Optimal Presentation of Quality Ratings: Application to Coarsened Automobile Crashworthiness Ratings

Siqi Liu
Brandeis University

Bhoomija Ranjan
Monash University

Benjamin Shiller
Brandeis University

PRELIMINARY AND INCOMPLETE

Do not cite

Abstract: Many ratings organizations intentionally coarsen ratings before public presentation, for example using a discrete star rating rather than a continuous rating. One explanation for this behavior is that coarsened ratings, when optimally designed, can increase incentives for quality provision, improving outcomes. However, if thresholds for higher ratings are poorly designed, either due to ineptitude or inherent uncertainty in firms' chosen responses, coarsening might worsen outcomes by masking useful information. We investigate the impact of coarsening in the context of car "crashworthiness" ratings. Specifically, we create a novel continuous crashworthiness rating, and estimate a random coefficient model of demand for safety. We then simulate the impacts of chosen coarsening schemes, optimal coarsening schemes, and a continuous crashworthiness ratings on welfare and fatalities.

Acknowledgements: We would like to thank Aiden Zhang for superb research assistance. We also would like Gary Feld and Leo Yon for helping us acquire needed data for this project. Lastly, we would like to thank import.io for providing a discounted rate for their web-scraping software.

Introduction

In addition to gathering, verifying, and providing relevant information to consumers, certification and rating organizations can choose how to package and present the information. Many organizations providing quality ratings intentionally coarsen information before public presentation. For example, the two U.S. organizations which evaluate crashworthiness of car models, the National Highway Traffic Safety Administration (NHTSA) and the Insurance Institute for Highway Safety (IIHS) employ a small number of discrete ratings. A natural question is whether intentionally coarsening the ratings further promotes their mission to reduce deaths, injuries, and economic losses from traffic accidents. In this paper, we empirically investigate the impact of coarsening safety quality ratings on welfare and fatalities, in the context of crashworthiness ratings.

It is an empirical question whether coarsening quality ratings improves outcomes. Ceteris paribus, continuous ratings provide more information to consumers than coarse ratings.[1] But, coarsening quality ratings can increase incentives for quality provision, as firms otherwise choosing quality levels slightly below a ratings threshold may want to increase quality beyond the threshold to be recognized as higher quality (Costrell, 1994; Dubey and Geanokoplos, 2010). We briefly motivate our empirical analyses with a hoteling-style theory model, showing, rather intuitively, that firms provide less (more) than efficient investments in safety provision when the marginal consumer values safety less (more) than the average consumer. Coarsening ratings may address these inefficiencies, at least in theory.

Even if coarse ratings are optimal in theory, certifiers may lack the information or expertise needed to design ratings to optimally encourage quality improvements. In the context of crashworthiness ratings, safety certifiers appear to use and follow the medical literature, but not the economics literature, when designing coarse rating thresholds. Even with the relevant expertise, certifiers may be unable to precisely model how manufacturers will respond to the ratings thresholds due to inherent uncertainties. The risks of poorly setting coarse rating thresholds may outweigh any theoretical gains from coarsening ratings.

A long and broad literature shows consumers and firms respond to quality ratings of other (non-safety) product features (Elfenbein et al., 2015; Hastings and Weinstein, 2008; Jin and Leslie, 2003;

---

[1] While bounded rationality might limit comprehension of continuous ratings, the results in Farrell et al. (2010) and Houde (2018) suggests consumers do meaningfully respond to continuous ratings, and that continuous ratings appear more informative even after accounting for mental processing limitations.

Luca, 2011; Tadelis and Zettelmeyer, 2011) --- for an overview of the quality certification literature, see Dranove and Jin (2010). We add to this literature by investigating the impacts of intentionally coarsening quality ratings.

Empirically, we focus on crashworthiness (i.e., safety quality) ratings for cars. We begin by constructing continuous crashworthiness ratings by relating the probability of driver death to the underlying continuous measures of dummy injury and cabin collapse from staged crash tests.

To address a potential selection issue typically ignored in the literature (e.g., Farmer, 2005; Harles and Hoffer, 2007) --- drivers that care more about safety might both drive more carefully and choose cars with better crashworthiness ratings, conflating the impacts of safety features and driving behaviors --- we include both continuous measures of crashworthiness (which are not observed by consumers) as well as the discrete crashworthiness measures as explanatory variables.[2,3][4] Intuitively, this approach is similar to a regression discontinuity design. Fatality risk presumably increases continuously in measures of injury risk and cabin collapse. But a discontinuity might exist at the threshold for a higher *reported* discrete rating, if driver selection to safer-rated cars is meaningful. By estimating the discontinuity, we can remove it (and the impact of driver selection) from our continuous crashworthiness ratings.

We then estimate a random coefficient discrete choice demand model as a function of car characteristics and the coarsened safety ratings that were available to consumers. We find...

Next, we compare outcomes with coarsened ratings, and with continuous ratings, by simulating car purchase decisions, as well as firm's investments in safety features, under continuous and coarsened safety ratings. We find …

---

[2] The opposite selection issue could hold as well, as risky drivers who are more likely to be involved in a serious accident might benefit more from and therefore choose cars with better crashworthiness ratings

[3] Observed safety features and ratings may directly impact chosen driving behaviors (Cohen and Einav, 2003; Peltzman, 1975) – one might drive more recklessly in a safer car. Controlling for the discrete safety ratings observable to consumers (when deciding which car to buy and how recklessly to drive) controls for such compensating behavior's impact on traffic fatalities (as well as controlling for driver selection).

[4] Some recent European studies attempt to limit endogeneity issues by focusing on the relative risks of injuries between occupants of each car involved in two-vehicle accidents (e.g., Kullgren et al., 2010). However, this strategy might underestimate the crashworthiness of higher-rated vehicles if safety features such as well-designed crumple zones which reduce decelerations limit injury risk of occupants in both cars, including the car collided with.

Theory

In this section, we first explore theoretically whether firms will voluntary exert efficient levels of effort and expenditures in improving safety, when continuous safety ratings are reported. We then discuss whether coarsened ratings might address inefficient investments in safety features.

Safety Provision under Continuous Ratings

Suppose a two-firm hoteling model with the firms located at the end points of the linear city of unit length. Horizontal differentiation and heterogeneous consumer tastes for non-safety features are represented by the distance between a consumer's location (x) along the linear city and the firm's location (0 or 1). Consumers are uniformly distributed along the linear city and experience disutility t per unit of travel distance from their location to firm' $j's$ location. Consumers also may have heterogeneous tastes for safety --- a consumer located at point x derives utility $v(x)$ from each additional unit of the firm's effort e in improving safety. We assume each firm j may invest in safety at Sutton variable fixed cost equal to $\frac{1}{4t}e_j^2$. Variable costs are zero.

With these assumptions, a consumer located at point x receives utility $s - tx - P_0 + v(x)\, e_0$ if buying from firm 0 (located at point zero). The consumer receives utility $s - t(1-x) - P_1 + v(x)\, e_1$ if instead buying from firm 1 (located at point 1). We assume s is large enough so that all consumers select one of the two products, as opposed to the outside good. The value of $\hat{x}$ denoting the location of the consumer who is indifferent between the two products is $\hat{x} = \frac{P_1 - P_0 + t + v(\hat{x})(e_0 - e_1)}{2t}$.

This framework is used to show that firms may overinvest or underinvest in safety, compared to the efficient level of investment. This possibility can be shown while restriction our attention to equilibria in which firms exert symmetric effort $(e_0 = e_1)$.

Firm 0's profits are given by $\pi_0 = P_0\hat{x} - \frac{1}{4t}e_0^2$, where $\hat{x}$ represent the indifferent consumer's location, and $\frac{1}{4t}e_0^2$ is the cost of improving safety. The first order condition of profits with respect to price yields $\frac{P_1 - 2P_0 + t + v(\hat{x})(e_0 - e_1)}{2t} = 0$. Under the symmetric safety investment assumption, one can rearrange yielding: $P_0 = \frac{P_1 + t}{2}$. Similarly, firm 1's optimal price is $P_1 = \frac{P_0 + t}{2}$. Equilibrium prices thus equal $P_0 = P_1 = t$.

Next, we consider necessary conditions for equilibrium safety efforts. The first order condition of firm 0's profits with respect to effort e_0 in improving safety yields $\frac{d\pi_0}{de_0} = P_0 \frac{v(\hat{x})+v'(\hat{x})(e_0-e_1)}{2t} - \frac{e_0}{2t} = 0$. Under the symmetric effort assumption, and noting $P_o = t$, this simplifies to $e_0 = e_1 = v(\hat{x})t$. Thus, a necessary condition for symmetric equilibria implies the amount of effort exerted by each firm to improve safety equals the *marginal consumer's* incremental valuation for safety $v(\hat{x})$ multiplied by the travel cost t.

Next, we investigate the efficient level of effort in improving safety, $\hat{e}$. Because the indifferent consumer is located at point 0.5 under the symmetric equilibria assumption, average travel costs are $\frac{t}{4}$. Assuming a unit mass of consumers, aggregate welfare, as a function of (symmetric) safety effort e, equals: $s - \frac{t}{4} + \int_{x=0}^{1} e\, v(x) f(x) dx - 2\frac{1}{4t}e^2 = s - \frac{t}{4} + e\,\bar{v} - 2\frac{1}{4t}e^2$, where $\bar{v}$ denotes the *average* valuation for safety across consumers. The first order condition (w.r.t. safety) yields efficient safety level $\hat{e} = \bar{v}t$.

Thus, necessary conditions under the symmetric effort assumption imply firms choose safety effort level equal to t times the *marginal* consumer's value for safety $(v(\hat{x})t)$, whereas the efficient safety level equals t times the *average* consumer's value for safety $(\bar{v}t)$. Therefore, if the marginal consumer values safety less than the average consumer, then the firm will provide less than the efficiently level of effort in improving safety. If instead the marginal consumer values safety more than the average consumer, then the firm will provide more than the efficiently level of effort in improving safety. Only when the marginal consumer has the same value for safety as the average consumer will the firm provide efficient effort in improving safety when a continuous measure of safety is reported.

Safety Accreditation

Now suppose a safety rating organization reports only a binary measure indicating whether some arbitrary safety threshold has been met or exceeded. In this case, firms have no incentive to exert effort beyond the threshold, because it will not be reported and thus is not observable by consumers. Likewise, firms have no incentive to exert non-zero effort below the safety threshold. Thus firms either exert safety effort exactly equal to threshold, or exert no effort at all.

Suppose the marginal consumer places less value on safety than the average consumer. Then firms exert more than the efficient level of effort into improving safety when a continuous safety measure is reported. A regulator could improve efficiency by reporting a binary measure of whether the efficient level of safety has been exceeded. Firms would not be recognized for exerting effort beyond the threshold, and thus would have no incentive to do so.

A more interesting question is whether a binary safety threshold could induce firms to invest more in safety compared to when a continuous measure is reported. Firms whose effort under continuous ratings would be near but still below the threshold after coarsening have a choice. If they increase their effort slightly, they will be recognized as high quality after ratings are coarsened. Otherwise, they will be pooled with and recognized as low quality after coarsening. Costrell (1994) shows such agents near the threshold increase effort when ratings of their effort are coarsened. However, agents lack incentives to exert effort beyond the coarse ratings threshold, since their additional effort would not be observed. Therefore, some agents who already exceeded the threshold may reduce their effort after ratings are coarsened. One can address this by adding additional discrete ratings for such agents to strive for. Therefore, ratings maximizing quality provision must be coarse, but not excessively coarse (Dubey and Geanokoplos, 2010).

Background

In the United States, there are two major institutions offering car safety ratings, the National Highway Traffic Safety Administration (NHTSA), a government agency, and Insurance Institute for Highway Safety (IIHS), an independent non-profit funded by insurers. The state mission of both is to improve vehicle safety.

A major and growing focus of both organizations is crash test safety ratings.[5] Starting in the 1990s, the sole NHTSA and IIHS crash test was a frontal collision into a fixed barrier.[6] The NHTSA and IIHS added side barrier tests in 1996, and 2003, respectively.[7] Based on concerns related to rollover deaths, in 2000 the NHTSA added a measure of the probability of rollover, and in 2009, the IIHS

[5] Timelines: A timeline of NHTSA's Safety Ratings Program(https://www.nhtsa.gov/ratings) , About the Institute, Milestones (http://www.iihs.org/iihs/about-us/milestones).
Crash test criterion: More Stars means safer cars (https://www.nhtsa.gov/ratings), About Our Tests (http://www.iihs.org/iihs/ratings/ratings-info/frontal-crash-tests)
[6] The NHTSA uses a fully head-on collision, the IIHS uses a moderate overlap collision where only 40% of the car's front hit the barrier.
[7] The NHTSA side-crash test involves a car-like object crashing into the vehicle, the IIHS involves a SUV/truck like vehicle that hits the side higher up.

added a roof strength test. In 2012, the IIHS added a more stringent frontal crash test where only 25% of the front on the driver side collides with the barrier, and in 2017 they expanded this test to the passenger side.[8] Both organizations note that comparisons of frontal collision test are only valid between cars of similar weight, whereas side and rollover tests are comparable across weight classes. Ratings are highly visible: both the IIHS and NHTSA ratings are prominently shown in car reviews on websites like ConsumerReports.com, Edmunds.com, and USNews.com.

The evolution of safety ratings provides strong suggestive evidence that manufacturers respond to ratings. Figure 1 depicts the average IIHS safety rating (on a four point discrete scale), by rating type, for different model years. Note that when new tests are introduced average ratings are initially rather low, but improve quickly, suggesting manufacturers respond. The introduction of the small overlap test in 2012 - only on the driver side - is perhaps even more informative. In response, manufacturers redesigned the structure of 97 different models for the US market, mostly by using stronger or thicker materials for the cabin. By 2016, three fourths subsequently scored "good," the best rating.[9] The IIHS then repeated the small overlap test on the passenger side for seven vehicles with "good" scores on the driver side. Only one scored "good" on the passenger side. This realization prompted the IIHS to introduce a passenger-side small overlap test in 2017. As Figure 1 shows, average scores on the passenger side were much lower than the scores on the analogous test on the driver's side, confirming the suspicion that manufacturers only strengthened parts of the car that they expected to be directly tested by independent ratings organizations.[10]

Crashworthiness Ratings Presentation

Both the NHTSA and IIHS report a discrete measure of safety for each test type. The NHTSA primarily rates each test type (e.g. side, front) based on five-point star scale, the IIHS on a four-point scale (poor, marginal, acceptable, good). Tests depend on continuous measures of injury severity to crash test dummies, and the IIHS tests also considers the cabin's ability to maintain its shape. A typical consumer could easily observe and comprehend reported discrete scores (see

[8] Some other tests (e.g. Headlights, and frontal impact prevention) have been introduced recently, but typically apply to optional equipment, and therefore apply to an uncertain fraction of a model's production.
[9] http://www.iihs.org/iihs/news/desktopnews/vehicles-with-good-driver-side-protection-may-leave-passengers-at-risk
[10] Among cars with a passenger ride rating in 2017, the average passenger side ratings on a four point scale (from 1=poor to 4=good) was 2.3, much lower than the corresponding driver side rating (3.4).

example in Figure 2), but understanding the many continuous scores underlying the discrete scores would be challenging and time-consuming. Furthermore, the continuous measures are hidden in a technical report, rather than on the main ratings page for a particular model.

The relationship between discrete and continuous scores is well-illustrated by the IIHS's side-impact crash test. A separate discrete rating (good, acceptable, marginal, or poor) is assigned to the vehicle structure and each of four separate injury regions. These scores from these five categories are combined to reach the side impact rating observed by consumers, using a demerits-based system described shortly.

The vehicle structure sub-rating in the side-impact crash test is based on a single continuous measurement, the final location of the cars central side pillar (known as the B-Pillar) relative to the driver's seat center line.[11] If the B-pillar remains at least 12.5 cm (~5 inches) from the centerline after being struck by 1500 kg (about 3300 lb) moving barrier at 50 kp/h (about 31 mph), the vehicle structure rating is "good." If the B-pillar collapses to between 5 and 12.4 cm of the driver's seat centerline, the rating is "acceptable." And it if collapses to between 5 and 0 cm of the centerline, the rating is "marginal." If the B-pillar crosses the centerline, implying a very serious impact with even the thinnest driver, the vehicle structure rating is "poor." See Figure 3.

The injury sub-ratings are intuitively similar. Each body region (head, neck, torso, pelvis/femur) depends on several continuous measures of force on the crash test dummy's body parts during the collision. Each individual measure is categorized into a discrete score (good, acceptable, marginal, or poor), depending on threshold values. The final discrete rating for a given body region is the worst sub-measure rating.[12]

Finally, the discrete vehicle structure rating and injury ratings for each body region are combined into a side-impact crash test rating using a demerit based system.[13] If the total demerits are six or less, a "good" score is received for the side crash test. If between eight and twenty demerits are received, an "acceptable" score is received for the side crash test. Etc. See Table 1. Ratings for

---

[11] "IIHS Side Impact Test Program – Rating Guidelines." http://www.iihs.org/media/2104caa9-7f7e-41fa-a4a5-af65f1cab89e/l9AWAw/Ratings/Protocols/current/side_impact_guide.pdf

[12] "Side Impact Crashworthiness Evaluation – Guidelines for Rating Injury Measures." http://www.iihs.org/media/ba19c647-c2e7-4341-8f12-fe84b0e68a21/9mhzGw/Ratings/Protocols/current/measures_side.pdf

[13] "Side Impact Crashworthiness Evaluation – Weighting Principles for Vehicle Ratings." http://www.iihs.org/media/eda4bd5a-06a8-4c8e-9b7e-a89ce7b822b0/snHVbw/Ratings/Protocols/current/side_impact_weighting.pdf

other crash tests (e.g. moderate overlap frontal collision), are constructed similarly.[14] In addition to reporting discrete scores for each type of crash test, the IIHS also offers a badge to cars that perform well on all crash tests, which they call *top safety picks*.[15] The badge is prominently featured, and they offer a separate list of cars attaining this badge.

The NHTSA's mapping of continuous measurements to discrete scores (stars) is less complicated. The NHTSA relies on crash test dummy injury measures, not explicitly incorporating structure deformation. The NHTSA's front crash test crashes the entire front of the car moving at 56.3 kilometers per hour into a fixed barrier. Before 2011, only two injury measures were used, the Head Injury Criteria (HIC), and the chest g-force. Awarded number of stars depended on total chance of serious injury, to either the head of chest. See Figure 4. NHTSA's side crash rating likewise depended on crash test dummy injury measures, to the chest. The NHTSA also rates a car's tendency to tip over. In 2011, to make the ratings more stringent, the NHTSA included additional injury measures, used more stringent thresholds along with a smaller crash test dummy, and introduced an additional side impact test which replicates side crashes with fixed objects like trees or utility poles.[16] They also started reporting an overall safety rating for each car in 2011. The NHTSA is in the process of overhauling the ratings again, and will soon incorporate tests of crash avoidance features.[17]

Discrete ratings based on continuous measurements allows one to separate the causal impact of vehicle design on fatality risk from the impact of safe drivers sorting to safer or less safe cars, thereby impacting a model's fatality rate. Consider the neck tension force measure. A continuous score anywhere between 0 and 2.1 kN receives a good discrete score.[18] However, scores near the higher end of this range, relative to scores near the lower end, might imply very different risks of serious injury in a collision at higher speed or with a heavier vehicle. Since differences in the continuous measurement in this range are not easily observable, consumers are not believed to sort based on them. Thus, comparing vehicles with the same discrete scores, we can infer a casual impact of continuous crash-test measurements on fatality risk.

---

[14] Comparing various protocols over time, it seem apparent that IIHS thresholds for continuous measurements have not changed over time.
[15] http://www.iihs.org/iihs/ratings/TSP-List
[16] https://tinyurl.com/y8dv8vcq
[17] https://tinyurl.com/y9c24oqk
[18] Even if a continuous measures passes a threshold, increasing demerits, the increase might not pass a demerits threshold, thus not impacting the final rating.

We might infer the impact of sorting of safe drivers by comparing fatality risks between cars with similar continuous measurements, but different observed ratings. Two cars near, but on other sides, of a threshold, presumably have negligible differences in crashworthiness but might have different discrete scores observed by consumers. This might lead drivers who care more about safety, and who might be inherently safer (or less safe) drivers, to sort into cars past the ratings threshold. If cars just surpassing a ratings threshold have much higher (or lower) fatality rates than cars just below the threshold (with a different discrete rating), we might attribute the bulk of these fatalities differences to driver composition, rather than vehicle design.

Data

Needs updating

The dataset used in this paper is a combination of data from a variety of sources. The discrete safety ratings and continuous measurements from staged crash tests were scraped from IIHS website, for model years 2005 to 2018. These data are merged with discrete ratings from the NHTSA (via their API), and car characteristics, including price, for base model cars from WARDS. The unit of observation for these data are model / model-year. We label this dataset the characteristics dataset. The characteristics data are merged with scraped data on production of cars made to United States specifications (and intended for sale in the U.S.) from NHTSA's Early Warning Reporting (EWR) Database, and nationwide driver fatalities from the FARS database.[19] The unit of observation for the production and fatalities data are model / model-year / quarter. Lastly, these data are merged with monthly sales data by model and weekly incentives (i.e. price discounts) from Automotive News's Data Center.[20]

Summary statistics/ observations / etc.

[19] We used autohotkey to scrape production data from https://www-odi.nhtsa.dot.gov/ewr/qb/index.cfm. Fatality data can be downloaded from https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars.
[20] http://www.autonews.com/section/datacenter

List all of the continuous measures in table? Maybe not.

Compare brands performance on new safety measures. Is there a measure of how safe people believe cars are? Is there data on that? Would be a nice comparison.

In addition to reporting discrete scores for each type of crash test, the IIHS also offers a badge to cars that perform well on all crash tests.[21] The badge has changed throughout the years. In 2006, they offered gold and silver badges. From 2007-2012, they offered only one badge type, denoted "top safety pick." In later years, they awarded "top safety pick" badges, and "top safety pick plus" badges. The requirements to receive a badge changed over the years. Generally, it required scoring well on each separate crash test, with some leniency for newly introduced crash test types. In later years, the top safety pick plus designation also depended on the performance of *optional* features, like special headlights and automated crash avoidance technologies. To accommodate discrepancies in the number and labeling of award types, we create a single indicator variable equaling one if the vehicle received any safety badge from IIHS. About 36% of the cars in the sample receive a badge.

Example of discontinuity I THINK WE SHOULD CUT THIS. It seems to add unnecessary confusion.

To illustrate the discontinuities, we focus on side impact test, and in particular, the torso sub-rating. Consider one of the three measurements nested in the torso sub-rating, the viscous criterion, a measure of the rate of deflection of ribs during the collision.[22] A continuous measurement below a value of one receives a "good" score for the viscous criterion sub-sub-rating. A measurement between one and 1.2 receives an "acceptable" score, between 1.2 and 1.4 a "marginal" rating, and above 1.4, a "poor" rating. Figure 4, which shows a scatter plot of viscous criterion measurements against the discrete overall side-impact ratings, illustrates the discontinuity. Note there is not a sharp discontinuity after the first threshold is passed. This is expected. Exceeding only the first

[21] http://www.iihs.org/iihs/ratings/TSP-List

[22] The viscous criterion equals the product of the rib velocity and deflection rate at a given point in time, normalized by the half with width of the chest in millimeters. The measures used is the maximum viscous criterion observed (over time). The IIHS cites a 5% risk of thoracic injury at a viscous criterion value of 1.0.

threshold yields a discrete score of “acceptable” for the viscous criterion sub-sub rating implying the torso sub-rating, equal to the worst of three nested sub-sub ratings (including the viscous criterion), is at best “acceptable.” However, a torso sub-rating of “acceptable” only incurs two demerits, not enough on its own to lower the overall side-impact rating seen by consumers below “good.” To lower the overall side-impact rating observed by consumers below good, the vehicle would need to incur at least 6 more demerits on other sub-ratings (vehicle structure, head, head and neck, femur/leg). See Table 1. A viscous criterion measurement exceeding 1.2, however, corresponds to a “marginal” viscous criterion sub-rating, implying the torso sub-rating can as best be “marginal.” A “marginal” torso rating incurs 10 demerits, singlehandedly lowering the overall side impact rating below “good.” As expected, all vehicles with a viscous continuous measure exceeding 1.2 receive a side-impact rating that is at best “acceptable.”

Figure 4 also illustrates the key variation in continuous measurements use to identify the causal impact of vehicle design on fatality risk. Note the wide range of continuous measurements among cars with an overall “good” side impact rating, and among cars with an overall “poor” impact side rating. Even within a given discrete rating, some vehicles impose a much larger force on a driver’s ribs in the collision test than others. Since the differences in continuous measurements are not observed across consumers, presumably they are uncorrelated with a driver’s value for safety, and thus tendency to have safer driving behaviors. We posit that the impact of discrete ratings on driver composition are accounted for, the residual relationship between continuous measurements and fatality rates can be interpreted as a causal relationship. Hence, our main identifying assumption is that after controlling for observed discrete rating, the continuous measurements from crash tests are uncorrelated with other factors that may independently impact fatality risks.

Safety Analysis

Our method is intuitively similar to the traditional regression discontinuity approach. But there are nuanced, and yet important, differences. First, our context includes multiple assignment variables which impact the discrete outcome (Papay et al., 2011), and the discontinuity is “fuzzy.” Second, we are not per se interested in the difference exactly at the threshold value of the forcing variable. Rather, we intend to control for average differences in driver traits (e.g. safe driving habits) across vehicles with different discrete ratings. Third, our goal is to control for discontinuous changes in fatality risk that occur at the threshold value of the forcing variable, not to attribute the discontinuity to a particular source. Hence threats to interpretation of the discontinuity (McCrary 2008) as driver composition effects are of no consequence in this context.

A potential concern is that consumer perceptions of brand safety might correlate with the continuous measurements from the IIHS crash tests. If so, consumers that care more about safety might sort into brands with better continuous safety measures, even when considering only cars with same reported discrete rating. If drivers that care more about safety are also inherently better (or worse) drivers, e.g. speed less, then we might conflate the direct impact of the continuous measurements (inherent vehicle safety) on fatality risk with differences due to driver composition. To address this concern, we add brand fixed effects to absorb the impact of driver sorting across brands by. The relationship between fatality risk and continuous measures is then identified by variation across different models from the same brand, and variation across different model-years for the same model.

Our regressions follow the form of:

$$\log\left(1 + fatalities_{js}\right) = h\left(x_{js}\right) + \sum_{r} B_r I\left(\mu_{r-1} < g\left(x_{js}\right) \leq \mu_r\right) + \phi \log\left(production_{js}\right) + \gamma_s + \epsilon_{js}$$

Where $h(x)$ denotes the continuous relationship between crash test measurements (x) and log fatalities of model j and model year s, and $I\left(\mu_{r-1} < g\left(x_{js}\right) \leq \mu_r\right)$ are binary variables indicating whether some function $g(x)$ of continuous variables x falls between adjacent thresholds, indicating r is the discrete safety rating observed by consumers. We also include controls for vehicle production and model year to control for the impact of vehicle count and duration of use on potential fatalities.

Consumers Reports surveyed about 1500 individuals a year from 2010 through 2014, asking, among other questions, which car brands they believed were the safest. Each year the top five brands were reported. The top five were fairly consistent. Just six brands (Ford, Honda, Mercedes-Benz, Subaru, Toyota, and Volvo), occupied the top five spots in all five years reported.

Model

References

Anderson, Michael L., and Maximilian Auffhammer. "Pounds that kill: The external costs of vehicle weight." Review of Economic Studies 81.2 (2013): 535-571.

Cohen, Alma, and Liran Einav. "The effects of mandatory seat belt laws on driving behavior and traffic fatalities." Review of Economics and Statistics 85.4 (2003): 828-843.

Costrell, Robert M. 1994. “A Simple Model of Educational Standards.” American Economic Review 84 (4): 956–71

Dubey, Pradeep, and John Geanakoplos. 2010. “Grading Exams: 100, 99, 98...or A, B, C?” Games and Economic Behavior 69 (1): 72–94.

Farmer, Charles M. "Relationships of frontal offset crash test results to real-world driver fatality rates." Traffic Injury Prevention 6.1 (2005): 31-37.

Farrell, Joseph, Janis K. Pappalardo, and Howard Shelanski. 2010. “Economics at the FTC: Mergers, Dominant-Firm Conduct, and Consumer Behavior.” Review of Industrial Organization 37 (4):263–77.

Hastings, Justine S., and Jeffrey M. Weinstein. "Information, school choice, and academic achievement: Evidence from two experiments." The Quarterly journal of economics 123.4 (2008): 1373-1414.

Houde, Sébastien. "How consumers respond to product certification and the value of energy information." The RAND Journal of Economics 49.2 (2018): 453-477. Fairly relevant

Jin, Ginger Zhe, and Phillip Leslie. 2003. “The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards.” Quarterly Journal of Economics 118 (2): 409–51.

Koppel, Sjaanie, et al. "How important is vehicle safety in the new vehicle purchase process?." Accident analysis & prevention 40.3 (2008): 994-1004.

Kullgren, Anders, Anders Lie, and Claes Tingvall. "Comparison between Euro NCAP test results and real-world crash data." Traffic injury prevention 11.6 (2010): 587-593.

Luca, Michael. 2011. “Reviews, Reputation, and Revenue: The Case of Yelp.com.” Harvard Business School Working Paper 12-016

Peltzman, Sam. "The effects of automobile safety regulation." Journal of political Economy 83.4 (1975): 677-725.

Tadelis, Steven, and Florian Zettelmeyer. 2011. “Information Disclosure as a Matching Mechanism: Theory and Evidence from a Field Experiment.” American Economic Review 105 (2): 886–905

Xiang Hui, Maryam Saeedi, Zeqian Shen, Neel Sundaresan 2010. Reputation and regulations: evidence from eBay. Management Science

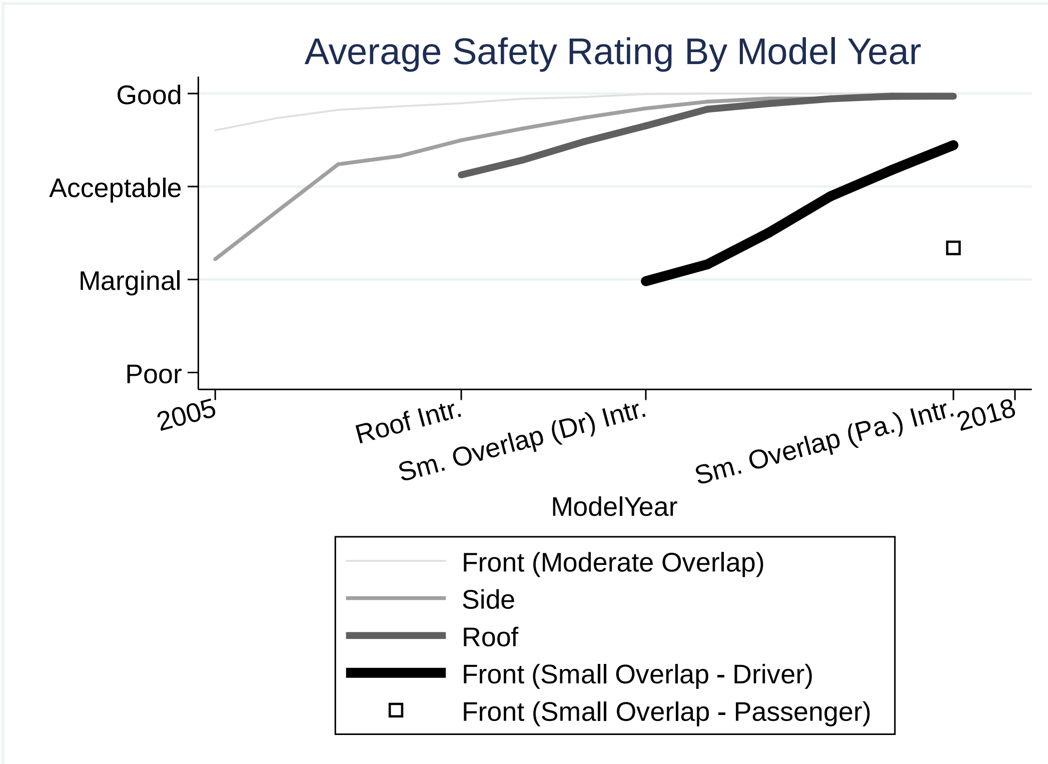
Average Safety Rating By Model Year
Good
Acceptable
Marginal
Poor
2005
Roof Intr.
Sm. Overlap (Dr) Intr.
Sm. Overlap (Pa.) Intr.
2018
ModelYear
Front (Moderate Overlap)
Side
Roof
Front (Small Overlap - Driver)
Front (Small Overlap - Passenger)

Figure 1

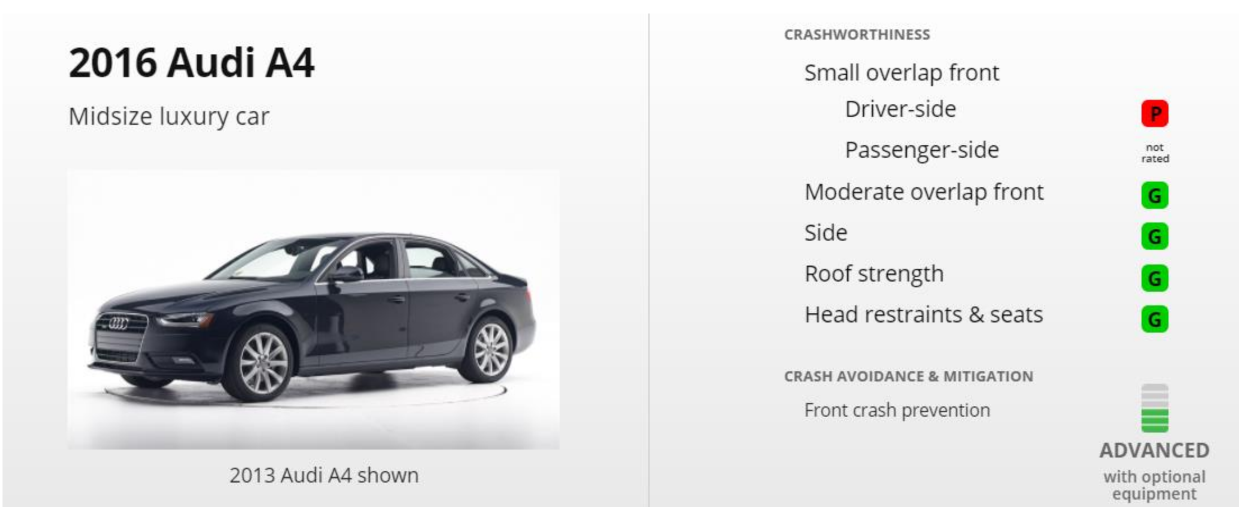
2016 Audi A4
Midsize luxury car
2013 Audi A4 shown
CRASHWORTHINESS
Small overlap front
Driver-side
P
Passenger-side
not rated
Moderate overlap front
G
Side
G
Roof strength
G
Head restraints & seats
G
CRASH AVOIDANCE & MITIGATION
Front crash prevention
ADVANCED
with optional equipment

Figure 2 – Example of safety ratings presented to consumers

Notes: This depiction is from a screenshot of IIHS's website for the 2016 Audi A4. See http://www.iihs.org/iihs/ratings/vehicle/v/audi/a4-4-door-sedan/2016.

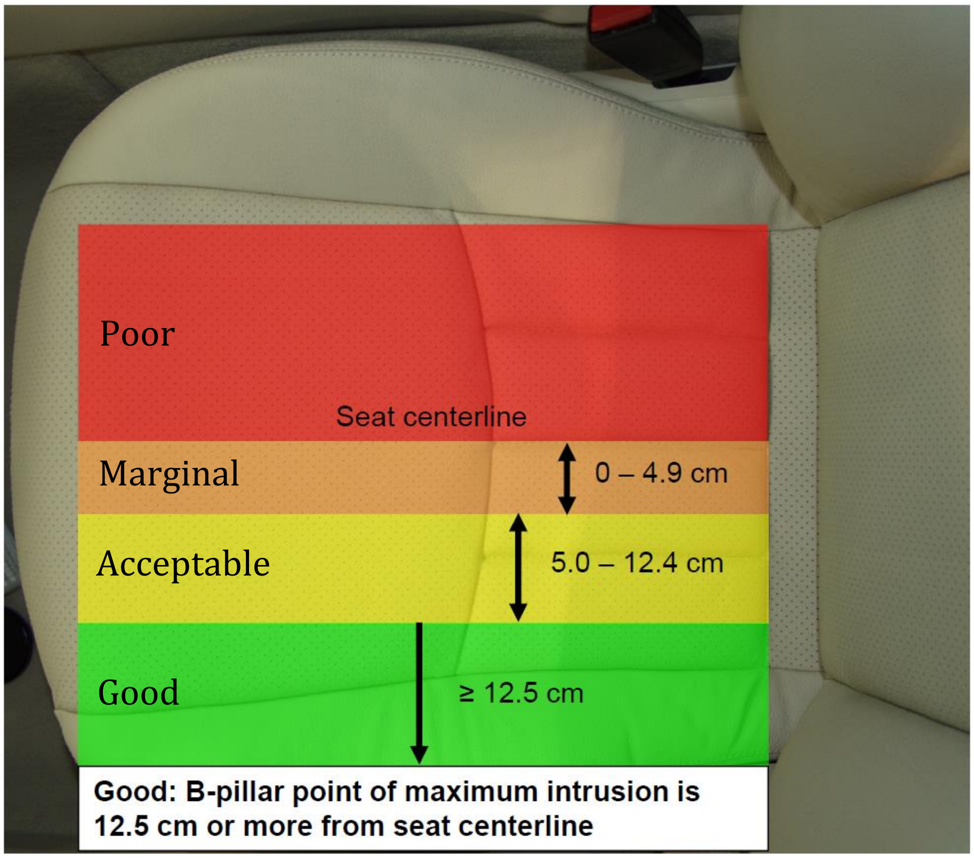

Figure 3 – Thresholds for B-Pillar Vehicle Structure Measure

Note: The B-pillar intrusion in cm is translated into a discrete score, based on how far the B-pillar is from the driver's seat center line. This picture was adapted from an illustration in IIHS's "IIHS Side Impact Test Program – Rating Guidelines." http://www.iihs.org/media/2104caa9-7f7e-41fa-a4a5-af65f1cab89e/l9AWAw/Ratings/Protocols/current/side_impact_guide.pdf

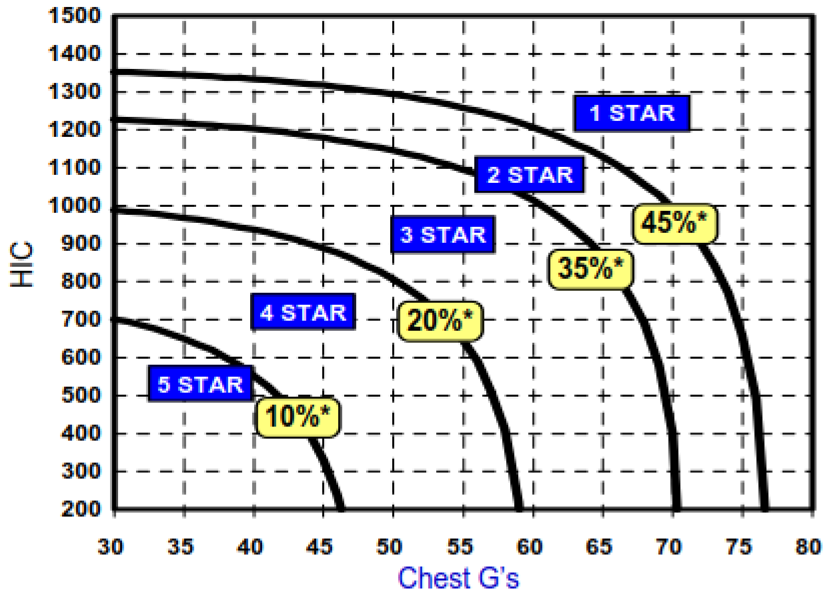

Figure 4 – Mapping from discrete injury measures to NHTSA's pre-2011 frontal crash rating

Notes: The HIC denotes the head injury criterion, and the Chest G's denotes the g-force applied to the dummy's chest in the staged collision. Source: "New Car Assessment Program (NCAP): Past, Present, and Future," https://tinyurl.com/y7tbj64l.

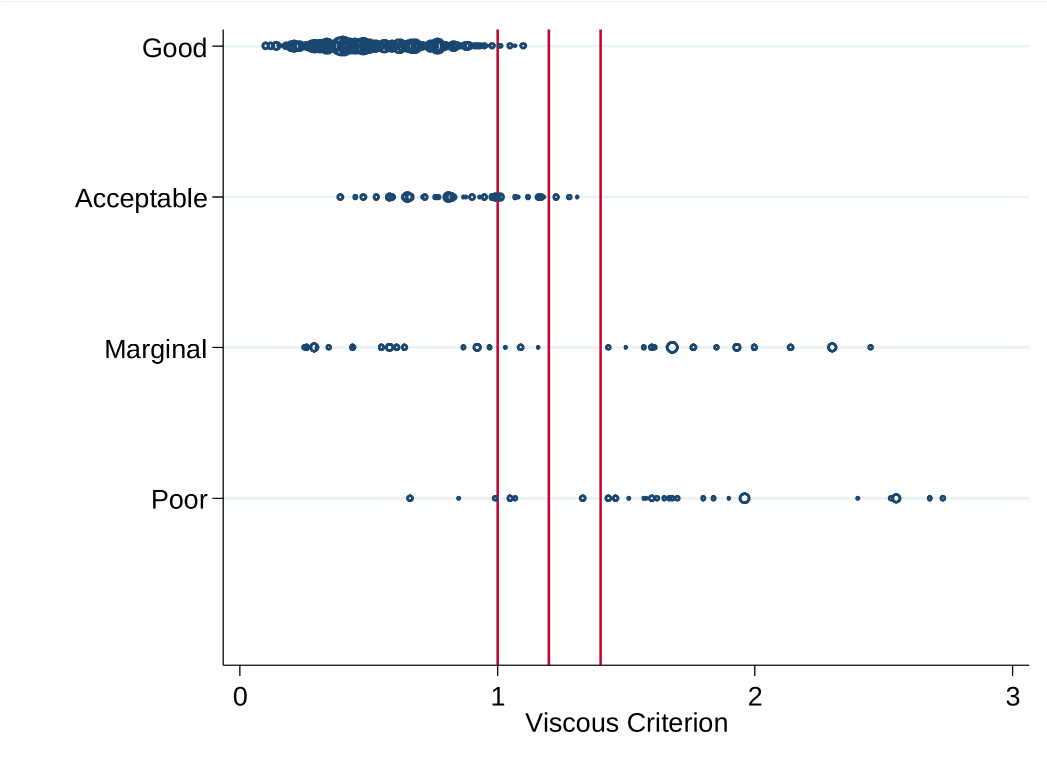

Figure 4 – Illustration of discontinuities

Notes: Figure 4 plots the side-impact rating (from IIHS) against one of the many continuous measurements the rating is based on, the viscous criterion. The viscous criterion equals the product of the rib velocity and deflection rate at a given point in time, normalized by the half with width of the chest in millimeters. The measures used is the maximum viscous criterion observed (over time). The IIHS cites a 5% risk of thoracic injury at a viscous criterion value of 1.0. The vertical lines denote the thresholds for the viscous criterion. If the first threshold is passed, the viscous rating counts an "acceptable" score, rather than the best score, "good." The second and third thresholds correspond to the cutoffs for "marginal" and "poor" scores. The torso rating equals the worst of three continuous measurements, including the viscous criterion. A viscous criterion measurement exceeding the first threshold guarantees the torso rating is at best "acceptable." A torso rating of "acceptable" incurs 2 demerits, less than the 8 demerits needed across all sub-ratings (vehicle structure, head, head and neck, torso, and femur/leg) to push the overall side impact ratings below the "good" rating. A viscous criterion measurement exceeding the second threshold, however, implies the torso rating is at best marginal, incurring at least 10 demerits, singlehandedly lowering the side impact rating below "good." Note that all of the vehicles with viscous criterion measurements exceeding the second threshold has at best an “acceptable” side-impact rating.

Table 1 – Mapping of sub-ratings to final observable side-impact crashworthiness rating (IIHS)

	Weighting of sub-ratings (demerits)			
	Good	Acceptable	Marginal	Poor
Vehicle Structure	0	2	6	10
Driver				
Head protection	0	2	4	10
Head and neck	0	2	10	20
Torso	0	2	10	20
Pelvis and left femur	0	2	6	10
Rear passenger				
Head protection	0	2	4	10
Head and neck	0	2	10	20
Torso	0	2	10	20
Pelvis and left femur	0	2	6	10
Overall Rating Cutoffs (demerits)	0-6	8-20	22-32	34+

Notes: This table is adapted from “Side Impact Crashworthiness Evaluation – Weighting Principles for Vehicle Ratings.” http://www.iihs.org/media/eda4bd5a-06a8-4c8e-9b7e-a89ce7b822b0/snHVbw/Ratings/Protocols/current/side_impact_weighting.pdf