

NBER WORKING PAPER SERIES

QUALITY DISCLOSURE AND CERTIFICATION:
THEORY AND PRACTICE

David Dranove
Ginger Zhe Jin

Working Paper 15644
<http://www.nber.org/papers/w15644>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2010

Accepted at the Journal of Economic Literature. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2010 by David Dranove and Ginger Zhe Jin. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Quality Disclosure and Certification: Theory and Practice
David Dranove and Ginger Zhe Jin
NBER Working Paper No. 15644
January 2010
JEL No. D18,D8,L15,L51

ABSTRACT

This essay reviews the theoretical and empirical literature on quality disclosure and certification. After comparing quality disclosure with other quality assurance mechanisms and describing a brief history of quality disclosure, we address three key theoretical issues: (i) Why don't sellers voluntarily disclose through a process of "unraveling?" (ii) When should government mandate disclosure? and (iii) Do certifiers necessarily report unbiased and accurate information? We further review empirical evidence on these issues, with a particular focus on healthcare, education, and finance. The empirical review covers quality measurement, the effect of third party disclosure on consumer choice and seller behavior, as well as the economics of certifiers.

David Dranove
Management & Strategy Department
Kellogg School of Management
Northwestern University
2001 Sheridan Road
Evanston, IL 60208-2001
d-dranove@kellogg.northwestern.edu

Ginger Zhe Jin
University of Maryland
Department of Economics
3105 Tydings Hall
College Park, MD 20742-7211
and NBER
jin@econ.umd.edu

Quality Disclosure and Certification: Theory and Practice

I. Introduction

A young couple expecting their first child might consult healthgrades.com hospital rankings to help choose where to deliver their baby. A year later, the couple decides they need an SUV and consults performance specifications provided by manufacturers and reads *Consumer Reports* to learn about reliability. Soon thereafter, the couple obtains test score results from several school districts to help choose where to raise their family. When their child is in high school they peruse *US News and World Reports* rankings of universities. Once their child is off to college, they plan for retirement by investing in AAA-rated corporate bonds and browse through Medicare's Nursing Home Compare to help plan for their parents' final years.

Literally from cradle to grave, consumers rely on quality disclosure to make important purchases. Although disclosure has a long history that we describe below, it has attracted considerable attention in the past few years, especially in the areas of healthcare, education, and finance. President Obama has made quality reporting a key component of his healthcare reform effort. President Bush's No Children Left Behind initiative relies on testing and disclosure to evaluate and, potentially, punish, underperforming public schools. Many states have similar programs. And much of the finger pointing for the recent crisis on Wall Street has been directed at corporate bond rating agencies that seemed to ignore systematic risk while giving firms clean bills of health. Many policy analysts in these and other industries believe that we need more and better disclosure.

In this essay, we review the theoretical and empirical literature on disclosure. Section I compares quality disclosure with other quality assurance mechanisms and offers a brief history of disclosure. In section II, we address three key theoretical issues: (i) Why don't sellers voluntarily disclose through a process of "unraveling?" (ii) How would government-mandated

disclosure differ from voluntary disclosure in costs and benefits? and (iii) Do certifiers necessarily report unbiased and accurate information? Section III discusses empirical evidence on disclosure with a particular focus on healthcare, education, and finance. We begin with a practical question: How is quality measured and reported? We then present evidence that unraveling often does not occur in practice, thereby creating a need for third party disclosure. We review whether third party disclosure helps consumers make better choices and whether it encourages sellers to improve quality. We also identify situations where sellers exploit private information so as to boost their ratings at the expense of consumers. We conclude the review of empirical evidence by examining the behavior of certifiers. Section IV concludes with suggestions for further research.

I.1 Disclosure versus other quality assurance mechanisms

We define quality disclosure as an effort by a certification agency to systematically measure and report product quality for a nontrivial percentage of products in a market. While we are mainly interested in third-party disclosure, we also include direct quality disclosure by sellers, provided that the disclosed information can be independently verified. This definition distinguishes disclosure from broader marketing efforts by sellers that do not contain verifiable product information. It also distinguishes disclosure from forums such as town squares, barber shops, or, more recently, Internet sites such as Angie's List where individuals share word-of-mouth reviews of local service providers without systematic editing and scoring. The latter distinction is admittedly blurry; ratings such as Amazon.com's customer reviews have elements of both a "town square" forum and a systematic report card.

Quality disclosure can take many forms. Sellers may voluntarily report product attributes. For example, a hospital may disclose that the majority of its medical staff is board certified. Or

an auto manufacturer may report performance specifications. An industry concerned about the lemons problem may establish a certification agency to collect and disseminate product information. Examples include the Joint Commission on the Accreditation of Healthcare Organizations (JCAHO), which reports the frequency of “sentinel events” (instances of poor quality) at member hospitals; and the Motion Picture Association of America, which is responsible for the familiar G/PG/R/M movie rating system. In these cases, sellers have the choice of disclosing or not disclosing quality information via the certification agency. Those that choose to disclose often pay a fee to cover the cost of certification.

Many industries face mandatory disclosure, whereby a regulatory body requires sellers to disclose certain product attributes in a standard format. In some cases, sellers must provide verifiable information to a designated agency (e.g. automobile manufacturers measure fuel economy and report the results to U.S. Environmental Protection Agency). In other cases, government officials inspect the product on site (e.g. a local health board inspects restaurant hygiene). Mandatory disclosure often focuses on health and safety issues and ignores other product attributes that might influence demand. For example, the U.S. Food and Drug Administration requires food manufacturers to report nutritional information but does not evaluate taste. In recent years, U.S. government agencies have expanded disclosure to include many other factors that can influence demand, including mortality rates for hospitals, on-time arrival rates for airlines, graduation rates for high schools, and consumer satisfaction with Medicare Advantage health insurance plans. There are similar disclosure requirements in many other nations. The targeted audience has also shifted from government officials who might fine or even shut down a business that failed inspection to the consumers whose demands will determine the fate of low scoring firms. By posting results online and publicizing them through

the media, government certifiers hope to ensure that consumers can access the disclosed information with little cost and in a timely manner.

In addition to industry-sponsored voluntary disclosure and government-enforced mandatory disclosure, many private third-party certifiers adopt disclosure regimes to satisfy market demand for quality information.¹ Examples include the Leapfrog Group's hospital quality ratings, Moody's bond ratings, *Consumer Reports'* evaluation of consumer products, and *U.S. News & World Report's* ranking of colleges. Some of these third-parties (e.g. Leapfrog) must obtain data directly from sellers and therefore require seller participation. Others may use public information (e.g., *U.S News*) to evaluate the products and do not require seller participation. In some cases, certifiers may be financially affiliated with sellers, introducing a conflict of interest. Stock analysts working for a brokerage firm that underwrites initial public offerings are often cited as an example of such conflict.

Aside from disclosure, there are many other well-known mechanisms for informing consumers about product attributes. We will call these “quality assurance” mechanisms, though in some cases they provide information about horizontal product attributes rather than vertical quality dimensions. Table 1 gives examples of the mechanisms used to help assure quality in a wide array of markets. All of these markets can be considered credence goods and many are experience goods, in that consumers may find it difficult to evaluate quality of all of these goods prior to purchase but may be able to assess quality of some of them after purchase.

Table 1 about here

As suggested by Table 1, brand and experience are perhaps the most common quality assurance mechanisms, but they are rarely sufficient. One limitation is that even with experience, consumers may find it difficult to link *ex-post* product failure with a product defect;

¹ Demand for quality information is usually stronger for credence goods because consumers have difficulty assessing their quality via search or experience.

think of a automobile owner establishing the reason for premature brake wear or whether a hospital patient determining whether the medical staff is responsible for an adverse outcome. Experience and word-of-mouth are also of limited value when products are infrequently purchased, such as open heart surgery and executive education. Disclosure has the potential to overcome these limitations because certifiers may have better expertise evaluating the product and they can aggregate experiences from many idiosyncratic consumers.

Branding, another common quality assurance mechanism, is usually initiated and maintained through the seller's marketing efforts. It is unclear whether branding acts as a “bond” in which the seller sinks an investment in branding to signal its high quality or whether branding makes it easier for consumers to recall their positive experiences when making repeat purchases.² In any event, consumers may find third-party disclosure more trustworthy than brands.

In some cases, sellers may offer warranties, especially if the value of the product is large relative to the cost to consumers of exercising the warranty. Thus, we see warranties for automobiles and televisions, but not for diapers or light bulbs. Warranties are also uncommon for professional services because consumers have difficulty gauging service quality even after consumption.³ Warranties for hospital care are almost unheard of, for example. Compared with disclosure, warranties often focus on narrow aspects of product performance, such as complete failure, and may not assure gradations of quality.

While most quality assurance mechanisms directly assure product quality, licensing focuses on inputs (e.g. training or staffing) rather than outputs. Licensing is usually done by a government agency, but some industries do their own credentialing. A good example is JCAHO hospital credentialing. Many insurers refuse to reimburse for services performed at non-

² See Bagwell (2007) for a summary of advertising literature.

³ As an exception, plaintiffs’ attorneys in some litigation cases work on a strict contingency basis.

credentialed hospitals. Sometimes government agencies may also establish a minimum quality standard that measures quality directly but does not differentiate quality above the minimum standard. Economists have long debated whether licensing or minimum quality standards serve to control entry, assure quality, or both.⁴ In comparison, disclosure does not have a direct impact on entry, though the disclosed information may motivate consumers to shy away from low quality products and eventually drive out low-quality sellers.

Another way to look at Table 1 is to identify the credibility and source of the quality assurance mechanism. Warranties and brands are offered and established by individual firms as a way to assure consumers of their own quality. Assuming they are enforceable, the effectiveness of warranties is self-explanatory. Brands have credibility because they are developed over time on the basis of experience and often require considerable expense to maintain. Industries often assure quality of member firms, through disclosure, credentialing, or lobbying for licensing laws. Although these may serve as entry barriers, they may also limit the ability of member firms to free ride off of the industry's overall positive reputation.⁵

Aside from disclosure by an industry group, certifying firms are usually independent of the individual firms they assess. The JCAHO may certify hospitals, but individual members do not otherwise provide industry-wide quality reports. An obvious explanation is the potential conflict of interest. One interesting exception occurs when financial analysts evaluate stock offerings in their own names, even though they are employed by investment banks involved in the offerings. This practice could endure if the analyst's own name is separable from the employer and the analyst develops a reputation of unbiasedness and accuracy.

To summarize, disclosure has three distinguishing features: First, disclosure systematically measures and disseminates information about product quality, which makes it

⁴ Stigler (1971).

⁵ Dranove (1988).

attractive when other mechanisms for quality assurance are inadequate and the value of quality information when aggregated across all consumers is large relative to the costs of information collection.⁶ Second, disclosure is usually conducted via third-party certifier(s) that identify themselves separately from manufacturers. This may give consumers an impression that the disclosed information is more trustworthy than seller advertising.⁷ Third, disclosure standardizes quality assessment so that results are readily comparable across sellers. Instead of granting the power of licensing to government officials, disclosure empowers consumer with information with the expectation that consumer choice will provide sufficient incentives to assure quality.

Disclosure both complements and substitutes for other quality assurance mechanisms. In lemons markets, disclosure provides more precise and comparable information than word of mouth, warranties and brand names. Positive reviews may be especially helpful to companies that lack a strong brand. The conventional wisdom is that strong reviews in *Consumer Reports* were critical to the successful 1970s invasion by Japanese automakers into the American car market. By the same token, negative reviews can bring down established brands, as occurred after Ralph Nader's *Unsafe at Any Speed* chronicled problems with the Chevrolet Corvair. Firms in lemons markets may even band together and voluntarily disclose quality as a way to prevent an Akerlof-style adverse selection death spiral.⁸ In the case of car safety, the 2000 mandated

⁶A glimpse at *Consumer Reports* and similar publications suggests that these factors are present in virtually all consumer goods markets where the goods are traded nationally or internationally, so that a single disclosure report can reach millions of potential consumers. Voluntary disclosure has traditionally been less common for local services where the costs of systematically collecting and disseminating information may be prohibitive relative to the size of the audience. The Internet may be reducing these costs, however.

⁷When producers self-disclose quantifiable quality information, consumers might infer that such information can be verified by third parties and is therefore trustworthy. Whether certifier-provided information is indeed more trustworthy than producer disclosure or consumer experience depends on certifier incentives, an active research topic we will review in details in Sections 4 and 5.

⁸Cutler and Zeckhauser (1997).

disclosure of rollover risks⁹ has fostered the set up of minimum performance standards for auto rollovers in 2005.¹⁰

I.2 A Brief History of Disclosure

Quality assurance has a long history. The term branding is derived from the practice of marking livestock that dates back as far as 2000 BC.¹¹ Averill Paints secured the first U.S. trademark (an eagle) in 1870 while Bass and Company (the brewer) and Lyle's Golden Syrup both claim to be Europe's oldest brand, sometime in the late 19th century.¹² Licensing in the United States can be traced to colonial days, when physicians had to obtain permission to practice from colonial governors.

Voluntary disclosure by industry participants emerged in the United States in the 19th century. The Chicago Board of Trade established a system for grading wheat (an example of voluntary disclosure) in 1848. In 1894, the National Board of Fire Underwriters established the Underwriters' Electrical Bureau (the predecessor to Underwriters Laboratories), which, in exchange for a fee, tested and reported on the safety of fittings and electrical devices. This gave high quality sellers a way to distinguish themselves from inferior competitors.

According to Fung, Graham and Weil (2007), U.S. government-mandated disclosure began with the 1906 Pure Food and Drug Act, which provided for inspection of meat products and monitoring of food and drug labeling. Since then, disclosure laws have spread to other markets. For example, the 1934 Securities and Exchange Act requires public companies to file unaudited financial statements quarterly and audited financial statements annually, the 1968

⁹Specified by the 2000 Transportation Recall Enhancement, Accountability, and Documentation Act.

¹⁰Specified by the 2005 Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users.

¹¹This information was obtained from Daye, D. and Van Auken, B., 2006, "History of Branding".

http://www.brandingstrategyinsider.com/2006/08/history_of_brand.html Searched 12/15/2008.

¹²Source: Wikipedia. http://en.wikipedia.org/wiki/Trademark#Oldest_trademarks and <http://en.wikipedia.org/wiki/Brand#History>

Truth in Lending Act requires clear disclosure of key terms and all costs associated with a lending contract), and the 1986 Emergency Planning and Community Right-to-Know Act produces EPA's Toxics Release Inventory Report. Other examples include the 1990 Nutritional Labeling and Education Act, and the hospital and doctor report cards adopted by New York and Pennsylvania in early 1990s.

The 1906 Pure Food and Drug Act was a response to Upton Sinclair's *The Jungle*, Samuel Hopkins Adams' *The Great American Fraud* and other accounts of the meat packing and patent medicines industries. Horrific accounts of "Thalidomide babies" led to the 1962 FDA Amendments.¹³ Despite these high profile examples, Wilson (1982) argues that mandatory disclosure laws are difficult to enact because the potential benefits are diffused among millions of individual consumers whereas the costs are concentrated among a few highly motivated sellers who can better capture the regulatory system. Graham (2002) gives three detailed examples of how public attention, industry lobbying, and political compromise shape mandatory disclosure laws.

Disclosure does not necessarily require legislation. Market driven, third-party disclosure first occurred in 1909 when John Moody issued bond ratings, followed quickly by Poor's Publishing in 1916 and Standard Statistics in 1922.¹⁴ The first issue of *Consumers' Union Reports* (the predecessor to *Consumer Reports*) appeared in May 1936 and featured evaluations of milk, breakfast cereals, soap, and stockings. The Internet has profoundly affected quality disclosure. Not only does the Internet facilitate the dissemination of quality information, it has spawned quality-rating websites such as cnet.com (consumer electronics), imdb.com (movie reviews), and tripadvisor.com (hotels). Rather than rely on experienced certifier(s) attesting to product quality, most of these web sites aggregate the experiences of individual consumers.

¹³ Thalidomide was a sleeping pill. Some pregnant women who used Thalidomide gave birth to infants with horrible deformities.

¹⁴The two companies merged in 1941, forming S&P, which was absorbed by McGraw-Hill in 1966.

I.3 Central Questions

The cursory history of disclosure raises numerous questions about the economics of disclosure:

- How do consumers respond to disclosure? Does the response depend on the source of the quality information (mandatory versus voluntary versus third party)? Does the response differ by the contents and presentation of the disclosed information?
- How do sellers respond to disclosure? Why do some sellers disclose but not others? Why do some industries disclose but not others? Do sellers improve quality after a disclosure system is in place? Does disclosure drive out low quality sellers?
- Do we need mandatory disclosure, or will the market provide sufficient quality assurance through voluntary or third-party disclosure?
- What is the economics of certifiers? Do they have incentives to be truthful and thorough? Does it matter if they collect revenue from sellers or buyers? How would competition, reputation, monitoring and the disclosure of conflicted interest affect certifier behavior?

In the remainder of this essay we review the theory and evidence on disclosure and certification. Most of the theoretical work focuses on the incentives for firms to voluntarily disclose quality and for certifiers to provide unbiased certification about product quality. Several empirical papers also address voluntary disclosure. Much of the empirical literature identifies challenges facing the practice of disclosure, ranging from measurement problems to unintended consequences and certifier bias. In the final section, we present some preliminary thoughts on the potential directions of future research. Our review is by no means exhaustive, nor do our examples cover all the industries that have adopted or attempted to adopt quality

disclosure in practice. Even so, we cite scores of studies; for easy reference Table 2 lists the citations by themes of insight.

Table 2 about here

II. Theory

The theory of quality disclosure can be divided into two strands. The first strand examines seller incentives to disclose quality information to uninformed buyers and often assumes that a third-party certifier can verify seller information. In this strand of literature, the main tension is between consumers who want more quality information to guide their choice of product and low-quality sellers who would like to hide in a pool of high-quality sellers. In addition to redistributing the gains from trade between sellers and buyers, quality disclosure may also result in efficiency gains if better information leads to a better sorting between consumers and products, encourages sellers to improve quality, or forces low-quality sellers to exit the market. In contrast, the second strand of literature puts certifiers under scrutiny. It emphasizes that the interest of certifiers may not be aligned with that of buyers, and that certifiers can manipulate the information flow to the public. This introduces a number of complications, because seller behavior is likely to change in response to certifier behavior and competition among certifiers could generate additional incentives for both sellers and certifiers.

Below we review the two strands of theory separately. In Sections II.1, we summarize the existing theories on seller incentive to voluntarily disclose quality information and then address the merits of mandatory disclosure. In Section II.2, we review theories regarding the role of third-party certifiers.

II.1 Seller disclosure

The best known theory of quality disclosure is the so-called “unraveling result.”¹⁵ The term “unraveling” refers to the process whereby the best quality firm is first to disclose as a way to distinguish itself from lower quality firms. Once the best firm discloses, the second best firm has the same incentive to disclose, and so forth until all but the worst firm discloses. According to Grossman (1981) and Milgrom (1981), if a seller possesses better information about product quality than consumers do and there is zero cost to verifiably disclose it, sellers will always disclose. This occurs because rational consumers will infer non-disclosure as having the lowest quality. It follows that sellers will voluntarily disclose quality unless consumers already have that information, implying that costly government-mandated disclosure is inefficient and non-necessary.

In reality, there are many markets in which voluntary disclosure is incomplete. This is not surprising, because the basic unraveling result requires several often strong assumptions:

- Products are vertically differentiated along a single, well-defined dimension of quality
- Sellers have complete and private information about their own product quality
- Disclosure is costless
- Monopoly or competitive market with no strategic interaction among competing sellers
- Consumers are willing to pay a positive amount for any enhancement of quality
- Consumers are homogeneous
- Consumers hold a rational expectation on the quality of non-disclosed products
- The distribution of available quality is public information

While any violation of these assumptions could lead to a failure of unraveling, theoretical research has focused on the problems posed by disclosure costs, market structure, and the role of consumers.

¹⁵The term “unraveling” is first used in Viscusi (1978) who provides an example in the context of labor markets.

Grossman and Hart (1980) and Jovanovic (1982) show that when disclosure is costly, only sellers with product quality above a specific threshold will disclose. Casual observation suggests that hospitals that are highly ranked by healthgrades.com and other rating services often advertise their rankings, while “average” hospitals remain silent.¹⁶ Though Jovanovic focuses on a market with a large number of sellers, it is easy to extend the logic to monopoly as in Grossman (1981) and Milgrom (1981) because disclosure incentives are driven by skeptical consumers instead of competition among sellers. Matthews and Postlewaite (1985) and Shavell (1994) show that if it is costly to acquire quality information, mandatory disclosure may motivate sellers to reduce information collection. For example, a drug company might limit studies of side effects if required to disclose all findings from such studies.

Several theories link disclosure incentives to market structure. Board (2008) shows that under certain conditions duopolists may fail to disclose quality even if disclosure cost is zero. The main intuition is that disclosure may intensify price competition and this can outweigh any consumer perceptions of inferior quality. Guo and Zhao (2008) demonstrate that the amount of information disclosed depends on whether the duopolists disclose simultaneously or sequentially. As compared to simultaneous disclosure, the leader discloses unambiguously less information while the follower may reveal less or more information depending on disclosure cost.

Unraveling requires consumers to play their part. Even if a third party verification agency rates quality, sellers may hide their ratings if consumers are unaware of them (Faure-Grimaud, Peyrache and Quesdada, 2007). For example, restaurants have not usually disclosed their health and sanitation reports until compelled by regulation.¹⁷ By the same token, unraveling may not occur if consumers do not pay attention to the available information, if

¹⁶ Most consumers are unaware of hospital report cards although newspaper accounts of report card scores do seem to improve awareness. See Dranove and Sfekas (2009) for further discussion.

¹⁷ See Napa News February 13, 2005 “Local restaurants skirt the law when it comes to telling diners about cleanliness and health.”

attentive consumers don't understand the disclosed content, or if consumers make naïve inferences about non-disclosure (Fishman and Hagerty, 2003; Hirshleifer, Lim and Teoh, 2004; Schwartz, 2005; Stivers, 2004). Under any of these conditions, lower quality sellers may not disclose because at least some consumers do not perceive non-disclosure as a signal of the lowest quality. This may further explain the lack of disclosure of hospital quality report card scores – patients may stubbornly believe that their health providers are above average even without disclosure.¹⁸ Unraveling may also fail if consumers have heterogeneous preferences for quality. Board (2008) shows that when duopolists fail to disclose quality, competition for heterogeneous consumers softens. Hotz and Xiao (2009) highlight the importance of consumer heterogeneity when products differ in one vertical attribute (quality) and one horizontal attribute (location). Under some configurations, providing consumers with more information may result in more elastic demand and more intensive price competition, which discourages both low and high quality firms from voluntarily disclosing their product quality.

Unraveling also assumes that consumers have perfect knowledge about the distribution of available quality. In some cases, disclosure can adversely shift the distribution of quality, thereby depressing consumer demand for the whole industry (Milgrom and Roberts, 1986). This explains why all cigarette manufacturers are reluctant to disclose the long-term harm of cigarettes, even if some cigarettes are less harmful than others.

Other reasons for the failure of full disclosure include (1) the standard of certification is unclear or endogenous (Harbaugh, Maxwell and Roussillon, 2007), (2) a seller with high measured quality at a given point in time may fear the obligation to disclose in the future when measured quality might be lower (for example due to mean regression) (Grubb, 2007)), and (3) high quality (and often non-profit) sellers may face capacity constraints and/or price regulations

¹⁸ Dranove (2008) calls this the “Lake Woebegone effect,” named for humorist Garrison Kiellor’s fictional town of Lake Woebegone, where “all of the children are above average.”

and therefore do not benefit from quality disclosure (Lizzeri and Gavazza, 2007). Some teaching hospitals have been reluctant to embrace report cards for the last reason.

Is it desirable to mandate seller disclosure? Market structure, unsophisticated consumers and heterogeneous preferences may all precipitate against voluntary disclosure. Many of the papers cited above argue that under these conditions mandatory disclosure laws can promote competition and raise consumer surplus, often at the expense of firm profits. Indeed, the failure of disclosure, often revealed in public disasters, has fostered a number of government mandates. But mandatory disclosure does not always raise social welfare. When non-disclosure is due solely to disclosure costs, Jovanovic (1982) shows that mandatory disclosure is socially excessive. Mandatory disclosure can also have unintended consequences, such as the aforementioned impact on seller effort in detecting quality (Matthews and Postlewaite, 1985). Mandatory disclosure may encourage “gaming” behavior that boost reported quality but actually reduce consumer welfare, as may be the case for hospital report cards that encourage providers to avoid the sickest patients (Dranove, Kessler, McClellan, and Satterthwaite, 2003) or result in rationing of high quality outputs because high quality suppliers (for example schools and hospitals) face a binding capacity constraint (Lizzeri and Gavazza, 2007). If there are multiple dimensions of product quality, mandatory disclosure on one dimension may encourage firms to invest in the disclosed dimension but cut back in other dimensions, leading to potential reduction in consumer welfare (Bar-Isaac, Caruana, and Cuñat 2008).

II.2 Third-party disclosure and the economics of certifiers

Third party disclosure can eliminate the need for government mandated disclosure if the certifier can provide precise and unbiased information about product quality. However, that condition is hard to meet, sometimes due to the noise in the data generating process and

sometimes due to conflict of interest. The theoretical literature has pinpointed how these problems inhibit third party disclosure, with a particular emphasis given to market and nonmarket mechanisms that might limit certifier conflict of interest.

Quality ratings based on consumer feedback provide a prominent example of noisy data. Even if we limit attention to products that consumers can easily evaluate after consumption (think of Zagat's rating of restaurant services and eBay's rating of seller service), consumer ratings may be noisy or biased because: (1) different consumers may use different criteria to measure quality and these criteria are often implicit and unstable; (2) those consumers who report quality may not represent all consumers (casual empiricism suggests that the most disgruntled consumers are overrepresented), (3) consumers may be reluctant to leave negative feedback in fear of retaliation in the future, and (4) consumer feedback is unverifiable, as consumers may offer feedback without ever having consumed the product and sellers may leave favorable reviews of their own products (while disparaging competitors).¹⁹

Researchers have offered solutions to problems inherent in consumer evaluations. Glazer, McGuire, Cao, and Zaslasky (2008) observe that reporting a simple average of consumer scores invites sellers to improve performance for the majority of consumers while ignoring product features that are costly to improve but only affect a small number of consumers. For example, health plan report cards may encourage insurers to improve prevention services but spend little on improving cancer care. They propose assigning utility weights on different consumer respondents in order to correct this problem. Alternatively, Miller, Resnick and Zeckhauser (2005) propose rewarding individuals whose ratings predict peer ratings.

¹⁹ Only half of eBay buyers leave feedback and very few (<1%) are negative (Resnick and Zeckhauser, 2002). Researchers have attributed the lack of negative feedback to consumer desires to be "nice" and fear of seller retaliation (see a detailed review in Dellarocus 2003).

Recent events have put a spotlight on the potential conflict of interest in certifiers. The Enron scandal raised questions about the veracity of firms that both audit financial statements and sell consulting services, ultimately leading to the downfall of Arthur Anderson.²⁰ In the wake of Enron, SEC disallowed accounting companies to perform audit and consulting services for the same client.²¹ The 2008 financial meltdown turned attention to bond ratings. The Securities and Exchange Commission requires that all public bonds receive a rating from a certified agency. The four agencies that are currently certified to rate bonds – Moody’s, Standard and Poor’s, Fitch Ratings, and Dominion Bond Rating Service – have access to detailed financial information about bond issuers and can provide valuable information to bond purchasers. However, a conflict of interest may arise because bond issuers select their rating agency and pay a fee for the rating service. This may motivate bond rating agencies to give excessively generous ratings in order to secure future rating business. Defenders of bond rating agencies have suggested that an agency’s reluctance to downgrade may instead reflect a longstanding policy of smoothing bond ratings over the peaks and valleys of the business cycle.²²

Can competition, reputation, or external monitoring mitigate the incentive problem of certifiers? The role of competition is ambiguous. On the positive side, theorists show that the information content of quality ratings can be enhanced if a monopoly certifier commits to a rating criterion before sellers choose their quality investment (Albano and Lizzeri 2001), if there is competition among certifiers along both price and rating criteria (Lizzeri, 1999; Hvide and Heifetz, 2001; Miao, 2006), or if consumers already possess some noisy information about

²⁰ For further discussion see Flegm (2005).

²¹ See the 2002 Sarbanes-Oxley Act, <http://www.sec.gov/rules/final/33-8220.htm>.

²² Beaver et al. (2006) argue that smoothing may reflect the wishes of large institutional investors, many of which have rules that require them to sell bonds that fall below investment grade. Because such sales can be costly and some downgraded bonds may revert to the mean, investors may prefer that bond ratings are smoothed. In addition to the four bond rating agencies, brokerage firms may issue their own debt reports.

product quality (Guerra, 2001). Except for perfect competition, the presence of multiple certifiers does not result in full information because noisy grading allows certifiers to extract more profits from low-quality sellers.²³ On the negative side, competition may even worsen the problem because the presence of multiple certifiers encourages sellers to shop around, especially when the application for certificate is non-transparent (Farhi, Lerner and Tirole 2008). In combination, certificate shopping and selective disclosure could create a systematic bias in the disclosed ratings even if each rating agency produces an unbiased estimate of true quality (Skreta and Veldkamp, 2009).

Like competition, reputation concerns do not always help correct the incentives of certifiers. Even if consumers can evaluate disclosed information, it may take a long time to distinguish honest error from strategic manipulation, which leads to an equilibrium where certifiers may first provide accurate information and then take advantage of this reputation in later periods (Benabou and Laroque, 1992). Motivated by the recent financial crisis, Mathis, McAndrews and Rochet (2009) show that reputation is sufficient to discipline credit rating agencies only when a large fraction of the agencies' income come from rating simple assets. The effectiveness of reputation becomes more doubtful in a market with naïve consumers. Accounting for certifiers' incentives to understate credit risk and security issuers' incentives to shop around, Bolton, Freixas, and Shapiro (2009) show that certifiers are more likely to inflate ratings when a larger fraction of consumers take ratings at face value or when the expected reputation costs of rating inflation is lower.

In some cases, reputation concerns may even drive certifiers to report biased information. For example, smog check inspectors may pass a failing car if a reputation for giving “easy”

²³ Albano and Lizzeri (2001) show that a monopoly certifier may choose to reveal full information if it can employ a non-linear pricing scheme depending on the certified quality. When the certifier is constrained to charge a flat fee for all certificates, it can implement a noisy grading criterion to achieve the same profit. In either case, the seller will under invest in quality as compared to the social optimal setting where all information is available free of charge.

passes increases future business (Hubbard, 1998). To address this problem, many local motor vehicle departments test smog check inspectors anonymously, serving the role of “certifier of certifiers”. Of course, this raises the problem of who certifies the certifier of certifiers. Another form of external monitoring is disclosing conflicts of interest of certifiers. Cain, Loewenstein and Moore (2005) argue that the disclosure can have perverse effects because consumers do not discount advice from biased certifiers as much as they should and the disclosure may lead certifiers to feel morally licensed and strategically encouraged to exaggerate their advice even further.

Besides competition, reputation and external monitoring, the fourth potential solution to the incentive problem of certifiers is isolating them from sellers. Intuitively, if certifiers can evaluate the product without seller consent and sell the ratings directly to final consumers, they should not have incentives to please sellers. However, this does not mean the certifier has full incentive to reveal unbiased information. For example, a financial analyst may bias a stock analysis due to career or reputation concerns (Scharfstein and Stein 1990, Ottaviani and Sorensen 2006). Even in the absence of conflict of interest, Durbin (2001) argues that certifiers may find it difficult to extract profits from information provided directly to consumers because uninformed consumers may infer quality from market prices and quantities (e.g., by observing a long line at a well-reviewed restaurant), limiting demand for the guidebooks.

The theoretical literature casts some doubt on the ability of third party certifiers to accurately measure quality and, on occasion, their incentives to truthfully disclose it. The former problem can be alleviated if the certifier knows the sources of noise in quality data. The latter is more problematic because all the usual mechanisms, including competition, reputation and external monitoring, do not necessarily correct the incentives of certifiers.

III. The Practice of Quality Disclosure

The theoretical literature demonstrates that voluntary disclosure, government mandates, and third party certifiers do not necessarily improve social welfare. For certification to enhance welfare, it is important to design quality-rating systems carefully, evaluate their effectiveness *ex post*, and improve system design based on theory and evidence. In this section, we will review a number of lessons that empirical researchers have learned from the practice of quality disclosure in education, health care, and other sectors. For reference, table 3 lists cited papers by industry.

Table 3 about here

We motivate the discussion in this section by considering hospital provider report cards. Patients often have little idea of a hospital staff's competence in diagnosis and surgery and would be hard pressed to obtain systematic data about quality. A hospital might disclose that a certain percentage of its staff is board certified. This is a relatively noisy indicator of quality, however. In contrast, a disclosing agency could easily gather data on patient outcomes such as mortality. Thus, hospital report cards seem like a natural arena for testing theories about disclosure. Yet most evaluations of hospital report cards bear only slightly on the theoretical issues described above; unraveling and the incentives of certifiers are not addressed at all. Instead, the empirical literature focuses on simple questions such as whether patients even pay attention to report cards. Other analyses focus on the statistical properties of hospital report cards, questioning whether they confound the quality of the hospital with unobservable differences in patient severity, thereby generating unreliable rankings and encouraging harmful selection behavior.

Bearing in mind the frequent disconnect between the issues that attract theorists and those that emerge in practice, we begin our review of the empirical literature in Section III.1 with a

discussion of quality measurement. Section III.2 examines the evidence on unraveling and, if voluntary disclosure is incomplete, the kinds of firms that are more likely to disclose. Section III.3 explores whether and how consumers respond to report cards and section III.4 considers providers' responses, including whether providers improve quality or try to game the system. We conclude in section III.5 by reviewing empirical papers that have examined certifier behavior in the financial industries.

III.1 Defining and Reporting Quality

Perhaps the most common approach to reporting quality is to compute average scores for one or more quality dimensions. For example, certifiers often report average mortality rates for a hospital or mean test scores for a school. A major problem with the "average" approach is precision; there often is not enough data to generate small confidence intervals around the reported scores. Consider evaluating hospital mortality. Medicare Hospital Compare recently identified hospitals whose mortality rates were statistical outliers. Because mortality is a rare event, confidence intervals were large and only 3 percent of the nation's hospitals were identified as having either high or low quality. Or consider school performance. There are typically fewer than 100 students in a given grade in a given U.S. public school. Kane and Staiger (2002) argue that as a result of this small numbers problem, the 95 percentage confidence interval of a school's mean score is as wide as the gap between the 25 and 75 percentiles of the score distribution. Not only does this imply that much of the score difference between two schools is likely due to sampling error, it also implies that the best and worst ranked schools are more likely to be small schools. Some certifiers report improvements in quality. These reports are compromised by mean reversion. Kane and Staiger (2002) argue that it is misleading to reward or punish schools depending on how their average scores have changed from one year to the next.

Large sample size can minimize sampling error but does nothing for mean reversion. Nor can a large sample size filter out confounding factors if many individuals in the sample share a common unobservable attribute and this attribute is correlated with the final score. For example, a simple average of patient mortality can underestimate the quality of a large teaching hospital because such a hospital tends to admit sicker patients. By the same logic, a school located in a high-income area may achieve better test scores because children from high-income families tend to have highly educated parents.

Although these problems seem to be well-known among certifiers, they are often ignored. For example, some health insurers emphasize vaccination rates in pediatrician report cards, even though these rates are known to vary with patient income and education. Many other certifiers do adjust raw quality scores using data on demographics and other relevant exogenous characteristics. Medicare computes risk-adjusted mortality rates for heart attack patients in two steps.²⁴ First, a hierarchical regression model is used to regress 30-day mortality of heart attack patients on age, gender and comorbidities. This regression yields the *predicted mortality rate* for a specific hospital given its own patient case mix, as well as an *expected mortality rate* that the same patients with the same characteristics would have should they be treated at an “average” hospital. Second, a hospital's risk-adjusted mortality rate is defined as [(actual mortality / expected mortality) * (U.S national unadjusted mortality rate of heart attack patients)]. To the extent that the regression model has controlled for all the health conditions that affect a patient's mortality risk, it permits a direct comparison of hospitals that treat patients with different severities.²⁵

²⁴http://www.hospitalcompare.hhs.gov/Hospital/Static/InformationforProfessionals_tabset.asp?activeTab=2&Language=English&version=default&subTab=3.

²⁵The algorithm claims to have a special control for small hospitals or small number of cases. See the website cited in the last footnote for more details.

Risk adjustment is no panacea. In a field as complex as medicine, the range of potential risk adjusters is vast, the availability of specific adjusters varies, and implications of choosing specific adjusters can be profound. Iezzoni (1997) obtains markedly different rankings by applying different risk adjusters to the same outcomes data. This suggests the need for uniform and complete risk adjusters. Unfortunately, the predictive power of mortality regressions remains low, suggesting that important risk adjusters are unavailable to certifiers. As more risk adjusters become available, Iezzoni's results suggest that rankings may change.

All of the problems mentioned above are magnified if quality is multi-dimensional or if outcomes are not readily tied to suppliers. Consider reporting the quality of urologists for the treatment of prostate cancer. Outcomes of interest include mortality, pain, incontinence and impotence. While theoretically possible to compute four quality measures, patients might find it difficult to compare them. Moreover, some outcomes such as incontinence may continue months or years after treatment, necessitating complex data collection. Similar issues arise in measuring quality in education, where the performance of a grade school teacher may not be apparent until the students are in high school or beyond.

III.2 Does unraveling generate full disclosure in practice?

Sellers often do not disclose their quality. Left to their own devices, hospitals did not report risk adjusted mortality, perhaps because it would be difficult for patients to verify and interpret the data. It is perhaps more difficult to explain why public schools have not voluntarily reported standardized test scores. Using salad dressing as an example, Mathios (2000) demonstrates that many producers of higher-fat salad dressing withheld fat information before the Nutrition Labeling and Education Act came into effect. Moreover, there remained large variation in fat content among the non-disclosing dressings and those with the highest fat levels

experienced a significant decline in sales after they were required to disclose. This experience suggests that unraveling may fail to occur even in markets with credible, low-cost mechanisms to disclose.

Theory predicts that firms are more likely to disclose if disclosure cost is lower, product quality is higher, or the expected benefits from disclosure are greater conditional on quality and disclosure cost. There is substantial evidence supporting these predictions: on the cost side, Bushee and Leuz (2005) find that regulation of disclosure information for firms that were traded on the over-the-counter bulletin board has increased disclosure costs and forced smaller firms out of the market. Using whether a seller uses professional software to post photos on eBay as a proxy for disclosure cost, Lewis (2009) finds that “low cost” sellers (those with access to professional software) post far more photos than average and that sellers post more photos after switching to professional software. On the benefit side, Francis, Khurana, and Pareira (2005) present evidence that firms in industries with greater external financing needs choose to disclose more financial information to the market, and an expanded disclosure policy for these firms leads to a lower cost of both debt and equity capital. Conversely, Leuz, Triantis, and Wang (2008) presents evidence that firms are more likely to go “dark” in the financial market because of poor future prospects, distress, and increased compliance costs after the Sarbanes-Oxley Act.

However, contrary to the common wisdom, even when the disclosed information is certified by a third-party, disclosing firms do *not necessarily* have higher quality than the non-disclosing firms. This may be because the non-disclosing firms already have a good reputation and therefore do not need certification. Edelman (2006) presents evidence that TRUSTe-certified websites are more than twice as likely to be untrustworthy as uncertified sites. He argues that this occurs because the online “trust” authority issues certifications without substantial verification of the actual trustworthiness of recipients.

In another example, Jin (2005) explores why only half of all health maintenance organizations (HMOs) voluntarily disclose quality via the National Committee on Quality Assurance (NCQA). Disclosure cost is definitely not the main story, as some reporting HMOs explicitly instruct NCQA to withhold the information from plan-identifiable publication. After controlling for cost and demand factors, Jin finds that the disclosure decision is likely driven by incentives to differentiate from competitors. She shows that early disclosers are more likely to operate in highly competitive markets but the average disclosure rate tends to be lower in such markets. These findings are consistent with product differentiation. The counter-intuitive relationship between competition and disclosure is not necessarily surprising; theorists have argued that zero-cost disclosure should unravel in a monopoly market (Grossman, 1981; Milgrom, 1981) but may not unravel in a differentiated duopoly (Board, 2008). Jin and Sorensen (2006) further show that the distribution of quality among reporting HMOs that authorize NCQA to disclose quality overlaps the distribution of quality among HMOs that do not authorize public disclosure, although the former do report better quality on average.

III.3 Does disclosure improve consumer choice?

One of the purported benefits of disclosure is that it facilitates better matches between consumers and products. Consumers may migrate towards higher quality sellers (“vertical sorting”) or to sellers whose product characteristics best meet their idiosyncratic needs (“horizontal sorting.”) Both vertical and horizontal sorting effect could substantially increase welfare even if product attributes remain unchanged.

All available evidence pertains to vertical sorting. For example, Ippolito and Mathios (1990) show that consumers switched to breakfast cereals with higher fiber content after producers were allowed to make health claims about fiber. Hastings and Weinstein (2008) find

that parents who responded to government-mandated information about public school quality by switching their children's enrollment chose schools scoring 0.5 student-level standard deviations above the schools their children left behind.

Several studies of vertical sorting examine health insurance and health provider report cards. Wedig and Tai-Seale (2002) examine health plan choices when plan report cards were introduced to federal employees in the mid 1990s. Scanlon et al. (2002) study how General Motor employees responded to the dissemination of health plan ratings in 1997, Beaulieu (2002) studies plan choices among Harvard employees, Jin and Sorensen (2006) examine how federal annuitants respond to publicized health plan ratings, and Dafny and Dranove (2008) focus on Medicare enrollees choices of Medicare managed care plans subsequent to the publication of *Medicare & You* quality rankings. In all these situations, higher ranked plans enjoy increases in market share. Most of these studies also consider plan choices prior to report cards, finding that consumers seem to have some knowledge of quality differences that report cards augment.

Quality disclosure may fail to affect demand if ratings are difficult to understand or provide irrelevant information. Disclosure may also fail to affect demand if ratings confirm what consumers already know about quality (Marshall et al. 2000). Following the 1990 introduction of New York's cardiovascular surgery report cards, Dranove and Sfekas (2008) show that higher ranked hospitals did not appear to gain significant market shares. This finding concurs with the previous literature (Schneider and Epstein, 1998; Romano and Zhou, 2004). However, Dranove and Sfekas also find that hospitals whose rankings differed from prior beliefs experienced a significant change in market share.

The endogeneity of voluntary disclosure poses a problem to researchers who may be unable to observe firm characteristics that are observed by consumers. Using instrumental variables to address the endogeneity of child care centers decisions to seek voluntarily

accreditation by the National Association for the Education of Young Children, Xiao (2007) finds that an accredited center tends to have lower unobservable (to the researcher) quality. Researchers who disregard endogeneity will underestimate consumer responses to accreditation. After correcting for the endogeneity bias, Xiao finds that consumers rely on both reputation and accreditation status for information and they respond less to accreditation for old firms.

Many of the aforementioned studies find heterogeneous consumer responses to quality information. For example, health plan ratings are more likely to affect individuals choosing a plan for the first time (Wedig and Tai-Seale 2002, Jin and Sorensen 2006), plan ratings are more effective in the areas where consumers had less information prior to the publication of quality measures (Dafny and Dranove, 2008), and quality reporting for fertility clinics has a greater effect in the states that mandate insurance coverage for the reported fertility treatment (Bundorf, Chun, Goda, and Kessler, 2008).

Consumer response to disclosure may differ by attention as well. In a study of earnings disclosures, Della Vigna and Pollet (2009) show that late day Friday announcements (a time where consumers arguably pay less attention to financial news) have a 15% lower immediate stock price response and a 70% higher delayed response as compared to announcements made in other weekdays.

Consumer response is sensitive to the reported measures of quality. Scanlon, Chernew, Sheffler, and Fenwick (2002) find that GM employees respond to overall quality indices but not to specific quality measures. Similarly, Dafny and Dranove (2008) find that the effect of health plan report cards on Medicare beneficiaries is driven by responses to consumer satisfaction scores, while other more objective quality measures rate did not affect enrollment decisions. Pope (2006) studies the effects of hospital rankings in *US News and World Reports*. He finds that changes in discrete rankings affected patient choice, even after controlling for continuous

quality. This array of findings suggest that consumers may have limited cognitive ability and therefore tend to focus on a subset of measures that are easier to understand. As a counter example, Bundorf, Chun, Goda, and Kessler (2008) find evidence of consumer sophistication when they evaluate the quality of fertility clinics. Clinics with a disproportionate share of young, relatively easy-to-treat patients were more likely to have high birth rates due to patient mix, yet such clinics are found to have lower market shares after the adoption of report cards, suggesting that consumers could see through the simple statistics.

A significant literature in finance and accounting assesses the financial impact of disclosure by measuring market price responses to a financial disclosure (by a firm itself or via a third-party certifier). For example, studies find that share values respond to changes in bond ratings, with some studies showing that the market responds more to bad news than to good news.²⁶ Such asymmetric response may explain why some credit rating agencies are reluctant to downgrade (Beaver, Shakespeare, and Soliman, 2006) and why firms tend to announce bad news on late Friday (Della Vigna and Pollet 2009). In a similar spirit, Greenstone, Oyer and Vissing-Jorgensen (2006) find strong market response when the 1964 Securities Acts Amendments extended the mandatory disclosure requirements for listed firms to large firms traded over the counter.

It is difficult to apply the event-study approach to quality disclosure of non-financial products because no centralized market exists to aggregate the disclosed information into a universal market price. Alternatively, Jin and Sorenson (2006) use demand estimates to monetize the value of disclosure. They let $u_i(j)$ denote individual i 's indirect utility from product j . Let A_i denote the product that individual i would choose if quality information is available, and B_i the product that would be chosen in the absence of the information. Then the dollar value of the

²⁶ See Klinger and Sarig (2000) for how security market responds to the refinement of Moody's credit rating system. This paper also reviews previous research on market response to changes in credit rating.

information to individual i is $[u_i(A_i) - u_i(B_i)]/\alpha_i$, where α_i is the marginal utility of a dollar. Under this definition, information is valueless (*ex post*) if it doesn't change the individual's choice. In other words, it does not count the psychological utility gain (loss) that people may derive from being told that their chosen product is rated high (low) even if that information would not have affected their choice.

Following this framework, Jin and Sorensen (2006) estimate that the publication of plan ratings only motivate 0.7 percent of federal annuitants to change their health plan choices, due to the enormous inertia in individual plan choice. For those individuals whose decisions are materially affected by the ratings, the value of the information is estimated to be \$160 per person per year. Averaged over all individuals in the sample, the value of the published scores is only \$1.11 per person. Adding a Bayesian learning structure to a similar random utility model, Chernew, Gowrisankaran, and Scanlon (2008) estimate that GM employees were willing to pay about \$330 per year (about 5 percent of premiums) to avoid one subpar performance rating, and the average value of the report card was about \$20 per employee per year. Finally, Dranove and Sfekas (2008) find that hospitals whose report card scores are two standard deviations below the expected score stood to lose \$1.4 million dollars in revenues annually.

Since the estimated value of information depends on the estimated marginal utility of a dollar (which is often obtained from the coefficient of price), this estimation is sensitive to the endogeneity of price. If price is correlated with unobserved plan quality, it tends to bias the estimates toward finding relatively inelastic demand, which in turn leads to an upward bias in the estimated dollar value of information. Again, this identification issue stresses the importance of controlling for the information that consumers already know before quality disclosure.

Estimating consumer response to quality information becomes more complicated if the industry is subject to price regulation or capacity constraints. School ratings provide an excellent

example. In light of the 2001 No Child Left Behind Act, many states require elementary and secondary schools to collect, submit, and publicize a report card describing school quality in a specific grade. Even if we put aside the measurement problems in school ratings, publicizing ratings could lead to very different consumer responses as compared to other markets for three reasons: first, most students obtain public education free of tuition; second, whether a public school is in a student's choice set often depends on student residence; and third, even within a student's choice set, most local schools are subject to capacity constraint and therefore students may not be able to attend their first-choice school. All three factors suggest that one cannot use the above utility framework to monetize the value of school ratings. Instead, researchers might prefer to examine other factors such as improvements in test scores (see references in Section III.4) or even real estate values (Figlio and Lucas, 2004).

To summarize, empirical studies have found evidence that consumers respond to quality disclosure when rankings differ from preconceptions. The nature of the response depends on whether the disclosed information is easy to access and understand, and whether consumers pay attention to disclosure.

III.4 Does disclosure improve quality?

If disclosure affects demand, the returns to quality should increase, as high quality firms increase sales, boost prices, or both.²⁷ This, in turn, should prompt sellers to raise quality. Several studies document this salutary provider response. Using detailed data before and after Los Angeles County adopted restaurant hygiene grade cards in 1998, Jin and Leslie (2003) find that after Los Angeles County posted restaurant hygiene grade cards in 1998, hospitalizations from food-borne diseases declined by 20 percent, largely because restaurants and consumers

²⁷Wimmer and Chezum (2003) and Dewan and Hsu (2004) have shown that certified goods have higher market price than non-certified goods.

preparing meals at home improved hygiene as opposed to consumers choosing more hygienic restaurants. Similarly, Bennear and Olmstead (2007) examine how Massachusetts drinking water suppliers responded to 1996 Amendments to the Safe Drinking Water Act, which mandated disclosure of contaminant levels. They find that larger utilities required to mail consumer confidence reports directly to customers reduced total violations by 30-44 percent and reduced more severe health violations by 40-57 percent.

Seller responses to disclosure can be heterogeneous. Focusing on the effect of India's Green Rating Project on the largest pulp and paper plants in India, Powers, Blackman, Lyon and Narain (2008) find that the GRP drove significant reductions in pollution loadings among dirty plants but not among cleaner ones. Moreover, plants located in wealthier communities were more responsive to GRP ratings, as were single-plant firms. Chen (2008) studies the effect of the Nursing Home Quality Initiative (NHQI), finding that lower quality nursing homes improved relative to high quality homes, with the most striking changes occurring in more competitive markets.

While disclosure appears to provide firms with incentives to improve quality, Bar-Isaac et al. (2008) point out that this may harm consumers if quality is multidimensional and only some dimensions are disclosed, as firms may boost reported quality but shirk on unreported quality. Lu (2009) notes that the NHQI collects data on a wide variety of quality dimensions but only reports a subset of this data. She finds evidence that after the introduction of the NHQI: (1) the proportion of effort allocated to unreported dimensions decreases; (2) quality improves insignificantly along the reported dimensions but deteriorates along the unreported ones; (3) there is no evidence that nursing homes increase quality-related inputs. These findings suggest that firms may respond to information disclosure by reallocating effort across dimensions of quality, with potentially no net benefit for consumers.

Because quality measures are often imperfect, it is possible to increase the reported quality measures without any improvement in actual quality. In particular, when reported quality depends on the characteristics of the consumer as well as the performance of the seller, sellers can improve performance by strategically selling to the “right” consumers. Hospital report cards provide an excellent example. In 1990-1992, New York and Pennsylvania adopted hospital and surgeon report card based on cardiovascular mortality rates. Although these mortality rates are risk adjusted, the adjustment is imperfect. Consequently, hospitals subject to mandatory report cards may have greater incentives to refuse to treat severely ill patients. The patients who inappropriately receive nonsurgical treatments may suffer greater long term health problems and potentially die due to treatment delays. Using national data on Medicare patients at risk for cardiac surgery, Dranove, Kessler, McClellan and Satterthwaite (2003) find that cardiac surgery report cards in New York and Pennsylvania led both to selection behavior by providers leading to higher levels of resource use and to worse health outcomes, particularly for sicker patients. They conclude that, at least in the short run, these report cards decreased patient and social welfare. As further evidence of selection, Werner and Asch (2005) find that the incidence of cardiac surgery for minority patients relative to white patients declined in New York subsequent to the introduction of report cards.

Similar concerns have been expressed about school report cards. Highlighted by the 2001 No Child Left Behind Act, both federal and state accountability laws require schools to report statistics of student performance as a measure of school quality. Most available studies find that accountability has had a positive effect on student outcomes (Carnoy and Loeb, 2002; Hanushek and Raymond, 2004; Jacob, 2005; Peterson and West, 2003). But the effect is not always attributable to disclosing school report cards to the public. Using student performance data across 42 states in 1993-2002, Hanushek and Raymond (2005) find that accountability laws

have a significant, positive effect on math and reading test scores, but the publication of school reports cards has zero impact. This suggests that quality improvement is more attributable to consequential accountability rather than disclosure *per se*.

Critics of school report cards express concerns about gaming. Schools may increase grade retention or place poor-performing students into special education so that they are not counted in school ratings. These gaming activities have been documented by Jacob (2005) for Chicago public schools, by Haney (2000), Deere and Strayer (2001) and Cullen and Reback (2006) in Texas, and by Figlio and Getzler (2006) in Florida. However, Hanushek and Raymond (2005) argue that state-specific evidence of gaming, often obtained by comparing special education placement rate immediately before and after the introduction of accountability, may be driven by national trends instead of strategic gaming of school ratings.

Changing the pool of subjects is not the only way to game a performance-based disclosure system. Other types of gaming include teaching to the test, extending test time, or blatant cheating. Using data from the Chicago Public Schools, Jacob and Levitt (2003) develop an algorithm to detect teacher cheating based on unexpected test score fluctuations and suspicious patterns of answers for students in a classroom. They find that severe teacher or administrator cheating on standardized tests occurs in at least 4 to 5 percent of elementary school classrooms annually. The observed cheating frequency increased after a 1996 accountability regulation holds schools accountable for low achievement in test scores.

In summary, empirical studies confirmed the theoretical arguments that quality disclosure has strength and pitfalls. On the positive side, there is fairly strong evidence from healthcare and finance that disclosure enables consumers to identify superior sellers. Evidence from education in this regard is less compelling. There is also some evidence from a variety of markets including healthcare that disclosure motivates sellers to improve quality. However, there is also

considerable evidence from healthcare and education that sellers have attempted to game the system at the expense of consumers, especially if the measured quality does not cover all dimensions of quality or does not adjust for characteristics of consumers that can affect the rankings. There is no consensus as to whether the benefits outweigh the costs.

III.5 Evidence on certifier behavior

Most evidence of certifier behavior focuses on (1) the content of the certified information and (2) competition among certifiers.

A certifier can manipulate both the bias and precision of reported information. It is easy to see how a certifier can bias information. Certifiers can also alter precision by intentionally adding noise to the signals they observe. The latter phenomenon has been highlighted in theoretical work predicting that a certifier may have incentive to adopt crude rating intervals (e.g. pass or fail) even if it observes the true product quality with zero cost (Lizzeri, 1999).

The most comprehensive studies of certifier bias and precision focus on financial analysts because their forecast of corporate earnings can be easily tested in the market. A large literature has documented that earnings forecasts are systematically overoptimistic and the extent of bias is predictable from publicly available information. Some attribute the bias to conflict of interest. For example, Michaely and Womack (1999) show that the “buy” recommendations made by analysts affiliated with the underwriting brokerage perform significantly worse than similar recommendations made by unaffiliated analysts. Their evidence suggests that underwriter analysts have a significant positive bias due to conflict of interest and, interestingly, the market does not recognize this bias to the full extent. Using the well-known hierarchy of brokerage firms as a proxy for career outcomes, Hong and Kubik (2003) find that optimistic analysts (relative to the consensus after controlling for forecast accuracy) are more likely to move up in career. This

suggests that career concern can be an important reason for the positive forecast bias. Lim (2001) presents evidence that rational analysts may choose to forecast with a positive bias so that the analyst can access better information from the management in the future and therefore improve her overall accuracy.

Studies of credit ratings suggest that SEC-certified bond ratings agencies are more conservative than non-certified agencies (Beaver, Shakespeare, and Soliman, 2006) and rating agencies revise grades only when they are unlikely to reverse their decision shortly afterwards (Loffler 2005). Concerns about such conservative behavior have prompted policy makers to encourage competition and transparency in order to raise the quality of rating services (SEC 2008). However, the theoretical prediction on competition is mixed: on the one hand, competition may motivate certifiers to provide refined information; on the other hand, it may also invite strategic certificate shopping and result in coarse ratings.

Even when reporting on the same set of firms, certifiers may not agree on the ratings. This fact has been documented in many markets but has received the most scrutiny in credit markets.²⁸ In bond ratings, different ratings from different agencies are often referred to as “split ratings.” Focusing on the comparison between Moody's and Standard & Poor's ratings, researchers have found that the market treats U.S. bonds with split ratings differently from the bonds with equal ratings and the bonds with only one of the two ratings (Thompson and Vaz, 1990; Cantor, Packer and Cole, 1997). This could be driven by (a) Moody's and Standard & Poor's using different rating criteria, or by (b) the bond issuers that select to be rated by both agencies are different from those choosing only one rating and the bonds with split ratings are systematically different the bonds with equal ratings.

²⁸ Differential ratings are documented in consumer products (Friedman 1990), health plan report cards (Scanlon, Chernew, Sheffler, and Fendrick, 1998) and college rankings (Pike, 2004).

To distinguish the two explanations, Cantor and Packer (1997) examine the factors driving the split ratings between Moody's, Standard & Poor's, and two other rating agencies that accept voluntary request for bond rating. They find limited evidence of selection bias, which implies that different rating agencies may use different rating criteria. Similarly, Doherty, Kartasheva and Phillips (2009) take S&P's entry as a natural experiment and show that S&P applies a more stringent rating standard than the incumbent (A.E. Best). As a result, better-than-average insurers within each A.E. Best's rating category are more likely to seek a second rating from S&P. Berger, Davies, and Flannery (2000) broaden the scope of professional certifiers to include both private certifiers and regulators. They use price and rating data to infer whether the government rating of a bank holding company causes a movement in Moody's rating of the same company, or vice versa. They find Granger-causality in both directions, suggesting that supervisors and bond rating agencies both acquire some information that aids the other group in forecasting changes in bank condition.

These studies use both price and rating data to infer differences across certifiers but they do not reveal the full structure of grading differentiation. To overcome this problem, Jin, Kato and List (2008) uses two field experiments to study three professional sports card certifiers. They find that the two new entrants adopt more precise signals and use finer grading cutoffs to differentiate themselves from the incumbent certifier. The measured grading cutoffs map consistently into prevailing market prices, suggesting that the market recognizes differences across multiple grading criteria.

Consistent with the theory, several studies suggest that competition among certifiers is not always helpful. Becker and Milbourn (2008) show that increased competition from Fitch's growth in the corporate bond rating market led to more issuer-friendly ratings and also less informative ratings. Similarly, Tan and Wang (2008) find that credit ratings are less stable when

more rating agencies rate a bond. Hubbard (2002) finds that auto smog emissions testers may issue favorable test results in order to cultivate future business.

Unlike for-profit certifiers who may have an incentive to understate problems in order to attract business, most government certifiers have little direct financial incentive to bias their ratings. But government certifiers face a different set of incentive issues: they may not be sufficiently rewarded for their effort, they may rely on subjective quality measures, and their personal preference may be in line with or against the interests of their clients (Prendergast 2007). Examining more than 1000 inspections by the Nuclear Regulatory Commissions, Feinstein (1989) finds considerable variation in the probability of detecting violations. Interestingly, detection rates increased sharply after the Three Mile Island accident of 1979. Similar inspector heterogeneity is also found in FDA inspection of drug manufacturing facilities (Macher, Mayo and Nickerson, 2008).

To summarize, empirical studies of certifiers have confirmed the theoretical insights: both public and private certification can be noisy and not fully revealing; in addition, competition among certifiers does not necessarily improve information as it may motivate certifiers to relax their rating criterion or encourage sellers to strategically shop for favorable ratings.

IV. Future research

Quality disclosure is an important tool for facilitating consumer purchases when other forms of quality assurance are inadequate. There are many examples in which quality disclosure has allowed consumers to find sellers who best meet their needs, including restaurants, education and healthcare. There is less evidence that sellers respond by boosting quality. Instead, most studies of seller responses seem to focus on gaming behavior that often harms consumers.

Research suggests that quality disclosure is a two-edged sword in other ways, with problems including measurement error, consumer misunderstanding, and inspector bias. It is difficult to state with confidence that disclosure in such important sectors as healthcare, education, or finance has unambiguously helped consumers.

Much additional research is required to help certifiers design optimal quality disclosure schemes. A well designed quality measure should be precise, inexpensive to generate, easy to understand, all while minimizing opportunities for sellers and certifiers to game the system. For example, a hospital quality report card might be more effective than a surgeon report card. One simple reason is that sample size is larger. Perhaps more importantly, many hospitals have the ability to allocate patients across surgeons and will be able to assign the toughest cases to the best surgeons. If surgeon quality was directly reported, then the best surgeons (indeed, all surgeons) might instead shun the toughest cases. Certification of hospitals raises unexplored issues concerning regulated industries including health care and education. If high quality sellers are unable to raise prices, how will they ration demand?

Optimal disclosure design will likely borrow from the literature on multitasking (Holmstrom and Milgrom, 1991). The effects of disclosing some dimensions of quality but not others is similar to the effects of directly rewarding some dimensions but not others (Lu, 2008). Thus, it will be important for disclosing organizations to consider substitution and complementarities in production, the organization of productive teams, and other factors that affect optimal contract design in agency relationships.

There is also a considerable gap in our understanding of certifier behavior, the importance of which was underscored by the 2008 subprime crisis. Theoretical work points to the potential benefits of regulating certifiers and empirical evidence is beginning to emerge on certifier bias and competition among credit rating agencies. Given the perception that

certification failed financial markets on a massive scale and the potential scope for regulation, work in this area is needed urgently.

Finally, while most existing studies have examined the short run consequences of quality disclosure, little is known about long run effects. Quality disclosure may drive out low quality firms, invite entry by high quality competitors, or encourage incumbents to improve quality. Even consumers who ignore disclosure can benefit from these responses, which may prove to be an important benefit of report cards.

VI. References

- Akerlof, G. (1970), "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, 84(3): 488-500.
- Albano, G. and A. Lizzeri (2001), "Strategic Certification and the Provision of Quality," *International Economic Review*.
- Bagwell, K. (2007), "The Economic Analysis of Advertising," in Mark Armstrong and Rob Porter (eds.), *Handbook of Industrial Organization*, Vol. 3, North-Holland: Amsterdam, 2007, 1701-1844.
- Bar-Isaac, H. Caruana, G., and V. Cuñat (2008), "Information gathering externalities in product markets" available at <http://pages.stern.nyu.edu/~hbar-isa/hbigcvccheap.pdf>.
- Becker, B. and T. Milbourn (2008), "Reputation and Competition: Evidence from the Credit Rating Industry," Harvard Business School Finance Working Paper No. 09-051. Available at SSRN: <http://ssrn.com/abstract=1278150>.
- Beaulieu, N. (2002), "Quality Information and Consumer Health Plan Choices," *Journal of Health Economics* 21(1): 43–63.
- Beaver, W., C. Shakespeare, and M. Soliman (2006), "Differential Properties in the Ratings of Certified vs. Non-Certified Bond Rating Agencies" *Journal of Accounting & Economics*, 42(3): 303-334.
- Benabou, R. and G. Laroque (1992), "Using Privileged Information to Manipulate Markets: Insiders, Gurus, and Credibility," *Quarterly Journal of Economics*, 107(3): 921-958.
- Berger, A., Davies, S. and M. Flannery (2000), "Comparing Market and Supervisory Assessments of Bank Performance: Who Knows What When?" *Journal of Money, Credit and Banking*, 32, Part 2, pp. 641-66.

- Benbear, L. S. and S. Olmstead (2009), "The Impacts of the 'Right to Know': Information Disclosure and the Violation of Drinking Water Standards," *Journal of Environmental Economics and Management*, forthcoming.
- Board, O. (2009), "Competition and Disclosure," *Journal of Industrial Economics*, forthcoming.
- Bolton, P., Freixas, X. and J. Shapiro (2009), "The Credit Ratings Game," NBER Working Paper W14712.
- Bundorf, M., Chun, N., Goda, G. and D. Kessler (2008), "Do Markets Respond to Quality Information? The Case of Fertility Clinics," *Journal of Health Economics*, forthcoming.
- Bushee, B.J. and C. Leuz (2005), "Economic consequences of SEC disclosure regulation: evidence from the OTC bulletin board" *Journal of Accounting & Economics* 39(2): 233-264.
- Cain, Loewenstein and Moore (2005), "The Dirt on Coming Clean: Perverse Effects of Disclosing Conflicts of Interest" *Journal of Legal Studies*, 34, January 2005.
- Carnoy, M. and S. Loeb (2002), "Does External Accountability Affect Student Outcomes? A cross-state Analysis," *Educational Evaluation and Policy Analysis*, 24(4):305-331.
- Cantor, R. and F. Packer (1997), "Differences of Opinion and Selection Bias in the Credit Rating Industry," *The Journal of Banking and Finance*, 21:1395-1417.
- _____, Packer, F. and Kevin Cole (1997), "Split Ratings and the Pricing of Credit Risk," *The Journal of Fixed Income*, December:72-82.
- Chen, M. (2008), "Minimum Quality Standards and Strategic Vertical Differentiation: An Empirical Study of Nursing Homes," Northwestern University PhD Dissertation.
- Chernew, M., Gowrisankaran, G. and D. Scanlon (2008), "Learning and the Value of Information: the Case of Health Plan Report Cards," *Journal of Econometrics*, 144: 156-74.

- Cullen, J. and R. Reback (2006), "Tinkering Toward Accolades: School Gaming Under a Performance Accountability System," NBER Working Paper W12286.
- Cutler, D. and R. Zeckhauser (1997), "Adverse Selection in Health Insurance," NBER Working Paper W6107.
- Dafny, L. and D. Dranove, (2008), "Do Report Cards Tell Consumers Anything They Don't Already Know? The case of Medicare HMOs," *RAND Journal of Economics*, 39(3):790-821.
- Daye, D. and B. Van Arlen (2006), "History of Branding,"
http://www.brandingstrategyinsider.com/2006/08/history_of_bran.html, searched 12/15/2008.
- Deere, D. and W. Strayer (2001), "Putting Schools to the Test: School Accountability, Incentives, and Behavior," Working Paper 113, Private Enterprise Research Center, Texas A&M University.
- Dellarocus, C. (2003), "The Digitization of Word-of-Mouth: Promise and Challenges of Online Feedback Mechanisms," *Management Science*, 49(10):1407-1424
- Della Vigna, S. and J. Pollet (2009), "Investor Inattention, Firm Reaction, and Friday Earnings Announcements," *Journal of Finance*, forthcoming.
- Dewan, S. and V. Hsu, (2004), "Adverse Selection in Reputations-Based Electronic Markets: Evidence from Online Stamp Auctions," *Journal of Industrial Economics* 52(4): 497-516.
- Doherty, N. A., Kartasheva, A. and R. Phillips (2009), "Competition among Rating Agencies and Information Disclosure" (February 13, 2009). Available at SSRN:
<http://ssrn.com/abstract=1101700>.
- Dranove, D. (1988), "Demand Inducement and the Physician Patient Relationship" *Economic Inquiry*, 26(2):281-98.

- _____, Kessler, D., McClellan, M. and M. Satterthwaite (2003), "Is More Information Better: The Effects of 'Report Cards' on Health Care Providers," *Journal of Political Economy*, 111: 555-88.
- _____ and A. Sfekas (2008), "Start Spreading the News: A Structural Estimate of the Effects of New York Hospital Report Cards," Northwestern Kellogg Working Paper.
- _____ (2008), *Code Red*, Princeton: Princeton University Press.
- Durbin, E. (2001), "Moody's or the Michelin Guide? Revealing Quality Through Private-sector Certification," Washington University Working Paper.
- Edelman, B. (2006), "Adverse Selection in Online 'Trust' Certifications," mimeo.
- Epstein, A. (1998), "Rolling Down the Runway – the Challenges Ahead for Quality Report Cards," *Journal of the American Medical Association*, 279(21): 1691-96.
- Farhi, E., Lerner, J. and J. Tirole (2008), "Fear Rejection? Tiered Certification and Transparency" NBER working paper #14457.
- Faure-Grimaud, A., Pevrache, E. and L. Quesada (2009), "The Ownership of Ratings," *RAND Journal of Economics*, forthcoming.
- Feinstein, J. (1989), "The Safety Regulation of U.S. Nuclear Power Plants: Violations, Inspections, and Abnormal Occurrences," *Journal of Political Economy*, 97(1): 115-54.
- Feng, L. (2009), "Multitasking, Information Disclosure and Product Quality: Evidence from Nursing Homes," University of Rochester, Working Paper.
- Figlio, D. and M. Lucas (2004), "What's in a Grade? School Report Cards and the Housing Market," *American Economic Review*, 94(3): 591-604.
- _____ and L. Getzler. (2006), "Accountability, Ability and Disability: Gaming the System?" NBER W9307, *Advances in Microeconomics*, T. Gronberg, ed., Amsterdam: Elsevier.

- Fishman, M. and C. Hagerty (2003), "Mandatory Versus Voluntary Disclosure in Markets with Informed and Uninformed Customers," *Journal of Law, Economics and Organization* 19(1).
- Flegm, E. (2005), "Accounting at a Crossroad" *CPA Journal* Accessed online on 10/6/2009.
- Francis, J.R.; I.K. Khurana and R. Pereira (2005), "Disclosure incentives and effects on cost of capital around the world" *Accounting Review* 80(4): 1125-1162.
- Friedman, M. (1990), "Agreement between Product Ratings Generated by Different Consumer Testing Organizations: A Statistical Comparison of 'Consumer Reports' and 'Which?' from 1957 to 1986," *Journal of Consumer Affairs*, 24(1):44-68.
- Fung, A., Graham, M. and D. Weil (2007), *Full Disclosure: the Perils and Promise of Transparency*, Cambridge University Press.
- Glazer, J., McGuire, T., Cao, Z. and A. Zaslavsky (2008), "Using Global Ratings of Health Plans to Improve the Quality of Health Care," *Journal of Health Economics*, 27(5):1182-1195.
- Graham, M. (2002), *Democracy by Disclosure: The Rise of Technopopulism*, Brookings Institution Press.
- Greenstone, M., Oyer, P. and A. Vissing-Jorgensen (2006), "Mandated disclosure, stock returns, and the 1964 Securities Acts Amendments" *Quarterly Journal of Economics*, May 2006.
- Grossman, S. and O. Hart (1980), "Disclosure Laws and Takeover Bids," *The Journal of Finance*, 35(2): 323-334.
- Grossman, S. (1981), "The Informational Role of Warranties and Private Disclosure about Product Quality," *Journal of Law and Economics*, 24: 461-489.
- Grubb, M. (2007), "Developing a Reputation for Reticence," MIT Sloan Working Paper, mimeo.

- Guerra, G. (2001), "Certification Disclosure and Informational Efficiency: A Case for Ordered Rankings of Levels," *University of Oxford Department of Economics Discussion Paper Series*.
- Guo, L. and Y. Zhao (2009), "Voluntary Quality Disclosure and Market Interaction," *Management Science*, forthcoming.
- Haney, W. (2000), "The Myth of the Texas Miracle in Education," *Education Policy Analysis Archives* 8(41).
- Hanushek, E. and M. Raymond, (2004), "The Effect of School Accountability Systems on the Level and Distribution of Student Achievement," *Journal of the European Economic Association*, 2(2-3):406-415.
- Hanushek, Eric A. and Margaret E. Raymond (2005), "Does School Accountability Lead to Improved Student Performance?" *Journal of Policy Analysis and Management*, 24(2): 297-327.
- Harbough, R., Maxwell, J and B. Roussillon (2007), "The Groucho Effect of Uncertain Standards," Indiana University, Working Paper.
- Hastings, J. and J. Weinstein (2008), "Information, School Choice and Academic Achievement: Evidence from Two Experiments," *Quarterly Journal of Economics*, 123(4):1329-1372.
- Hirshleifer, D., Lim, S. and S. Teoh (2004), "Disclosure to an Audience with Limited Attention," Ohio State, Working Paper.
- Holmström, B. and P. Milgrom (1991), "Multitask Principal-agent Analyses: Incentives Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics and Organization*, 7:24-52.
- Hong, H. and J. Kubik (2003), "Analyzing the Analysts: Career Concerns and Biased Earnings Forecasts" *Journal of Finance*, 58(1): 313-351.

- Hotz, V. and M. Xiao (2009), "Strategic Information Disclosure: The Case of Multi-Attribute Products with Heterogeneous Consumers," Working Paper.
- Hubbard, T. (1998), "An Empirical Investigation of Moral Hazard in the Vehicle Inspection Market," *RAND Journal of Economics*, 29(2):406-426.
- _____ (2002), "How Do Consumers Motivate Experts? Reputational Incentives in an Auto Repair Market," *Journal of Law and Economics*, 45(2): 437-468.
- Hvide, H. and A. Heifetz (2001), "Free-Entry Equilibrium in a Market for Certifiers," Norwegian School of Economics, Working Paper.
- Iezzoni, L. (1997), "The Risks of Risk Adjustment," *Journal of American Medical Association* 278(19):1600-7.
- Ippolito, P. and A. Mathios (1990), "Information, Advertising and Health Choices: A Study of the Cereal Market," *RAND Journal of Economics*, 21(3): 459-480.
- Jacob, B. (2005), "Accountability, Incentives and Behavior: Evidence from School Reform in Chicago," *Journal of Public Economics*, 89(5-6): 761-796.
- _____ and S. Levitt (2003), "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating," *Quarterly Journal of Economics*, 118(3).
- Jin, G. and P. Leslie (2003), "The Effects of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards," *Quarterly Journal of Economics*, 118(2).
- _____ (2005), "Competition and Disclosure Incentives: An Empirical Study of HMOs," *RAND Journal of Economics*, 36(1):93-112.
- _____ and A. Sorensen (2006), "Information and Consumer Choice: The Value of Publicized Health Plan Ratings," *Journal of Health Economics*, 26(2).
- _____, A. Kato and J. List (2008), "That's News to Me! Information Revelation in Professional Certification Markets," *Economic Inquiry*, forthcoming.

- _____ and A. Kato (2006), "Price, Quality and Reputation: Evidence From an Online Field Experiment," *RAND Journal of Economics*, 37(4).
- Jovanovic, B. (1982), "Truthful Disclosure of Information," *Bell Journal of Economics*, 13:36-44.
- Kane, T., and D. Staiger (2002), "The Promise and Pitfalls of Using Imprecise School Accountability Measures," *Journal of Economic Perspectives* 16(4): 91-114.
- Kliger, D. and O. Sarig (2000), "The Information Value of Bond Ratings" *Journal of Finance*, 55(6): 2879-2902.
- Leuz, C.; A. Triantis; and T.Y. Wang (2008), "Why do firms go dark? Causes and economic consequences of voluntary SEC deregistrations" *Journal of Accounting & Economics*: 45(2-3): 181-208.
- Lewis, G. (2009), "Asymmetric Information, Adverse Selection and Online Disclosure: The Case of eBay Motors," Harvard University, Working Paper.
- Lim, T. (2001), "Rationality and Analysts' Forecast Bias" *Journal of Finance*, 56(1): 369-385.
- Lizzeri, A. (1999), "Information Revelation and Certification Intermediaries," *The RAND Journal of Economics*.
- _____ and A. Gavazza (2007), "The Perils of Transparency in Bureaucracies," *American Economic Review, Papers and Proceedings*.
- Loffler, G., (2005), "Avoiding the Rating Bounce: Why Rating Agencies are Slow to React to New Information" *Journal of Economic Behavior and Organization*, 56(3): 365-81.
- Macher, J., Mayo, J. and J. Nickerson (2008), "Exploring the Information Asymmetry Gap: Evidence from FDA Regulation," Working Paper.

- Marshall, M., Shekelle, P., Leatherman, S. and R. Brook (2000), "The public release of Performance Data: What Do we Expect to Gain? A Review of the Evidence," *Journal of the American Medical Association*, 283:1866-1874.
- Mathios, A. (2000), "The Impact of Mandatory Disclosure Regulations on Health Choices: An Analysis of the Salad Dressing Market," *Journal of Law and Economics*, 43: 651-678.
- Mathis, J., Mc Andrews, J. and J. Rochet (2009), "Rating the Raters: Are Reputation Concerns Powerful Enough to Discipline Rating Agencies?" forthcoming *Journal of Monetary Economics*.
- Mattews, S. and A. Postlewaite (1985), "Quality Testing and Disclosure," *The RAND Journal of Economics* , 16(3): 328-340.
- Miao, C. (2009), "Competition in Quality Standards" *Journal of Industrial Economics* Volume LVII, Notes, March 2009, published at www.essex.ac.uk/jindec/
- Michaely, R. and K. (1999), "Conflict of Interest and the Credibility of Underwriter Analyst Recommendations" *The Review of Financial Studies*, 12(4): 653-686.
- Milgrom, P. (1981), "Good News and Bad News: Representation Theorems and Applications," *The Bell Journal of Economics*, 12: 380-391.
- _____ and J. Roberts (1986), Relying on the information of interested parties," *RAND Journal of Economics*, 17:18-32.
- Miller, N., Resnick, P. and R. Zechhauser (2005), "Eliciting Information Feedback: The Peer-Prediction Method," *Management Science*, 51(9):1359-1373.
- Ottaviani, M. and P. Sorensen (2006) "Professional Advice." *Journal of Economic Theory* 126(1): 120-142.
- Peterson, P. and West, M., eds. (2003), "No child left behind? The Politics and Practice of Accountability," Washington, DC: Brookings.

- Pope, D. (2006), "Reacting to Rankings: Evidence from 'America's Best Hospitals and Colleges'," University of California, Berkeley, PhD Dissertation.
- Pike, G. (2004), "Measuring Quality: A Comparison of US News Rankings and NSSE Benchmarks," *Research in Higher Education*, 45(2).
- Powers, N., Blackman, A., Lyon, T. and U. Narain (2008), "Does Disclosure Reduce Pollution? Evidence from India's Green Ratings Project," University of Michigan, Working Paper.
- Prendergast, C. (2007), "The Motivation and Bias of Bureaucrats" *American Economic Review* 97(1): 180-196.
- Resnick, P. and R. Zeckhauser (2002), "Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System," in *The Economics of the Internet and E-Commerce*, M. Bayes, ed., Amsterdam: Elsevier Science, 11: 127-157.
- Romano, P. and H. Zhou (2004), "Do Well-publicized Risk-adjusted Outcomes Reports Affect Hospital Volume?" *Medical Care*, 42(4):367-377.
- Scanlon, D., Chernew, M., Sheffler, S. and A. Fendrick (1998), "Health Plan Report Cards: Exploring Differences in Plan Ratings," *Journal on Quality Improvement*, 24(1): 5-20.
- _____, McLaughlin, G. and G. Solon (2002), "The Impact of Health Plan Report Cards on Managed Care Enrollment," *Journal of Health Economics*, 21(1):19-41.
- Scharfstein, D. and J. Stein (1990), "Herding Behavior and Investment," *American Economic Review*, 80(3): 465-479.
- Schneider, E. and A. Epstein (1998), "Use of Public Performance Reports: A Survey of Patients Undergoing Cardiac Surgery," *Journal of the American Medical Association*, 279(20): 1638-1642.
- Schwartz, A. (2008), "How Much Irrationality Does the Market Permit?" *Journal of Legal Studies*, Vol. 37.

- Securities and Exchange Commission (2008), "Annual Report on Nationally Recognized Statistical Rating Organizations." available at <http://www.sec.gov/divisions/marketreg/ratingagency/nrsroannrep0608.pdf>.
- Shavell, S. (1994), "Acquisition and Disclosure of Information Prior to Sale," *RAND Journal of Economics*, 25(1): 20-36.
- Skreta, V. and L. Veldkamp (2009), "Ratings Shopping and Asset Complexity: A Theory of Ratings Inflation" NBER working paper, W14761.
- Stigler, G. (1971), "The Theory of Economic Regulation," *The Bell Journal of Economics*, Spring: 3–21.
- Stivers, A. (2004), "Unraveling of Information: Competition and Uncertainty," *B.E. Journals: Topics in Theoretical Economics*, 4(1).
- Tan, W. and F. Wang (2008), "Credit Rating Agency Interaction and Credit Rating Quality," mimeo.
- Thompson, G. and P. Vaz (1990), "Dual Bond Ratings: A Test of the Certification Function of Rating Agencies," *The Financial Review*, 25(3): 457-471.
- Viscusi, W. (1978), "A Note on 'Lemons' Markets with Quality Certification," *The Bell Journal of Economics*, 9(1): 277-279.
- Wilson, J. (1982), *The Politics of Regulation*, New York: Basic Books.
- Wimmer, B., and B. Chezum (2003), "An Empirical Examination of Quality Certification in a 'Lemons Market'," *Economic Inquiry*, 41(2): 279-91.
- Wedig, G. and M. Tai-Seale (2002), "The Effect of Report Cards on Consumer Choice in the Health Insurance Market," *Journal of Health Economics* 21(6):1031–1048.
- Werner, R. and D. Asch (2005), "The Unintended Consequences of Publicly Reporting Quality Information," *Journal of the American Medical Association*, 293(10):1239-44.

Xiao, M. (2007), "Is Quality Certification Effective? Evidence from the Childcare Market,"
University of Arizona, Working Paper.

Table 1: Quality Assurance Mechanisms Used in Various Markets

	Brand	Experience /Word of mouth	Warranties	Industry-sponsored Voluntary Disclosure	Third-party Disclosure	Government-mandated Disclosure	Licensing
Airlines	X	X			X	X	
Appliances	X	X	X		X		
Automobiles	X	X	X		X	X ²⁹	
Consumer Electronics	X	X	X		X		
Hospitals	X	X		X	X	X ³⁰	X
Lawyers	X	X					X
Movies	X	X		X	X		
Plumbers		X					X
Restaurants	X	X			X	X ³¹	
Universities	X	X		X ³²	X		X

²⁹Fuel economy standards and other safety standards.

³⁰Several states and the federal Medicare program publish quality report cards.

³¹Notably, health and safety inspections.

³²A number of regional and national accreditation agencies accredit universities, colleges, and vocational programs for post-secondary education.

Table 2: List of cited papers by themes of insight

Themes	Citations
Theory: voluntary vs. mandatory disclosure	
Unraveling results	Grossman (1981); Milgram (1981); Jovanovic (1982); Viscusi (1978)
Failure of unraveling: seller-side reasons	Board (2008); Guo and Zhao (2008); Jovanovic (1982); Matthews and Postlewaite (1985); Shavell (1994)
Failure of unraveling: buyer-side reasons	Fishman and Hagerty (2003); Hirshleifer, Lim and Teoh (2004); Hotz and Xiao (2009); Milgrom and Roberts (1986); Schwartz (2005); Stivers (2004)
Failure of unraveling: other reasons	Grubb (2007); Harbaugh, Maxwell and Roussillon (2007); Lizzeri and Gavazza (2007)
Consequence of mandatory disclosure	Bar-Isaac, Caruana, and Cuñat (2008); Jovanovic (1982); Lizzeri and Gavazza (2007); Matthews and Postlewaite (1985)
Theory: the economics of certifiers	
Quality measurement	Glazer, McGuire, Cao and Zaslavsky (2008); Miller, Resnick and Zeckhauser (2005)
Certifier competition and information content of quality certificates	Albano and Lizzeri (2001); Faure-Grimaud, Peyrache and Quesdada (2007); Guerra (2001); Hvide and Heifetz (2001); Farhi, Lerner and Tirole (2008); Lizzeri (1999); Miao (2006); SEC (2008); Skreta and Veldkamp (2009) ;
Reputation and other mechanisms that discipline certifier behavior	Benabou and Laroque (1992); Bolton, Freixas and Shapiro (2009); Cain, Loewenstein and Moore (2005); Durbin (2001); Mathis, McAndrews and Rochet (2008); Ottaviani and Sorensen (2006); Scharfstein and Stein (1990)
Practice on quality disclosure	
Quality measurement	Dellarocus (2003); Iezzoni (1997); Kane and Staiger (2002)
Who volunteers to disclose?	Bushee and Leuz (2005); Edelman (2006); Francis et al. (2005); Jin (2005); Jin and Sorensen (2006); Leuz et al. (2008); Lewis (2009); Mathio (2000)
Consumer response to quality disclosure	Beaulieu (2002); Bundorf, Chun, Goda and Kessler (2008); Chernew, Gowrisankara and Scanlon (2008); Dafny and Dranove (2008); Della Vigna and Pollet (2009); Dranove and Sfekas (2008); Figlio and Lucas (2004); Greenstone, Oyer and Vissing-Jorgensen (2006); Hastings and Weinstein (2008); Ippolito and Mathios (1990); Jin and Sorensen (2006); Marshall, Shekelle, Leatherman and Brook (2000); Pope (2006); Romano and Zhou (2004); Scanlon et al. (2002); Schneider and Epstein (1998); Xiao (2007); Wedig and Tai-Seale (2002)

Seller response to quality disclosure	Benneer and Olmstead (2007); Carnoy and Loeb (2002); Chen (2008); Cullen and Reback (2006); Deere and Strayer (2001); Dranove et al (2003); Figlio and Getzler (2006); Haney (2000); Hanushek and Raymond (2004); Hanushek and Raymond (2005); Jacob (2005); Jacob and Levitt (2003); Jin and Leslie (2003); Peterson and West (2003); Powers et al (2008); Lu (2009); Werner et al. (2005)
Certifier bias, heterogeneity and competition	Beaver, Shakespeare and Soliman (2006); Becker and Milbourn (2008); Berger, Davies and Flannery (2000); Cantor, Packer and Cole (1997); Cantor and Packer (1997); Doherty, Kartasheva and Phillips (2009); Feinstein (1989); Friedman (1990); Hong and Kubik (2003); Hubbard (1998); Hubbard (2002);); Jin, Kato and List (2008); Kliger and Sarig (2000); Lim (2001); Loffler (2005); Macher, Mayo and Nickerson (2008); Michaely and Womack (1999); Pike (2004); Scanlon, Chernew, Sheffler, and Fendrick (1998); Tan and Wang (2008); Thompson and Vaz (1990)
Political forces behind quality disclosure	Wilson (1982); Graham (2002); Fung, Graham and Weil (2007)

Table 3: List of cited empirical papers by industry

Industry/Market Segment	Citations
Education	Carnoy and Loeb (2002); Cullen and Reback (2006); Figlio and Getzler (2006); Figlio and Lucas (2004); Hanushek and Raymond (2004); Hastings and Weinstein (2008); Jacob (2005); Jacob and Levitt (2003); Kane and Staiger (2002); Peterson and West (2003); Pike (2004); Xiao (2007)
Food, beverage, and food service	Becker et al., (2009); Benneer and Olmstead (2007); Jin and Leslie (2003); Mathios (2000); Ippolito and Mathios (1990)
Finance	Becker and Milbourn (2008); Beaver, Shakespeare and Soliman (2006); Berger et al. (2000); Bushee and Leuz (2005); Cantor and Packer (1997); Cantor et al. (1997); Doherty, Kartasheva and Phillips (2009); Farhi, Lerner and Tirole (2008); Francis et al. (2005); Greenstone, Oyer and Vissing-Jorgensen (2006); Hong and Kubik (2003); Kliger and Sarig (2000); Loffler (2005); Lim (2001); Leuz et al. (2008); Michaely and Womack (1999); Ottaviani and Sorensen (2006); Scharfstein and Stein (1990); SEC (2008); Tan and Wang (2008); Thompson and Vaz (1990)
Health care	Beaulieu (2002); Bundorf et al. (2008); Chen (2008); Dafny and Dranove (2008); Dranove et al (2003); Dranove and Sfekas (2008); Iezzoni (1997); Jin (2005); Jin and Sorensen (2006); Lu (2009); Macher et al. 2008); Pope (2006); Romano and Zhou (2004); Scanlon et al. (1998); Scanlon et al. (2002); Schneider and Epstein (1998); Wedig and Tai-Seale (2002); Werner et al. (2005)
Internet businesses	Dewan and Hsu (2004); Edelman (2006); Lewis (2009)
Miscellaneous	Feinstein (1989); Friedman (1990); Hubbard (2002); Powers et al (2008)
Sports	Jin, Kato and List (2008); Wimmer and Chezum (2003)