Doctors and the Cash Economy: Does Compassion Count?

RUNNING HEAD: Doctors and the Cash Economy

Louis R. Nemzer, Ph.D.
Fulbright Distinguished Scholar, Biology Chair[1]
Associate Professor [2, 3]

Florence Neymotin, Ph.D.
Fulbright Distinguished Scholar, Entrepreneurship Chair[4]
Associate Professor[5, 6]

Declarations of interest: none

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

[1] Carleton University, 1125 Colonel By Drive, Ottawa, ON, K1S 5B6
[2] Nova Southeastern University, 3301 College Ave., Fort Lauderdale, FL, USA, 33314,
[3] louisiqn@gmail.com
[4] Carleton University, 1125 Colonel By Drive, Ottawa, ON, K1S 5B6
[5] Nova Southeastern University, 3301 College Ave., Fort Lauderdale, FL, USA, 33314
[6] (Corresponding Author) neymotin@gmail.com

Doctors and the Cash Economy: Does Compassion Count?

Abstract: We examine how the patient reviews and scores for internists and family doctors in the states of California and Florida differ based upon status as a concierge doctor. Data are drawn from Healthgrades.com and Vitals.com, two of the largest providers of online reviews. A machine learning sentiment analysis is used to determine the predictors of concierge status and numerical patient ratings. We find that concierge doctors are most highly associated with technical health care words, while non-concierge doctors are more highly associated with interpersonal "bedside manner" skills. We further determine that technical factors have a stronger association with patient scores for all doctors. The present work represents a first step towards understanding the patient experience aspect of quality of care as related to the growing field of concierge medicine. It is also the first we are aware of to employ sentiment analysis in this context.

Keywords: Concierge Medicine; Interpersonal Skills; Sentiment Analysis; Patient Reviews; Machine Learning

JEL Codes: I111; I130; Y80

1. INTRODUCTION

As the cost of medical school and physician student debt keeps growing, so have the costs of health care generally (Dai and Taiyur, 2018). At the same time, the advantages of primary care work compared with other medical subspecialties have become negative (Okie, 2012; Zabar et al., 2019). Students have increasingly moved away from primary care specialization due to falling returns to their time, and the simultaneous increased pressure for efficiency and to provide treatment to as many patients as quickly as possible (Hartzbrand and Groopman, 2016). Rather than spending time with each patient and providing a thorough and precise model of care, doctors feel forced to provide faster and "thinner" levels of treatment (Tai-Seale and McGuire, 2011). Concierge care - in which patients pay a retainer to their doctor in return for prioritized attention - has emerged as a potential solution for both health care providers and patients, and it is the focus of the present analysis.

While uncertainty surrounds the medical profession generally as to the future course of medicine and of primary care in particular in the face of this issue (Okie, 2012), a few points are beyond dispute. First, the shortage of primary care physicians in any one area relates to lower levels of population health in that area (Zabar et al., 2019), and for that, a solution needs to be found. Second, physicians do seem to respond to financial incentives, as doctors are increasingly

interested and moving towards concierge medicine, or “cash-based” care systems. They are doing so to alleviate financial pressure and as an opportunity to see fewer patients for longer and more involved stretches of time (Carnahan, 2006; Hartzbrand and Groopman, 2009; Hartzbrand and Groopman, 2016). So far, this has been particularly true of family doctors and internists, but this model of care is also beginning to reach other specialties (Calhoun, 2015).

In concierge care, an individual is asked to pay an up-front, yearly fee to obtain service from their doctor. These fees vary, with some as high as $3,000 or $13,000 per year, while other models are geared more towards middle- or lower-income individuals, costing closer to $600 or $1,500 per year (Brown, 2009; Carnahan, 2006). Interestingly, the low-cost models seem to be growing with the increased interest in personalized medicine and care (Basik et al., 2011; Calhoun, 2015; Del Puente et al., 2015; Tetreault et al., 2014; Wieczner, 2013). Given concierge care’s apparent status as a non-luxury economic good, the low-cost model of care appears to have made some inroads (Sack, 2009).

The models also vary based upon whether the doctor is willing to take insurance fully or partially—some doctors have moved to concierge care for the express purpose of avoiding insurance claims and paperwork entirely (Carnahan, 2006)—as well as whether the doctor started out as concierge or transitioned to it over time. In the latter case, there are additional questions regarding how patients who were already served by the doctor are treated after the transition of the practice (Carnahan, 2006).

This new model of care also raises many potential legal issues, such as accepting Medicare in the practice, equal treatment of patients in split practices (concierge and non-concierge), rules for transitioning patients when the practice changes (French et al., 2010), and intertwined relationships with specialists, just to cite a few examples. (Carnahan, 2006). Nevertheless, so far it has been allowed to continue its growth path (Carnahan, 2006). If concierge medicine, or “direct primary care”, maintains its increasing levels of awareness and growth (AAFP, 2015), the implications of its presence do need to be understood.

One might also wonder how such care should be evaluated (Carnahan, 2006). To this end, the current work employs online patient reviews from the websites Vitals.com and Healthgrades.com, two of the largest providers of online physician ratings. We focus on internists and family doctors in the states of California and Florida. While these reviews may or may not correlate with quality of care (Gray et al, 2015), particularly in the United States

(Greaves et al., 2012), and some are certainly untrustworthy (Kadry et al., 2011; Shukla et al, 2018), people do seem to pay attention to these online reviews in choosing their doctor—although maybe not as much as in choosing a durable good, like a car (Hanauer et al., 2014; Sharma et al., 2016; Wallace et al., 2014).

Furthermore, while prospective patients may only consider to the numeric portion of the review (Alodadi and Zhou, 2016; Kadry et al., 2011), this is also very highly related to, and predicted by, the specific semantics used in the written portion of the review (Alemi et al., 2012; Greaves et al, 2013). Finally, while these reviews might not relate to "quality of care" per se, they do relate to the patient experience portion of care evaluations (Gray et al., 2015). As such, understanding how reviews vary, and using them to decipher patterns of patient experience, is an important facet of understanding patterns in the evolving field of concierge care.

What we find is that patients of concierge doctors use more "technical" and "system-related" words to describe their care, as contrasted with non-concierge doctors' patients, who focus much more on "interpersonal" factors. We find that other predictors of either concierge status or of online scores are generally consistent with the previous literature. For example, we find that women have worse patient ratings (Wallace and Paul, 2016), and concierge doctors appear to try to look more prestigious by using their middle or extra (Jr/Sr, I/II/III) names compared with non-concierge doctors.

Our empirical approach, which focused on a machine learning sentiment analysis, represents a growing field in economics (Oster, 2018). While machines and programs cannot build the model (Heckman and Singer, 2017), their use has evolved into a field that deciphers patterns particularly important for policy and predictions (Bajari et al., 2015; Mullainathan and Spies, 2017). In this sense, as an emerging field ripe for policy intervention, and whose future is entirely unknown, concierge care is a perfect target for machine learning (Krumholz, 2014).

Although ours is not a controlled sample, and it relies on patients who both "chose" concierge/non-concierge care, as well "chose" to fill out reviews, we feel that the current work is a crucial first step to understanding the evolving aspects of concierge care. As the very first study that we are aware of to undertake an analysis of the semantics and ratings of concierge physicians employing a machine learning sentiment analysis, we feel that this work will be a crucial stepping stone to later policy analyses and choices.

2. METHODS

2.1 Data

In March and April of 2019, we gathered information from Vitals.com and Healthgrades.com to create a database with 550 doctors and 2,673 patient reviews. Of these, 152 were concierge doctors, with an aggregate 852 reviews, who were listed on the MDVIP.com website. The remaining 400 were non-concierge doctors who received the other 1,821 reviews. Current review data were coded and hand-checked to ensure accuracy. We chose to focus on MDVIP doctors because they are one of the largest concierge doctor networks available for public viewing and sorting (GAO, 2005; French et al., 2010). They are based in Boca Raton, FL (Calhoun, 2015) and previously commissioned a study of their doctors and patients to show that health care is superior for these patients, and moreover, that this is not due to selection (O'Brien, 2013). This is despite their special efforts to select good doctors with high ratings and expensive patients (Brown, 2009). We also focused on MDVIP because most of their doctors are in the US states of interest where concierge care is seen as very large and growing (Calhoun, 2015; Carnahan, 2006)—namely, Florida and California. It is possible that a few of the other doctors in the sample are also concierge-based, but we believe our randomized efforts would be unlikely to find such doctors, given that concierge rates are still generally low at this time.

Due to the prevalence of concierge doctors in these two professions, observations were limited to doctors listed as either specializing in "family medicine" or "internal medicine" (GAO, 2005), and located in either California or Florida. The same cities within these two states were sampled for the non-concierge doctors as for the concierge doctors listed in the MDVIP database. Since the non-concierge group was potentially quite large, we simply used an alphabetical structure to randomly choose which doctors to include within a specific city.

Because we had a very specific and select dataset of doctors, we employed techniques to ensure the most complete information possible. We initially used information from Vitals.com as well as Healthgrades.com, filling in information from one database for the same doctor to amend gaps from the other when necessary. In terms of the individual reviews, we favored Healthgrades.com when possible, since it usually had a more complete exposition of patient reviews. Due to the possibility that concierge doctors were selected for inclusion based on age and initial abilities, we only chose non-concierge doctors who were between forty and seventy

years old, and whose ratings were at least 3.5 on the five-point rating scale. Robustness checks determined that this sampling restriction did not bias our conclusions.

In addition to the doctor's full name, city, state, age, and gender, their medical school, location of their medical residency, and listed languages spoken were potentially employed in the analysis as control factors.[7]

Specifically, gender was included in a binary fashion, while age was included continuously. Language was tracked by total fluency as well as Spanish-speaking ability. The doctor had an "extra name" if they had any of a middle name or initial, a Jr./Sr., or a "I," "II," or "III after their name. Medical schools were ranked from 1-100 based on U.S. News 2019 Medical Research School rankings.[8] Schools outside the United States or whose rank was lower than 100 were similarly grouped together. Schools were then recoded, with the top-ranked school in the list (Harvard Medical School) receiving a "score" of 100, and the unranked, non-U.S., or below 100 ranked schools received a score of 0 where SCORE=101-RANK. We believe this generally captures the concept of name-recognition and school rankings for US-based institutions.

We attempted to further control for the "rank" of medical residencies and employ this as a control, however, it was apparent that residencies are generally not chosen based on school ranking nationally. Instead, we found that doctors involved in the medical residency matching system more generally consider a multitude of other factors such as proximity, specialization, or a mentor at that institution, etc. Therefore, we felt that creating comparable choices and rankings between individuals would be questionable at best.

Finally, since we did not want any one doctor's reviews to have a disproportionate weight in the machine learning sentiment analysis, we only included the first 1-19 patient reviews for any specific doctor.[9]

2.2 Summary Statistics

---

[7] Our choice of controls was partially motivated by authors such as Gao et al.(2012) who found that ratings are typically higher for doctors who are younger, and went to better medical schools.

[8] Due to issues in differentiating rankings, we grouped together the US News rankings of 93-120 at the same ranking of 100 for all relevant schools. For more on the rankings, see: https://www.usnews.com/best-graduate-schools/top-medical-schools/research-rankings

[9] Less than 5% of the doctors had more than 19 reviews, and the numbers sharply dropped off to only three doctors with more than 30 reviews and one outlier doctor with 167 reviews.

Table 1 displays the means and counts and stratifies by concierge status in the dataset. It is apparent that gender and state of residence are balanced between the concierge and the non-concierge doctors in the data, as we would expect from a randomization procedure. Our selection by age and rating is also apparent from the relatively higher rating of the non-concierge doctors and their slightly younger ages. We did not, however, select on school, as we were not entirely comfortable with employing the US News ranking as a selector (even though this appears to be the best possible choice), and we can see that schools may be slightly better in the concierge sample of doctors. We also did not select on family medicine or internal medicine to mirror the concierge sample, and so our relatively equal set (51% family medicine) for non-concierge doctors, varies from the concierge sample where internists are more prevalent. Interestingly, the concierge doctors speak fewer languages overall—with zero indicating English-only speakers—as well as being less likely to speak Spanish in particular. They are also extremely likely to use an extra name (80%), whether it is a middle, Jr/Sr, or I/II/III, relative to the non-concierge sample, where this almost never occurs (3%).

Table 1: Summary Statistics

	Non-Concierge	*Concierge*	*Total*
Age	59.33	60.48	59.63*
Female	0.26	0.22	0.24
Rating	4.54	4.34	4.49**
School Score	26.57	30.76	27.72+
Florida	0.61	0.55	0.59
Family Med.	0.51	0.39	0.47**
# Languages	0.75	0.43	0.66**
Spanish	0.37	0.27	0.34**
Extra Name	0.03	0.80	0.24**

Note: Means are listed, and N=400 for column 1 of almost all variables while N=152 for column 2 of almost all variables. For 1-sided T-tests of means, + denotes significance at the 10% level, * denotes significance at the 5% level and ** denotes significance at the 1% level.

We interpret a number of these differences as signals of perceived quality that could have been selected for by MDVIP in building their network of doctors. Extra names, higher school scores, and older doctors (particularly males) may be perceived as being more prestigious. In

contrast, the lack of language ability in the concierge sample was not immediately explainable, but we believe this may relate to socioeconomic status in the population and ability to pay.

Taken together, Table 1 describes a concierge group of doctors who are less diverse in language ability, more "prestigious"-seeming, but not necessarily having higher ratings. We continue this analysis in a more rigorous fashion in the sections that follow.

3. RESULTS

3.1 Non-Semantic Regressions

Prior to the semantic analysis, we first consider the effects of various controls on both the likelihood of doctors presenting as concierge in the sample, and the effect of controls on the doctor's ratings.

For doctor i, we have:

$$Outcome_i = f(Demog_i, Specialty_i, Geog_i, Prestige_i, Lang_i, Specialty_i, NumReviews_i)$$

Where *Demog* represents the demographic characteristics of gender and age; *Specialty* is limited to a binary for whether the individual is a family doctor versus an internist; *Prestige* is represented by whether the doctor uses an extra name as well as the US News score of their medical school; *Lang* represents the ability of the doctor to speak non-English languages, and in particular Spanish; *Numreviews* is the number of reviews for each doctor, with possible values of 1-19. *Outcome* can either be a binary for concierge status—in which case we run a Logistic regression, or it is a continuous 1-5 average doctor rating—in which case we use an Ordinary Least Squares (OLS) regression. We additionally control for either concierge status when ratings are the outcome, or ratings when concierge is the outcome variable.

Table 2: Relationship between Controls and Doctor Ratings

	(1)	(2)	(3)	(4)	(5)
Female	-0.107	-0.109	-0.113	-0.123	-0.124
	(2.48)*	(2.57)*	(2.64)**	(2.90)**	(2.92)**
Age	0.058	0.031	0.031	0.030	0.012
	(1.60)	(0.84)	(0.85)	(0.84)	(0.31)
Age ^{2}	-0.000	-0.000	-0.000	-0.000	-0.000
	(1.54)	(0.73)	(0.74)	(0.76)	(0.26)
Fam. Doc	-0.082	-0.099	-0.100	-0.097	-0.103
	(2.24)*	(2.72)**	(2.74)**	(2.68)**	(2.83)**
Florida	-0.021	-0.026	-0.027	-0.000	-0.012

	(0.58)	(0.68)	(0.71)	(0.00)	(0.29)
Extra Nm		-0.210	-0.215	-0.202	-0.082
		(4.80)**	(4.86)**	(4.59)**	(1.09)
US News		-0.000	-0.000	-0.000	-0.000
		(0.35)	(0.30)	(0.38)	(0.41)
# Lang			-0.008	-0.000	-0.007
			(0.30)	(0.01)	(0.24)
Spanish			-0.030	-0.039	-0.034
			(0.65)	(0.86)	(0.74)
# Rev.				-0.012	-0.011
				(2.89)**	(2.68)**
Concierge					-0.146
					(1.98)*
Constant	2.803	3.610	3.625	3.723	4.300
	(2.60)**	(3.37)**	(3.37)**	(3.49)**	(3.89)**
R^2	0.03	0.07	0.07	0.09	0.09

N=540; * $p<0.05$; ** $p<0.01$

Table 2 displays results of the OLS regression where physician ratings are the outcome variable, and controls are added in a progressive fashion. We see from Table 2 that women, family doctors, doctors with more reviews, and – interestingly – concierge doctors have lower average ratings. While it is possible that some of this reflects our sample selection, or the fact that certain individuals choose to go into concierge work, we cannot rule out the possibility that there is a causal relationship of how doctors are viewed by patients, or how certain doctors choose to present their practices. Furthermore, the fact that women have lower reviews is in line with the literature, and the fact that internists receive higher reviews may also be expected. We see that having an extra name is initially related to ratings, however, including concierge status diminishes the effects of extra name to insignificance. Our other measure of status, school scores, does not appear to matter. It is also important to note that state of residence, age, and languages spoken, have no bearing on patient reviews. While we expected state of residence to have little impact, we were surprised that polylingual, and particularly Spanish-speaking doctors, received no boost or downgrade. Age making no difference was also heartening, as it implies a more even playing field than we might have expected.

Table 3: Relationship between Controls and Concierge Status

	(1)	(2)	(3)	(4)	(5)
Female	-0.050	0.018	-0.098	-0.015	-0.118

	(0.20)	(0.04)	(0.23)	(0.03)	(0.26)
Age	-1.196	-1.444	-1.581	-1.579	-1.553
	(5.47)**	(4.30)**	(4.47)**	(4.50)**	(4.37)**
Age ^{2}	0.010	0.012	0.013	0.013	0.013
	(5.58)**	(4.10)**	(4.27)**	(4.31)**	(4.21)**
Fam. Doc	-0.569	-0.716	-0.815	-0.854	-0.955
	(2.68)**	(1.87)+	(2.10)*	(2.16)*	(2.36)*
Florida	-0.324	-0.967	-1.073	-1.314	-1.341
	(1.56)	(2.41)*	(2.64)**	(3.06)**	(3.08)**
Extra Name		5.562	5.586	5.618	5.568
		(12.25)**	(11.98)**	(11.84)**	(11.67)**
US News Score		0.002	0.001	0.000	-0.001
		(0.31)	(0.17)	(0.05)	(0.13)
# Lang.			-0.662	-0.690	-0.647
			(2.33)*	(2.38)*	(2.21)*
Span. Lang.			0.463	0.590	0.555
			(0.96)	(1.20)	(1.12)
# Reviews				0.085	0.078
				(2.37)*	(2.10)*
Doc Rating					-0.797
					(1.70)+
Constant	33.739	41.397	45.979	45.331	48.099
	(5.23)**	(4.28)**	(4.48)**	(4.45)**	(4.58)**
Pseudo R^2					

N=540; * $p<0.05$; ** $p<0.01$

In Table 3, we rerun the same analysis, but employ concierge status as the outcome variable in our Logistic regression. The results are consistent with the idea that older doctors are more likely to go into concierge work, as are internists. Concierge doctors are also much less likely to speak additional languages—perhaps because of their target clientele, and there is a slight bias towards California versus Florida. Perhaps most importantly, we find that, while concierge doctors have more reviews, they tended to have worse ratings (significant at the 10% level). Also important, gender does not relate to concierge status, nor does school rank or Spanish-language ability.

Taken together, Tables 2 and 3 paint a picture of more elite “sounding” doctors going into concierge work, but not necessarily receiving better reviews. School ranking doesn’t matter either to patients in their reviews or to doctors in how they choose to go into concierge work. Gender doesn’t affect the choice to be a concierge doctor, but it does impact ratings. To determine the effect of semantics, we next employ a machine learning sentiment analysis.

3.2 Machine Learning and Semantics

We use a machine learning sentiment analysis to determine which words are most related with a physician's concierge status and numerical score. First, each doctor's patient text reviews are concatenated, and the word frequencies are tabulated. To identify the most critical words, the word counts for each physician are converted into vectors and normalized using "term frequency–inverse document frequency" (tf-idf) weighting. The purpose of tf-idf is to soften the impact of common words that generally do not provide as much meaningful information. The tf–idf value increases based on the number of times a certain word appears for each physician but is reduced by the fraction of physicians in the total corpus who received that word:

$$Weighted\ Value = \frac{Frequency\ in\ a\ physician's\ reviews}{1 + \log[Number\ of\ physicians/Physicans\ that\ received\ that\ word]}$$

To ensure significance, the included words must be associated with at least 25 physicians, and basic "stop words" are excluded. This yields 328 unique words. We then run either a Linear regression when the outcome is physician ratings, or a Logistic regression when the outcome is concierge status. Only words with a significance of p = 0.05 or better are retained. In all cases, controls are employed as in column 4 of Table 3. We also vary whether we include the other outcome variable as a control, with very similar effects of semantics either way.

Table 4 displays the effects of the words that reach statistical significance (0.05 or better) in affecting physician reviews. A color-coding method is employed, with reddest being the most negative coefficients and greenest being the most positive coefficients.

In the terminology of the Consumer Assessment of Healthcare Providers and Systems' Clinician and Group Survey[10], the systems and technical aspects of the visit are extremely important in affecting patient perceptions. Words such as "phone," "staff," and "appointments" point to procedural aspects of the visit. There are also a great number of words relating to time—"wait," "times," "hours," "year," "weeks," "time," "longer"—in keeping with the literature noting that patient waiting times are crucial for doctor reviews (Ko et al., 2019). This may also relate with the length of time that doctors spend in the room with a patient (Hartzband and

[10] Others in the literature also employ this terminological breakdown (Alemi et al., 2012; Paul et al., 2013; Wallace et al., 2014).

Groopman, 2016; Okie, 2012; Quigley et al., 2013; Tai-Seale and McGuire, 2011). Interestingly, apart from a handful of words, such as “rude”, “love”, and perhaps “knowledgeable” and “exceptional”, very few of the words relate to interpersonal skills. We understand these results as indicating that, in deciding on their rating, patients use fewer words related to physician interpersonal skills.[11]

[11] We also found (results available upon request) that having an extra name and having more reviews once again had a negative impact on physician rating. In this semantic analysis, our other controls no achieved statistical significance at conventional levels.

Table 4: Effect of Semantics on Reviews								
	p	F	coeff			p	F	coeff
plan	0.036	4.437	-2.536		staff	0.000	13.061	0.072
little	0.000	17.148	-2.412		patients	0.027	4.949	0.093
specialist	0.000	15.344	-2.320		worth	0.048	3.935	0.133
exceptional	0.001	11.087	-1.906		room	0.025	5.057	0.233
answer	0.022	5.289	-1.813		make	0.036	4.413	0.241
times	0.000	26.689	-1.754		knowledgeable	0.007	7.438	0.271
return	0.017	5.703	-1.714		love	0.007	7.340	0.432
follow	0.018	5.675	-1.593		wait	0.000	15.777	0.482
way	0.005	8.070	-1.441		away	0.027	4.906	0.490
hospital	0.010	6.637	-1.258		appointments	0.034	4.524	0.656
appointment	0.000	17.038	-1.187					
experience	0.013	6.228	-1.032					
coming	0.011	6.445	-1.024					
extra	0.006	7.527	-1.001					
having	0.025	5.027	-0.959					
symptoms	0.012	6.299	-0.913					
physicians	0.002	9.524	-0.900					
hours	0.004	8.208	-0.880					
visits	0.006	7.524	-0.796					
rude	0.000	16.535	-0.744					
old	0.022	5.252	-0.730					
year	0.028	4.875	-0.601					
weeks	0.043	4.110	-0.563					
time	0.013	6.265	-0.549					
provide	0.043	4.134	-0.526					
stay	0.032	4.635	-0.517					
office	0.000	27.907	-0.484					
longer	0.009	6.949	-0.467					
months	0.030	4.762	-0.456					
medical	0.001	10.781	-0.419					
left	0.006	7.574	-0.333					
phone	0.001	11.289	-0.331					
takes	0.048	3.919	-0.281					
patient	0.039	4.289	-0.214					
change	0.039	4.294	-0.092					
second	0.010	6.736	-0.006					

Finally, we determine which words relate to concierge status employing a Logistic framework. Due to its binary nature, we can consider the words with negative coefficients as

associated with non-concierge doctors and ones with positive coefficients as associated with concierge doctors.

Our results in Table 5 show that non-concierge doctor reviews include many *bedside manner* interpersonal-skill words ("attentive," "questions," "explained," "explains," "grateful," "thorough," "comfortable," "helped," "saw," "recommend," "thank," "concerns") and only a few relating to systems and technical aspects ("clean," "easy). In contrast, most of the review words used for concierge doctors relate to systems and technical aspects ("getting," "times," "phone," "longer," "spend," "medicine," "specialist," "follow," "money," "medical," "months," "years," "going," "staff," "nurse," "office"), and only a small number relate to interpersonal skills ("care," "rude"), or a general assessment of doctor quality ("outstanding," "wonderful").[12] To sum up, patients reviewing non-concierge doctors focus on bedside manner above all else, while patients reviewing concierge doctors are more focused on the technical and systems aspects of their care.

Since concierge care is still generally considered as more expensive, it is reasonable to assume that patients will focus both on the significantly increased cost and on what they feel they are getting for their money. Concierge practices often promise many of these technical aspects, such as increased (or sometime 24/7/365) phone access to the doctor, greatly reduced appointment waiting times, and longer visits. If this is the premise of their care, and perhaps even their reason for going, it is reasonable to see these patients focused on technical and system aspects of care.[13]

---

[12] Quigley et al. (2015) found that "respect" seemed to matter the most to patients, although this varied by specialty. Our work is in keeping with this result in a negative fashion. Similarly, Doing-Harris et al. (2016) found that unhappy patients were most likely to use words such as "doctor", "feel ", "appointment", "rude", and "symptoms". These were words that were similar to the list for concierge doctors below. Gao et al. (2012) also found that most negative review variation was due to patients being upset by the office staff and technical aspects of their visit, very much in keeping with our results for non-concierge doctors below.

[13] Many of the words have to do with "diagnosis," "medicine," or "specialist". We read a number of these reviews in detail, and it appears that some of these people have been previously misdiagnosed, or were unable to see the specialist they wanted, or they were happy with the relationship with the specialist and the concierge doctor. This would also point to a different subpopulation choosing to have concierge care

Table 5: Effect of Semantics on Concierge

	p	F	Coeff			p	F	Coeff
recommend	0.001	10.184	-0.760		needed	0.021	5.400	0.015
thank	0.020	5.416	-0.514		little	0.018	5.627	0.028
concerns	0.007	7.274	-0.507		getting	0.023	5.224	0.122
attentive	0.033	4.567	-0.489		times	0.005	7.931	0.126
easy	0.022	5.295	-0.354		told	0.024	5.111	0.131
questions	0.009	6.867	-0.279		mother	0.045	4.049	0.131
explained	0.011	6.567	-0.278		phone	0.000	14.831	0.138
grateful	0.028	4.830	-0.268		wonderful	0.010	6.688	0.162
helped	0.009	6.887	-0.242		longer	0.000	13.938	0.163
saw	0.024	5.133	-0.221		spend	0.000	16.021	0.189
clean	0.045	4.034	-0.199		extra	0.001	11.724	0.202
comfortable	0.015	5.945	-0.150		medicine	0.010	6.631	0.203
explains	0.047	3.960	-0.134		specialist	0.004	8.434	0.208
thorough	0.016	5.798	-0.108		money	0.008	7.151	0.212
					change	0.048	3.932	0.224
					asked	0.029	4.792	0.231
					follow	0.015	5.949	0.238
					hospital	0.005	8.056	0.266
					medical	0.048	3.935	0.269
					old	0.015	5.971	0.270
					diagnosis	0.002	9.991	0.281
					months	0.025	5.059	0.295
					physicians	0.002	10.089	0.319
					years	0.032	4.603	0.345
					necessary	0.045	4.035	0.355
					outstanding	0.009	6.871	0.356
					manner	0.031	4.686	0.364
					going	0.004	8.242	0.375
					wife	0.021	5.387	0.375
					doctor	0.011	6.583	0.383
					husband	0.007	7.279	0.398
					care	0.007	7.204	0.420
					staff	0.001	11.331	0.424
					spent	0.044	4.081	0.428
					nurse	0.013	6.187	0.438
					think	0.000	14.439	0.469
					physician	0.023	5.213	0.479
					office	0.036	4.423	0.546
					rude	0.001	11.084	0.564
					problem	0.010	6.636	0.633

4. DISCUSSION

Primarily employing a machine learning sentiment analysis framework, we have made substantial inroads in understanding the semantic differences between reviews given to concierge and to non-concierge doctors, as well as to understanding how the words chosen in these reviews relate to overall perceptions of care. Notice that patient perception of care is not necessarily related to quality of care, particularly in the United States (Alemi et al., 2012; Gray et al., 2015; Greaves et al., 2012), however, it is probably a good way to begin understanding how patient experience relates with concierge care via patient ratings (Doing-Harris et al., 2016; Gray et al., 2015; Paul et al., 2013). Considering the expanding role of concierge medicine in the United States health care system and its likely continued growth, understanding its relationship with any facets of patient care is a crucial aspect of understanding the evolution of our system of care.

We would further suggest, in the context of improving business models, that if concierge care is being touted as helping systems and technical aspects of care, our results imply that this may be what patients focus on in their reviews. If, however, more interpersonal aspects of care matter to patients in choosing their doctors from online reviews, then this marketing strategy on the part of concierge networks deserve additional consideration and thought.

References:

AAFP Survey on Direct Primary Care 2015

Alemi, F., Torii, M., Clementz, L., & Aron, D. C. (2012). Feasibility of real-time satisfaction surveys through automated analysis of patients' unstructured comments and sentiments. Quality Management in Health Care, 21(1), 1-11.

Alodadi, N., & Zhou, L. (2016). Predicting the helpfulness of online physician reviews. Working Paper

Bajari, P., Nekipelov, D., Ryan, S. P., & Yang, M. (2015). Machine learning methods for demand estimation. American Economic Review, 105(5), 481-485.

Basik, M., Balducci, L., & Bregni, M. Personalized medicine: A comprehensive approach in oncology. Journal of Personalized Medicine, 9, 89-90.

Brown, J. (2009, March 18). 'Boutique' Care Just What Some Doctors Ordered. The Denver Post. Retrieved from http://www.denverpost.com/breakingnews/ci_11936690

Calhoun, S. L. (2015). Concierge Medicine: Incorporating components into your healthcare system. Working Paper

Carnahan, S. J. (2006). Law, medicine, and wealth: Does concierge medicine promote health care choices, or is it a barrier to access? Stanford Law and Policy Review, 17, 121-164.

del Puente, A., Lombardi, V., & Esposito, A. (2015). Personalized medicine is the evidence-based medicine saved in the right horizon. Journal of Medicine and the Person, 13, 91-95.

Dai, T., & Tayur, S. (2018). Healthcare operations management: A snapshot of emerging research. Manufacturing & Service Operations Management, Forthcoming

Doing-Harris, K., Mowery, D. L., Daniels, C., Chapman, W, W., & Conway, M. (2016). Understanding patient satisfaction with received healthcare services: A natural language processing approach. AMIA Annual Symposium Proceedings, 524-533

French, M. T., Homer, J. F., Klevay, S., Goldman, E., Ullmann, S. G., & Kahn, B. E. (2010). Is the United States ready to embrace concierge medicine? Population Health Management, 13(4), 177-182.

GAO: United States Government Accountability Office Report to Congressional Committees. (2005). Physician services: Concierge care characteristics and considerations for Medicare. GAO 05-929, 1-45

Gao, G. G., McCullough, J. S., Agarwal, R., & Jha, A. K. (2012). A changing landscape of physician quality reporting: Analysis of patients' online ratings of their physicians over a 5-year period. Journal of Medical Internet Research, 14(1), 1-11.

Gray, B. M., Vandergrift, J. L., Gao, G., McCullough, J. S., & Lipner, R. S. (2015). Website ratings of physicians and their quality of care. JAMA Internal Medicine, 175(2), 291-293.

Greaves, F., Pape, U. J., King, D., Darzi, A., Majeed, A.,…, & Millett, C. (2012). Associations between web-based patient ratings and objective measures of hospital quality. Archives of Internal Medicine, 172(5), 435-436.

Greaves, F., Ramirez-Cano, D., Millett, C., Darzi, A., & Donaldson, L. (2013). Use of Sentiment Analysis for Capturing patient experience from free-text comments posted online. Journal of Medical Internet Research, 15(11), e239.

Hanauer, D. A., Zheng, K., Singer, D. C., Gebremariam, A., & Davis, M. M. (2014). Public awareness, perception, and use of online physician rating sites. JAMA, 311(7), 734-735.

Hartzband, P., & Groopman, J. (2016). Medical Taylorism. New England Journal of Medicine, 374(2), 106-108.

Hartzband, P., & Groopman, J. (2009). Money and the changing culture of medicine. New England Journal of Medicine, 360(2), 101-103.

Heckman, J. J., & Singer, B. (2017). Abducting economics. American Economic Review, 107(5), 298-302.

Kadry, B., Chu, L. F., Kadry, B., Gammas, D., & Macario, A. (2011). Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. Journal of Medical Internet Research, 13(4), 1-10.

Ko, D. G., Mai, F., Shan, Z., & Zhang, D. (2019). Operational efficiency and patient-centered health care: A view from online physician reviews. Journal of Operational Management, 1-27. DOI: 10.1002/joom.1028

Krumholz, H. M. (2014). Big data and new knowledge in medicine: The thinking, training, and tools needed for a learning health system. Health Affairs, 33(7), 1163-1170.

Mullainathan, S., & Spies, J. (2017). Machine learning: An applied econometric approach. Journal of Economic Perspectives, 31(2), 87-106.

O'Brien, E. (2013, January 17). Why Concierge Medicine Will Get Bigger. MarketWatch. Retrieved from http://www.marketwatch.com/story/why-conciergemedicine-will-get-bigger-2013-01-17

Okie, S. (2012). The evolving primary care physician. New England Journal of Medicine, 366(20), 1849-1853.

Oster, E. (2018). Diabetes and diet: Purchasing behavior change in response to health information. American Economic Journal: Applied Economics, 10(4), 308-348.

Paul, M. J., Wallace, B. C., & Dredze, M. (2013). What affects patient (dis)satisfaction? Analyzing online doctor ratings with a joint topic-sentiment model. Expanding the Boundaries of Health Informatics Using Artificial Intelligence: Papers from the AAAI 2013 Workshop, 53-58.

Quigley, D. D., Elliott, M. N., Farley, D. O., Burkhart, Q., Skootsky, S. A., & Hays, R. D. (2013). Specialties differ in which aspects of doctor communication predict overall physician ratings. Journal of General Internal Medicine, 29(3), 447-454.

Sack, K. (2009, May 10). Despite Recession, Personalized Health Care Remains in Demand. The New York Times. Retrieved from http://www.nytimes.com/2009/05/11/health/policy/11concierge.html?_r=0

Sharma, R. D., Tripathi, S, Sahu, S. K., Mittal, S., & Anand, A. (2016). Predicting online doctor ratings from user reviews using convolutional neural networks. International Journal of Machine Learning and Computing, 6(2), 149-154.

Shukla, A. D., Wang, W., Gao, G., & Agarwal, R. (2018). Catch me if you can—Detecting fraudulent online reviews of doctors using deep learning. SSRN Working Paper

Tai-Seale, M., & McGuire, T. (2011). Time is up: Increasing shadow price of time in primary care office visits. Health Economics, 21(4), 457-476.

Tetreault, M. (2014, February 20). Concierge Medicine's Best Kept Secret, the Price. The Concierge Medicine Research Collective. Retrieved from https://conciergemedicineresearch.wordpress.com/2014/02/20/concierge-medicines-bestkept-secret-the-price-revised/

Wallace, B. C., & Paul, M. J. (2016). 'Jerk' or "Judgmental"? Patient perceptions of Male versus Female physicians in online reviews. Working Paper

Wallace, B. C., Paul, M. J., Sarkar, U., Trikalinos, T. A., & Dredze, M. (2014). A large-scale quantitative analysis of latent factors and sentiment in online doctor reviews. Journal of the American Medical Information Association, 21, 1098-1103.

Wieczner, J. (2013, November 10). Pros and cons of concierge medicine. Wall Street Journal online. Retrieved from https://www.wsj.com/articles/pros-and-cons-of-concierge-medicine-1383766139

Zabar, S., Wallach, A., & Kalet, A. (2019). The future of primary care in the United States depends on payment reform. JAMA Internal Medicine, 179(4), 515-516.