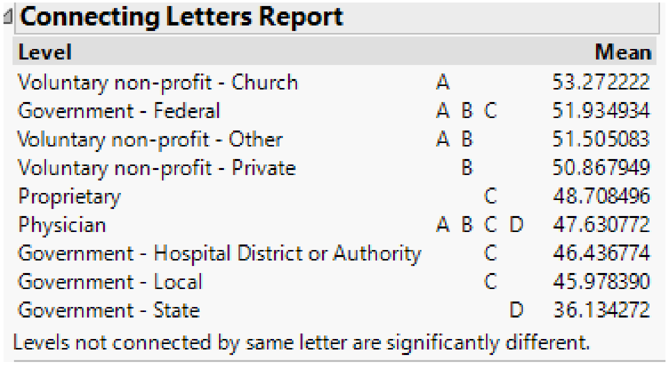


Figure 7: ANOVA of Hospital Performance by Ownership Type

profit – Private, Proprietary, Physician, Government - Hospital District or Authority, Government – Local, and Government – State (Figure 7) (Ordered Differences Report Appendix 22). Voluntary non-profit – Church hospitals were found to perform the highest on average (mean = 54) compared to Physician-owned (mean = 48) and Government Hospitals (Table 27). Voluntary non-profit – Private hospitals performed relatively moderately on average (mean = 50.8). These differences were found to be significant at the 0.0001 level.

Table 27: Average Hospital Performance Broken Down by Ownership Type



Connecting Letters Report

Level		Mean
Voluntary non-profit - Church	A	53.272222
Government - Federal	A B C	51.934934
Voluntary non-profit - Other	A B	51.505083
Voluntary non-profit - Private	B	50.867949
Proprietary	C	48.708496
Physician	A B C D	47.630772
Government - Hospital District or Authority	C	46.436774
Government - Local	C	45.978390
Government - State	D	36.134272

Levels not connected by same letter are significantly different.

Data were also available regarding whether each hospital was part of larger health system. Affiliation with a health system was not associated with significant differences in average overall hospital quality scores, but significant variation was found when looking at domain scores for Coordination and Centeredness (Table 28). Hospitals that were part of a system had higher average scores on Coordination, which is somewhat intuitive; large systems would intuitively have larger management structures, thereby improving information sharing and the establishment of standard operating procedures. However, hospitals that were part of a system were found to have lower average scores on Patient Centeredness/Satisfaction using data from HCAHPS surveys. This is also a logical finding. Larger organizations are often regarded as less intimate and more bureaucratic. This is potentially an oversimplification, but nonetheless the directionality of this relationship is in line with prevailing expectations.

Table 28 Hospital System Affiliation and Coordination and Centeredness Domain Scores:

Missing Rows 1894

Oneway Anova

Summary of Fit

Rsquare	0.03626
Adj Rsquare	0.035928
Root Mean Square Error	15.83693
Mean of Response	50.13946
Observations (or Sum Wgts)	2904

t Test

1-0

Assuming equal variances

Difference	6.37089	t Ratio	10.44919
Std Err Dif	0.60970	DF	2902
Upper CL Dif	7.56639	Prob > |t|	<.0001*
Lower CL Dif	5.17540	Prob > t	<.0001*
Confidence	0.95	Prob < t	1.0000

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Part of System	1	27384.63	27384.6	109.1855	<.0001*
Error	2902	727845.73	250.8		

Missing Rows 1254

Oneway Anova

Summary of Fit

Rsquare	0.074076
Adj Rsquare	0.073814
Root Mean Square Error	16.03976
Mean of Response	50
Observations (or Sum Wgts)	3544

t Test

1-0

Assuming equal variances

Difference	-9.135	t Ratio	-16.8335
Std Err Dif	0.543	DF	3542
Upper CL Dif	-8.071	Prob > |t|	<.0001*
Lower CL Dif	-10.199	Prob > t	1.0000
Confidence	0.95	Prob < t	<.0001*

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Part of System	1	72902.74	72902.7	283.3663	<.0001*
Error	3542	911263.93	257.3		
C. Total	3543	984166.67			

Hospitals were also classified according to their academic status, and an ANOVA analysis revealed that teaching hospitals had significantly lower average overall quality scores, Safety, Timeliness, and Centeredness scores, but actually performed significantly higher on measures of Coordination (Appendix 23). This finding conflicts with public belief that academic hospitals are of a higher quality.

Explaining this trend is somewhat challenging. One approach would be to question the validity of risk adjustment techniques employed by CMS in the making of indicator scores. Academic hospitals tend to treat sicker patients; if risk adjustment techniques are unable to account for this higher average severity, it would make sense that academic hospitals appeared to perform poorer. Yet, if these risk adjustment techniques *are* indeed sound, one potential explanation could be that academic hospitals – while highly advanced and capable of treating complex/rare diseases – might not perform as well when treating the *majority* of cases (which do not require such advanced techniques). In other words, an academic hospital might have surgeons trained in the most advanced surgical techniques with the newest technologies. But these surgeons would inherently spend less time (and have less experience) handling more routine cases. Academic hospital attending physicians also spend a large percentage of their time performing duties outside clinical practice, such as teaching and conducting research. Not to mention, a certain percentage of cases at academic hospitals are handled by new residents with virtually no years of experience. Thus, the assumption that academic hospitals are superior appears unfounded; the prestige and brand recognition of these facilities does not ensure that a higher quality product is being offered.

Some of the most interesting findings related to financial metrics. For example, bivariate regressions were conducted relating overall hospital quality to (1) market share (%), (2) Herfindahl-Hirschman Index (HHI – a measure of market competitiveness), (3) hospital bed count, (4) and hospital revenue.

These results were rather startling. To preface, the capitalistic model of the U.S. healthcare system would predict that competition improves quality and lowers cost. Therefore, hospitals in more competitive markets (a lower HHI value) would be forced to improve the quality of their services in order to maintain patient occupancy levels. This was not found to be the case. In fact, a weak (but statistically significant) linear model was fitted in the opposite direction (Appendix 27). Hospitals in highly competitive environments were found to perform poorer on average than hospitals in less competitive markets. Moreover, there was no relationship between market share and overall hospital performance (Appendix 27). This, too, conflicts with prevailing economic theory, which would predict that hospitals with a higher market share obtained greater control of the market because they offered products of higher quality. Perhaps more startling, a weak but significant *negative* relationship was found between hospital revenue and overall hospital quality; that is to say, hospitals with the highest revenues tended to offer lower quality products (Appendix 26).

These findings add further evidence to the claim that the U.S. healthcare system possesses a number of market failures that undermine the effectiveness of selective pressures on hospitals. There does not appear to be any evidence to the claim that higher quality hospitals obtain a higher market share or higher revenues as a result of their performance. Likewise, it does not appear that patients currently have the tools necessary to choose amongst competing facilities based upon quality.

In other words, none of our current assumptions hold water. The way we choose healthcare is not in line with the statistics on hospital quality. The hospitals that are rewarded are often not the best (and in some instances are visibly worse than their competitors). Such evidence calls into question the validity of our capitalistic health care model and begins to explain why the U.S. performs so lowly on international comparisons. Yet, we need further research examining the mediating variables of overall hospital quality; if only a small percentage of the variation in quality can be explained by factors such as HHI or revenue, than what other unexplored characteristics can explain the variation amongst U.S. facilities?

II. FINAL CONCLUSIONS

This research culminated in a model for rating and disseminating information on hospital quality in an online format. However, as discussed in the four chapters of this thesis, there are a limitless number of approaches that can be taken for assessing the quality of healthcare facilities.

One of the obvious initial barriers is establishing an accepted theoretical framework for healthcare quality. The Focus Groups conducted made progress in this area; however, no all-encompassing model could be extracted. Instead, what became clear was a set of tentative guidelines.

In many instances, these three focus groups had overlapping findings. For example, Focus Group Type One illustrated that trust is an important factor for consumers when deciding which sources of information to rely upon when assessing healthcare facilities. Because little data are often available, consumers often use referrals from trusted family and friends to make decisions regarding where to receive care. Likewise, Focus Group Type Three supported this finding, only from a different angle. In the discussion on website coloration, subjects strongly suggested that catchy pictures and bright colors be eliminated, for they would undermine the professional, academic, and trustworthy branding of the site. Participants explained that the user needs to be convinced that the site is a legitimate source; thus, coloration plays a role in conveying trust.

However, there were also instances where the findings between the different focus group types posited challenges. For example, Focus Group One illustrated that subjects often have varying definitions of healthcare quality. They ascribe different weights to similar domains often based upon the medical scenario’s severity or estimated treatment duration. This suggests that a feature which allows users to determine their own domain weights prior to overall-score aggregation would be of use to subjects with varying ideological preferences. Yet, in Focus Group Type Three, when this feature was tested, subjects found it mathematically confusing and therefore unrealistic to implement.

While certain results were easy to reconcile and others were challenging, the combination of the findings from each of these focus groups provides insight into how a consumer-centric healthcare rating website could be completed. Everything from defining a theoretical framework, assessing the granularity of data availability, and debating the merits of various user interfaces has been explored. Here I will now discuss relevant areas for future research.

First, Focus Group Type One showed that subjects place a greater weight on outcomes domains in high severity/high acuity care. Conversely, subjects place a greater weight on processes in low severity/chronic (or long-term) care. While one way to establish the weightings of domains for various healthcare facility types and services would be to hold individual focus groups for every possible medical scenario rated, this would be extremely costly and time consuming. Perhaps a more effective approach would be to leverage the findings from Focus Group One. That is to say, the appropriate percentage weight ascribed to outcomes or processes could be mathematically estimated by creating an index of service risk and an index of service length. Because desired consumer weights tend to vary across these variables, such indices could be used to create a sliding scale for domain weights. Such statistical approaches are beyond the scope of this text, but the development of a mathematical model that accounts for illness severity and length of care period, while imperfect, seems like a more attainable approach for estimating desired weights.

Second, Chapter Four illustrated an attempt at rating CMS-registered U.S. hospitals. The chosen methodology and website design were explained in depth and integrated the findings from Focus Groups One, Two, and Three. To test the validity of this model, later studies should have consumers assess the ease-of-use and trustworthiness of this website in comparison to existing web-based (Yelp, Hospital Compare, HealthGrades, etc.) and experience-based (word of mouth and references).

Implications for Government, Healthcare Facilities, and Insurance Companies:

The U.S. healthcare system is under pressure. With rising national expenditures and low health outcomes, U.S. citizens are increasingly looking towards the government, healthcare facilities, clinicians, and insurance companies for reform. Yet, how these needed improvements will be introduced and who is responsible for implementing them is unclear. The objective of this paper was to illustrate a path; a path for how American consumers can be empowered to exert further pressure on the healthcare industry and control the direction of healthcare improvements. While patient choice in the current system might be little more than an illusion, if realized, patients would have the ability to reward top performers and penalize facilities failing to meet ethical standards of care.

To realize this aim, a few steps must occur. For starters, the government (specifically CMS) must exert further pressure on healthcare facilities to comply and report all data annually. As it stands, many facilities fail to report on a number of crucial indicators, making quality evaluation (as I realized) often impossible. Because data are self-reported, it is possible that facilities choose only to send in metrics that present them favorably whilst hiding metrics on which they performed poorly. Such manipulation of the system can only be prevented if a strict mandate for total data compliance is put into effect. Without near-universal data availability, the vision of an informed healthcare consumer is undermined.

Compliance could be attained through a number of mechanisms; for example, financial penalties for failing to report could be put into effect. Secondly, a more thorough approach would be for an independent third party to be tasked with collecting data on-site. This is the method used by the National Surgical Quality Improvement Program (NSQIP), whereby NSQIP hires its own clinicians to work within facilities and collect data using random sampling techniques. Introducing a third party, while obviously more logistically complex and costly, is the surest way to prevent the perverse incentives for data manipulation seen under self-reporting.

Also, the government should make efforts to revise its quality domains and weightings to center upon more patient-centric definitions. Arriving at a patient-centric definition was the objective of Focus Group One, and while more research must be conducted in this area, the provided guidelines serve as a useful starting place to that effect. The idea of an agreed upon patient-centric definition of quality also applies to healthcare facilities, clinicians, and insurance companies. Until there is a universal understanding of what we are trying to achieve, it will be near-impossible for these actors to implement directed reform efforts. That means establishing not only agreed upon domains, but also agreed upon indicators to populate those domains.

Importantly, CMS is the only insurance organization currently attempting to publically rate healthcare facilities through data collection. The long-term objective should be to establish an all payer database, whereby participating insurance companies collect data on the quality their patients receive at in-network healthcare facilities and then centralize this information into a singular database. This would be extremely useful, not only to consumers (who would then have more data on which to analyze facility quality), but also to insurers – who could use these evaluations to determine where to expand coverage networks. It is in the interest of insurers to offer plans with networks that include only facilities of the highest quality; i.e. the healthier the beneficiaries, the lower the cost of covering them.

It is important to recognize the legitimate concerns that healthcare facilities might have regarding this line of research. Any attempts at rating healthcare facilities are often viewed with skepticism; healthcare is complex, rating it is hard, and facilities often question the legitimacy of these evaluations. Perhaps more concerning is the unknown – i.e. how would their facility rank if evaluated? The disruptive potential of this future work is of major concern to healthcare facilities; for example, the American Hospital Association has publically objected to (and lobbied against) CMS star ratings and the public reporting of data since its initiation in 2011.

Some of these concerns are legitimate. Currently, most ratings agencies evaluate hospitals using widely varied approaches. Yelp, LeapFrog, Consumer Reports, Health Grades, and Hospital Compare each have unique methods for realizing the same objective. These differences undermine the underlying aim of providing transparency to the consumer. If each rating system provides contrasting advice regarding quality, clarity is lost for *both* consumers and the healthcare facilities subjected to ratings. We must keep in mind that these ratings are not only going to be used by patients. The long-term objective is to incentivize facilities to initiate quality improvement programs (QIPs) along the domains/indicators included in this rating system. The domains/indicators included should therefore represent the aspects of healthcare patients care most about improving. Thus, it essential to first achieve clarity regarding what is meant by healthcare quality.

While this research provided initial answers to the questions of data granularity and data presentation, the most crucial question (how to define quality?) must be explored in further detail in future work. In addition, this research also highlights the need for further opportunities for patient education. The development of a consumer-choice healthcare ratings website is useful to that effect, but efforts should extend beyond this. Websites such as WebMD – while they include their own flaws – enable patients to better understand their illnesses. This could be useful for combatting patients' perceived competency barriers to judging healthcare quality. On a related note, patient education more generally is vital for promoting quality lifestyle and disease-management techniques.

Appendix One

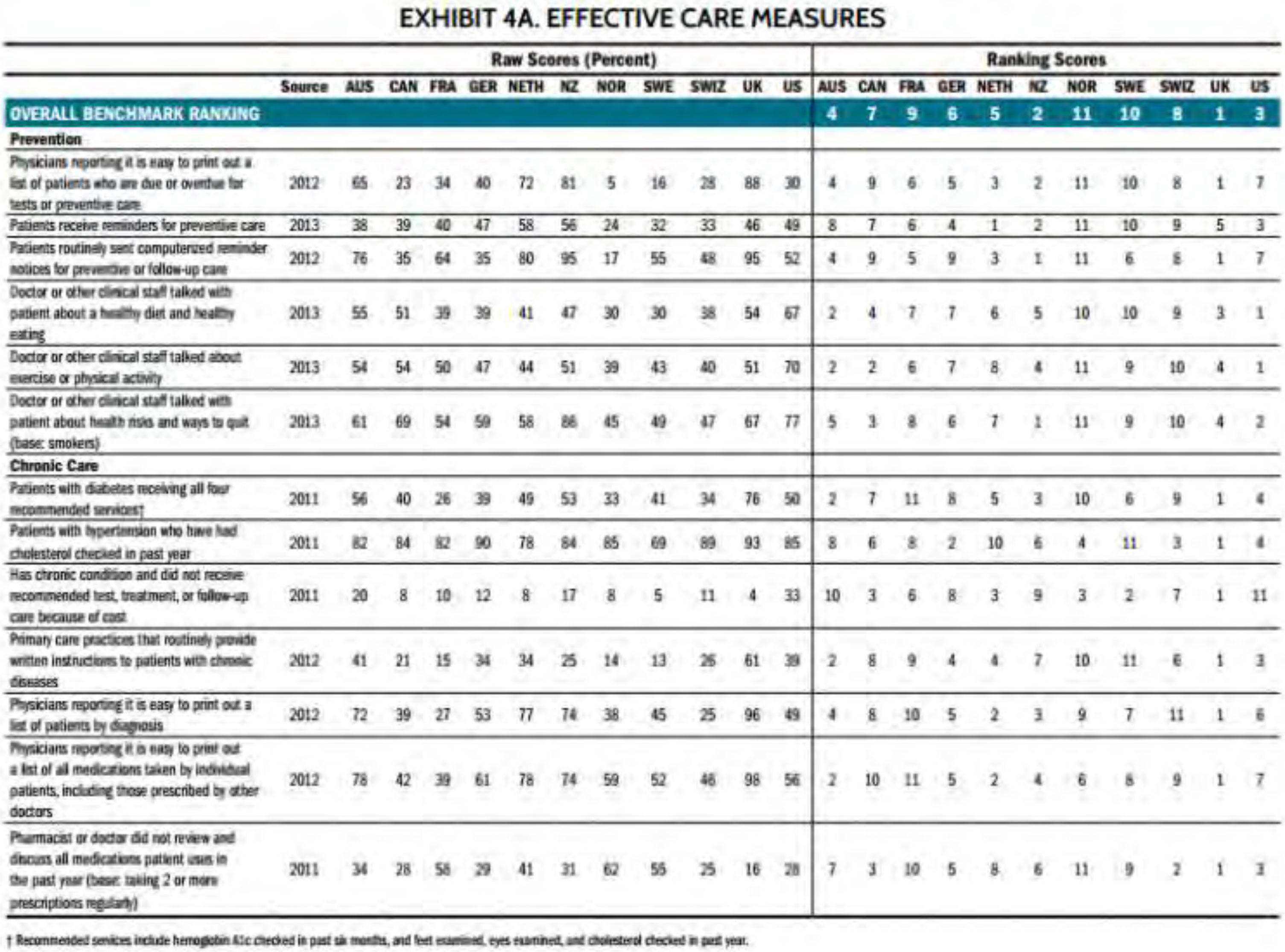


EXHIBIT 4A. EFFECTIVE CARE MEASURES

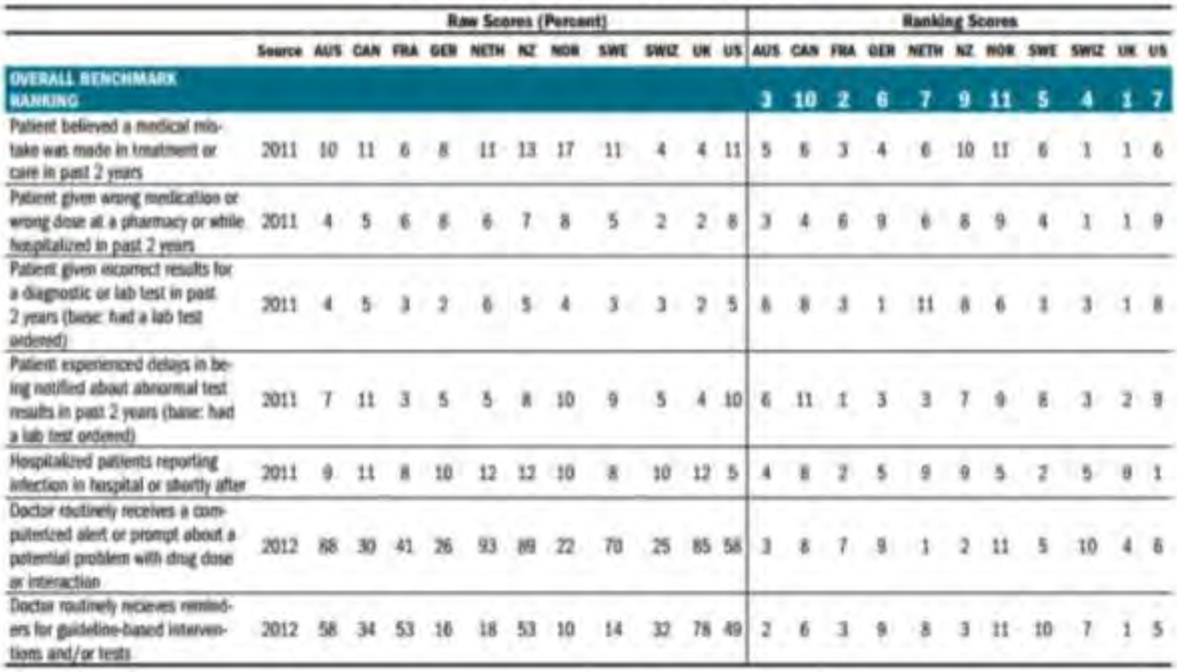
		Raw Scores (Percent)											Ranking Scores										
	Source	AUS	CAN	FRA	GER	NETH	NZ	NOR	SWE	SWIZ	UK	US	AUS	CAN	FRA	GER	NETH	NZ	NOR	SWE	SWIZ	UK	US
OVERALL BENCHMARK RANKING													**4**	**7**	**9**	**6**	**5**	**2**	**11**	**10**	**8**	**1**	**3**
Prevention																							
Physicians reporting it is easy to print out a list of patients who are due or overdue for tests or preventive care	2012	65	23	34	40	72	81	5	16	28	88	30	4	9	6	5	3	2	11	10	8	1	7
Patients receive reminders for preventive care	2013	38	39	40	47	58	56	24	32	33	46	49	8	7	6	4	1	2	11	10	9	5	3
Patients routinely sent computerized reminder notices for preventive or follow-up care	2012	76	35	64	35	80	95	17	55	48	95	52	4	9	5	9	3	1	11	6	8	1	7
Doctor or other clinical staff talked with patient about a healthy diet and healthy eating	2013	55	51	39	39	41	47	30	30	38	54	67	2	4	7	7	6	5	10	10	9	3	1
Doctor or other clinical staff talked about exercise or physical activity	2013	54	54	50	47	44	51	39	43	40	51	70	2	2	6	7	8	4	11	9	10	4	1
Doctor or other clinical staff talked with patient about health risks and ways to quit (base: smokers)	2013	61	69	54	59	58	86	45	49	47	67	77	5	3	8	6	7	1	11	9	10	4	2
Chronic Care																							
Patients with diabetes receiving all four recommended services†	2011	56	40	26	39	49	53	33	41	34	76	50	2	7	11	8	5	3	10	6	9	1	4
Patients with hypertension who have had cholesterol checked in past year	2011	82	84	82	90	78	84	85	69	89	93	85	8	6	8	2	10	6	4	11	3	1	4
Has chronic condition and did not receive recommended test, treatment, or follow-up care because of cost	2011	20	8	10	12	8	17	8	5	11	4	33	10	3	6	8	3	9	3	2	7	1	11
Primary care practices that routinely provide written instructions to patients with chronic diseases	2012	41	21	15	34	34	25	14	13	26	61	39	2	8	9	4	4	7	10	11	6	1	3
Physicians reporting it is easy to print out a list of patients by diagnosis	2012	72	39	27	53	77	74	38	45	25	96	49	4	8	10	5	2	3	9	7	11	1	6
Physicians reporting it is easy to print out a list of all medications taken by individual patients, including those prescribed by other doctors	2012	78	42	39	61	78	74	59	52	46	98	56	2	10	11	5	2	4	6	8	9	1	7
Pharmacist or doctor did not review and discuss all medications patient uses in the past year (base: taking 2 or more prescriptions regularly)	2011	34	28	58	29	41	31	62	55	25	16	28	7	3	10	5	8	6	11	9	2	1	3

† Recommended services include hemoglobin A1c checked in past six months, and feet examined, eyes examined, and cholesterol checked in past year.

Appendix 1: Commonwealth Fund Effectiveness Variables for Country Comparisons (Commonwealth Fund, 2014).

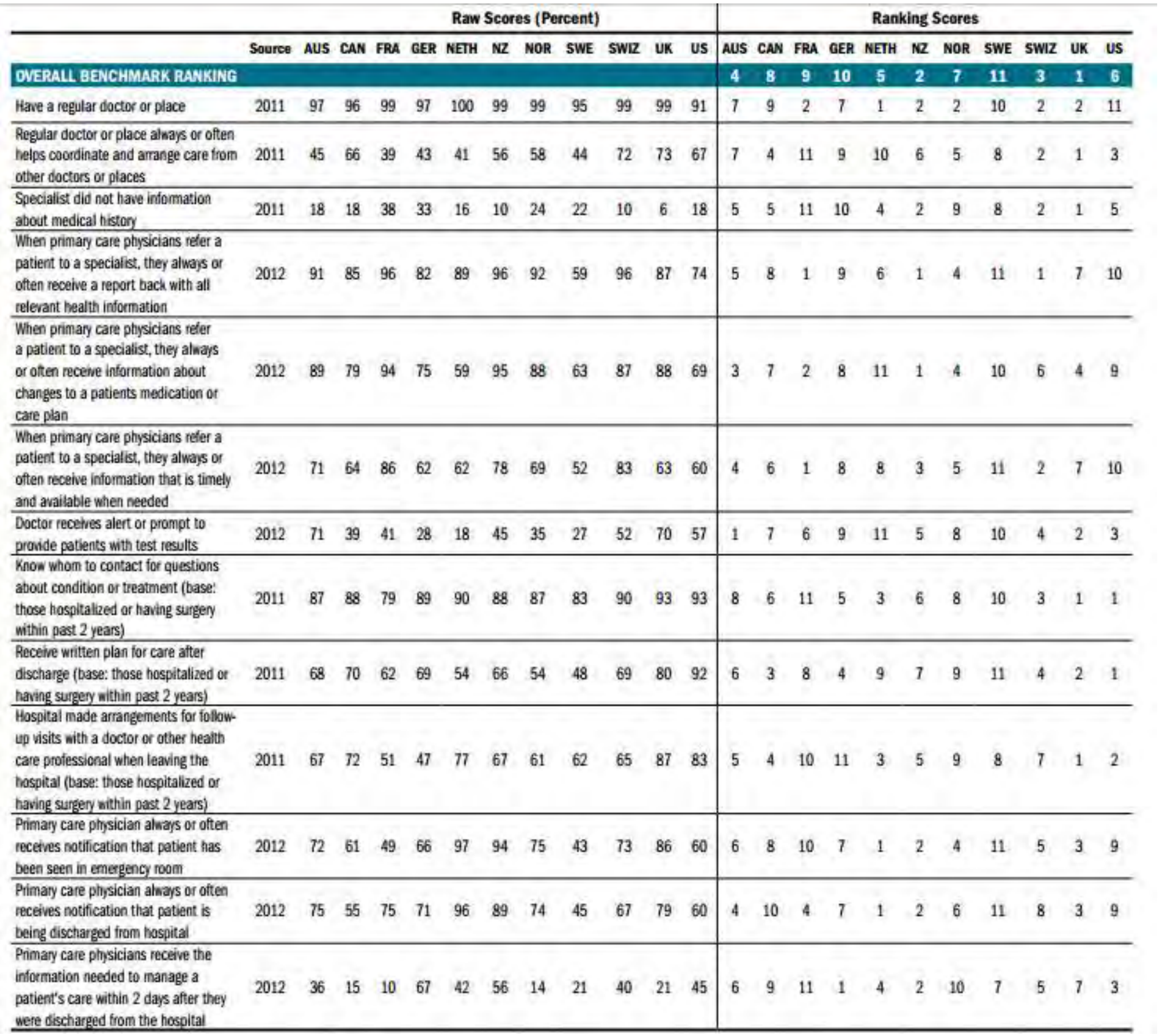
Appendix 2

EXHIBIT 4B. SAFE CARE MEASURES



		Raw Scores (Percent)											Ranking Scores										
	Source	AUS	CAN	FRA	GER	NETH	NZ	NOR	SWE	SWIZ	UK	US	AUS	CAN	FRA	GER	NETH	NZ	NOR	SWE	SWIZ	UK	US
OVERALL BENCHMARK RANKING													3	10	2	6	7	9	11	5	4	1	7
Patient believed a medical mistake was made in treatment or care in past 2 years	2011	10	11	6	8	11	13	17	11	4	4	11	5	6	3	4	6	10	11	6	1	1	6
Patient given wrong medication or wrong dose at a pharmacy or while hospitalized in past 2 years	2011	4	5	6	8	6	7	8	5	2	2	8	3	4	6	9	6	8	9	4	1	1	9
Patient given incorrect results for a diagnostic or lab test in past 2 years (base: had a lab test ordered)	2011	4	5	3	2	6	5	4	3	3	2	5	6	8	3	1	11	8	6	3	3	1	8
Patient experienced delays in being notified about abnormal test results in past 2 years (base: had a lab test ordered)	2011	7	11	3	5	5	8	10	9	5	4	10	6	11	1	3	3	7	9	8	3	2	9
Hospitalized patients reporting infection in hospital or shortly after	2011	9	11	8	10	12	12	10	8	10	12	5	4	8	2	5	9	9	5	2	5	9	1
Doctor routinely receives a computerized alert or prompt about a potential problem with drug dose or interaction	2012	68	30	41	26	93	89	22	70	25	85	58	3	8	7	9	1	2	11	5	10	4	6
Doctor routinely recieves reminders for guideline-based interventions and/or tests	2012	58	34	53	16	18	53	10	14	32	78	49	2	6	3	9	8	3	11	10	7	1	5

Appendix 3



		Raw Scores (Percent)											Ranking Scores										
	Source	AUS	CAN	FRA	GER	NETH	NZ	NOR	SWE	SWIZ	UK	US	AUS	CAN	FRA	GER	NETH	NZ	NOR	SWE	SWIZ	UK	US
OVERALL BENCHMARK RANKING													**4**	**8**	**9**	**10**	**5**	**2**	**7**	**11**	**3**	**1**	**6**
Have a regular doctor or place	2011	97	96	99	97	100	99	99	95	99	99	91	7	9	2	7	1	2	2	10	2	2	11
Regular doctor or place always or often helps coordinate and arrange care from other doctors or places	2011	45	66	39	43	41	56	58	44	72	73	67	7	4	11	9	10	6	5	8	2	1	3
Specialist did not have information about medical history	2011	18	18	38	33	16	10	24	22	10	6	18	5	5	11	10	4	2	9	8	2	1	5
When primary care physicians refer a patient to a specialist, they always or often receive a report back with all relevant health information	2012	91	85	96	82	89	96	92	59	96	87	74	5	8	1	9	6	1	4	11	1	7	10
When primary care physicians refer a patient to a specialist, they always or often receive information about changes to a patients medication or care plan	2012	89	79	94	75	59	95	88	63	87	88	69	3	7	2	8	11	1	4	10	6	4	9
When primary care physicians refer a patient to a specialist, they always or often receive information that is timely and available when needed	2012	71	64	86	62	62	78	69	52	83	63	60	4	6	1	8	8	3	5	11	2	7	10
Doctor receives alert or prompt to provide patients with test results	2012	71	39	41	28	18	45	35	27	52	70	57	1	7	6	9	11	5	8	10	4	2	3
Know whom to contact for questions about condition or treatment (base: those hospitalized or having surgery within past 2 years)	2011	87	88	79	89	90	88	87	83	90	93	93	8	6	11	5	3	6	8	10	3	1	1
Receive written plan for care after discharge (base: those hospitalized or having surgery within past 2 years)	2011	68	70	62	69	54	66	54	48	69	80	92	6	3	8	4	9	7	9	11	4	2	1
Hospital made arrangements for follow-up visits with a doctor or other health care professional when leaving the hospital (base: those hospitalized or having surgery within past 2 years)	2011	67	72	51	47	77	67	61	62	65	87	83	5	4	10	11	3	5	9	8	7	1	2
Primary care physician always or often receives notification that patient has been seen in emergency room	2012	72	61	49	66	97	94	75	43	73	86	60	6	8	10	7	1	2	4	11	5	3	9
Primary care physician always or often receives notification that patient is being discharged from hospital	2012	75	55	75	71	96	89	74	45	67	79	60	4	10	4	7	1	2	6	11	8	3	9
Primary care physicians receive the information needed to manage a patient's care within 2 days after they were discharged from the hospital	2012	36	15	10	67	42	56	14	21	40	21	45	6	9	11	1	4	2	10	7	5	7	3

Appendix 4

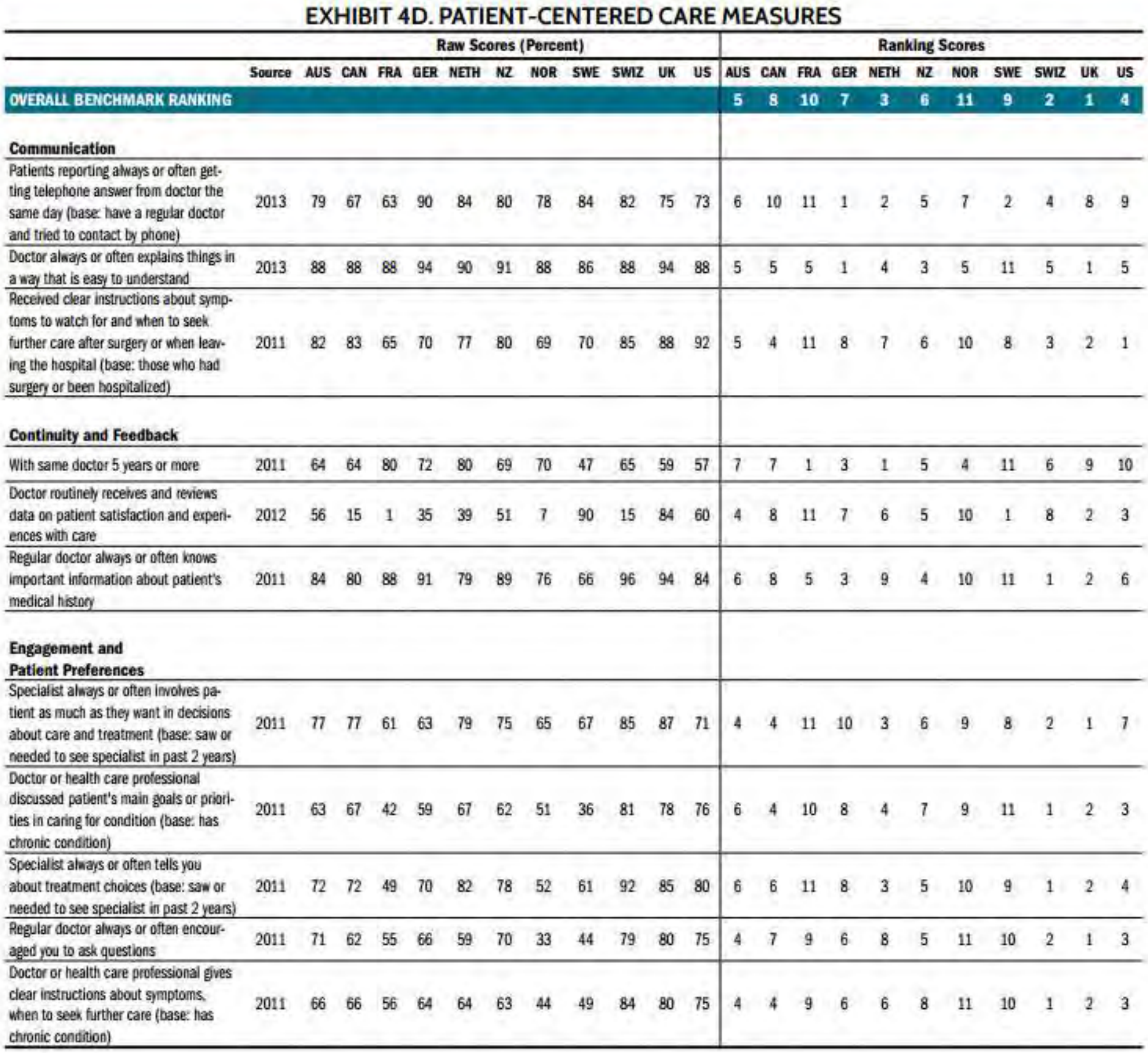


EXHIBIT 4D. PATIENT-CENTERED CARE MEASURES

		Raw Scores (Percent)											Ranking Scores										
	Source	**AUS**	**CAN**	**FRA**	**GER**	**NETH**	**NZ**	**NOR**	**SWE**	**SWIZ**	**UK**	**US**	**AUS**	**CAN**	**FRA**	**GER**	**NETH**	**NZ**	**NOR**	**SWE**	**SWIZ**	**UK**	**US**
OVERALL BENCHMARK RANKING													**5**	**8**	**10**	**7**	**3**	**6**	**11**	**9**	**2**	**1**	**4**
Communication																							
Patients reporting always or often getting telephone answer from doctor the same day (base: have a regular doctor and tried to contact by phone)	2013	79	67	63	90	84	80	78	84	82	75	73	6	10	11	1	2	5	7	2	4	8	9
Doctor always or often explains things in a way that is easy to understand	2013	88	88	88	94	90	91	88	86	88	94	88	5	5	5	1	4	3	5	11	5	1	5
Received clear instructions about symptoms to watch for and when to seek further care after surgery or when leaving the hospital (base: those who had surgery or been hospitalized)	2011	82	83	65	70	77	80	69	70	85	88	92	5	4	11	8	7	6	10	8	3	2	1
Continuity and Feedback																							
With same doctor 5 years or more	2011	64	64	80	72	80	69	70	47	65	59	57	7	7	1	3	1	5	4	11	6	9	10
Doctor routinely receives and reviews data on patient satisfaction and experiences with care	2012	56	15	1	35	39	51	7	90	15	84	60	4	8	11	7	6	5	10	1	8	2	3
Regular doctor always or often knows important information about patient's medical history	2011	84	80	88	91	79	89	76	66	96	94	84	6	8	5	3	9	4	10	11	1	2	6
Engagement and Patient Preferences																							
Specialist always or often involves patient as much as they want in decisions about care and treatment (base: saw or needed to see specialist in past 2 years)	2011	77	77	61	63	79	75	65	67	85	87	71	4	4	11	10	3	6	9	8	2	1	7
Doctor or health care professional discussed patient's main goals or priorities in caring for condition (base: has chronic condition)	2011	63	67	42	59	67	62	51	36	81	78	76	6	4	10	8	4	7	9	11	1	2	3
Specialist always or often tells you about treatment choices (base: saw or needed to see specialist in past 2 years)	2011	72	72	49	70	82	78	52	61	92	85	80	6	6	11	8	3	5	10	9	1	2	4
Regular doctor always or often encouraged you to ask questions	2011	71	62	55	66	59	70	33	44	79	80	75	4	7	9	6	8	5	11	10	2	1	3
Doctor or health care professional gives clear instructions about symptoms, when to seek further care (base: has chronic condition)	2011	66	66	56	64	64	63	44	49	84	80	75	4	4	9	6	6	8	11	10	1	2	3

Appendix 5

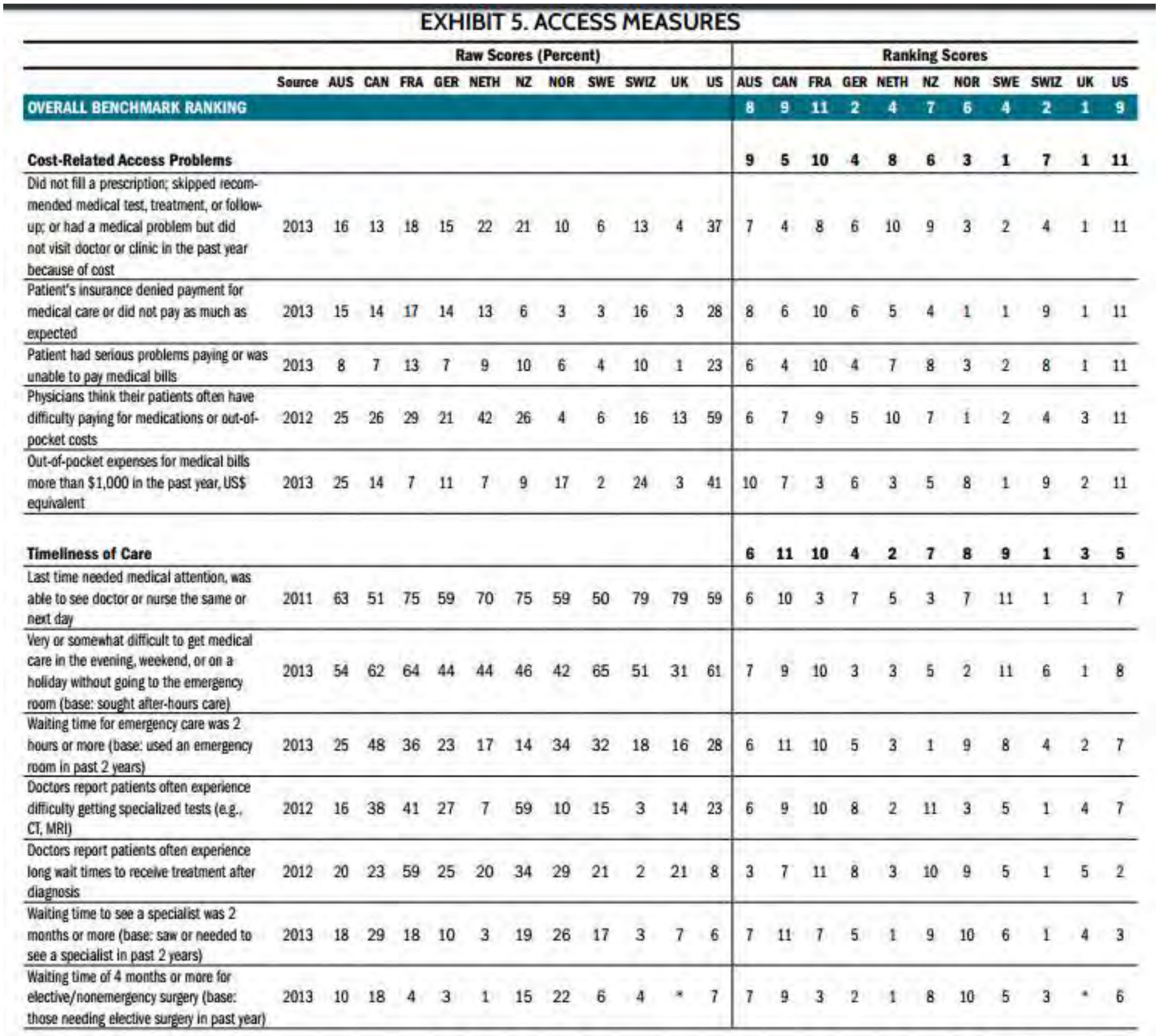


EXHIBIT 5. ACCESS MEASURES

		Raw Scores (Percent)											Ranking Scores										
	Source	AUS	CAN	FRA	GER	NETH	NZ	NOR	SWE	SWIZ	UK	US	AUS	CAN	FRA	GER	NETH	NZ	NOR	SWE	SWIZ	UK	US
OVERALL BENCHMARK RANKING													**8**	**9**	**11**	**2**	**4**	**7**	**6**	**4**	**2**	**1**	**9**
Cost-Related Access Problems													**9**	**5**	**10**	**4**	**8**	**6**	**3**	**1**	**7**	**1**	**11**
Did not fill a prescription; skipped recommended medical test, treatment, or follow-up; or had a medical problem but did not visit doctor or clinic in the past year because of cost	2013	16	13	18	15	22	21	10	6	13	4	37	7	4	8	6	10	9	3	2	4	1	11
Patient's insurance denied payment for medical care or did not pay as much as expected	2013	15	14	17	14	13	6	3	3	16	3	28	8	6	10	6	5	4	1	1	9	1	11
Patient had serious problems paying or was unable to pay medical bills	2013	8	7	13	7	9	10	6	4	10	1	23	6	4	10	4	7	8	3	2	8	1	11
Physicians think their patients often have difficulty paying for medications or out-of-pocket costs	2012	25	26	29	21	42	26	4	6	16	13	59	6	7	9	5	10	7	1	2	4	3	11
Out-of-pocket expenses for medical bills more than $1,000 in the past year, US$ equivalent	2013	25	14	7	11	7	9	17	2	24	3	41	10	7	3	6	3	5	8	1	9	2	11
Timeliness of Care													**6**	**11**	**10**	**4**	**2**	**7**	**8**	**9**	**1**	**3**	**5**
Last time needed medical attention, was able to see doctor or nurse the same or next day	2011	63	51	75	59	70	75	59	50	79	79	59	6	10	3	7	5	3	7	11	1	1	7
Very or somewhat difficult to get medical care in the evening, weekend, or on a holiday without going to the emergency room (base: sought after-hours care)	2013	54	62	64	44	44	46	42	65	51	31	61	7	9	10	3	3	5	2	11	6	1	8
Waiting time for emergency care was 2 hours or more (base: used an emergency room in past 2 years)	2013	25	48	36	23	17	14	34	32	18	16	28	6	11	10	5	3	1	9	8	4	2	7
Doctors report patients often experience difficulty getting specialized tests (e.g., CT, MRI)	2012	16	38	41	27	7	59	10	15	3	14	23	6	9	10	8	2	11	3	5	1	4	7
Doctors report patients often experience long wait times to receive treatment after diagnosis	2012	20	23	59	25	20	34	29	21	2	21	8	3	7	11	8	3	10	9	5	1	5	2
Waiting time to see a specialist was 2 months or more (base: saw or needed to see a specialist in past 2 years)	2013	18	29	18	10	3	19	26	17	3	7	6	7	11	7	5	1	9	10	6	1	4	3
Waiting time of 4 months or more for elective/nonemergency surgery (base: those needing elective surgery in past year)	2013	10	18	4	3	1	15	22	6	4	*	7	7	9	3	2	1	8	10	5	3	*	6

Appendix 6

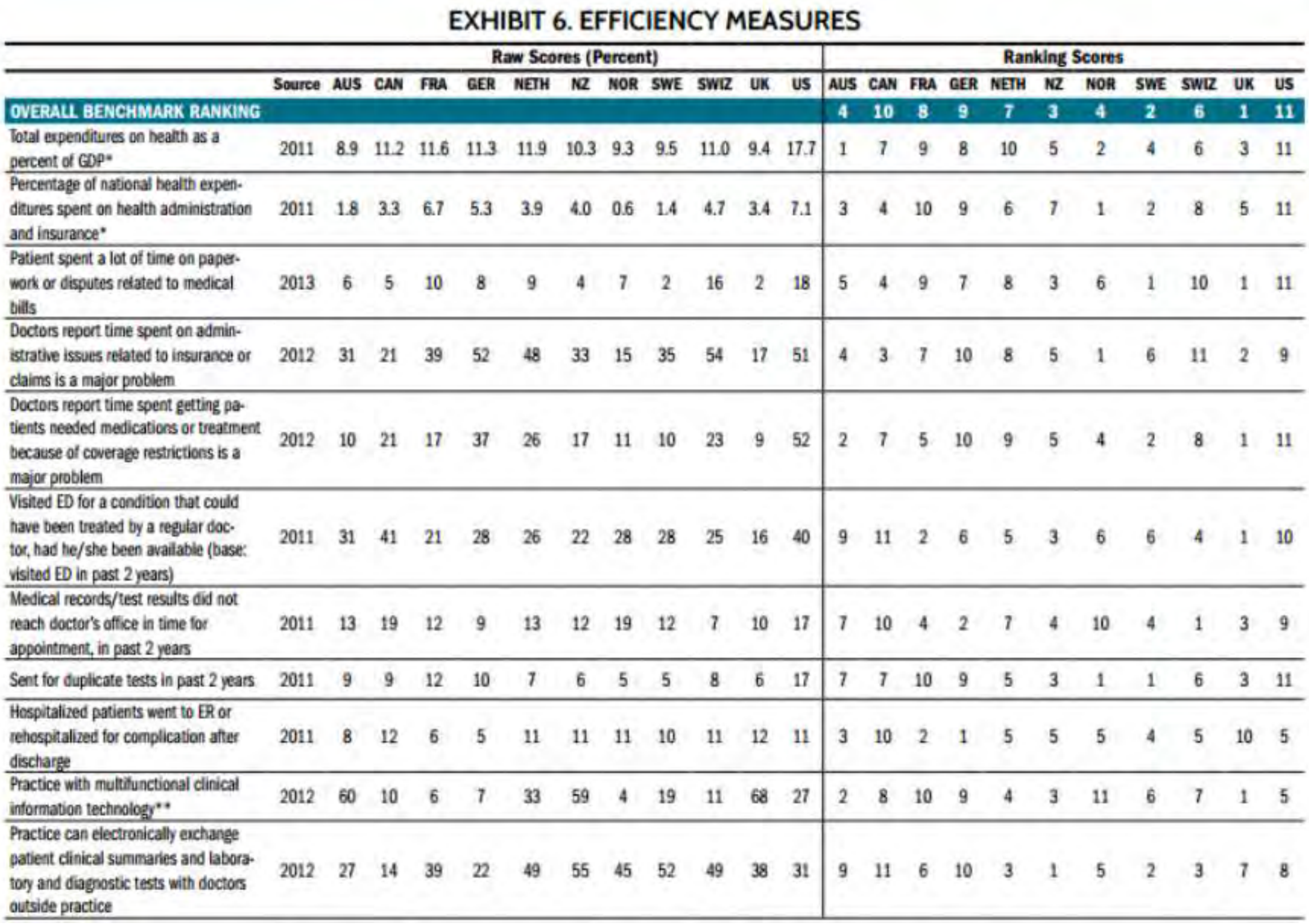


EXHIBIT 6. EFFICIENCY MEASURES

		Raw Scores (Percent)											Ranking Scores										
	Source	AUS	CAN	FRA	GER	NETH	NZ	NOR	SWE	SWIZ	UK	US	AUS	CAN	FRA	GER	NETH	NZ	NOR	SWE	SWIZ	UK	US
OVERALL BENCHMARK RANKING													4	10	8	9	7	3	4	2	6	1	11
Total expenditures on health as a percent of GDP*	2011	8.9	11.2	11.6	11.3	11.9	10.3	9.3	9.5	11.0	9.4	17.7	1	7	9	8	10	5	2	4	6	3	11
Percentage of national health expenditures spent on health administration and insurance*	2011	1.8	3.3	6.7	5.3	3.9	4.0	0.6	1.4	4.7	3.4	7.1	3	4	10	9	6	7	1	2	8	5	11
Patient spent a lot of time on paperwork or disputes related to medical bills	2013	6	5	10	8	9	4	7	2	16	2	18	5	4	9	7	8	3	6	1	10	1	11
Doctors report time spent on administrative issues related to insurance or claims is a major problem	2012	31	21	39	52	48	33	15	35	54	17	51	4	3	7	10	8	5	1	6	11	2	9
Doctors report time spent getting patients needed medications or treatment because of coverage restrictions is a major problem	2012	10	21	17	37	26	17	11	10	23	9	52	2	7	5	10	9	5	4	2	8	1	11
Visited ED for a condition that could have been treated by a regular doctor, had he/she been available (base: visited ED in past 2 years)	2011	31	41	21	28	26	22	28	28	25	16	40	9	11	2	6	5	3	6	6	4	1	10
Medical records/test results did not reach doctor's office in time for appointment, in past 2 years	2011	13	19	12	9	13	12	19	12	7	10	17	7	10	4	2	7	4	10	4	1	3	9
Sent for duplicate tests in past 2 years	2011	9	9	12	10	7	6	5	5	8	6	17	7	7	10	9	5	3	1	1	6	3	11
Hospitalized patients went to ER or rehospitalized for complication after discharge	2011	8	12	6	5	11	11	11	10	11	12	11	3	10	2	1	5	5	5	4	5	10	5
Practice with multifunctional clinical information technology**	2012	60	10	6	7	33	59	4	19	11	68	27	2	8	10	9	4	3	11	6	7	1	5
Practice can electronically exchange patient clinical summaries and laboratory and diagnostic tests with doctors outside practice	2012	27	14	39	22	49	55	45	52	49	38	31	9	11	6	10	3	1	5	2	3	7	8

Appendix 7

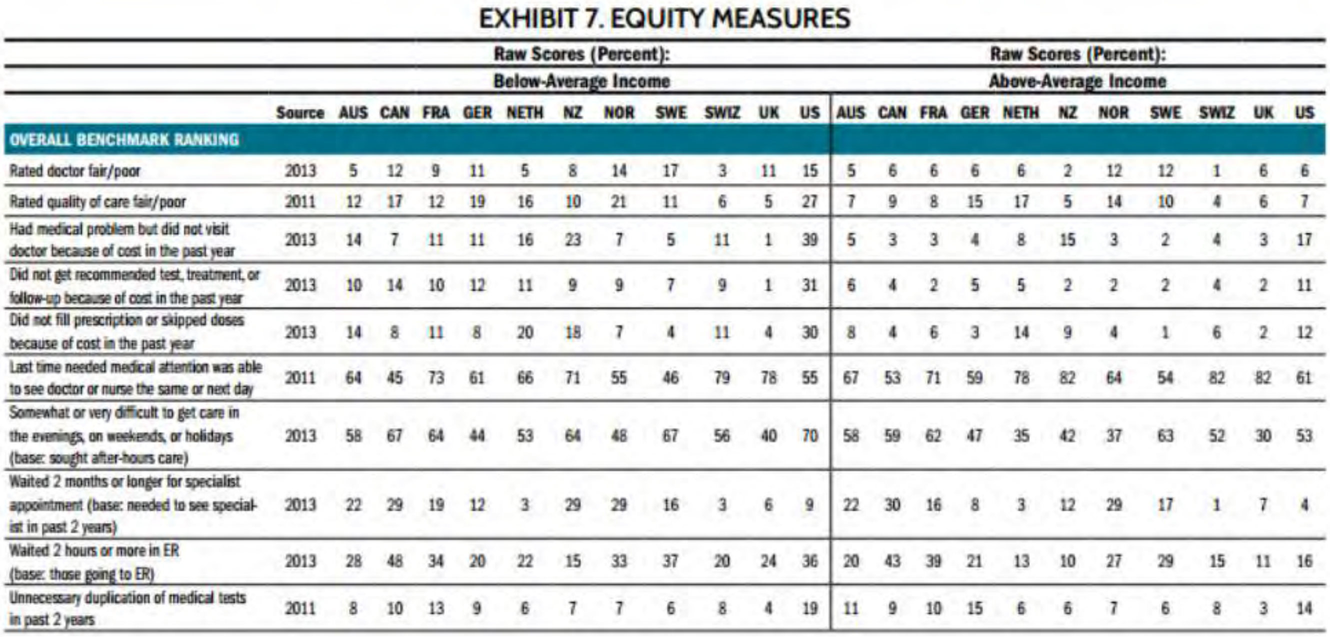


EXHIBIT 7. EQUITY MEASURES

		Raw Scores (Percent): Below-Average Income											Raw Scores (Percent): Above-Average Income										
	Source	**AUS**	**CAN**	**FRA**	**GER**	**NETH**	**NZ**	**NOR**	**SWE**	**SWIZ**	**UK**	**US**	**AUS**	**CAN**	**FRA**	**GER**	**NETH**	**NZ**	**NOR**	**SWE**	**SWIZ**	**UK**	**US**
OVERALL BENCHMARK RANKING																							
Rated doctor fair/poor	2013	5	12	9	11	5	8	14	17	3	11	15	5	6	6	6	6	2	12	12	1	6	6
Rated quality of care fair/poor	2011	12	17	12	19	16	10	21	11	6	5	27	7	9	8	15	17	5	14	10	4	6	7
Had medical problem but did not visit doctor because of cost in the past year	2013	14	7	11	11	16	23	7	5	11	1	39	5	3	3	4	8	15	3	2	4	3	17
Did not get recommended test, treatment, or follow-up because of cost in the past year	2013	10	14	10	12	11	9	9	7	9	1	31	6	4	2	5	5	2	2	2	4	2	11
Did not fill prescription or skipped doses because of cost in the past year	2013	14	8	11	8	20	18	7	4	11	4	30	8	4	6	3	14	9	4	1	6	2	12
Last time needed medical attention was able to see doctor or nurse the same or next day	2011	64	45	73	61	66	71	55	46	79	78	55	67	53	71	59	78	82	64	54	82	82	61
Somewhat or very difficult to get care in the evenings, on weekends, or holidays (base: sought after-hours care)	2013	58	67	64	44	53	64	48	67	56	40	70	58	59	62	47	35	42	37	63	52	30	53
Waited 2 months or longer for specialist appointment (base: needed to see specialist in past 2 years)	2013	22	29	19	12	3	29	29	16	3	6	9	22	30	16	8	3	12	29	17	1	7	4
Waited 2 hours or more in ER (base: those going to ER)	2013	28	48	34	20	22	15	33	37	20	24	36	20	43	39	21	13	10	27	29	15	11	16
Unnecessary duplication of medical tests in past 2 years	2011	8	10	13	9	6	7	7	6	8	4	19	11	9	10	15	6	6	7	6	8	3	14

Appendix 8

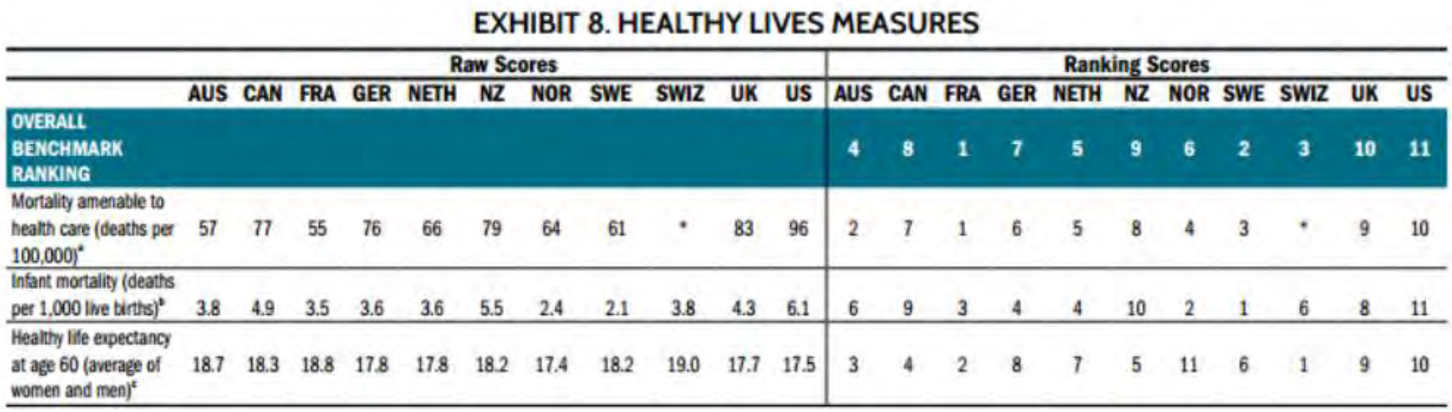
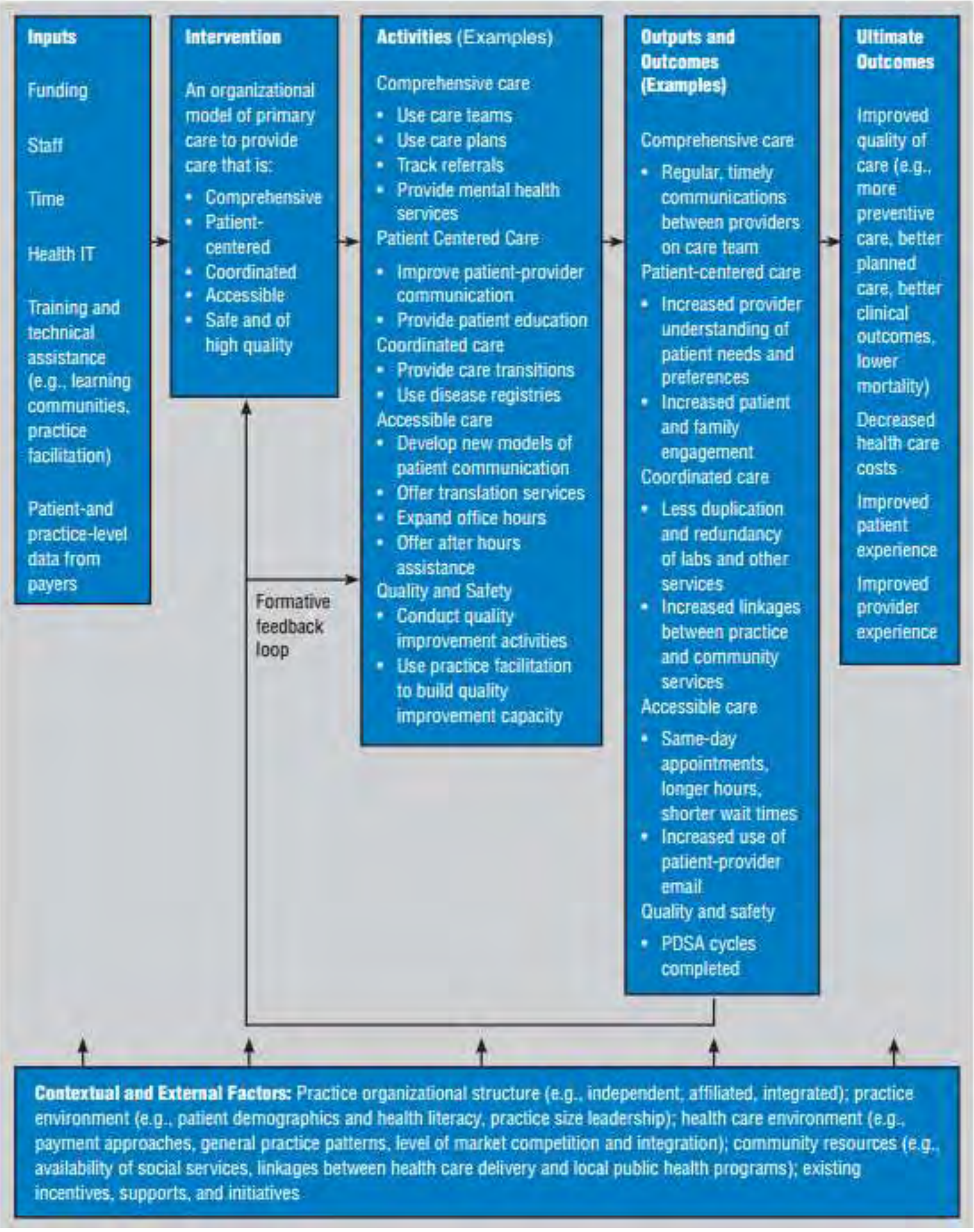


EXHIBIT 8. HEALTHY LIVES MEASURES

	Raw Scores											Ranking Scores										
	AUS	CAN	FRA	GER	NETH	NZ	NOR	SWE	SWIZ	UK	US	AUS	CAN	FRA	GER	NETH	NZ	NOR	SWE	SWIZ	UK	US
OVERALL BENCHMARK RANKING												**4**	**8**	**1**	**7**	**5**	**9**	**6**	**2**	**3**	**10**	**11**
Mortality amenable to health care (deaths per 100,000)[a]	57	77	55	76	66	79	64	61	*	83	96	2	7	1	6	5	8	4	3	*	9	10
Infant mortality (deaths per 1,000 live births)[b]	3.8	4.9	3.5	3.6	3.6	5.5	2.4	2.1	3.8	4.3	6.1	6	9	3	4	4	10	2	1	6	8	11
Healthy life expectancy at age 60 (average of women and men)[c]	18.7	18.3	18.8	17.8	17.8	18.2	17.4	18.2	19.0	17.7	17.5	3	4	2	8	7	5	11	6	1	9	10

Appendix 10



Inputs

- Funding
- Staff
- Time
- Health IT
- Training and technical assistance (e.g., learning communities, practice facilitation)
- Patient-and practice-level data from payers

Intervention

An organizational model of primary care to provide care that is:

- Comprehensive
- Patient-centered
- Coordinated
- Accessible
- Safe and of high quality

Activities (Examples)

Comprehensive care

- Use care teams
- Use care plans
- Track referrals
- Provide mental health services

Patient Centered Care

- Improve patient-provider communication
- Provide patient education

Coordinated care

- Provide care transitions
- Use disease registries

Accessible care

- Develop new models of patient communication
- Offer translation services
- Expand office hours
- Offer after hours assistance

Quality and Safety

- Conduct quality improvement activities
- Use practice facilitation to build quality improvement capacity

Outputs and Outcomes (Examples)

Comprehensive care

- Regular, timely communications between providers on care team

Patient-centered care

- Increased provider understanding of patient needs and preferences
- Increased patient and family engagement

Coordinated care

- Less duplication and redundancy of labs and other services
- Increased linkages between practice and community services

Accessible care

- Same-day appointments, longer hours, shorter wait times
- Increased use of patient-provider email

Quality and safety

- PDSA cycles completed

Ultimate Outcomes

- Improved quality of care (e.g., more preventive care, better planned care, better clinical outcomes, lower mortality)
- Decreased health care costs
- Improved patient experience
- Improved provider experience

Contextual and External Factors: Practice organizational structure (e.g., independent, affiliated, integrated); practice environment (e.g., patient demographics and health literacy, practice size leadership); health care environment (e.g., payment approaches, general practice patterns, level of market competition and integration); community resources (e.g., availability of social services, linkages between health care delivery and local public health programs); existing incentives, supports, and initiatives

Appendix 11

INFORMED CONSENT FORM

My name is Christopher Di Capua, and I am student at Union College in Schenectady, NY. I am inviting you to participate in a research study, which is part of my senior thesis in Sociology under the direction of Professor Melinda Goldner. Involvement in the study is voluntary, so you may choose to participate or not. A description of the study is written below.

I am interested in learning about consumer definitions of healthcare quality and how web-based healthcare data can be used to help patients choose amongst competing healthcare facilities. You will be asked to partake in a focus group where you will be introduced to 3 separate prototype websites. I will ask you a series of questions about the websites' displays and content to better understand patient information preferences. This will take approximately 30-40 minutes. There are no foreseeable risks to participating in this study. However, if you no longer wish to continue, you have the right to withdraw from the study, without penalty, at any time.

Your responses will be held confidential but not anonymous. This means that your name and responses will be linked in data file(s) retained by the researcher, but the researcher promises not to divulge this information. At no time will you be asked to state your name on video and you will be referred to in all written reports by a pseudonym number that can in no way be used to identify you. Following the completion of this study, all video and audio recordings will be erased to ensure the confidentiality of your responses.

Even though all aspects of the study may not be explained to you beforehand (e.g., the entire purpose of the study), during the debriefing session you will be given additional information about the study and have the opportunity to ask questions.

By signing below, you indicate that you understand the information above, and that you wish to participate in this research study.

________________________________ ____

Participant Signature Printed Name Date

You may consent to having your focus group recorded via video camera or microphone or you may decline. Please sign your initials by the appropriate statement below to indicate these wishes.

__ I consent to being recorded via video camera or microphone (please circle all that apply)
__ I do not consent to being recorded via video camera or microphone.

Demographics Sheet Questionnaire

1. Your year of birth: ____

2. Your sex. (Circle number) 1. Male 2. Female

3. Your race or ethnicity. (Circle all that apply) 1. White, not Hispanic

2. Black 3. Hispanic 4. Native American

5. Asian or Pacific Islander 6. Other (specify) ___________________

4. Your present marital status. (Circle number) 1. Never married

2. Married 3. Divorced 4. Separated 5. Widowed

5. Number of children you have in each age group. (If none, write 0)

_____ Under five years _____ 5-13 _____ 14-18 _____ 19 and over

6. Which is the highest level of education that you have completed? (Circle number)

1. No high school 2. Some High School 3. Completed High School (or equivalent)

4. Some college 5. Completed College 6. Some graduate work

7. Completed a graduate degree (specify degree) ______________________________

7. Are you presently: (Circle number)

1. Employed (specify your occupation) ____

2. Unemployed 3. Full-time homemaker 4. Student

5. Retired (specify your occupation before retirement) __________________________

Appendix 12

Focus Group One Task One:
Briefing (to be given orally):
Today we are going to discuss the idea of healthcare quality. We are trying to gain an understanding of what characteristics you think are particularly important from providers and healthcare facilities when you are seeking out care services. To start, we would like you to take 5 minutes to brainstorm the factors that you think describe a quality healthcare facility. There are no incorrect answers. Once the 5-minute period has expired, we will discuss your choices to try and abstract an extended list.

Handout:
Please list the factors that you believe describe a quality healthcare facility.

A.

B.

C.

D.

E.

F.

G.

H.

I.

J.

K.

L.

M.

N.

O.

P.

Q.

Appendix 13

Task Two:

Briefing: Now that we have discussed factors you believe influence healthcare quality, I want to introduce the idea that healthcare facilities in the United States and abroad are frequently rated. Sometimes ratings are used purely to produce aggregate descriptions of an entire nation's healthcare system. In other instances, measurements are tied directly to a facility's accreditation status or payment. A few of the government or private/not-for-profit organizations that assess healthcare quality include, the World Health Organization, the Organization for Economic Cooperation and Development, the National Quality Forum, Institute of Medicine, and the U.S. Federal Government's Agency for Healthcare Research and Quality. Below I have listed a sample of healthcare quality domains sourced from these organizations.

Handout:
Circle any number of these healthcare quality domains that you think are particularly important, place an X next to any that you do not believe should be used to assess healthcare quality, and place a "?" next to any that you believe are unclear.

Effectiveness – The extent to which planned outcomes are achieved as a result of healthcare services

Safety – Avoiding harm to patients from the care intended to help them

Coordination – The extent to which care is organized by the multiple providers overseeing a patient to facilitate the appropriate delivery of services

Patient-Centeredness – Care that is respectful of and responsive to patient preferences, needs, and values

Timeliness of Care – Reducing waits and sometimes harmful delays for those who receive care

Affordability

Efficiency – Avoiding waste, including waste of equipment, supplies, ideas, and energy

Equity – Providing care that does not vary in quality because of personal characteristics such as gender, ethnicity, geographic location, and socioeconomic status

Health Expenditures – Reducing the cost of healthcare services

Patient Satisfaction Ratings – Whether a patient is content with the healthcare they receive from a provider

Adherence to Evidence-Based Guidelines/Appropriateness of Services - The extent to which providers provide services in line with universally agreed upon clinical principles

Appendix 14

Task Three:

Briefing: Now that you have had the opportunity to brainstorm characteristics of healthcare quality and discuss some commonly utilized quality domains, you will now create a new list. This time, choose any 5 domains from the previous two lists. Imagine that you feel sick and are seeking out healthcare services. In this ideal world, you have unlimited information on every type of healthcare facility. You are deciding where to seek out health services and want to go to the facility that will provide you with the highest quality of care. Information on which 5 domains would be the most useful for you when making such a decision? Write your list below. Note: if you believe you do not need 5 domains, your list can be shorter (but no longer!).

Handout:
Most Important Domains

1.
2.
3.
4.
5.

Now, I want you to try an exercise called the Budget Allocation Process. It is designed to help you determine the relative *importance* of each domain. Assume you have $100 dollars to spend on your 5 (or shorter) domains. How much 'money' would you spend on each (note: spending more money on a domain signifies a greater level of importance). Remember, you cannot spend more than $100 in total! Write the budgeted dollar amount next to each item on your above list.

Handout:
Budget for Most Important Domains

	Domain	Budget
1.		$
2.		$
3.		$
4.		$
5.		$_____
		$100

Appendix 15

Task Four:

Briefing:

Now, it is important to recognize that these "domains" are really umbrella terms designed to describe an aspect of healthcare quality that should apply to multiple types of healthcare facilities and multiple types of conditions.

For example, the World Health Organization (WHO) includes "Safety" as a domain. It would make sense that safety is important in both nursing homes and in hospitals. Also, the WHO argues that administering safe care is important when treating any kind of condition, whether it be cancer or the common cold.

However, when using domains to assess the quality at a facility, the individual *measurements* within each domain change depending upon what kinds of services are being assessed. For example, in a nursing home, safety may be measured by looking at certain variables such as rates of resident falls or the percentage of patients that develop pressure ulcers (i.e. bed soars). However, in a hospital surgery department, safety may be measured by looking at rates of surgical infections or avoidable blood clots. Regardless, the point is that domains serve as *categories that describe a certain aspect of quality across multiple facility or treatment types* while many different measurable variables can be organized into a domain.

Domains are designed to stay constant for multiple types of healthcare services while the individual variables that make up a domain may change depending upon the type of facility or service being measured. For example, while the WHO uses the safety domain for both Nursing Homes and Hospitals, the variables included in the safety domain are different for each facility.

This brings us to the last important question: do you think your 5 chosen domains apply to all healthcare facilities? Let's answer this through a scenario. Assume you have one relative who needs a knee replacement surgery and one friend that is in need of a primary care office to treat a bad case of the flu. If you had unlimited information and were tasked with choosing healthcare facilities for your relative and friend, would the same 5 domains still apply? If not, what domains would you choose for the knee replacement and what domains would you choose for the primary care? (List below).

Second, if you decide that your current 5 domains *do* apply to both scenarios, would the relative weights (budget allocations out of $100) be different for each? If so, write the dollar amounts allocated to each domain in the two scenarios below.

Handout:

Domains for Knee Replacement	Budget
1	
2	
3	
4	
5	

Domains for Primary Care	Budget
1	
2	
3	
4	
5	

Appendix 16

Focus Group Type Two: Assessing Granularity and Data Presentation

Briefing:

Healthcare facilities in the United States and abroad are frequently rated. Sometimes ratings are used purely to produce aggregate descriptions of an entire nation's healthcare system. In other instances, measurements are tied directly to a facility's accreditation status or payment. A few of the government or private/not-for-profit organizations that assess healthcare quality include, the World Health Organization, the Organization for Economic Cooperation and Development, the National Quality Forum, Institute of Medicine, and the U.S. Federal Government's Agency for Healthcare Research and Quality.

When assessing healthcare quality, traditional measurement systems organize variables into what are called domains. For example, the World Health Organization's (WHO) Quality Framework includes 6 Domains: (1) effectiveness, (2) efficiency, (3) accessibility, (4) acceptable/patient-centeredness, (5) equitability, and (6) safety.

Domains are really umbrella terms designed to describe an aspect of healthcare quality that should apply to multiple types of healthcare facilities. Within each domain, a variety of indicators can be used and aggregated to create a de facto measurement score for that domain. For example, let's assume we are discussing the WHO's "safety" domain for nursing home services. While the domain may be safety, two of the *indicators* included in the domain may be (1) rates of resident falls and (2) the percentage of patients that develop pressure ulcers (i.e. bed soars). By aggregating measurements from a variety of indicators, domain 'scores' can be established.

In my thesis project, me and a few peers from Union College are using government data to construct a public website that patients can use to assess the quality of different healthcare facilities when trying to find one that meets their needs. Today, we will be looking through one mock design for this website, which I will show you on the above monitor.

Task One:

There is one principal objective of this focus group; that is to determine how to handle displaying information on *indicators.*

In the above prototype, the domains for hospital care have all been the same. Likewise, the same indicators have been categorized into each domain. Thus, the hospital ratings would be identical for all prototype versions. However, we are now trying to decide what information to provide on the indicators that *make up* each domain.

Version One: All indicators are listed and are organized by domain

Version Two: All indicators are listed, are organized by domain, and include raw measurement score

Version Three: All indicators are listed, are organized by domain, and include raw score as well as a 0-
10 standard score

Which version do you prefer and why?

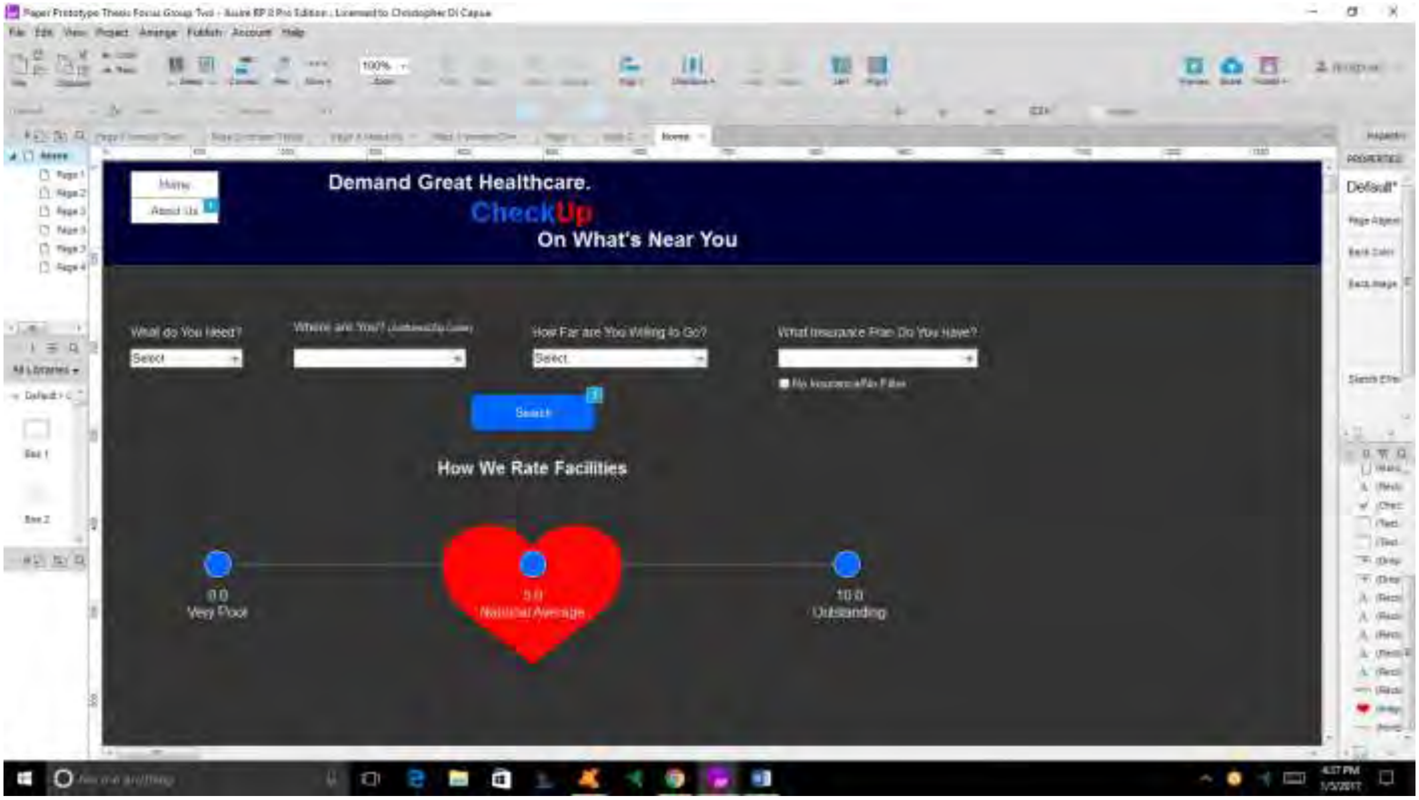
Researcher's guiding questions (not to be handed out to subjects but worked into discussion if appropriate):

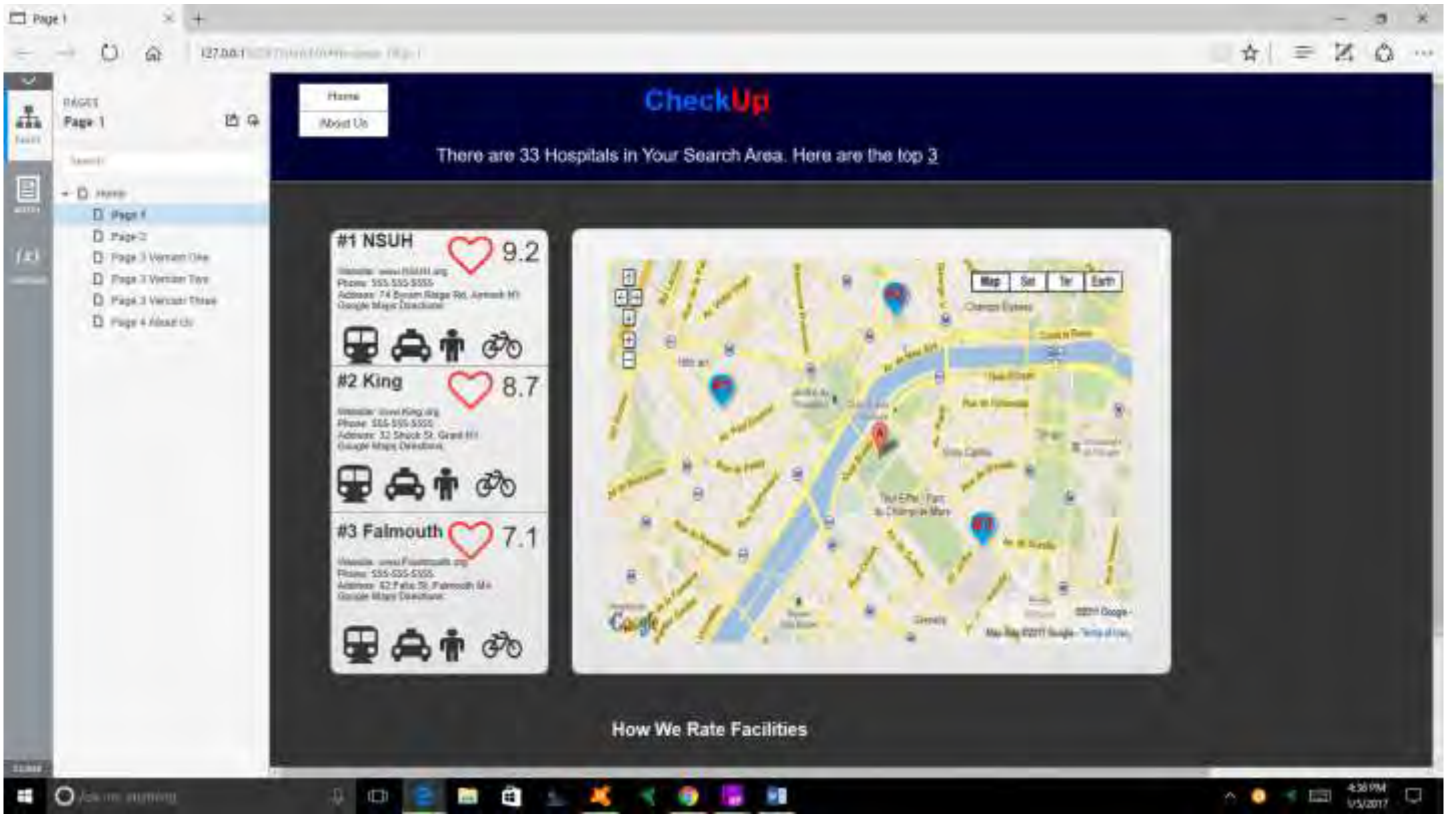
1. Are the concepts of a raw score and a standardized score clear? Should this be defined on this page?
2. Do you think that a raw measurement score could be misleading? Why or why not?
3. Do you think that listing indicators without providing scores withholds potentially valuable information?
4. Are there any downsides to providing so many scores?
5. Is there any other information you think would be useful to provide for each indicator?
    a. IF subjects can't think of any, suggest: (1) number of cases that make up the raw score (i.e. sample size) or (2) date collected

Debriefing:

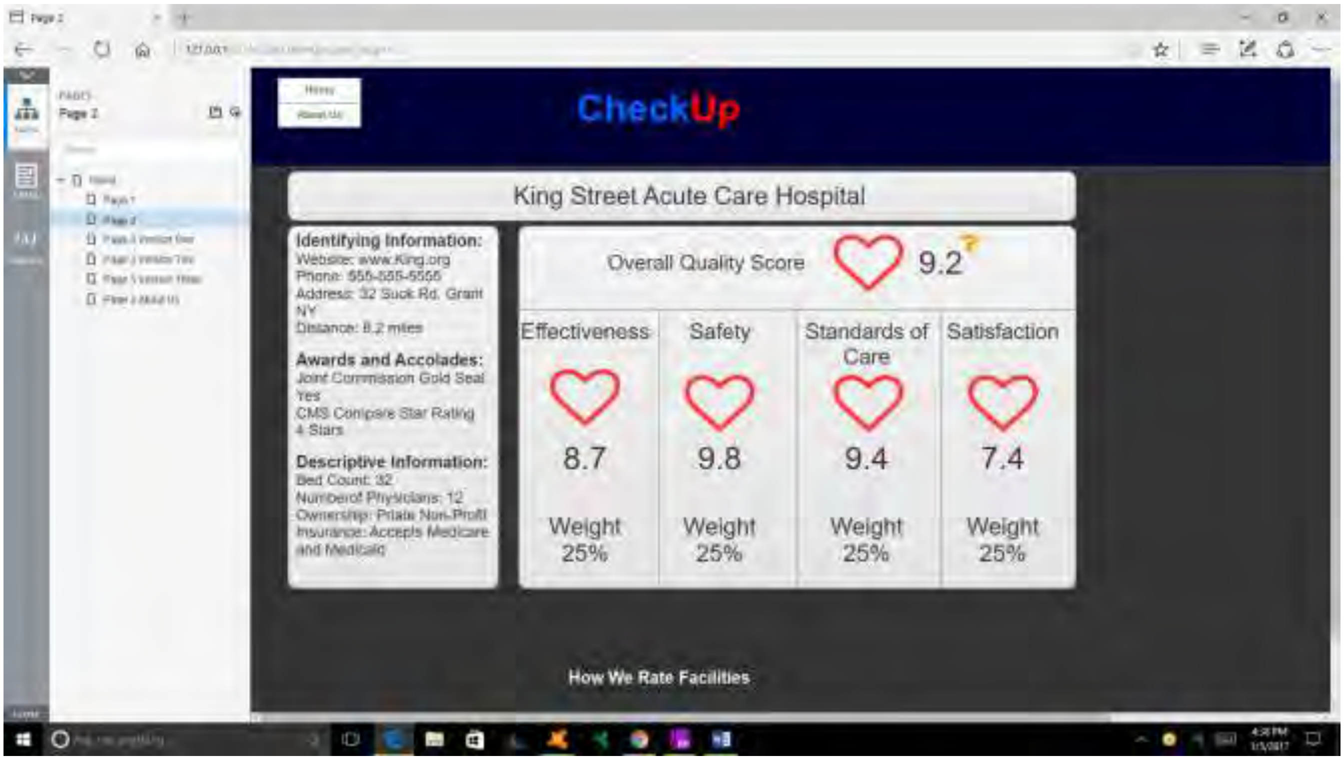
As explained, the purpose of my thesis project is to design and construct a website that patients can use to assess the quality of healthcare facilities when choosing among competing locations. However, for the website to be of use, the information therein must be presented in such a way that is both thorough and understandable. In this focus group, we looked at multiple methods for presenting information on the indicators used to rate facilities. This information will be used to influence the eventual website design. Thank you for all of your input!

AXURE PROTOTYPE MODELS OF WEBSITE FOR FOCUS GROUP #2 (on next page)





There are 33 Hospitals in Your Search Area. Here are the top 3



CheckUp

King Street Acute Care Hospital

Identifying Information:
Website: www.King.org
Phone: 555-555-5555
Address: 32 Suck Rd. Grant NY
Distance: 8.2 miles

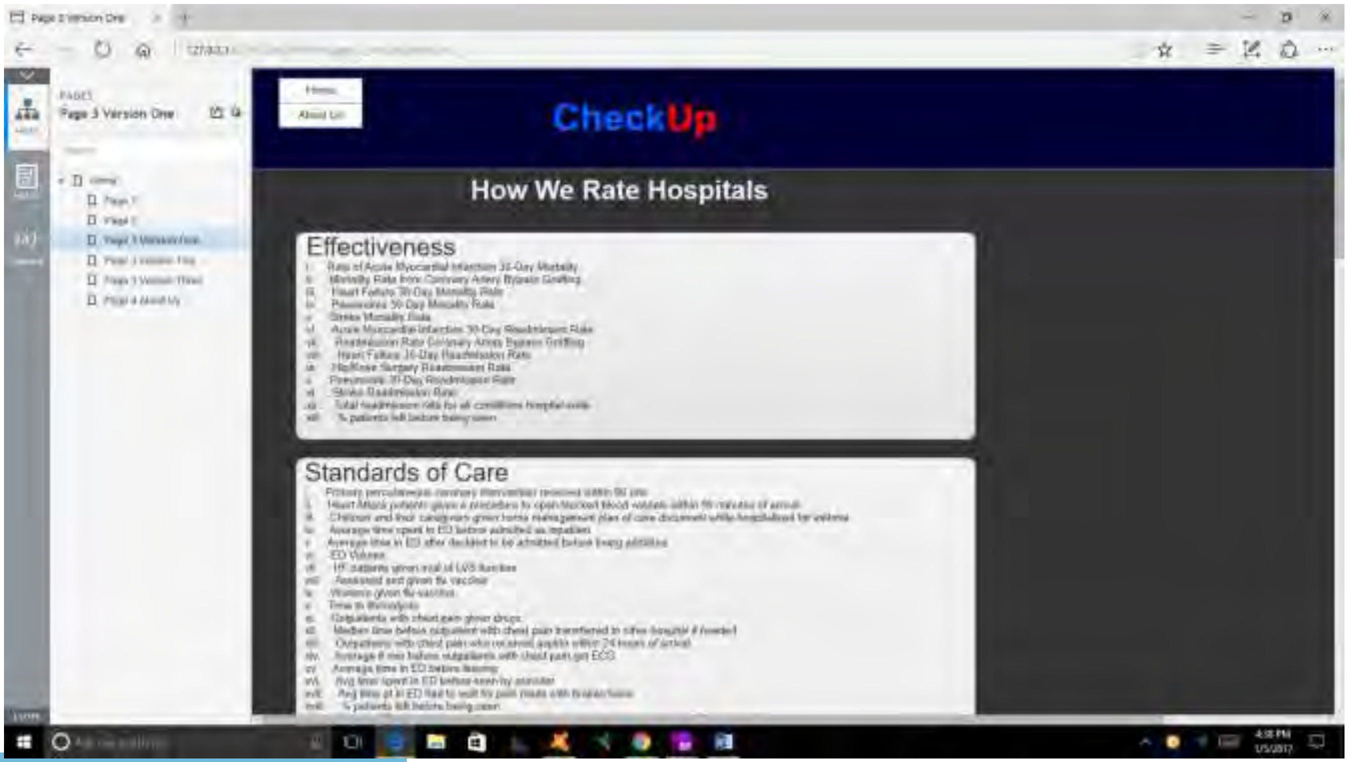
Awards and Accolades:
Joint Commission Gold Seal Yes
CMS Compare Star Rating 4 Stars

Descriptive Information:
Bed Count: 32
Numberof Physicians: 12
Ownership: Priate Non-Profit
Insurance: Accepts Medicare and Medicaid

Overall Quality Score 9.2

Effectiveness	Safety	Standards of Care	Satisfaction
8.7	9.8	9.4	7.4
Weight 25%	Weight 25%	Weight 25%	Weight 25%

How We Rate Facilities

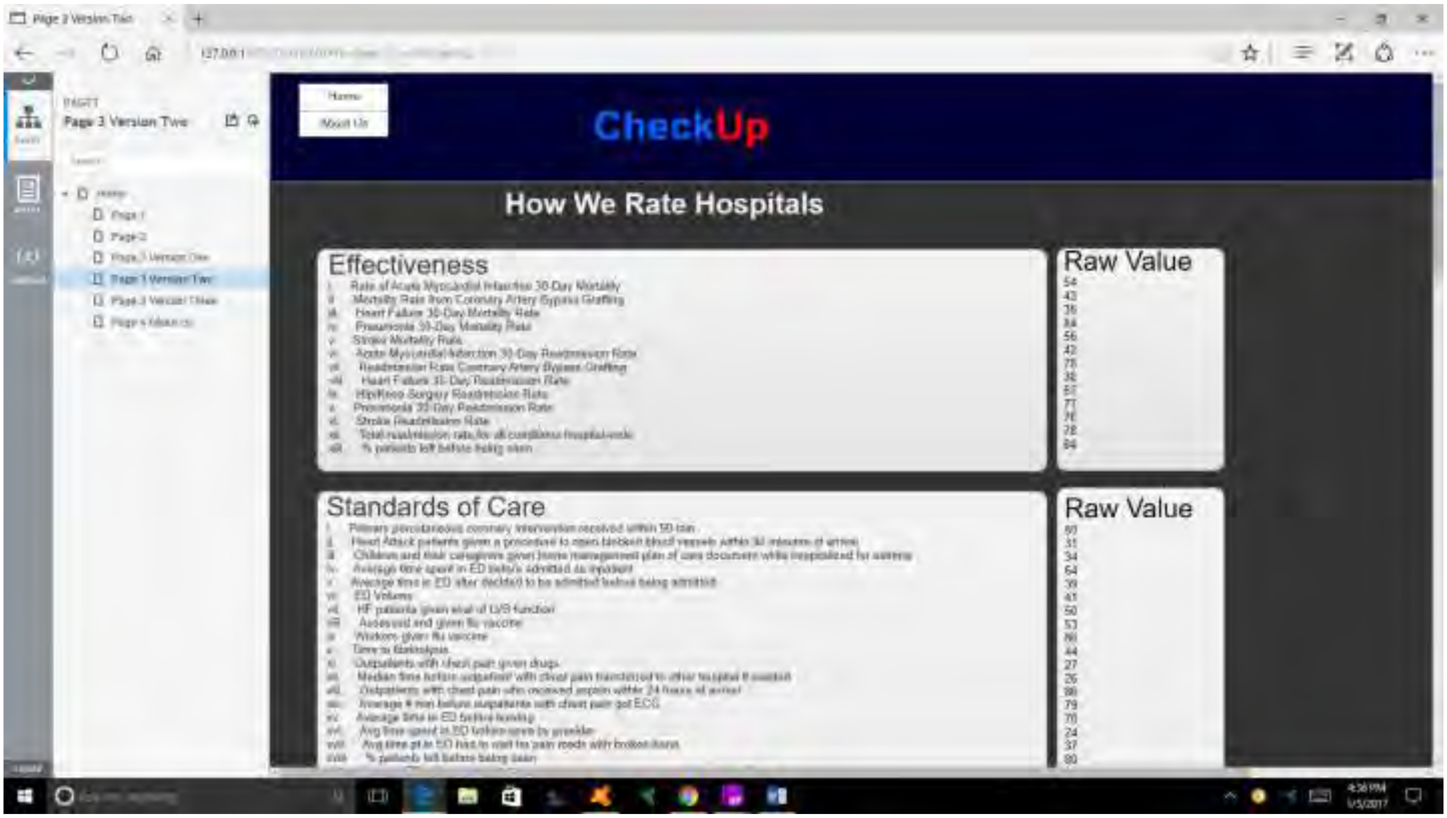


CheckUp

How We Rate Hospitals

Effectiveness
i. Rate of Acute Myocardial Infarction 30-Day Mortality
ii. Mortality Rate from Coronary Artery Bypass Grafting
iii. Heart Failure 30-Day Mortality Rate
iv. Pneumonia 30-Day Mortality Rate
v. Stroke Mortality Rate
vi. Acute Myocardial Infarction 30-Day Readmission Rate
vii. Readmission Rate Coronary Artery Bypass Grafting
viii. Heart Failure 30-Day Readmission Rate
ix. Hip/Knee Surgery Readmission Rate
x. Pneumonia 30-Day Readmission Rate
xi. Stroke Readmission Rate
xii. Total readmission rate for all conditions hospital-wide
xiii. % patients left before being seen

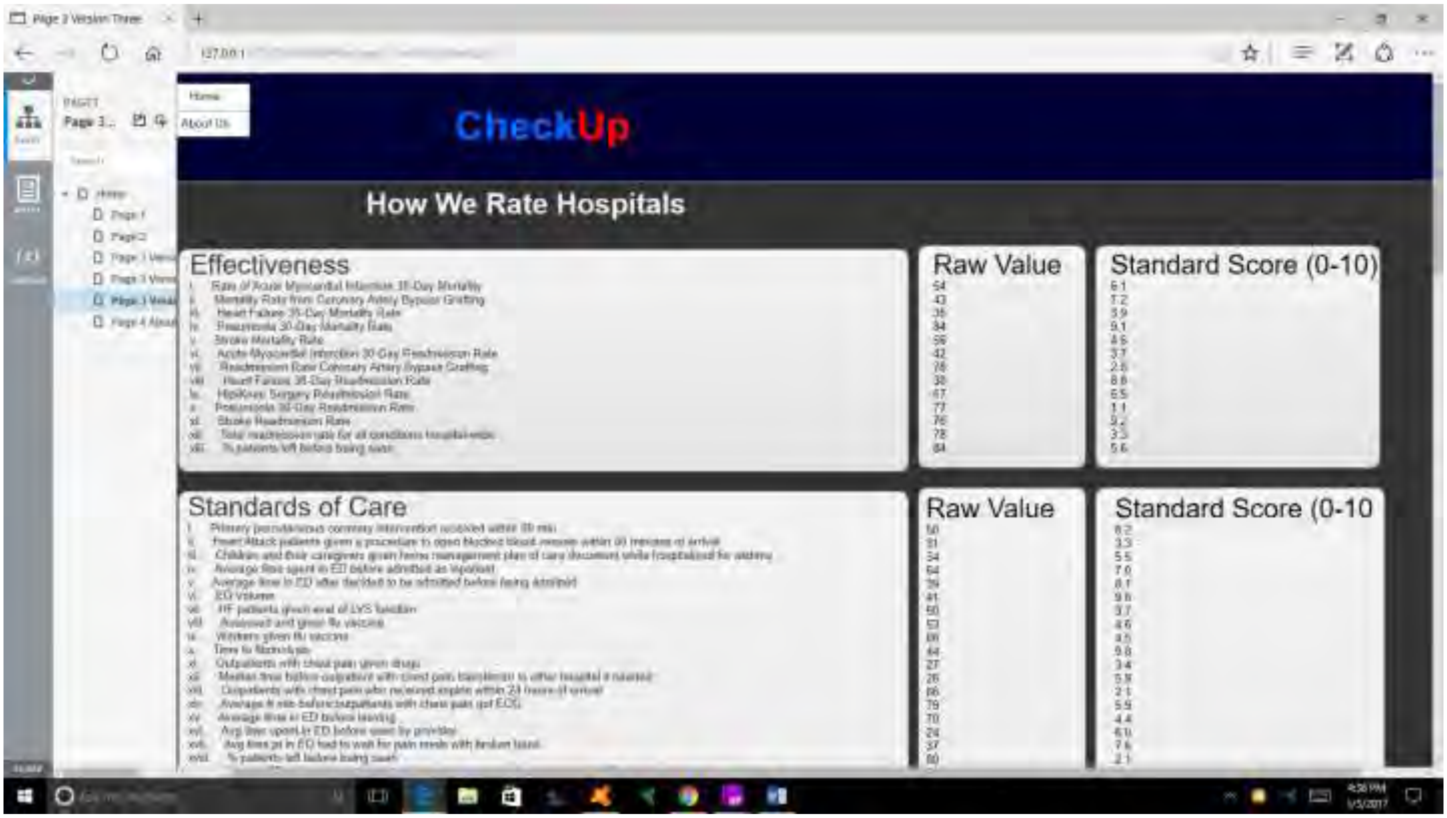
Standards of Care
i. Primary percutaneous coronary intervention received within 90 min
ii. Heart Attack patients given a procedure to open blocked blood vessels within 90 minutes of arrival
iii. Children and their caregivers given home management plan of care document while hospitalized for asthma
iv. Average time spent in ED before admitted as inpatient
v. Average time in ED after decided to be admitted before being admitted
vi. ED Volume
vii. HF patients given eval of LVS function
viii. Assessed and given flu vaccine
ix. Workers given flu vaccine
x. Time to thrombolysis
xi. Outpatients with chest pain given drugs
xii. Median time before outpatient with chest pain transferred to other hospital if needed
xiii. Outpatients with chest pain who received aspirin within 24 hours of arrival
xiv. Average # min before outpatients with chest pain got ECG
xv. Average time in ED before leaving
xvi. Avg time spent in ED before seen by provider
xvii. Avg time pt in ED had to wait for pain meds with broken bone
xviii. % patients left before being seen



How We Rate Hospitals

Effectiveness	Raw Value
i. Rate of Acute Myocardial Infarction 30-Day Mortality	54
ii. Mortality Rate from Coronary Artery Bypass Grafting	43
iii. Heart Failure 30-Day Mortality Rate	36
iv. Pneumonia 30-Day Mortality Rate	84
v. Stroke Mortality Rate	56
vi. Acute Myocardial Infarction 30-Day Readmission Rate	42
vii. Readmission Rate Coronary Artery Bypass Grafting	78
viii. Heart Failure 30-Day Readmission Rate	38
ix. Hip/Knee Surgery Readmission Rate	67
x. Pneumonia 30-Day Readmission Rate	77
xi. Stroke Readmission Rate	76
xii. Total readmission rate for all conditions hospital-wide	78
xiii. % patients left before being seen	84

Standards of Care	Raw Value
i. Primary percutaneous coronary intervention received within 90 min	60
ii. Heart Attack patients given a procedure to open blocked blood vessels within 30 minutes of arrival	31
iii. Children and their caregivers given home management plan of care document while hospitalized for asthma	34
iv. Average time spent in ED before admitted as inpatient	64
v. Average time in ED after decided to be admitted before being admitted	39
vi. ED Volume	41
vii. HF patients given eval of LVS function	50
viii. Assessed and given flu vaccine	53
ix. Workers given flu vaccine	86
x. Time to fibrinolysis	44
xi. Outpatients with chest pain given drugs	27
xii. Median time before outpatient with chest pain transferred to other hospital if needed	26
xiii. Outpatients with chest pain who received aspirin within 24 hours of arrival	86
xiv. Average # min before outpatients with chest pain got ECG	79
xv. Average time in ED before leaving	70
xvi. Avg time spent in ED before seen by provider	24
xvii. Avg time pt in ED had to wait for pain meds with broken bone	37
xviii. % patients left before being seen	80



How We Rate Hospitals

Effectiveness	Raw Value	Standard Score (0-10)
i. Rate of Acute Myocardial Infarction 30-Day Mortality	54	6.1
ii. Mortality Rate from Coronary Artery Bypass Grafting	43	7.2
iii. Heart Failure 30-Day Mortality Rate	36	3.9
iv. Pneumonia 30-Day Mortality Rate	84	9.1
v. Stroke Mortality Rate	56	4.6
vi. Acute Myocardial Infarction 30-Day Readmission Rate	42	3.7
vii. Readmission Rate Coronary Artery Bypass Grafting	78	2.8
viii. Heart Failure 30-Day Readmission Rate	38	8.8
ix. Hip/Knee Surgery Readmission Rate	67	6.5
x. Pneumonia 30-Day Readmission Rate	77	1.1
xi. Stroke Readmission Rate	76	9.2
xii. Total readmission rate for all conditions hospital-wide	78	3.3
xiii. % patients left before being seen	84	5.6

Standards of Care	Raw Value	Standard Score (0-10
i. Primary percutaneous coronary intervention received within 90 min	50	8.2
ii. Heart Attack patients given a procedure to open blocked blood vessels within 30 minutes of arrival	31	3.3
iii. Children and their caregivers given home management plan of care document while hospitalized for asthma	34	5.5
iv. Average time spent in ED before admitted as inpatient	64	7.0
v. Average time in ED after decided to be admitted before being admitted	39	8.1
vi. ED Volume	41	9.6
vii. HF patients given eval of LVS function	50	3.7
viii. Assessed and given flu vaccine	53	4.6
ix. Workers given flu vaccine	86	4.5
x. Time to fibrinolysis	44	9.8
xi. Outpatients with chest pain given drugs	27	3.4
xii. Median time before outpatient with chest pain transferred to other hospital if needed	26	5.8
xiii. Outpatients with chest pain who received aspirin within 24 hours of arrival	86	2.1
xiv. Average # min before outpatients with chest pain got ECG	79	5.9
xv. Average time in ED before leaving	70	4.4
xvi. Avg time spent in ED before seen by provider	24	6.0
xvii. Avg time pt in ED had to wait for pain meds with broken bone	37	7.6
xviii. % patients left before being seen	80	2.1

Appendix 17

Focus Group Three: Alternative User Interface Designs

Briefing:

Healthcare facilities in the United States and abroad are frequently rated. Sometimes ratings are used purely to produce aggregate descriptions of an entire nation's healthcare system. In other instances, measurements are tied directly to a facility's accreditation status or payment. A few of the government or private/not-for-profit organizations that assess healthcare quality include, the World Health Organization, the Organization for Economic Cooperation and Development, the National Quality Forum, Institute of Medicine, and the U.S. Federal Government's Agency for Healthcare Research and Quality.

In my thesis project, me and a few peers from Union College are using government data to construct a public website that patients can use to assess the quality of different healthcare facilities when trying to find one that meets their needs. Today, we will be looking through a few mock designs for this website.

Task One:

Let's start with a basic model. Look at the above monitor while I walk you through a design for a fake healthcare facility rating website.

Task Two:

That website had five pages: (1) the Search Page, where the user input their location, facility type, etc. (2) the List Page, where facilities in the search area were listed and organized by rating, (3) the Facility Page, where more information was provided on an individual facility, (4) an Indicators Page, where the indicators that were used to construct the facilities' scores were listed, and (5) an About Us Page, which described the purpose of the website.

We will now go through a few different versions of these pages and will discuss different design options. First, let's take a look at five versions of page 1, the Search Page. The versions are shown on the above monitor.

All of the pages have the exact same content; however, each is a different color.

Researcher Guiding Questions (not to be handed out to participants)

1. Which color scheme do you prefer? Do you think color matters when using a website? Are there any special considerations to keep in mind when choosing a color scheme for a healthcare website?
2. How do you feel about the picture included in the search page? Did it make the page appear more appealing? Was it distracting?
3. What is your opinion of using check boxes or dropdown menus to organize the search page?

Task Three:

There are multiple scales that could be used to rate healthcare facilities. Here are four different versions on the above monitor.

Version One uses a 0-10 point scale, with 5 being the national average, 0 being extremely poor and 10 being outstanding.

Version Two uses a 0-100 point scale, with 50 being the national average, 0 being extremely poor and 100 being outstanding.

Version Three uses an A-F scale, like in school classes, with an F being extremely poor and an A being outstanding.

Version Four uses a 0-5 stars scale, with 0 stars being extremely poor and 5 stars being outstanding.

Researcher Guiding Questions (not to be handed out to participants)

1. Which of these scales do you prefer? Do you find any misleading? Intuitive?

Task Four:

Page Two is the List Page. That is, it lists out the facilities within the search area. Two versions are shown here. Version One includes a map that shows your current location and the location of each facility labeled by rank number. Version Two excludes the map, instead choosing to merely list the travel distance next to each facility's name and score.

Researcher Guiding Questions (not to be handed out to participants)

1. Which design do you prefer and why?
2. Do you think facilities should be ordered by rank or by distance?

Task Five:

Here are another two versions of Page Two, the List Page. Looking back at the original base design, remember that a facility's overall score is determined through a weighted average of the domain scores. For example, in the base design, Pretend Hospital received an overall score of 8.5 out of 10. In the four domains, its safety score was 8.0, its effectiveness score was 9.0, its satisfaction score was 9.4, and its affordability score was 7.6. The overall score of 8.5 was calculated by averaging the four domain scores.

However, in a weighted average, the weight (i.e. importance) of each domain can be modified. If safety was given a weight of 10%, effectiveness a weight of 40%, satisfaction a weight of 30%, and affordability a score of 10%, then the overall score would end up being 8.75. In other words, the weight of each domain affects the overall score.

In Version One of Page Two, the List Page, the weights of each domain are set in stone at 25% for each of the four. But, in Version Two, while the default is 25% each, the user is given the ability to change the weights of each domain as long as the total weight adds up to 100%. The page would then re-load, this time organizing the facilities according to their new rating.

Do you think this is a useful feature? Is it confusing? Which do you prefer and why?

Task Six:

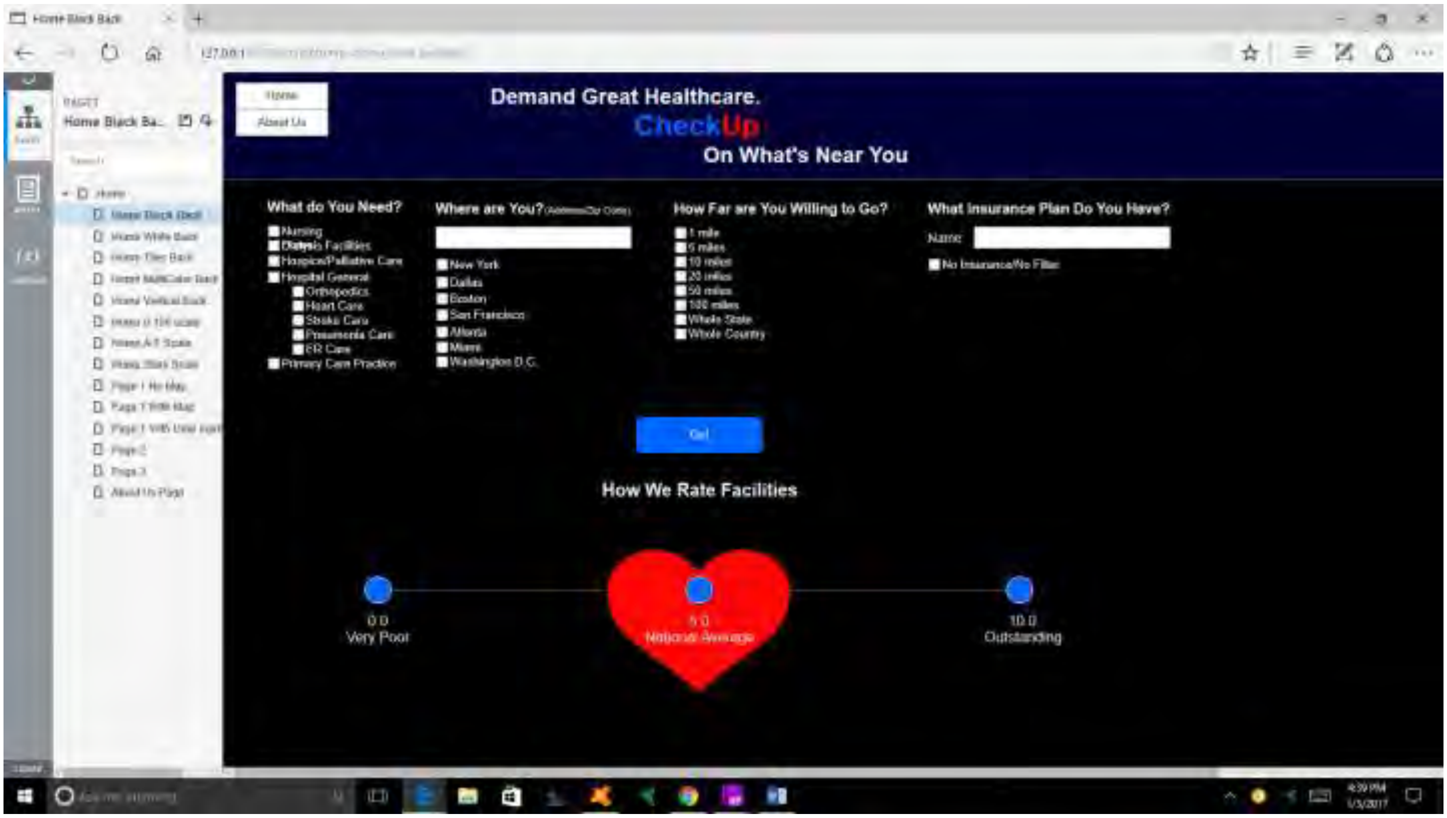
Concluding Questions: Are there any other functions that you think would be useful to include in a healthcare facility rating website?

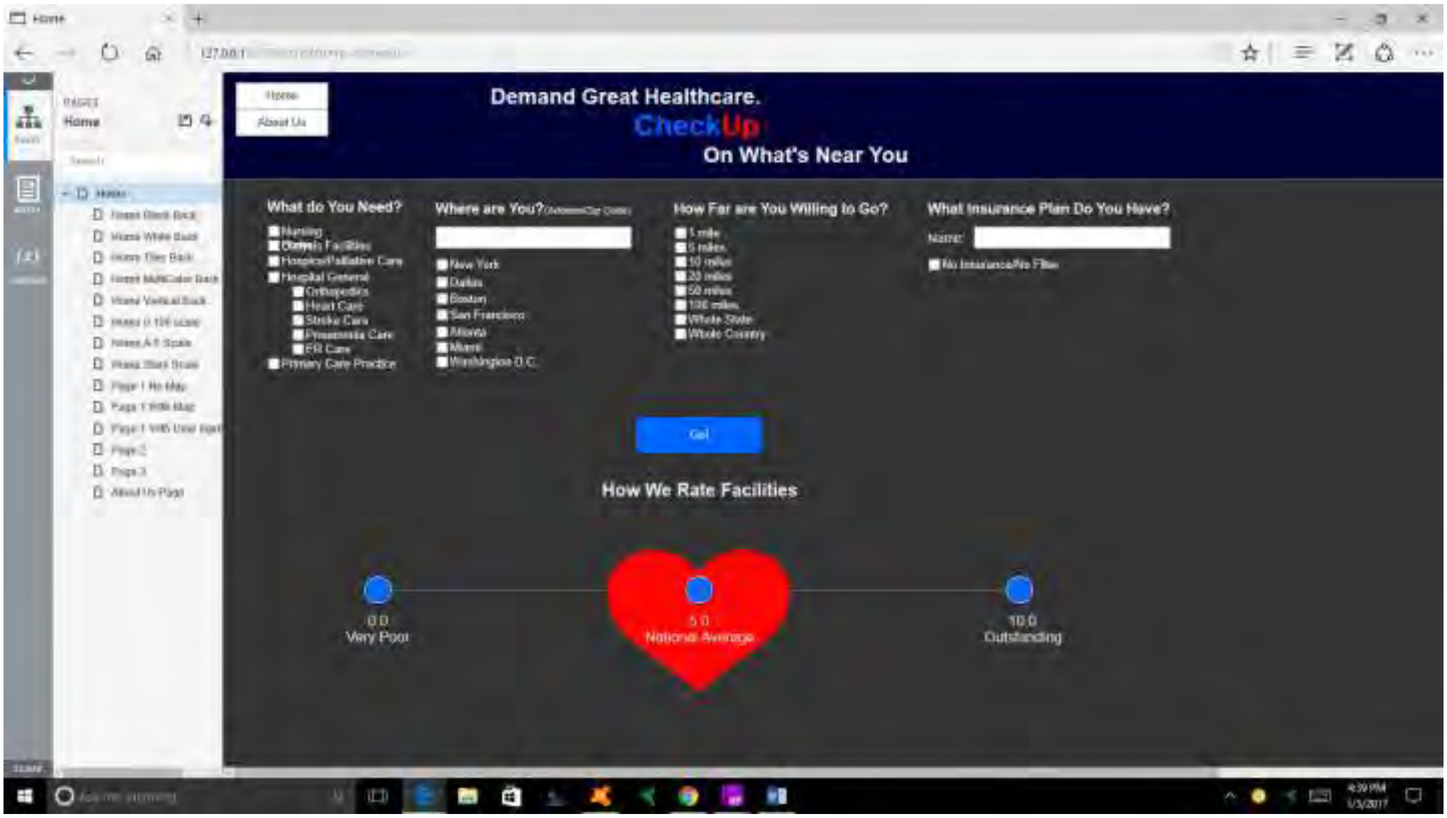
Do you have any other suggestions for how any of these pages could be formatted?

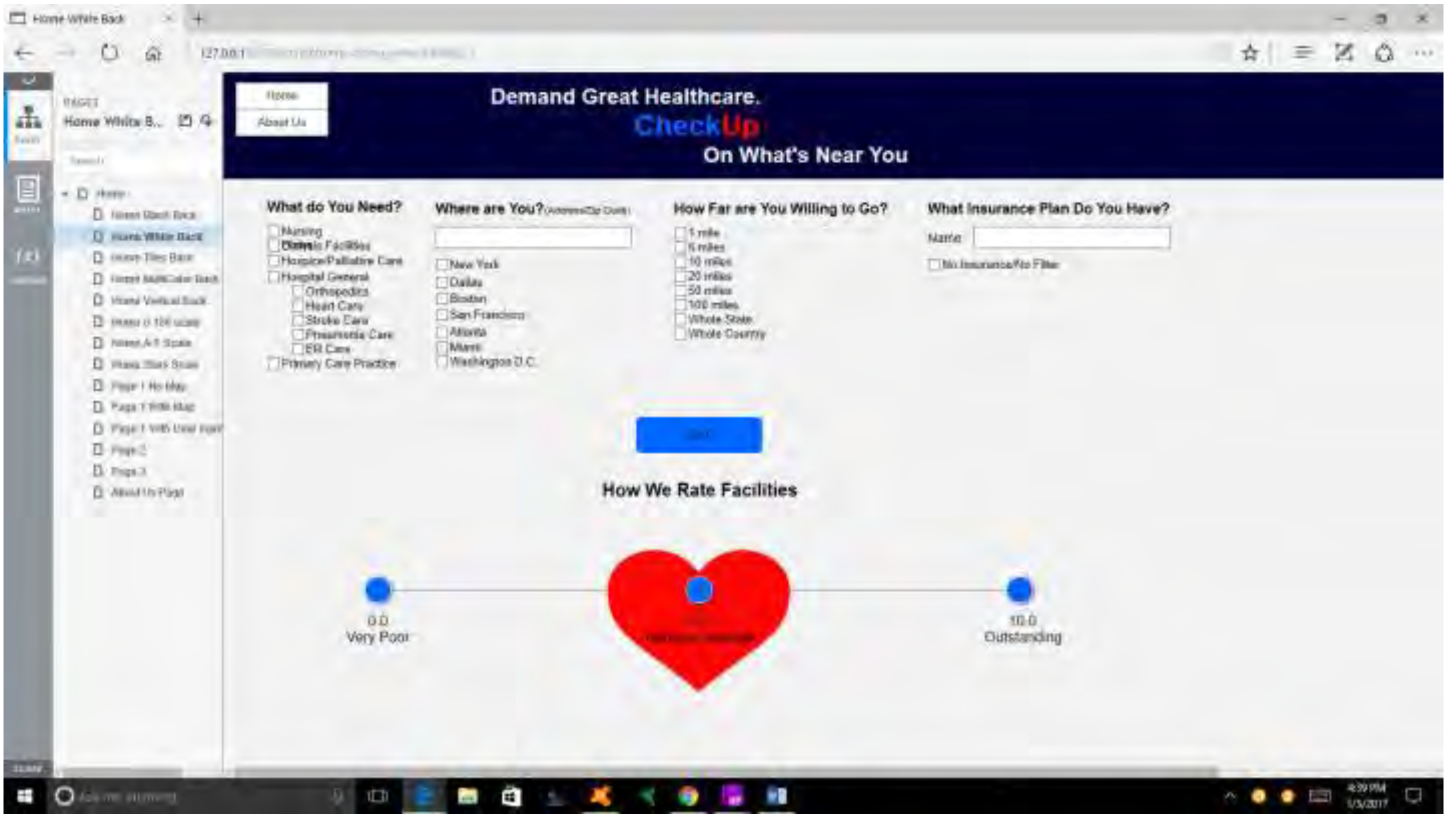
Debriefing:

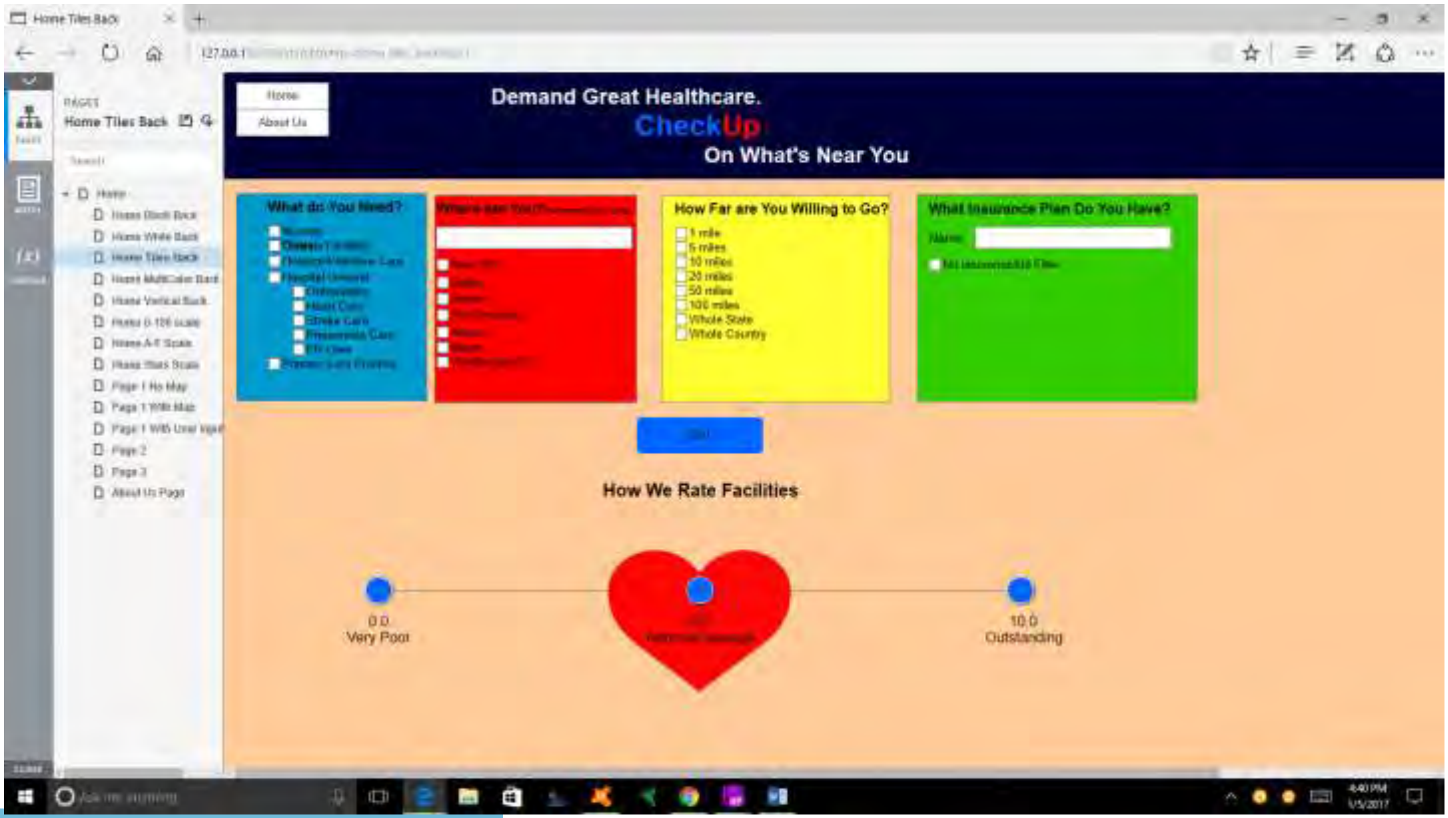
As explained, the purpose of my thesis project is to design and construct a website that patients can use to assess the quality of healthcare facilities when choosing among competing locations. However, for the website to be of use, the information therein must be presented in such a way that is both thorough and understandable. In this focus group, we looked at multiple methods for presenting information on the indicators used to rate facilities. This information will be used to influence the eventual website design. Thank you for all of your input!

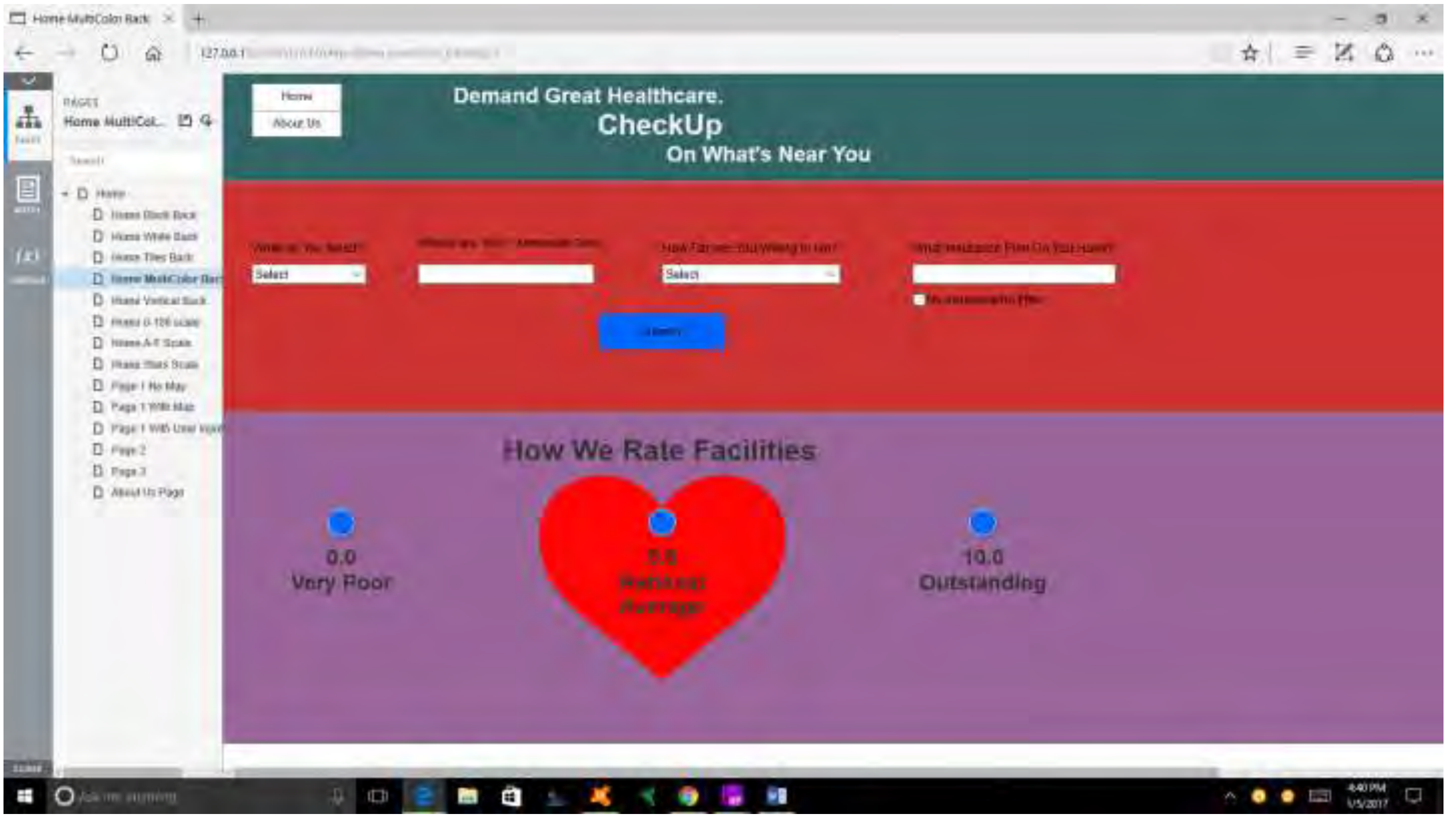
AXURE PROTOTYPE WEBSITE MODELS FOR FOCUS GROUP #3 (on next page)

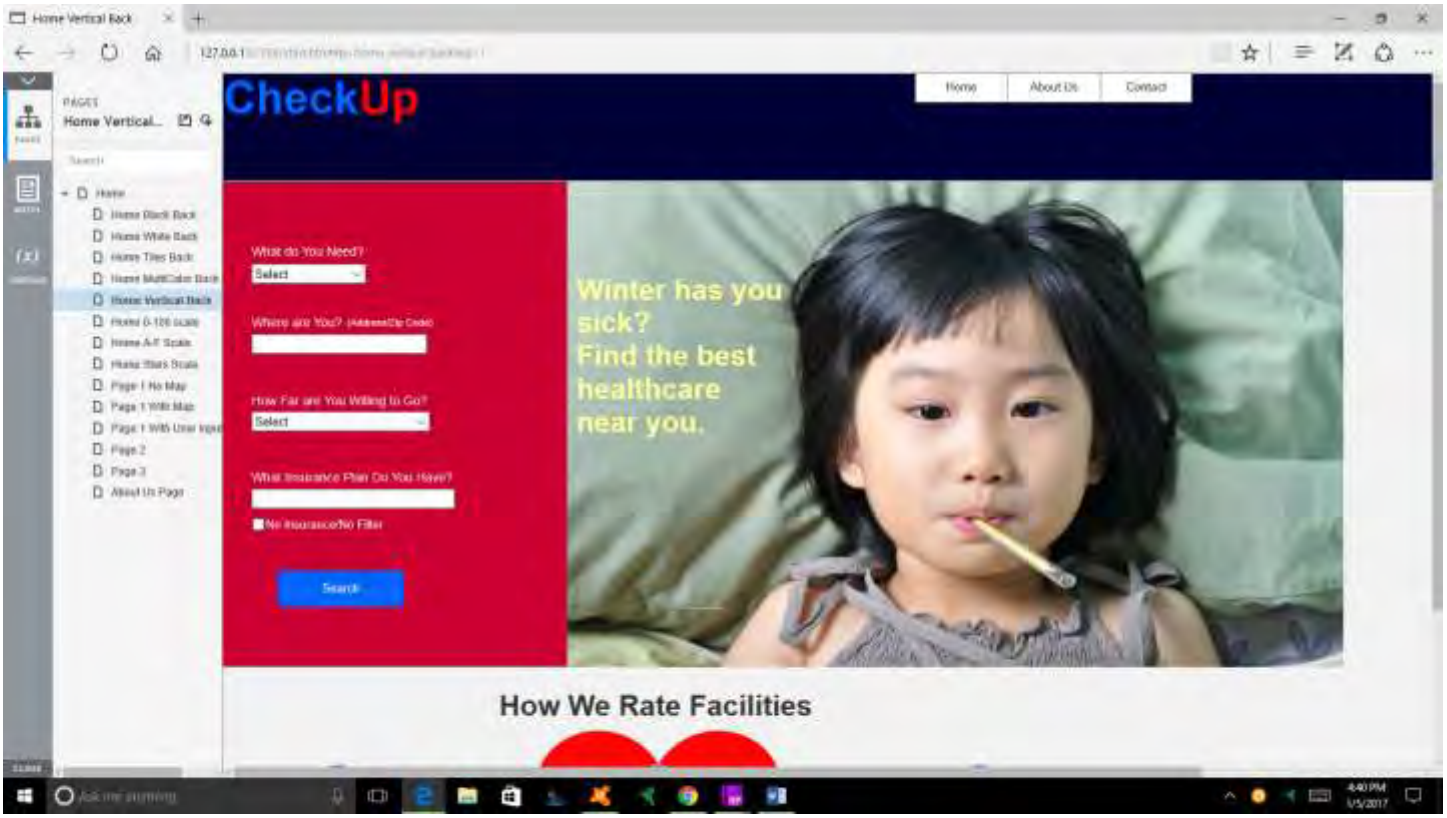

Home
About Us
Demand Great Healthcare.
CheckUp
On What's Near You
What do You Need?
☐ Nursing
☐ Dialysis Facilities
☐ Hospice/Palliative Care
☐ Hospital General
☐ Orthopedics
☐ Heart Care
☐ Stroke Care
☐ Pneumonia Care
☐ ER Care
☐ Primary Care Practice
Where are You? (Address/Zip Code)
☐ New York
☐ Dallas
☐ Boston
☐ San Francisco
☐ Atlanta
☐ Miami
☐ Washington D.C.
How Far are You Willing to Go?
☐ 1 mile
☐ 5 miles
☐ 10 miles
☐ 20 miles
☐ 50 miles
☐ 100 miles
☐ Whole State
☐ Whole Country
What Insurance Plan Do You Have?
Name:
☐ No Insurance/No Filter
Go!
How We Rate Facilities
0.0
Very Poor
5.0
National Average
10.0
Outstanding

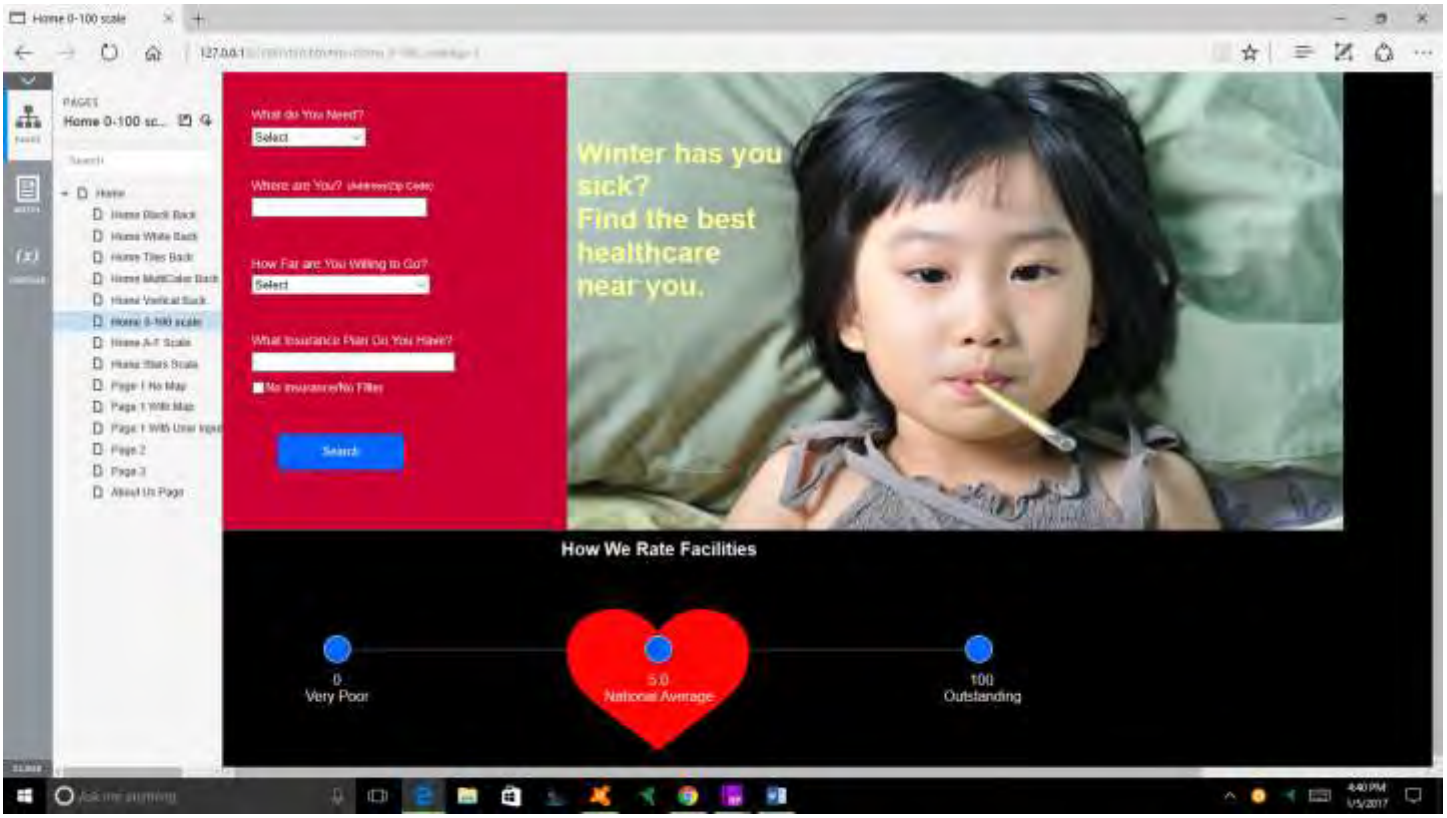

Home
About Us
Demand Great Healthcare.
CheckUp
On What's Near You
What do You Need?
☐ Nursing
☐ Dialysis Facilities
☐ Hospice/Palliative Care
☐ Hospital General
☐ Orthopedics
☐ Heart Care
☐ Stroke Care
☐ Pneumonia Care
☐ ER Care
☐ Primary Care Practice
Where are You? (Address/Zip Code)
☐ New York
☐ Dallas
☐ Boston
☐ San Francisco
☐ Atlanta
☐ Miami
☐ Washington D.C.
How Far are You Willing to Go?
☐ 1 mile
☐ 5 miles
☐ 10 miles
☐ 20 miles
☐ 50 miles
☐ 100 miles
☐ Whole State
☐ Whole Country
What Insurance Plan Do You Have?
Name:
☐ No Insurance/No Filter
Go!
How We Rate Facilities
0.0
Very Poor
5.0
National Average
10.0
Outstanding

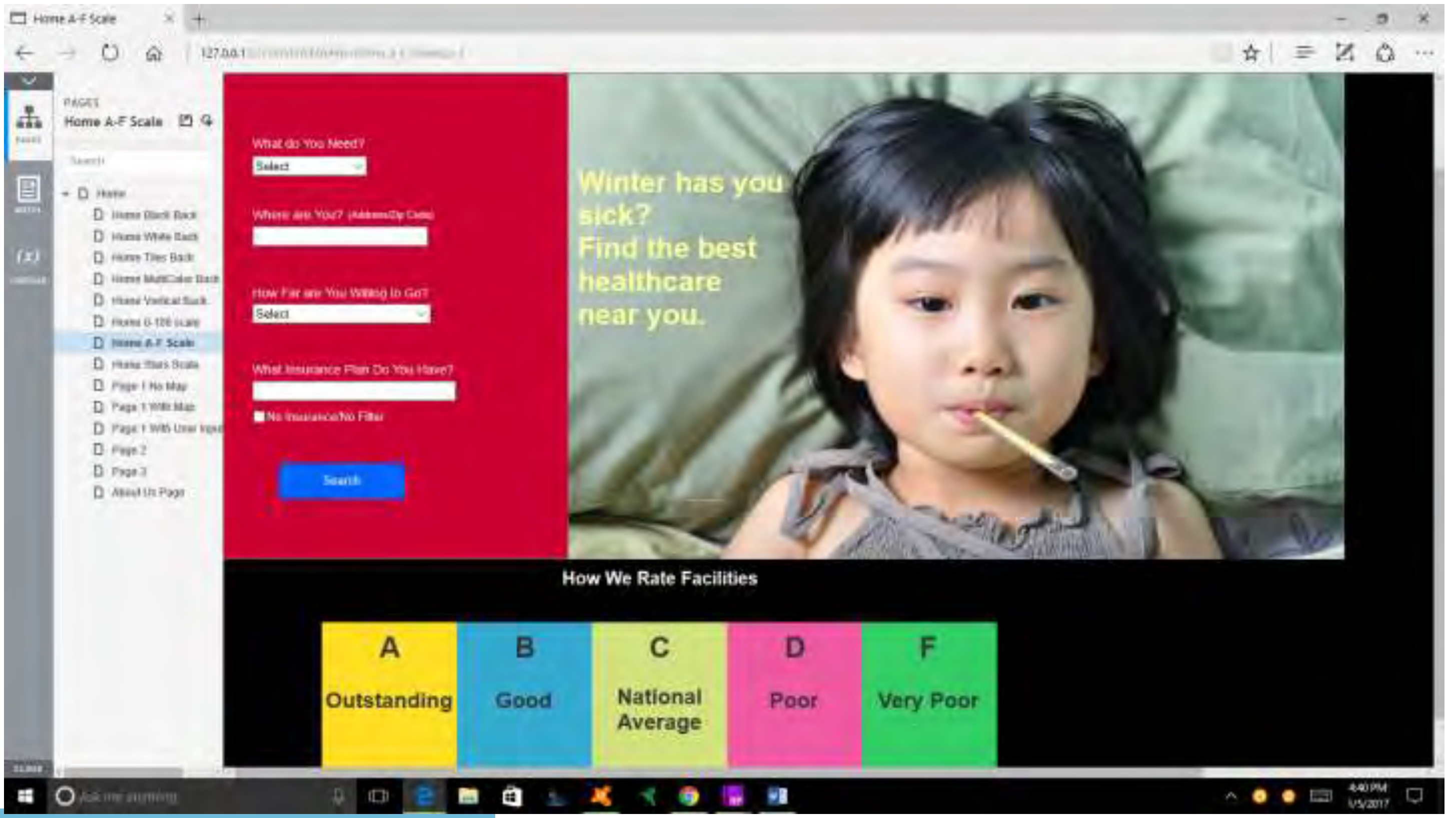

Home White Back
Home
About Us
Demand Great Healthcare.
CheckUp
On What's Near You
What do You Need?
☐ Nursing
☐ Dialysis Facilities
☐ Hospice/Palliative Care
☐ Hospital General
☐ Orthopedics
☐ Heart Care
☐ Stroke Care
☐ Pneumonia Care
☐ ER Care
☐ Primary Care Practice
Where are You? (Address/Zip Code)
☐ New York
☐ Dallas
☐ Boston
☐ San Francisco
☐ Atlanta
☐ Miami
☐ Washington D.C.
How Far are You Willing to Go?
☐ 1 mile
☐ 5 miles
☐ 10 miles
☐ 20 miles
☐ 50 miles
☐ 100 miles
☐ Whole State
☐ Whole Country
What Insurance Plan Do You Have?
Name
☐ No Insurance/No Filter
How We Rate Facilities
0.0
Very Poor
10.0
Outstanding

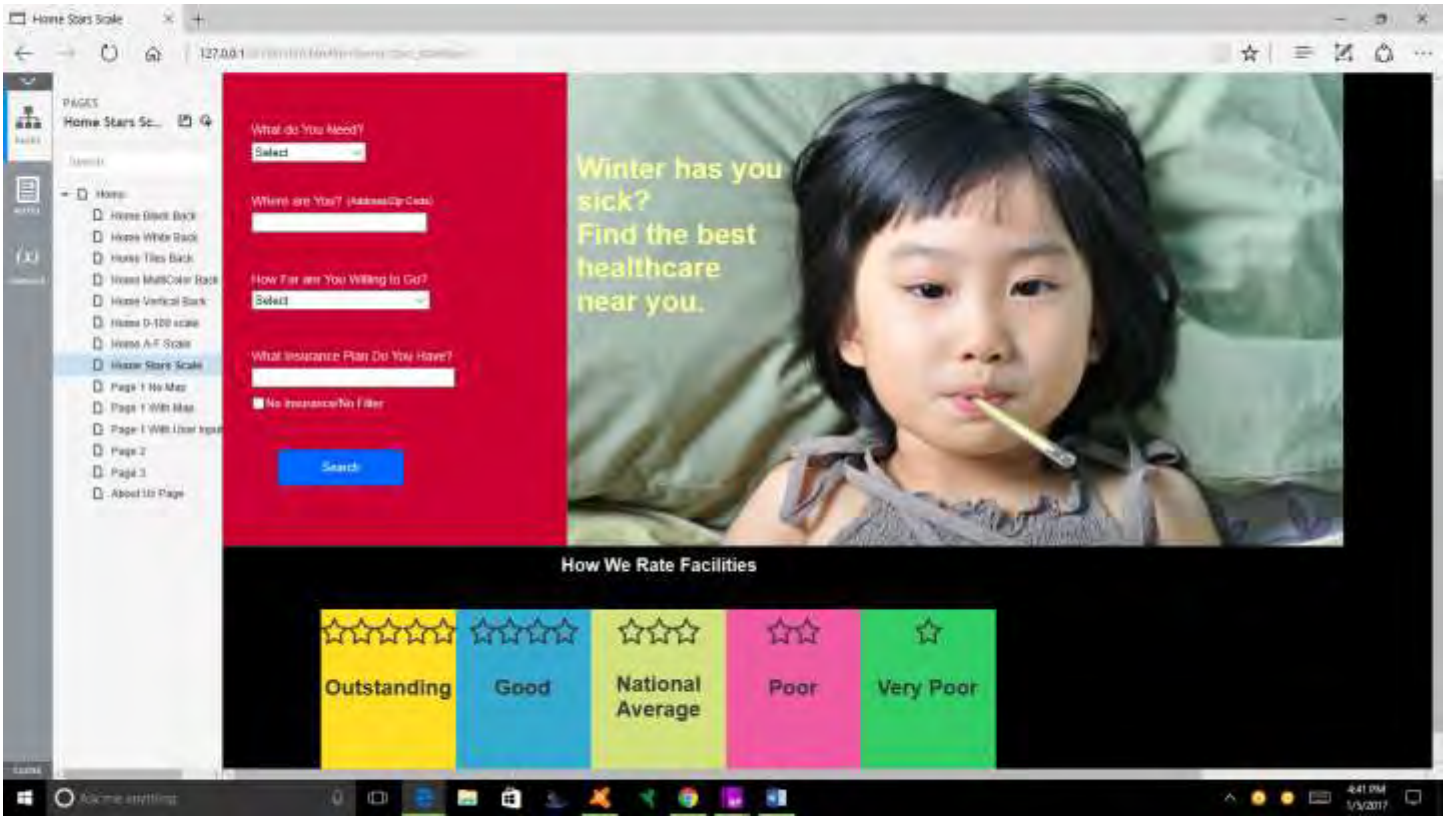

Home Tiles Back
Home
About Us
Demand Great Healthcare.
CheckUp
On What's Near You
What do You Need?
Where are You?
How Far are You Willing to Go?
☐ 1 mile
☐ 5 miles
☐ 10 miles
☐ 20 miles
☐ 50 miles
☐ 100 miles
☐ Whole State
☐ Whole Country
What Insurance Plan Do You Have?
Name
How We Rate Facilities
0.0
Very Poor
10.0
Outstanding


Home MultiColor Back
PAGES
Home MultiCol...
Home
About Us
Demand Great Healthcare.
CheckUp
On What's Near You
Select
Select
How We Rate Facilities
0.0
Very Poor
5.0
Average
10.0
Outstanding


Home Vertical Back
PAGES
Home Vertical...
CheckUp
Home
About Us
Contact
What do You Need?
Select
Where are You? (Address/Zip Code)
How Far are You Willing to Go?
Select
What Insurance Plan Do You Have?
No Insurance/No Filter
Search
Winter has you sick?
Find the best healthcare near you.
How We Rate Facilities


Home 0-100 scale
PAGES
Home 0-100 sc...
What do You Need?
Select
Where are You? (Address/Zip Code)
How Far are You Willing to Go?
Select
What Insurance Plan Do You Have?
No Insurance/No Filter
Search
Winter has you sick? Find the best healthcare near you.
How We Rate Facilities
0
Very Poor
5.0
National Average
100
Outstanding


Home A-F Scale
PAGES
Home A-F Scale
What do You Need?
Select
Where are You? (Address/Zip Code)
How Far are You Willing to Go?
Select
What Insurance Plan Do You Have?
No Insurance/No Filter
Search
Winter has you sick? Find the best healthcare near you.
How We Rate Facilities
A
Outstanding
B
Good
C
National Average
D
Poor
F
Very Poor



What do You Need?

Where are You? (Address/Zip Code)

How Far are You Willing to Go?

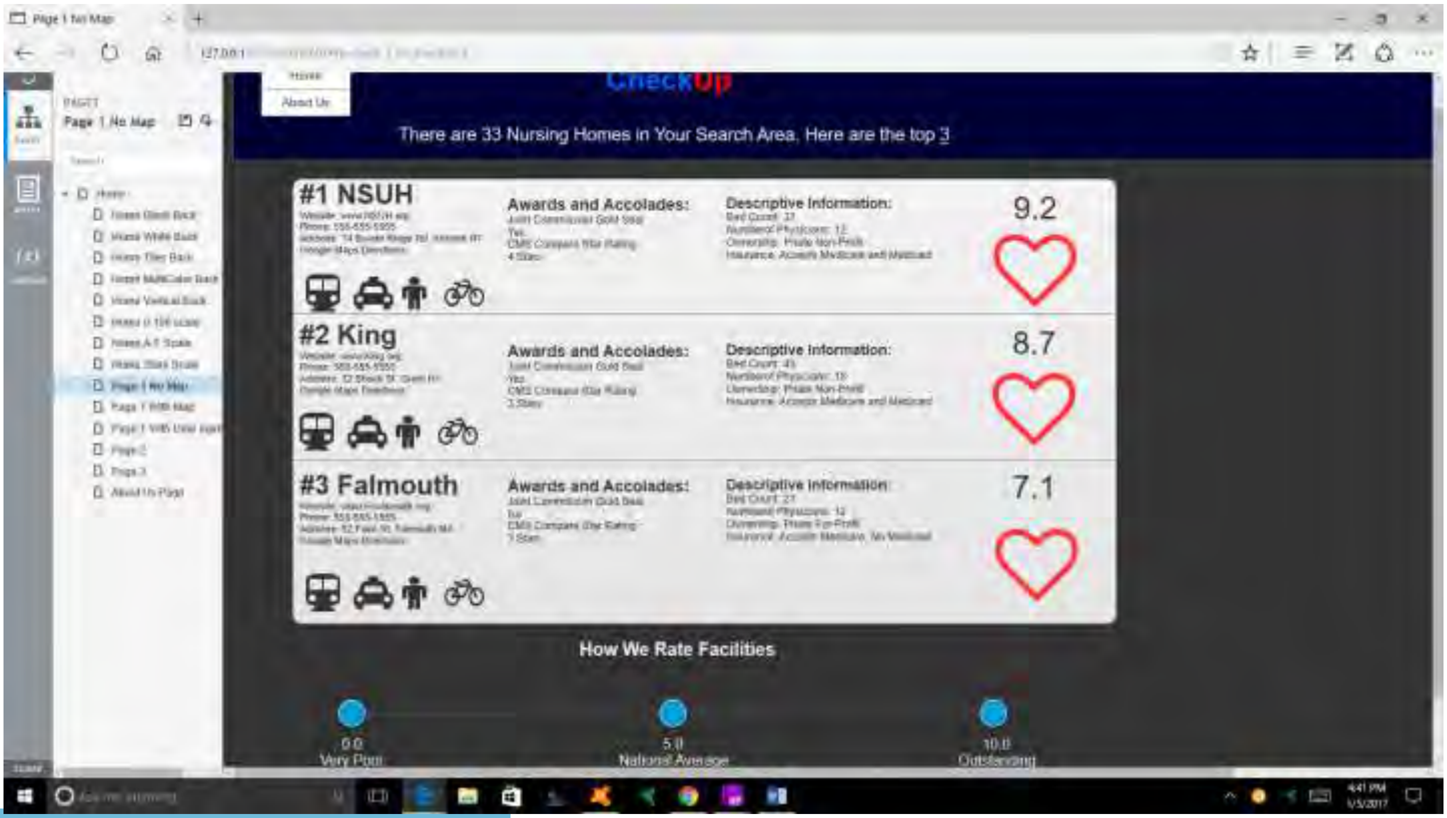
What Insurance Plan Do You Have?

☐ No Insurance/No Filter

Winter has you sick?
Find the best healthcare near you.

How We Rate Facilities

☆☆☆☆☆	☆☆☆☆	☆☆☆	☆☆	☆
Outstanding	Good	National Average	Poor	Very Poor



There are 33 Nursing Homes in Your Search Area. Here are the top 3

#1 NSUH

Awards and Accolades:

Descriptive Information:

9.2

#2 King

Awards and Accolades:

Descriptive Information:

8.7

#3 Falmouth

Awards and Accolades:

Descriptive Information:

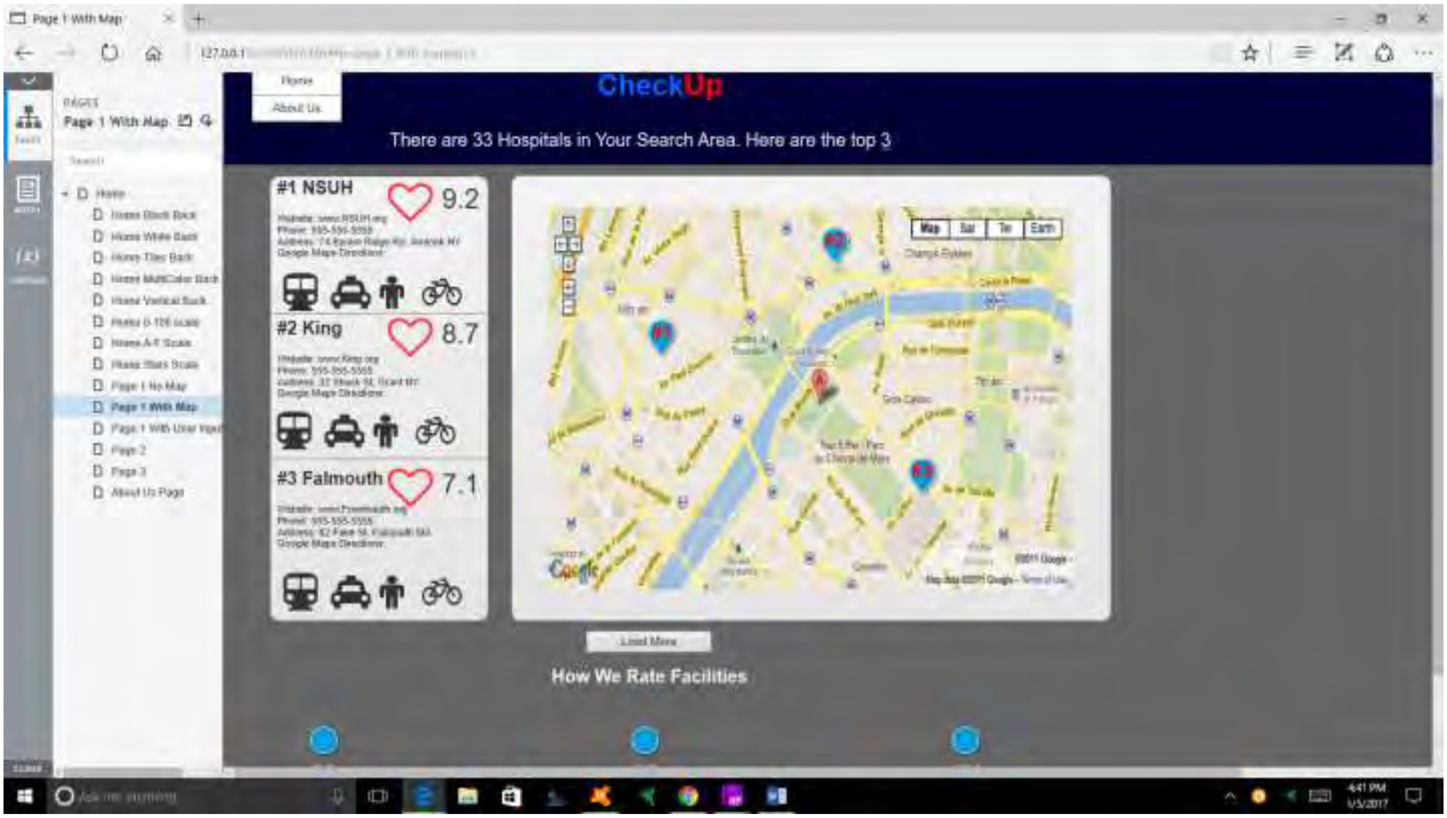
7.1

How We Rate Facilities

0.0 Very Poor

5.0 National Average

10.0 Outstanding



There are 33 Hospitals in Your Search Area. Here are the top 3

#1 NSUH 9.2
Website: www.NSUH.org
Phone: 555-555-5555
Address: 74 Byram Ridge Rd. Armonk NY
Google Maps Directions:

#2 King 8.7
Website: www.King.org
Phone: 555-555-5555
Address: 32 Shack St. Grant NY
Google Maps Directions:

#3 Falmouth 7.1
Website: www.Freemouth.org
Phone: 555-555-5555
Address: 62 Fake St. Falmouth MA
Google Maps Directions:

How We Rate Facilities

There are 33 Hospitals in Your Search Area. Here are the top 3

Overall scores are calculated by averaging four domains

Effectiveness	Patient Satisfaction	Safety	Standards of Care
25%	25%	25%	25%

Change the weights of each domain based upon your preferences, and then press GO to recalculate. Remember, the total must add up to 100%.

#1 NSUH 9.2
Website: www.NSUH.org
Phone: 555-555-5555
Address: 74 Byram Ridge Rd. Armonk NY
Google Maps Directions:

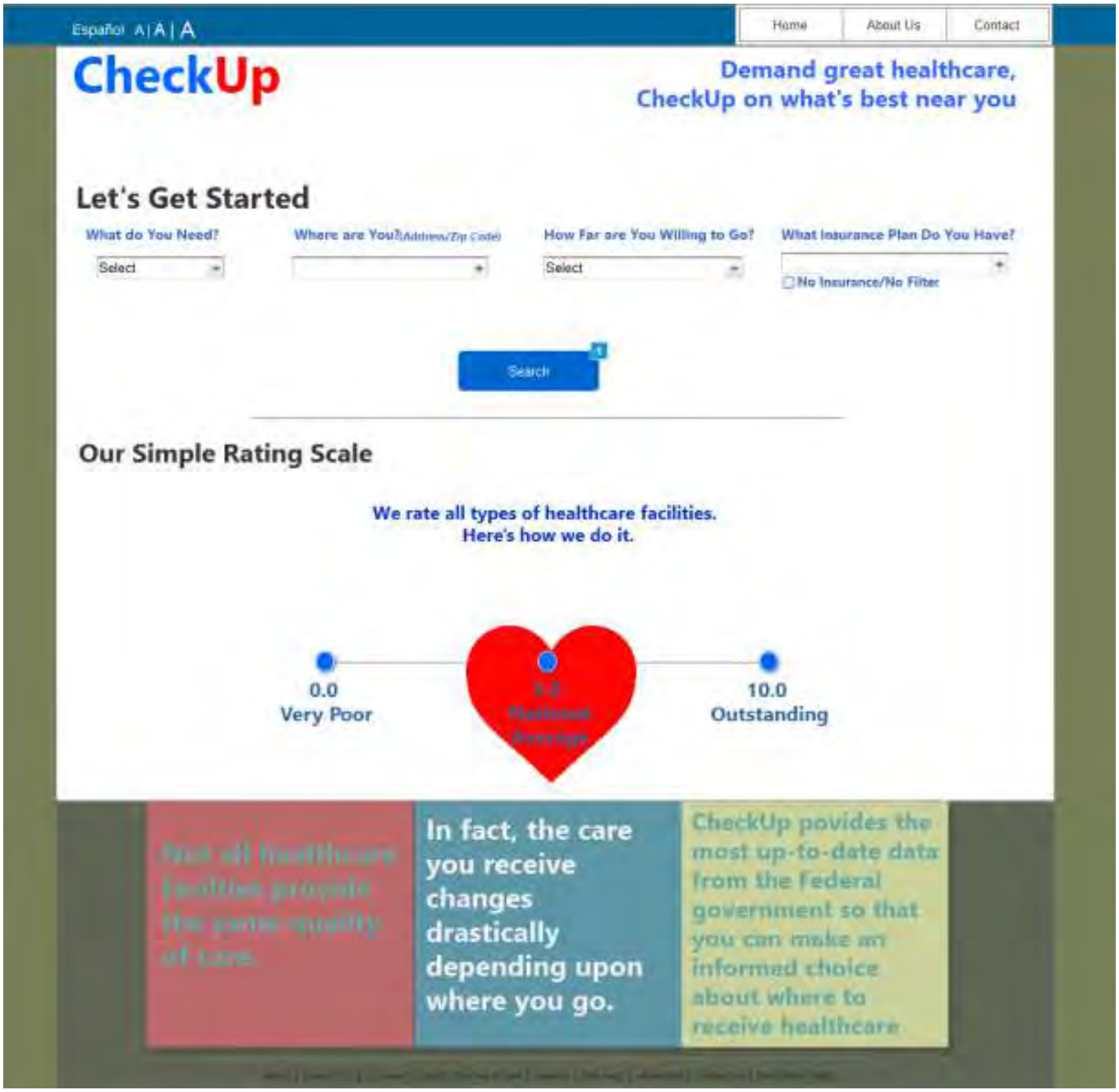
#2 King 8.7
Website: www.King.org
Phone: 555-555-5555
Address: 32 Shack St. Grant NY
Google Maps Directions:

Appendix 18 – Correlation Matrix of Effectiveness Indicators

	Acute MI 30-Day Mortality Rate Wins Z Score	Death Rate COPD Wins Z Score	HF 30 day mortality Wins Z Score	Pneumonia 30-Day Mortality Wins Z Score	Death Rate for Stroke Patients Wins Z Score	Acute MI 30 day readmission Wins Z Score	Unplanned Readmission Rate COPD Wins Z Score	HF 30 day readmission rate Wins Z Score	Rate of Readmission Hip/Knee Replacement Wins Z Score	Pneumonia 30 day Readmission Rate Wins Z Score	Readmission rate Stroke patients Wins Z Score	Readmission after Discharge Hospital-Wide Wins Z Score
Acute MI 30-Day Mortality Rate Wins Z Score	1	0.1877	0.2713	0.2227	0.2063	0.0673	0.0095	0.0061	-0.0141	0.0022	-0.0337	0.0031
Death Rate COPD Wins Z Score	0.1877	1	0.3568	0.3044	0.2593	-0.09	-0.0055	-0.1361	-0.1	-0.0776	-0.1538	-0.1416
HF 30 day mortality Wins Z Score	0.2713	0.3568	1	0.3822	0.2788	-0.1099	-0.0606	-0.1335	-0.1098	-0.1226	-0.1875	-0.1733
Pneumonia 30-Day Mortality Wins Z Score	0.2227	0.3044	0.3822	1	0.2216	-0.0117	0.0495	-0.0087	-0.0413	0.0174	-0.0642	-0.024
Death Rate for Stroke Patients Wins Z Score	0.2063	0.2593	0.2788	0.2216	1	-0.0839	-0.0637	-0.1266	-0.0599	-0.0705	-0.1108	-0.1029
Acute MI 30 day readmission Wins Z Score	0.0673	-0.09	-0.1099	-0.0117	-0.0839	1	0.2505	0.3575	0.1771	0.2567	0.2964	0.3668
Unplanned Readmission Rate COPD Wins Z Score	0.0095	-0.0055	-0.0606	0.0495	-0.0637	0.2505	1	0.3642	0.1423	0.3546	0.2694	0.457
HF 30 day readmission rate Wins Z Score	0.0061	-0.1361	-0.1335	-0.0087	-0.1266	0.3575	0.3642	1	0.1453	0.3837	0.3567	0.5392
Rate of Readmission Hip/Knee Replacement Wins Z Score	-0.0141	-0.1	-0.1098	-0.0413	-0.0599	0.1771	0.1423	0.1453	1	0.1526	0.1685	0.2746
Pneumonia 30 day Readmission Rate Wins Z Score	0.0022	-0.0776	-0.1226	0.0174	-0.0705	0.2567	0.3546	0.3837	0.1526	1	0.2614	0.4581
Readmission rate Stroke patients Wins Z Score	-0.0337	-0.1538	-0.1875	-0.0642	-0.1108	0.2964	0.2694	0.3567	0.1685	0.2614	1	0.4105
Readmission after Discharge Hospital-Wide Wins Z Score	0.0031	-0.1416	-0.1733	-0.024	-0.1029	0.3668	0.457	0.5392	0.2746	0.4581	0.4105	1

Appendix 19

Search Page



Demand great healthcare,
CheckUp on what's best near you

Let's Get Started

Our Simple Rating Scale

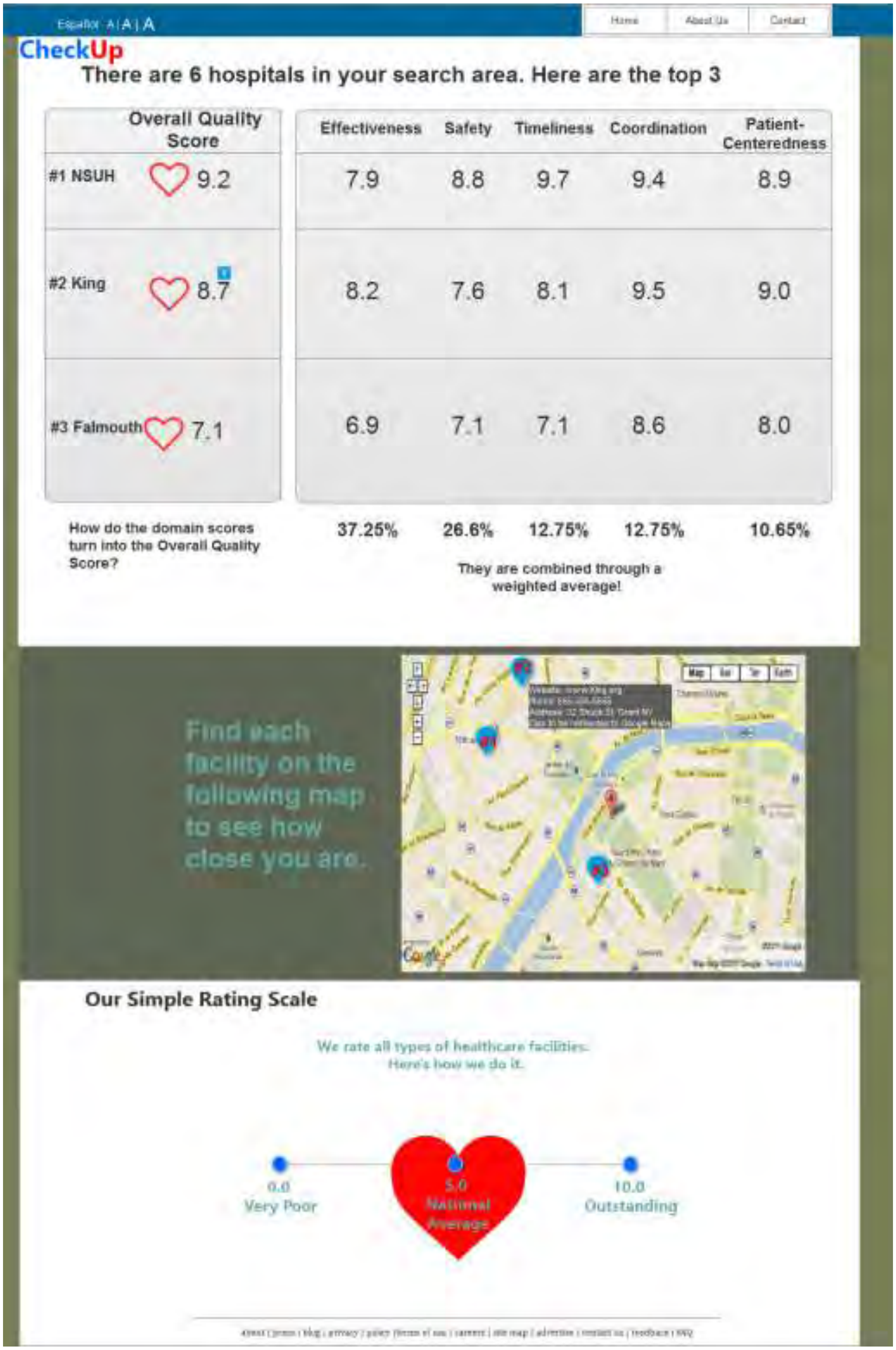
We rate all types of healthcare facilities.
Here's how we do it.

Not all healthcare facilities provide the same quality of care.

In fact, the care you receive changes drastically depending upon where you go.

CheckUp povides the most up-to-date data from the Federal government so that you can make an informed choice about where to receive healthcare

List Page



CheckUp

There are 6 hospitals in your search area. Here are the top 3

	Overall Quality Score	Effectiveness	Safety	Timeliness	Coordination	Patient-Centeredness
#1 NSUH	9.2	7.9	8.8	9.7	9.4	8.9
#2 King	8.7	8.2	7.6	8.1	9.5	9.0
#3 Falmouth	7.1	6.9	7.1	7.1	8.6	8.0
How do the domain scores turn into the Overall Quality Score?		37.25%	26.6%	12.75%	12.75%	10.65%

They are combined through a weighted average!

Find each facility on the following map to see how close you are.

Our Simple Rating Scale

We rate all types of healthcare facilities.
Here's how we do it.

Facility Page

Demand great healthcare,
CheckUp on what's best near you

Here's some more information on King Street Acute Care Hospital

Identifying Information:
Website: www.King.org
Phone: 555-555-5555
Address: 32 That Rd, Grant NY
Distance: 8.2 miles

Awards and Accolades:
Joint Commission Gold Seal:
Yes
CMS Compare Star Rating
4 Stars

Descriptive Information:
Bed Count: 721
Numberof Physicians: 126
Ownership: Private Non-Profit
Insurance: Accepts Medicare and Medicaid

In the News:

Cardiologist Cindy Grines to leave DMC for New York health system

Northwell Health celebrates birthday of 100-year-old volunteer

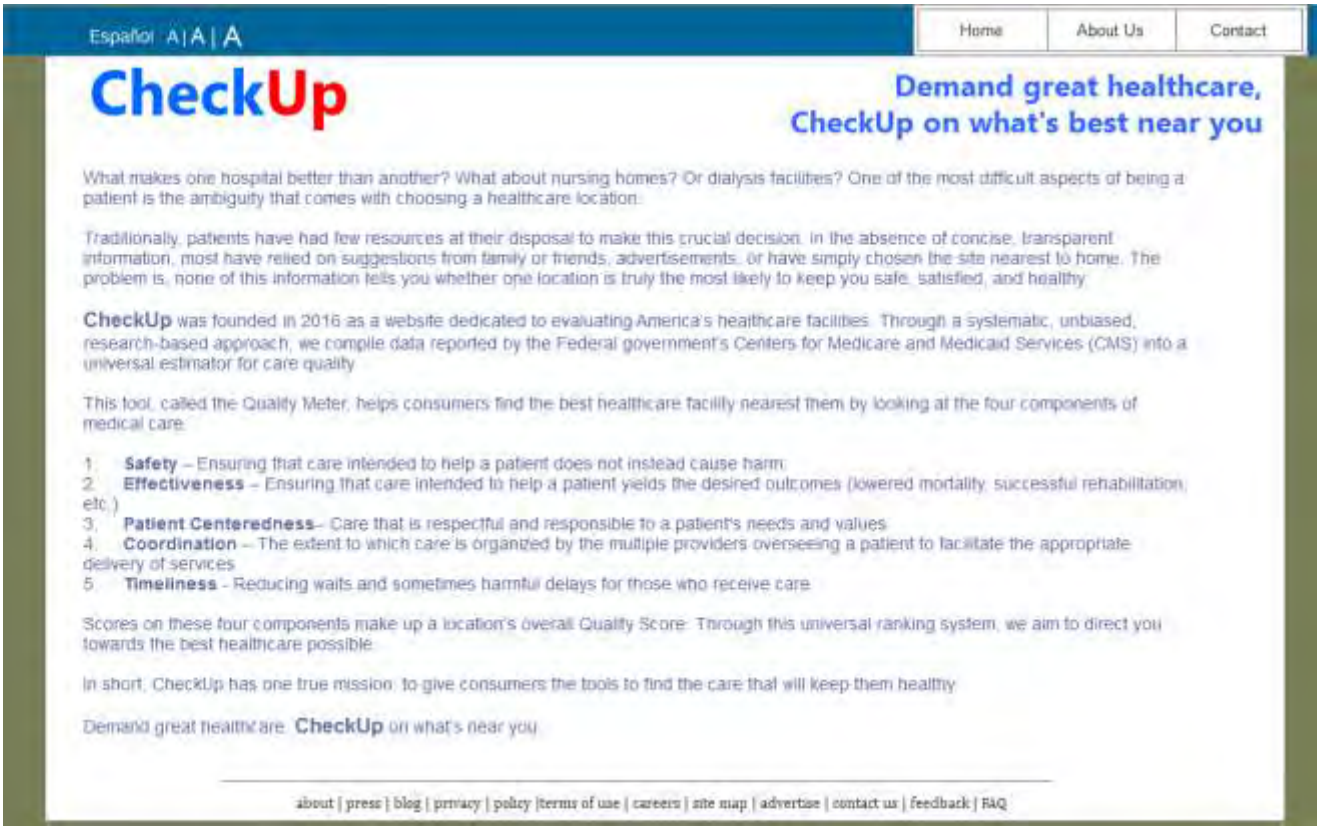
Williamsburg Gets Northwell Health-GoHealth Urgent Care Clinic

CheckUp

King Hospital's scored an 8.7 overall. See how it did on the individual measures that made up that overall score.

Indicators Page

About Us Page



CheckUp

Demand great healthcare,
CheckUp on what's best near you

What makes one hospital better than another? What about nursing homes? Or dialysis facilities? One of the most difficult aspects of being a patient is the ambiguity that comes with choosing a healthcare location.

Traditionally, patients have had few resources at their disposal to make this crucial decision. In the absence of concise, transparent information, most have relied on suggestions from family or friends, advertisements, or have simply chosen the site nearest to home. The problem is, none of this information tells you whether one location is truly the most likely to keep you safe, satisfied, and healthy.

CheckUp was founded in 2016 as a website dedicated to evaluating America's healthcare facilities. Through a systematic, unbiased, research-based approach, we compile data reported by the Federal government's Centers for Medicare and Medicaid Services (CMS) into a universal estimator for care quality.

This tool, called the Quality Meter, helps consumers find the best healthcare facility nearest them by looking at the four components of medical care.

1. **Safety** – Ensuring that care intended to help a patient does not instead cause harm.
2. **Effectiveness** – Ensuring that care intended to help a patient yields the desired outcomes (lowered mortality, successful rehabilitation, etc.)
3. **Patient Centeredness**– Care that is respectful and responsible to a patient's needs and values.
4. **Coordination** – The extent to which care is organized by the multiple providers overseeing a patient to facilitate the appropriate delivery of services.
5. **Timeliness** - Reducing waits and sometimes harmful delays for those who receive care.

Scores on these four components make up a location's overall Quality Score. Through this universal ranking system, we aim to direct you towards the best healthcare possible.

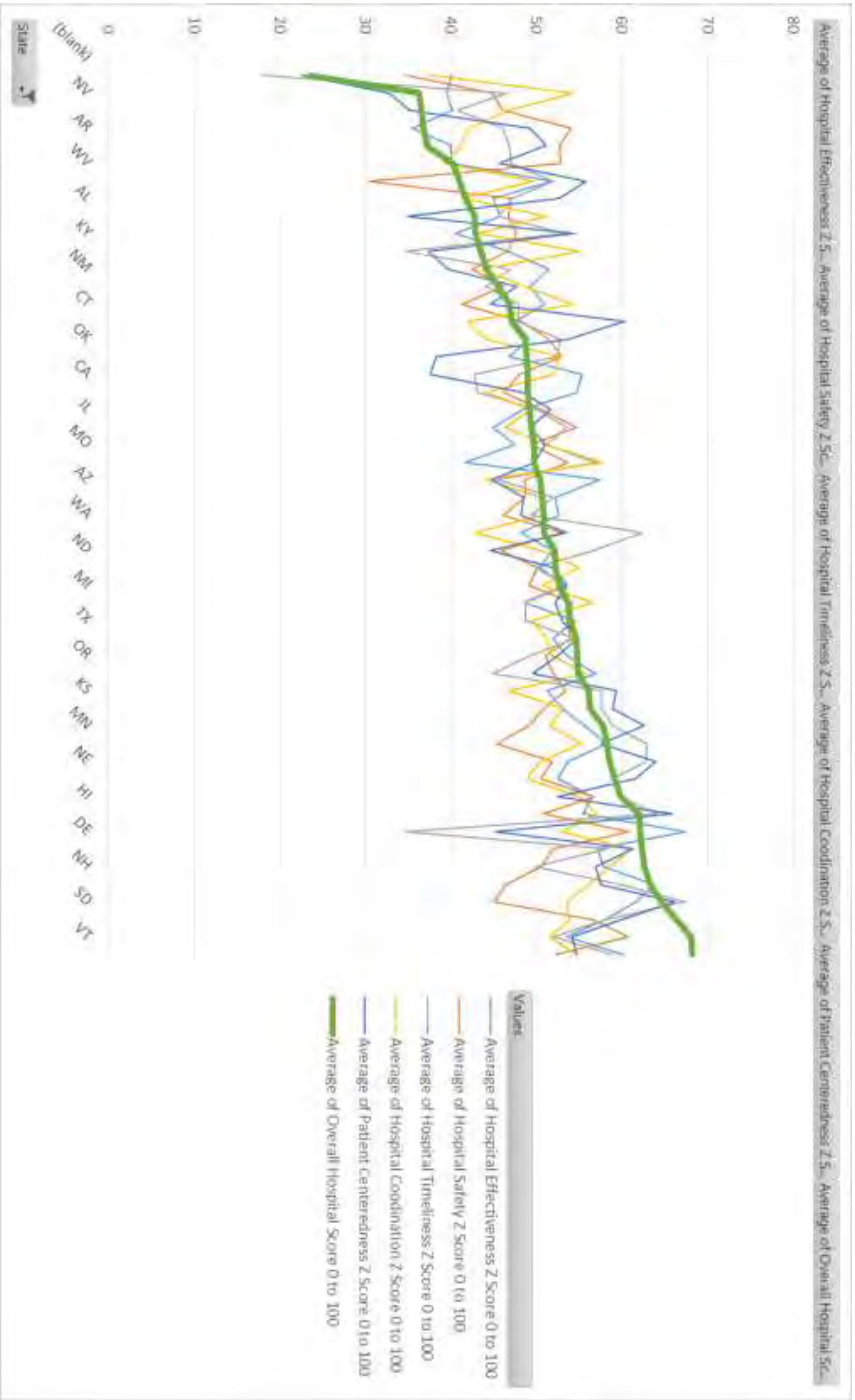
In short, CheckUp has one true mission: to give consumers the tools to find the care that will keep them healthy.

Demand great healthcare. **CheckUp** on what's near you.

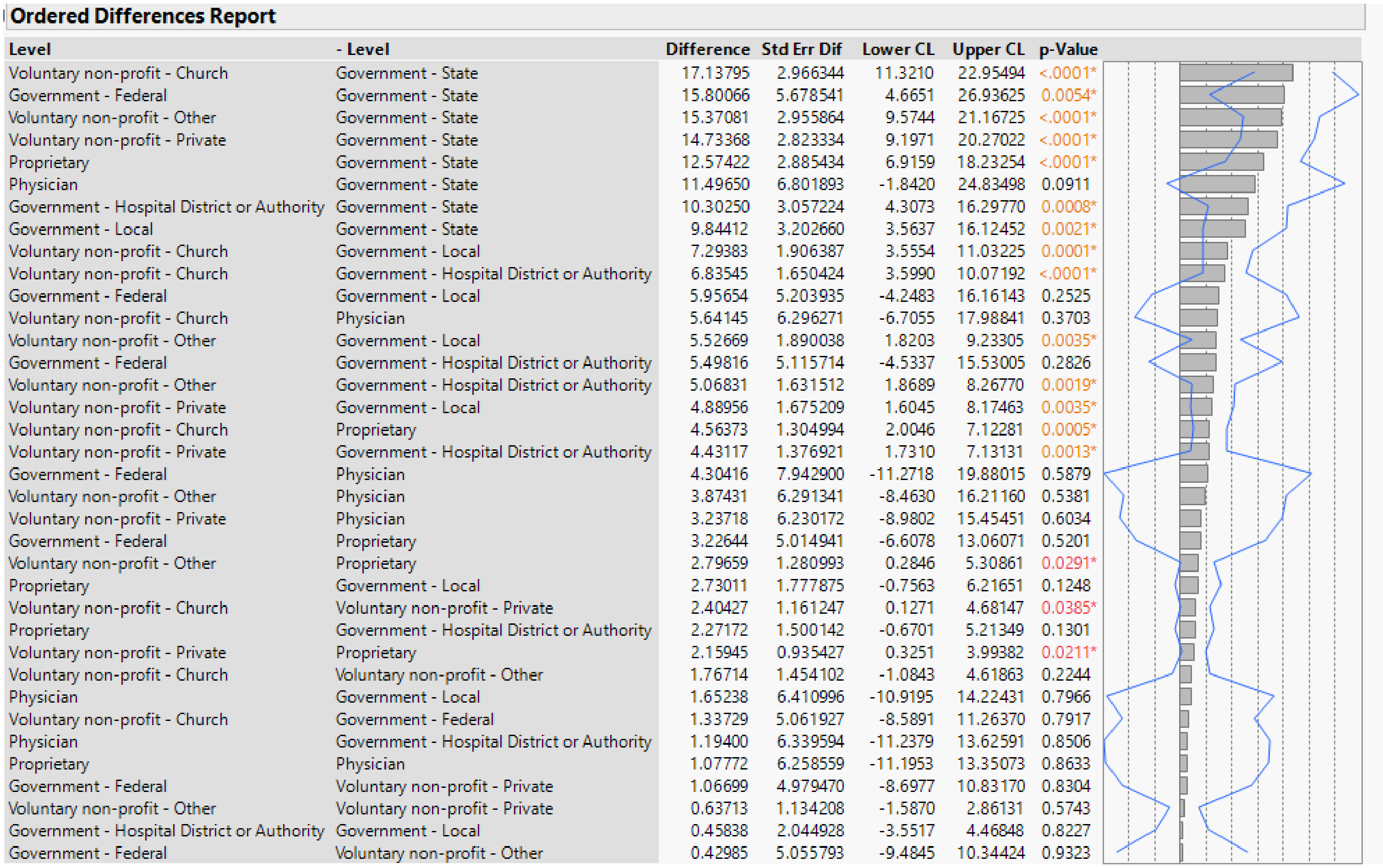
Appendix 20

Row Labels	Average of Hospital Effectiveness Z Score 0 to 100	Average of Hospital Safety Z Score 0 to 100	Average of Hospital Timeliness Z Score 0 to 100	Average of Hospital Coodination Z Score 0 to 100	Average of Patient Centeredness Z Score 0 to 100	Average of Overall Hospital Score 0 to 100
(blank)						
DC	40.1	34.8	18.0	37.7	22.6	23.6
NV	39.5	44.8	46.6	54.3	32.8	36.3
NY	40.2	46.1	40.8	46.7	35.2	36.6
AR	35.6	54.1	46.2	42.1	49.3	36.8
MS	40.0	52.7	46.9	41.3	51.2	37.2
WV	40.1	52.8	47.1	39.3	45.7	40.3
AK	51.9	30.5	52.0	49.7	55.8	41.0
AL	45.0	47.1	47.2	41.4	52.7	41.8
NJ	45.8	46.7	45.6	51.2	34.9	42.8
KY	40.6	47.7	53.3	43.1	54.4	42.9
MD	46.8	46.6	34.7	55.1	37.4	43.5
NM	51.5	42.5	47.0	43.7	39.7	44.2
GA	49.3	46.3	44.1	48.6	47.7	45.5
CT	50.9	41.3	47.9	54.3	44.7	46.9
LA	47.6	48.3	47.9	42.0	60.3	47.2
OK	49.1	51.5	52.9	44.2	53.7	48.7
FL	46.9	53.0	52.4	52.8	38.3	49.0
CA	55.4	48.0	42.8	52.0	37.7	49.0
RI	54.8	46.2	43.2	44.3	48.5	49.0
IL	48.4	50.6	50.2	49.4	51.7	49.2
TN	44.9	54.7	53.6	46.4	49.2	49.5
MO	47.5	49.3	51.2	51.1	51.1	49.7
VA	41.7	53.7	56.9	57.8	49.8	49.8
AZ	57.5	48.8	45.6	44.0	44.9	50.5
PA	52.3	48.7	52.0	50.6	48.6	50.7
WA	52.7	46.1	49.4	50.8	48.2	50.9
WY	48.2	53.0	62.4	42.7	53.6	50.9
ND	51.9	45.9	56.7	50.5	44.7	52.1
SC	50.4	50.1	49.9	55.0	52.0	52.3
MI	53.2	49.2	53.5	50.6	53.5	52.7
NC	48.6	54.2	48.8	56.7	52.2	53.7
TX	54.6	53.6	48.7	49.6	53.1	54.0
OH	52.0	54.0	54.5	51.2	54.9	54.7
OR	53.6	51.1	49.3	51.9	52.5	54.8
MA	57.0	52.1	45.0	55.8	49.6	54.9
KS	51.3	53.5	57.7	46.8	59.4	56.1
IN	53.1	50.9	57.7	53.0	59.0	56.3
MN	55.6	49.1	59.0	51.6	62.7	57.9
UT	58.2	45.4	62.9	55.3	57.6	58.3
NE	53.7	51.9	62.8	50.6	64.0	58.5
IA	52.6	50.5	59.0	49.0	61.6	59.2
HI	56.6	56.7	59.6	55.0	52.6	59.8
WI	55.5	50.8	64.2	57.1	65.8	62.1
DE	67.4	60.8	34.7	53.0	45.2	62.1
ME	57.1	52.0	58.3	61.1	61.4	62.3
NH	57.9	50.5	49.8	58.7	57.0	62.6
CO	63.7	46.5	60.0	56.1	57.6	63.3
SD	61.6	45.1	67.3	53.6	66.2	64.6
ID	58.0	56.7	58.1	53.9	59.9	66.3
VT	53.5	60.7	52.0	51.5	54.2	68.1
MT	60.2	52.2	58.9	54.8	54.7	68.3
Grand Total	**50.1**	**50.0**	**50.4**	**50.3**	**50.0**	**50.0**

Appendix 21 – Ascending Avg. Overall Scores by State



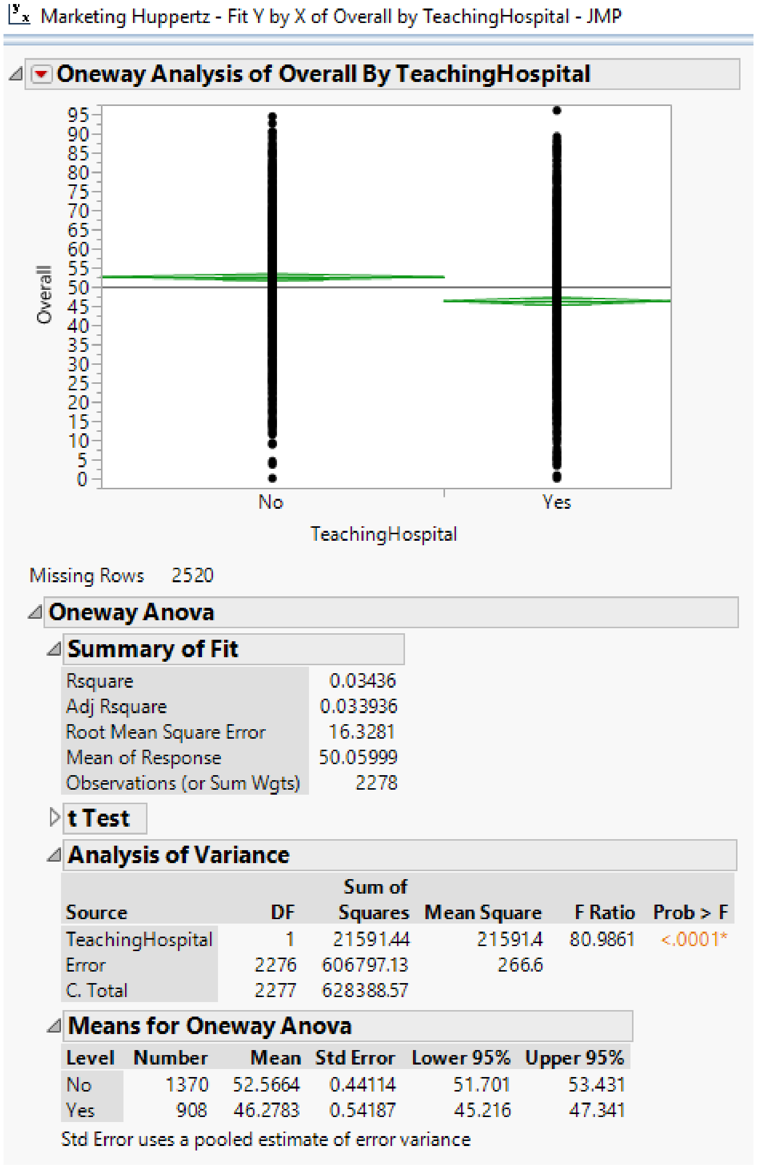
Appendix 22 – Ordered Differences Report Hospital Performance by Ownership Type



Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
Voluntary non-profit - Church	Government - State	17.13795	2.966344	11.3210	22.95494	<.0001*
Government - Federal	Government - State	15.80066	5.678541	4.6651	26.93625	0.0054*
Voluntary non-profit - Other	Government - State	15.37081	2.955864	9.5744	21.16725	<.0001*
Voluntary non-profit - Private	Government - State	14.73368	2.823334	9.1971	20.27022	<.0001*
Proprietary	Government - State	12.57422	2.885434	6.9159	18.23254	<.0001*
Physician	Government - State	11.49650	6.801893	-1.8420	24.83498	0.0911
Government - Hospital District or Authority	Government - State	10.30250	3.057224	4.3073	16.29770	0.0008*
Government - Local	Government - State	9.84412	3.202660	3.5637	16.12452	0.0021*
Voluntary non-profit - Church	Government - Local	7.29383	1.906387	3.5554	11.03225	0.0001*
Voluntary non-profit - Church	Government - Hospital District or Authority	6.83545	1.650424	3.5990	10.07192	<.0001*
Government - Federal	Government - Local	5.95654	5.203935	-4.2483	16.16143	0.2525
Voluntary non-profit - Church	Physician	5.64145	6.296271	-6.7055	17.98841	0.3703
Voluntary non-profit - Other	Government - Local	5.52669	1.890038	1.8203	9.23305	0.0035*
Government - Federal	Government - Hospital District or Authority	5.49816	5.115714	-4.5337	15.53005	0.2826
Voluntary non-profit - Other	Government - Hospital District or Authority	5.06831	1.631512	1.8689	8.26770	0.0019*
Voluntary non-profit - Private	Government - Local	4.88956	1.675209	1.6045	8.17463	0.0035*
Voluntary non-profit - Church	Proprietary	4.56373	1.304994	2.0046	7.12281	0.0005*
Voluntary non-profit - Private	Government - Hospital District or Authority	4.43117	1.376921	1.7310	7.13131	0.0013*
Government - Federal	Physician	4.30416	7.942900	-11.2718	19.88015	0.5879
Voluntary non-profit - Other	Physician	3.87431	6.291341	-8.4630	16.21160	0.5381
Voluntary non-profit - Private	Physician	3.23718	6.230172	-8.9802	15.45451	0.6034
Government - Federal	Proprietary	3.22644	5.014941	-6.6078	13.06071	0.5201
Voluntary non-profit - Other	Proprietary	2.79659	1.280993	0.2846	5.30861	0.0291*
Proprietary	Government - Local	2.73011	1.777875	-0.7563	6.21651	0.1248
Voluntary non-profit - Church	Voluntary non-profit - Private	2.40427	1.161247	0.1271	4.68147	0.0385*
Proprietary	Government - Hospital District or Authority	2.27172	1.500142	-0.6701	5.21349	0.1301
Voluntary non-profit - Private	Proprietary	2.15945	0.935427	0.3251	3.99382	0.0211*
Voluntary non-profit - Church	Voluntary non-profit - Other	1.76714	1.454102	-1.0843	4.61863	0.2244
Physician	Government - Local	1.65238	6.410996	-10.9195	14.22431	0.7966
Voluntary non-profit - Church	Government - Federal	1.33729	5.061927	-8.5891	11.26370	0.7917
Physician	Government - Hospital District or Authority	1.19400	6.339594	-11.2379	13.62591	0.8506
Proprietary	Physician	1.07772	6.258559	-11.1953	13.35073	0.8633
Government - Federal	Voluntary non-profit - Private	1.06699	4.979470	-8.6977	10.83170	0.8304
Voluntary non-profit - Other	Voluntary non-profit - Private	0.63713	1.134208	-1.5870	2.86131	0.5743
Government - Hospital District or Authority	Government - Local	0.45838	2.044928	-3.5517	4.46848	0.8227
Government - Federal	Voluntary non-profit - Other	0.42985	5.055793	-9.4845	10.34424	0.9323

Appendix 23



Marketing Huppertz - Fit Y by X of Overall by TeachingHospital - JMP

Oneway Analysis of Overall By TeachingHospital

Missing Rows 2520

Oneway Anova

Summary of Fit

Rsquare	0.03436
Adj Rsquare	0.033936
Root Mean Square Error	16.3281
Mean of Response	50.05999
Observations (or Sum Wgts)	2278

t Test

Analysis of Variance

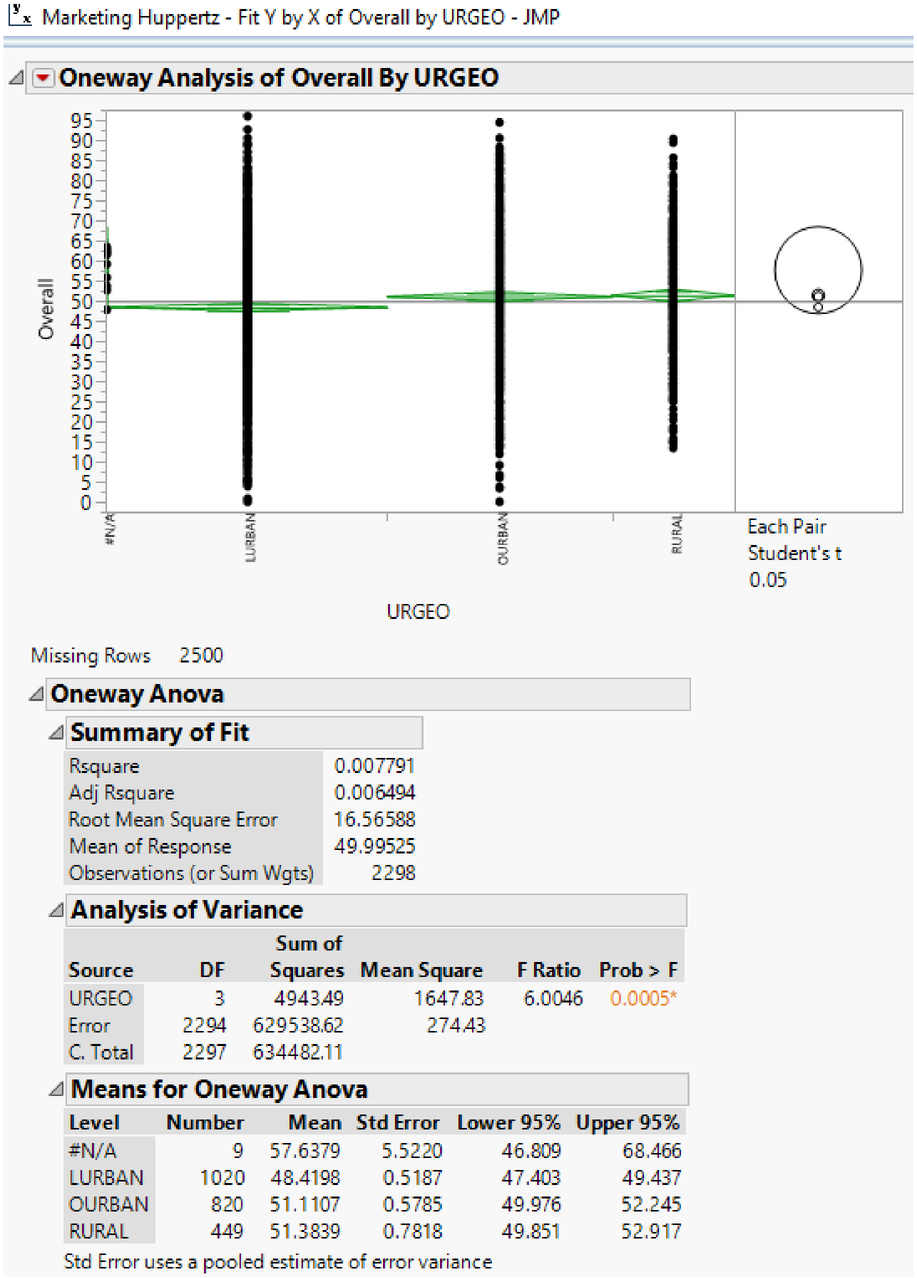
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
TeachingHospital	1	21591.44	21591.4	80.9861	<.0001*
Error	2276	606797.13	266.6		
C. Total	2277	628388.57			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
No	1370	52.5664	0.44114	51.701	53.431
Yes	908	46.2783	0.54187	45.216	47.341

Std Error uses a pooled estimate of error variance

Appendix 24



Marketing Huppertz - Fit Y by X of Overall by URGEO - JMP

Oneway Analysis of Overall By URGEO

Missing Rows 2500

Oneway Anova

Summary of Fit

Rsquare	0.007791
Adj Rsquare	0.006494
Root Mean Square Error	16.56588
Mean of Response	49.99525
Observations (or Sum Wgts)	2298

Analysis of Variance

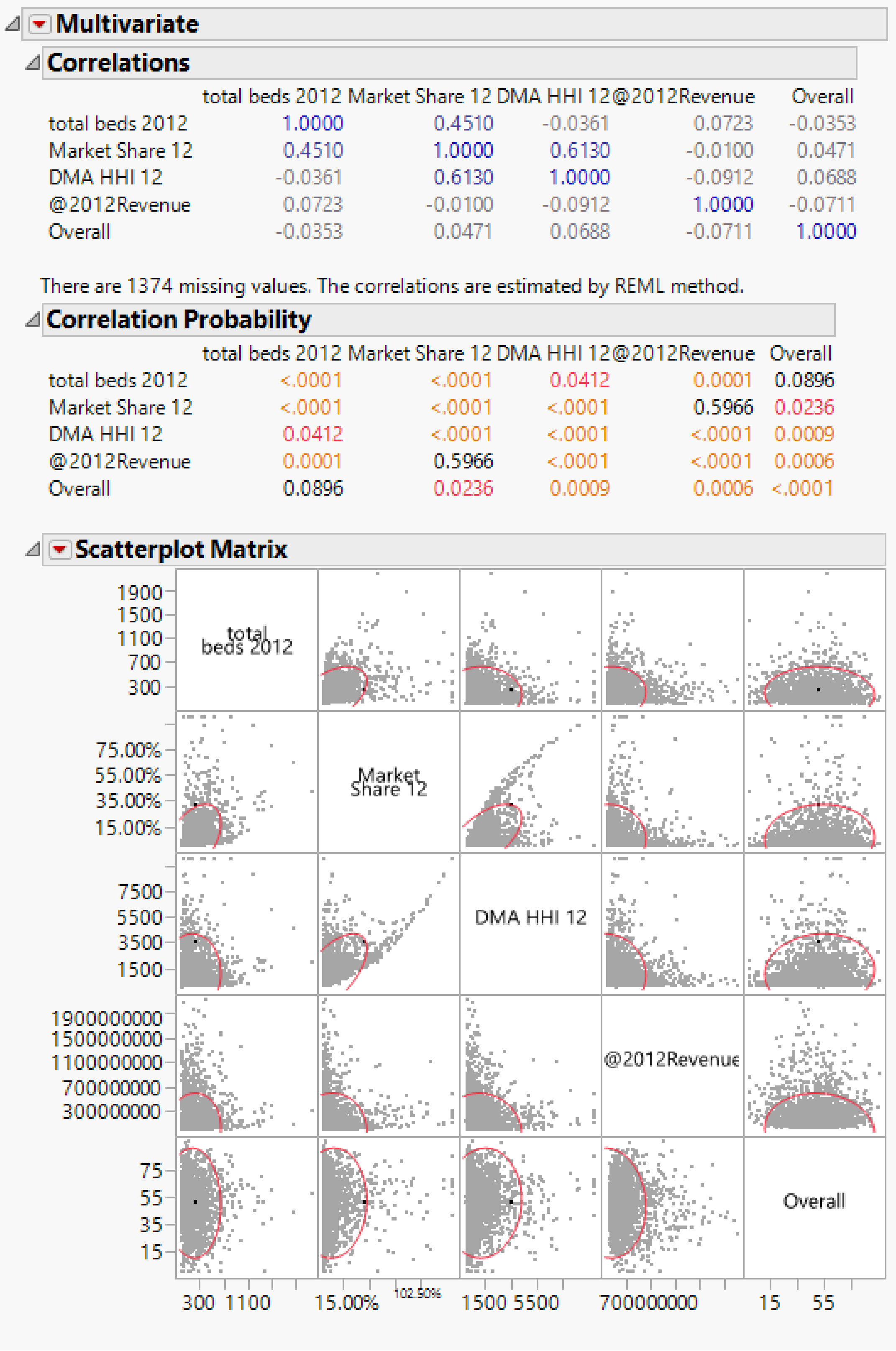
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
URGEO	3	4943.49	1647.83	6.0046	0.0005*
Error	2294	629538.62	274.43		
C. Total	2297	634482.11			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
#N/A	9	57.6379	5.5220	46.809	68.466
LURBAN	1020	48.4198	0.5187	47.403	49.437
OURBAN	820	51.1107	0.5785	49.976	52.245
RURAL	449	51.3839	0.7818	49.851	52.917

Std Error uses a pooled estimate of error variance

Appendix 25



Multivariate

Correlations

	total beds 2012	Market Share 12	DMA HHI 12	@2012Revenue	Overall
total beds 2012	1.0000	0.4510	-0.0361	0.0723	-0.0353
Market Share 12	0.4510	1.0000	0.6130	-0.0100	0.0471
DMA HHI 12	-0.0361	0.6130	1.0000	-0.0912	0.0688
@2012Revenue	0.0723	-0.0100	-0.0912	1.0000	-0.0711
Overall	-0.0353	0.0471	0.0688	-0.0711	1.0000

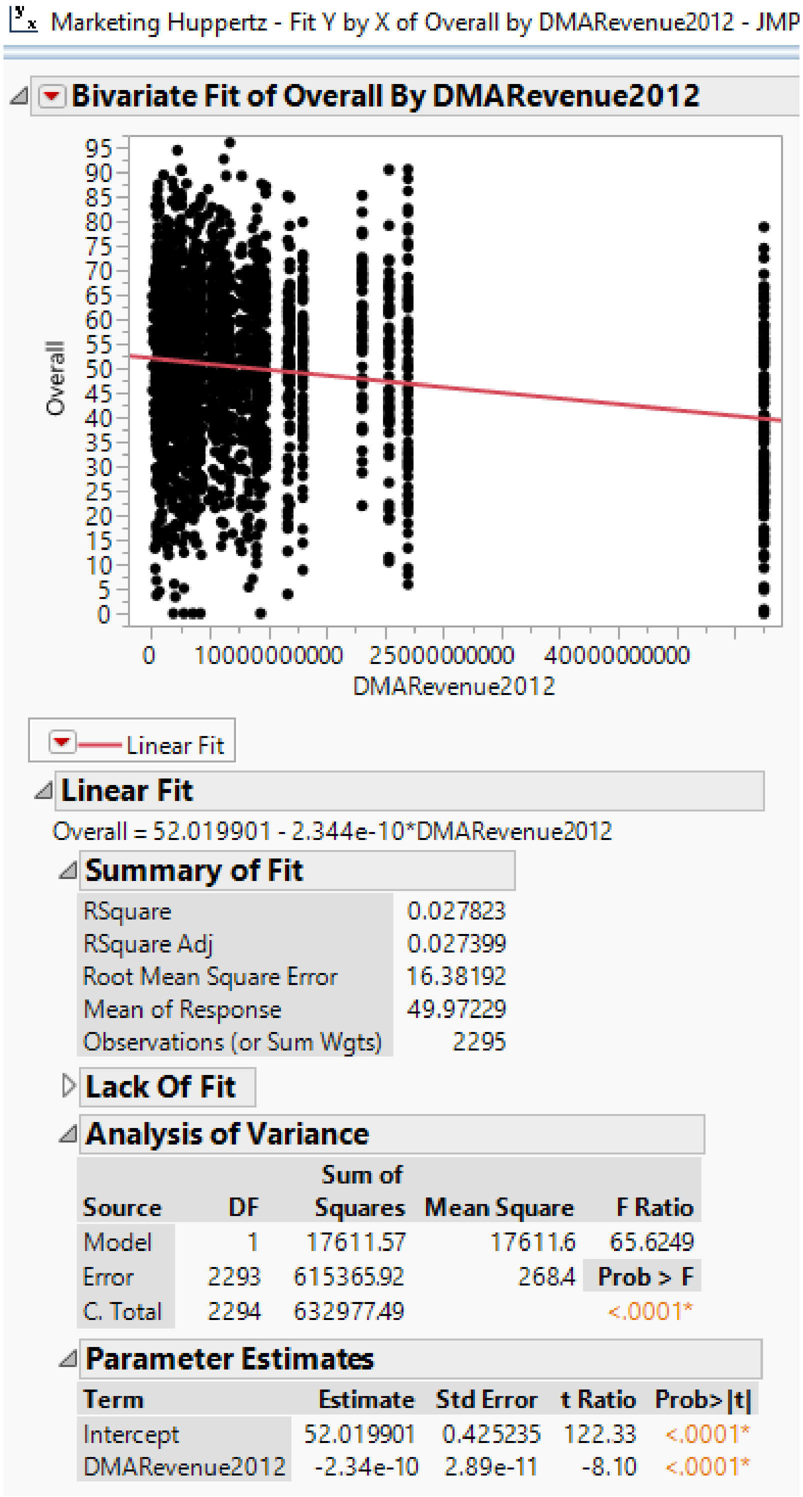
There are 1374 missing values. The correlations are estimated by REML method.

Correlation Probability

	total beds 2012	Market Share 12	DMA HHI 12	@2012Revenue	Overall
total beds 2012	<.0001	<.0001	0.0412	0.0001	0.0896
Market Share 12	<.0001	<.0001	<.0001	0.5966	0.0236
DMA HHI 12	0.0412	<.0001	<.0001	<.0001	0.0009
@2012Revenue	0.0001	0.5966	<.0001	<.0001	0.0006
Overall	0.0896	0.0236	0.0009	0.0006	<.0001

Scatterplot Matrix

Appendix 26



Marketing Huppertz - Fit Y by X of Overall by DMARevenue2012 - JMP

Bivariate Fit of Overall By DMARevenue2012

Linear Fit

Overall = 52.019901 - 2.344e-10*DMARevenue2012

Summary of Fit

RSquare	0.027823
RSquare Adj	0.027399
Root Mean Square Error	16.38192
Mean of Response	49.97229
Observations (or Sum Wgts)	2295

Lack Of Fit

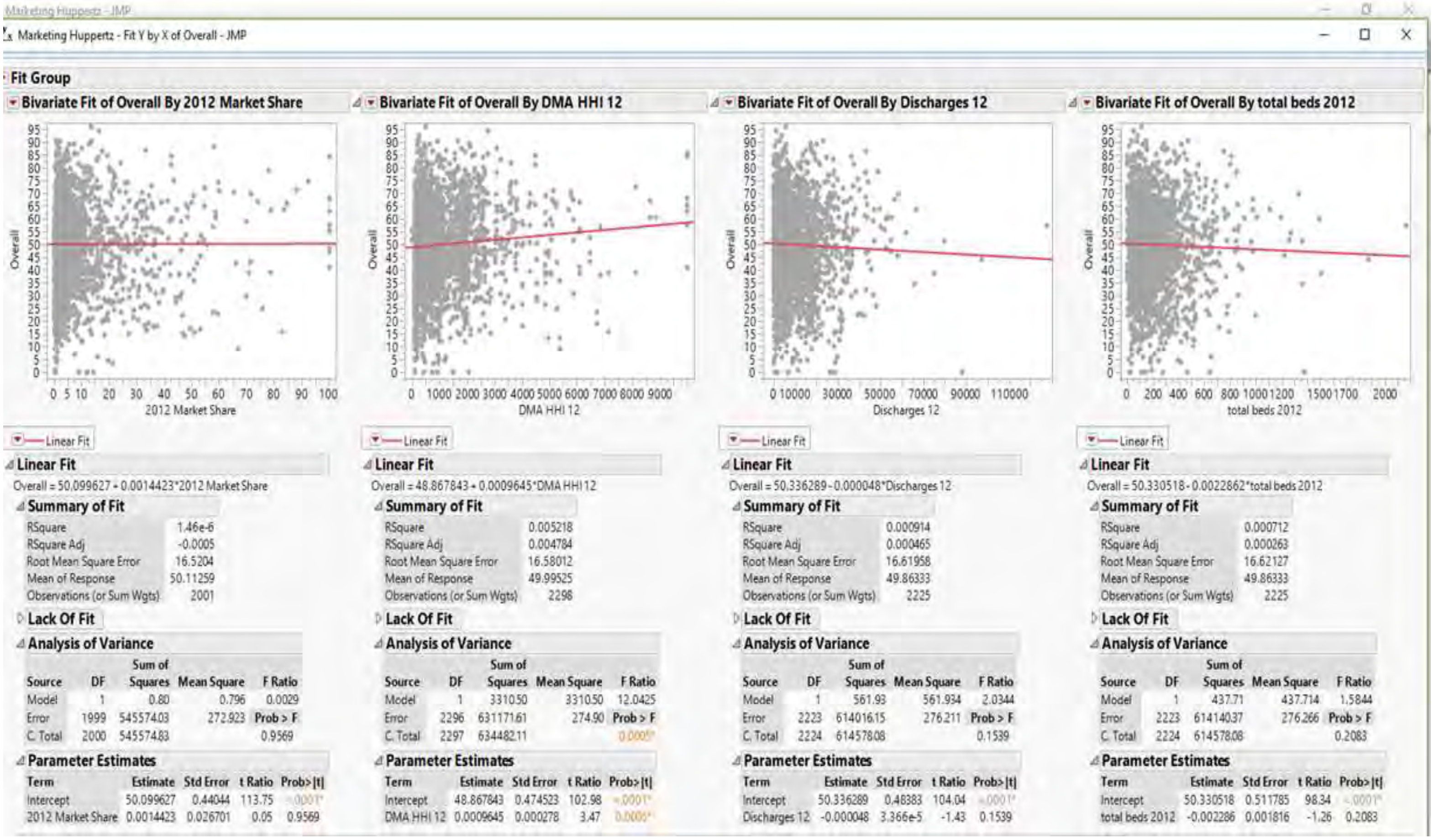
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	17611.57	17611.6	65.6249
Error	2293	615365.92	268.4	**Prob > F**
C. Total	2294	632977.49		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	52.019901	0.425235	122.33	<.0001*
DMARevenue2012	-2.34e-10	2.89e-11	-8.10	<.0001*

Appendix 27



Fit Group

Bivariate Fit of Overall By 2012 Market Share

Linear Fit

Overall = 50.099627 + 0.0014423*2012 Market Share

Summary of Fit

RSquare	1.46e-6
RSquare Adj	-0.0005
Root Mean Square Error	16.5204
Mean of Response	50.11259
Observations (or Sum Wgts)	2001

Lack Of Fit

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	0.80	0.796	0.0029
Error	1999	545574.03	272.923	Prob > F
C. Total	2000	545574.83		0.9569

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	50.099627	0.44044	113.75	<.0001*
2012 Market Share	0.0014423	0.026701	0.05	0.9569

Bivariate Fit of Overall By DMA HHI 12

Linear Fit

Overall = 48.867843 + 0.0009645*DMA HHI 12

Summary of Fit

RSquare	0.005218
RSquare Adj	0.004784
Root Mean Square Error	16.58012
Mean of Response	49.99525
Observations (or Sum Wgts)	2298

Lack Of Fit

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	3310.50	3310.50	12.0425
Error	2296	631171.61	274.90	Prob > F
C. Total	2297	634482.11		0.0005*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	48.867843	0.474523	102.98	<.0001*
DMA HHI 12	0.0009645	0.000278	3.47	0.0005*

Bivariate Fit of Overall By Discharges 12

Linear Fit

Overall = 50.336289 - 0.000048*Discharges 12

Summary of Fit

RSquare	0.000914
RSquare Adj	0.000465
Root Mean Square Error	16.61958
Mean of Response	49.86333
Observations (or Sum Wgts)	2225

Lack Of Fit

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	561.93	561.934	2.0344
Error	2223	614016.15	276.211	Prob > F
C. Total	2224	614578.08		0.1539

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	50.336289	0.48383	104.04	<.0001*
Discharges 12	-0.000048	3.366e-5	-1.43	0.1539

Bivariate Fit of Overall By total beds 2012

Linear Fit

Overall = 50.330518 - 0.0022862*total beds 2012

Summary of Fit

RSquare	0.000712
RSquare Adj	0.000263
Root Mean Square Error	16.62127
Mean of Response	49.86333
Observations (or Sum Wgts)	2225

Lack Of Fit

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	437.71	437.714	1.5844
Error	2223	614140.37	276.266	Prob > F
C. Total	2224	614578.08		0.2083

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	50.330518	0.511785	98.34	<.0001*
total beds 2012	-0.002286	0.001816	-1.26	0.2083

References

To Err is Human: Building a Safer Health System. 1999. Washington, D.C.: Institute of Medicine.

Crossing the Quality Chasm: A New Health System for the 21st Century. 2001. Washington, D.C.: Institute of Medicine.

The National Quality Forum: A National Framework for Healthcare Quality. 2002. Washington, D.C.: NQF.

Quality of Health Care: What is it, Why is it Important, and How Can it Be Improved in California's Workers' Compensation Program? 2003. Santa Monica, CA: RAND Corporation.

"Quality of Care: A Process for Making Strategic Choices in Health Systems."2006. *World Health Organization*.

The Essential Guide to Health Care Quality. 2011. Washington, D.C.: NCQA.

"Consumer Reports Health."2016. Retrieved October 11, 2016. (https://www.google.com/webhp?sourceid=chrome-instant&ion=1&espv=2&ie=UTF-8#q=consumer%20reports%20hospital).

"HealthGrades."2016. Retrieved October 10, 016. (HealthGrades.com).

"HealthyPeople 2020 Social Determinants of Health."2016. *HealthyPeople.Gov*.

"Hospital Compare."2016. Washington, D.C.: Centers for Medicare and Medicaid Services, Retrieved September 10, 2016. (https://www.medicare.gov/hospitalcompare/search.html?).

"LeapFrog Hospital Ratings."2016. Retrieved October 20, 2016. (LeapFrogGroup.org).

"PCMH incentive, recognition, and accreditation programs."2016. American Academy of Family Physicians, Retrieved October 20, 2016. (http://www.aafp.org/practice-management/transformation/pcmh/recognition.html).

"U.S. News and World Report Hospital Rating."2016. , Retrieved October 5, 2016. (http://health.usnews.com/best-hospitals/rankings).

"Vitals."2016, Retrieved November 1, 2016. (vitals.com).

"Yelp Doctors."2016. Retrieved September 20, 2016. (https://www.yelp.com/c/nyc/physicians).

"ZocDoc."2016. Retrieved October 29, 2016. (ZocDoc.com).

Agoritsas, T, A Lubbeke, L Schiesari, and T V Perneger. 2008. "Assessment of Patient's Tendency to Give a Positive Or Negative Rating to Healthcare." *British Medical Journal Quality and Safety* 18(3):374-379.

Anderson, Roger, Angela Barbara, and Steven Feldman. 2007. "What Patients Want: A Content Analysis of Key Qualities that Influence Patient Satisfaction." *Medical Practice Management*: 255-261.

------. 2007. "What Patients Want: A Content Analysis of Key Qualities that Influence Patient Satisfaction." *Medical Practice Management*: 255-261.

Andreassen, Hege and Marianne Trondsen. 2010. "The Empowered Patient and the Sociologist." *Social Theory and Health* 8:280-287.

Arah, Onyebuchi, Gert Westert, Jeremy Hurst and Niek Klazinga. 2006. "A Conceptual Framework for the OECD Health Care Quality Indicators Project." *International Journal for Quality in Health Care*: 5-13.

Attree, M. 2001. "Patients' and Relatives' Experiences and Perspectives of 'Good' and 'Not so Good' Quality Care." *J Adv Nurs* 33(4):456-466.

Backman, Chantal, Saskia Vanderloo, and Alan John Forster. 2016. "Measuring and Improving Quality in University Hospitals in Canada: The Collaborative for Excellence in Healthcare Quality." *Elsevier Health Policy*: 1-5.

Barr, Donald. 2016. *Introduction to US Health Policy, 4th Edition.*Baltimore, MD: Johns Hopkins University Press.

Bastian, Hilda. 2008. "Health Literacy and Patient Information: Developing the Methodology for a National Evidence-Based Health Website." *Elsevier Patient Education and Counseling* 73:551-556.

Batarseh, Feras and Eyad Abdel Latif. 4. "Assessing the Quality of Service using Big Data Analytics with Application to Healthcare." *2016*(13):24.

Berkowitz, Eric. 2010. *The Essentials of Health Care Marketing, 3rd Edition.*Mississauga, Ontario: Jones & Bartlett Learning, LLC.

Berwick, Donald, Thomas Nolan and John and Whittington. 2008. "The Triple Aim: Care, Health and Cost." *Healthaffairs* 27(3):759-769.

Blumenthal, David and Sara R. Collins. 2014. "Health Care Coverage Under the Affordable Care Act - A Progress Report." *New England Journal of Medicine* 371:275-281.

Carayon, Pascale, Tosha Wetterneck, Joy Revera-Rodriguez, Ann Hundt, Peter Hoonakker, Richard Holden, and Ayse Gurses. 2013. "Human Factor Systems Approach to Healthcare Quality and Patient Safety." *Elsevier* 45:14-25.

Chang, John, Ron Hays, Paul Shekelle, Catherine MacLean, David Solmon, David Reuben, Carol Roth, Caren Kamberg, John Adams, and Roy Young. 2006. "Patients' Global Ratings of their Health Care are Not Associated with the Technical Quality of their Care." *Annals of Internal Medicine* 144:655-672.

Chien, Alyna and Meredith Rosentha. 2013. "Medicare's Physician Value-Based Payment Modifier - Will the Tectonic Shift Create Waves?" *New England Journal of Medicine* 369(22):2076-2078.

Chiew, Thiam Kian and Siti Salwa Salin. 2003. "Webuse: Website Usability Evaluation Tool." *Malaysian Journal of Computer Science* 16(1):47-57.

Coulter, Angela. 2005. "What do Patients and the Public Want from Primary Care?" *Bmj* 331:1199.

Cutilli, Carolyn. 2010. "Seeking Health Information: What Sources do Your Patients use?" *Orthopedic Nursing* 29(3):214-219.

Detsky, Allan. Journal of the American Medical Association. "What Patients Really Want from Healthcare." *2011* 306(22):2500-2501.

Donabedian, Avedis. 1965. "Evaluating the Quality of Medical Care." *The Millbank Quarterly* 83(4):691-729.

Donaldson, Molla. 1999. *Measuring the Quality of Health Care.*Washington D.C.: Institute of Medicine.

Emanuel, Ezekiel, and Linda Emanuel. 1992. "Four Models of the Physician-Patient Relationship." *Jama* 267(16):2221-2226.

Enthoven, A. C. 2014. *Theory and Practice of Managed Competition in Health Care Finance.*Elsevier Science.

Eysenbach, Gunther. 2001. "Evidence-Based Patient Choice and Consumer Health Informatics in the Internet Age." *Journal of Medical Internet Research* 3(2):19-27.

Eysenback, Gunther. 2002. "Empirical Studies Assessing the Quality of Health Information for Consumes on the World Wide Web: A Systematic Review." *Journal of the American Medical Association* 287(20):2691.

Eysenback, Gunther and Thomas Diepgen. 2001. "The Rol Eof e-Health and Consumer Health Informatics for Evidence-Based Patient Choice into the 21st Century." *Elsevier* 19:11-17.

Friedman, Thomas. 2005. *The World is Flat: A Brieft History of the Twenty-First Century.*New York, New York: Farrar, Straus and Grioux.

Gehrke, Dave and Efraim Turban. 1999. "Determinants of Successful Website Design: Relative Importance and Recommendations for Effectiveness." *Hawaii Conference on System Sciences* 32:1-8.

Gosfield AG and JL Reinersten. 2005. "The 100,000 Lives Campaign: Crystallizing Standards of Care for Hospitals." *Health Affairs* 24(6):1560-1570.

Grol, Richard, Michel Wensing, Jan Mainza, Pedro Ferreirab, Hilary Hearnshawc, Per Hjortdahld, Frede Olesena, Mats Ribackee, Tomi Spenserf, and Joachim Szecsenyig. 2016. "Patients' Priorities with Respect to General Practice Care: An International Comparison." *Oxford Journals* 16(1):4-11.

Haas-Wilson, Deborah. 2001. "Arrowand the Information Market Failure in Healthcare: The Changing Content and Sources of Health Care Information." *Journal of Health Politics, Policy and Law* 26(5):1031-1044.

Henwood, Flis, Sally Wyatt, Angie Hart and Julie and Smith. 2003. "'Ignorance is Bliss Sometimes': Constraints on the Emergence of the 'Informed Patient' in the Changing Landscape of Health Information." *Sociology of Health & Wellness* 25(6):589-607.

Holahan, John and Rebecca Peters. 2014. *The Launch of the Affordable Care Act in Selected States: Insurer participation, Competition, and Premiums.* Washington, D.C.: Urban Institute.

Kelley, Edward and J. Hurst. 2006. "OECD Health Working Papers no. 23 Health Care Quality Indicators Project." *OECD Publishing*.

Koh, Howard, Donald Berwick, Carolyn Clancy, Cynthia Baur, Cindy Brach, and Linda Harriss. 2012. "New Federal Policy Initiatives to Boost Health Literacy can Help the Nation Move Beyond the Cycle of Costly 'Crisis Care'." *Health Affairs* 31(2):434-443.

Lawrence D. Brown. 2003. "Comparing Health Systems in Four Countries: Lessons for the United States." *Am J Public Health* 93(1):52-56.

Lin, Hsiu-Fen. 2010. "The Impact of Website Qualty Dimensions on Customer Satisfaction in the B2C E-Commerce Contextt." *Total Quality Management & Business Excellence* 18(4):363-378.

Malat, Jennifer. 2001. "Social Distance and Patient's Rating of Healthcare Providers." *Journal of Health and Social Behavior* 42(4):360-372.

Martin, Graham, Lorna McKee, and Mary Dixon-Woods. 2015. "Beyond Metrics? Utilizing "Soft Intelligence' for Healthcare Quality and Safety." *Elsevier Social Science and Medicine* 142:19-26.

Mascie-Taylor and Karim E. 2003. "The Burden of Chronic Disease." *Pubmed Ncbi* 12(302):1921-1922.

Matthew, Austin, Guy D'Andrea, John Birkmeyer, Lucian Leape, Arnold Milstein, Peter Pronovost, and Patrick Romano. 2014. "Safety in Numbers: The Development of Leapfrog's Composite Patient Safety Score for U.S. Hospitals." *Journal of Patient Safety* 10(1):64-71.

McHugh, Matthew and Amy Witoski. 2012. "Nurse Reported Quality of Care: A Measure of Hospital Quality." *Research in Nursing & Health* 35(6):566-575.

Mitchell, PH, S Ferketich and BM Jennings. 1998. "Quality Health Outcomes Model. American Academy of Nursing Expert Panel on Quality Health Care." *Image J Nurs Sch* 30(1):43-46.

Mosadeghrad, Ali M. 2012. "A Conceptual Framework for Quality of Care." *Mat Soc Med.* 24(4):251-261.

Nayak, Jaxman, Lee Priest, Ian Stuart-Hamilton, and Allan White. 2006. "Website Design Attributes for Retrieving Health Information by Older Adults: Ann Aplication of Architectual Criteria." *Universal Access in the Information Society* 5(2):170-179.

Nelson, Judith, Kathleen Puntillo, Peter Pronovost, Amy Walker, Jennifer McAdam, Debra Ilaoa, and Joan Penrod. 2010. "In their Own Words: Patients and Families Define High-Quality Palliative Care in the Intensive Care Unit." *Critical Care Medicine* 38(3):808-818.

Nielsen-Bohlman, Lynn, Allison Panzer and David Kindig. 2004. *Health Literacy: A Prescription to End Confusion.*Washington, D.C.: Institute of Medicine of the National Academies.

Oberlander, Jonathan. 2014. "Between Liberal Aspirations and Market Forces: Obamacare's Precarious Balancing Act." *Journal of Law, Medicine, & Ethics*:431-441.

O'Connor, Annette, John E. Wennberg, France Legare, Hilary A. Llewellyn-Thomas, Benjamin Moulton, Karen Sepucha, Andrea Sodano, and Jaime King. 2007. "Toward the 'Tipping Point': Decision Aids and Informed Patient Choice." *Health Affairs* 26(3):716-725.

Pak, Richard, Margaux Price, Jason Thatcher. 2009. "Age-Sensitive Design of Online Health Information: Comparative Usability Study." *Journal of Medical Internet Research* 11(4):45.

Park, Geun-wan, Yongse Kim, Kwangtae Park, Anurag Agarwal. 2016. "Patient-Centric Quality Assessment Framework for Healthcare Services." *Elsevier*.

Park, H., P.J. Murray, C. Delaney. 2006. *Consumer-Centered Computer-Supported Care for Healthy People.*Amsterdam, Netherlands: IOS Press.

Pennbridge, Julia, Rita Moya, and Lakeshia Rodrigues. 1999. "Questionnaire Survey of California Consumers' use and Rating of Sources of Health Care Information Including the Internet." *Western Journal of Medicine* 171:302-307.

Petre, Marian, Shailey Minocha and Dave Roberts. 2011. "Usability Beyond the Website: An Empirically-Grounded e-Commerce Evaluation Instrument for the Total Customer Experience." *Behavior and Information Technology* 25(2):189-203.

Quanstrum, Kerianne and Rodney Hayward. 2010. "Lessons from the Mammography Wars." *New England Journal of Medicine* 363:1076-1079.

Rademakers, Jany, Diana Delnoij, Dolf de Boer. 2010. "Structure, Process Or Outcome: Which Contributes most to Patients' overall Assessment of Healthcare Quality?" *British Medical Journal Quality and Safety* 21(4).

Reinhardt, Uwe. 2013. "The Disruptive Innovation of Price Transparency in Health Care." *Journal of the American Medical Association* 310(18):1927.

Rezaeenour, Amini. 2016. "Ranking Healthcare Centers using Fuzzy Analytic Hierarchy Process and TOPSIS: Iranian Experience." *International Journal of Applied Operational Research* 6(1):25-39.

Richards, Bernard and Andrew Colman, and Robert Hollingsworth. 1998. "The Current and Future Role of the Internet in Patient Education." *International Journal of Medical Informatics* 50:279-285.

Ringel, Jeanne, Susan Hosek, Ben Vollaard, and Sergej Mahnovski. 2010. "The Elasticity of Demand for Health Care: A Review of the Literature and its Application to the Military Health System." *RAND Health National Defense Research Institute*:1-50.

Rothenfluh, Fabia, Evi Germeni, and Peter Schulz. 2016. "Consumer Decision-Making Based on Review Websites: Are there Dfferences between Choosing a Hotel and Choosing a Physician?" *Journal of Medical Internet Research* 18(6):129.

Saltelli, Andrea. 2007. "Composite Indicators between Analysis and Advocacy." *Social Indicators Research* 81(1):65-77.

Schang, Laura, Yrjana Hynninen, Alec Morton, and Ahti Salo. 2015. "Developing Robust Composite Measures of Healthcare Quality - Ranking Intervals and Dominace Relations for Scottish Health Boards." *Elsevier* 162:59-67.

Schroeder, Steven and William First. 2013. "Phasing Out Fee-for-Service Payment." *New England Journal of Medicine* 368:2029-2032.

Shahian, David, Sean O'Brien, Sharon-Lise T. Normand, Eric Peterson, and Fred Edwards. 2009. "Association of Hospital Coronary Artery Bypass Volume with Processes of Care, Mortality, Morbidity, and the Society of Thoracic Surgeons Composite Quality Score." *Acquired Cardiovascular Disease* 139(2):273-281.

Sofaer, Shoshanna, Christine Crofton, Elizabeth Goldstein, Elizabeth Koy, and Jenny Crabb. 2005. "What do Consumers Want to Know about the Quality of Care in Hospitals?" *Health Serve Res* 40(6):2018-2036.

Stvilia, Besiki, Lori Mon, and Yong Jeong Yi. 2009. "A Model for Online Consumer Health Information Quality." *Journal of the American Society for Information Science and Technology* 60(9):1781-1791.

Tullis, Thomas and Jacqueline N. Stetson. 2004. *A comparison of questionnaires for assessing website usability.* Boston, MA: UPA.

VanLare, Jordan and Patrick Conway. 2012. "Value-Based Purchasing - National Programs Move from Volume to Value." *New England Journal of Medicine* 367(4):292-295.

Verhoef, Lise, Tom Van de Belt, Lucien JLPG Engelen, Lisette Schoonhoven, and Rudolf, Kool. 2014. "Social Media and Rating Sites as Tools to Understanding Quality of Care: A Scoping Review." *Journal of Medical Internet Research* 16(2):56.

Vogus, Timothy and Laura E. McClelland. 2015. "When the Customer is the Patient: Lessons from Healthcare Research on Patient Satisfaction and Service Quality Ratings." *Elsevier* 26:37-49.

Williams, Peter, David Nicholas, Paul Huntington, and Fiona McLean. 2002. "Surfing for Health: User Evaluation of a Health Information Website." *Blackwell Science Ltd* 19:98-108.

Williams, Susan, John Weinman, Jeremy Dale, and Stanton Newman. 2016. "Patient Expectations: What do Primary Care Patients Want from the GP and how Far does Meeting Expectations Affect Patient Satisfaction?" *Oxford Journals* 33(5):193-201.