The Recall Decision Exposed: Auto Recall Timing and Process Data Set

Preprint *in* Manufacturing & Service Operations Management · December 2021

DOI: 10.1287/msom

3 authors, including:

Vivek Astvansh
Indiana University Bloomington

Matt Josefy
Indiana University Bloomington

The Recall Decision Exposed: Auto Recall Timing and Process Data Set

Forthcoming, *Manufacturing & Service Operations Management*

Article: https://pubsonline.informs.org/doi/epdf/10.1287/msom.2022.1085
Web Appendix: https://pubsonline.informs.org/doi/suppl/10.1287/msom.2022.1085/suppl_file/msom.2022.1085.sm1.pdf
Data File: https://pubsonline.informs.org/doi/suppl/10.1287/msom.2022.1085/suppl_file/msom.2022.1085.sm2.xls

This Version (Accepted by the Journal): December 9, 2021

Vivek Astvansh
Assistant Professor, Department of Marketing
Associate Director of Research, Center for Education and Research in Retail,
Fellow, Center for Corporate Governance,
Kelley School of Business, Indiana University Bloomington

Affiliate, Environmental Research Institute, Indiana University
1309 E 10th St, Bloomington, IN 47405, United States
812-855-8953 | astvansh@iu.edu

George P. Ball
Associate Professor, Department of Operations & Decision Technologies
Weimer Faculty Fellow
1309 E 10th St, Bloomington, IN 47405, United States
812-856-0625 | gpball@indiana.edu

Matthew Josefy
Assistant Professor, Department of Management & Entrepreneurship
1309 E 10th St, Bloomington, IN 47405, United States
812-855-3683 | mjosefy@iu.edu

The Recall Decision Exposed: Automobile Recall Timing and Process Data Set

Problem Definition. There is a concerted effort across multiple academic disciplines to understand the recall decision-making process. Specifically, what steps does a manufacturer take following a product defect discovery and resulting in the product recall decision? This effort has often been limited to case studies within a particular manufacturer largely due to the absence of consistent and comparable data across firms.

Methodology/Results. This data paper provides a foundation for future research on recall decisions by processing and coding textual disclosures on 2,120 recalls initiated in the United States by 27 automobile manufacturers from 2009-2018. For each recall, the data set provides the time the firm took to make the recall decision by comparing the defect awareness date to the recall decision date, whether the recall is associated with a supplier, the number of events in the recall decision-making process, and the date and description of each event.

Managerial Implications. Not only can this data enhance product recall research by providing key recall decision-making variables unavailable in related research, an additional indication of the value of our data set also comes from National Highway Traffic Safety Administration (NHTSA), the automobile regulator in the United States. We held discussions with a senior leader at the NHTSA's Recall Management Division related to this data set. This discussion revealed that the NHTSA does not have these data in an analyzable form and that they would be interested in using our data set for its reports, such as the NHTSA's biennial reports to the U.S. Congress. This signal suggests that regulators, as well as researchers, practitioners, and other safety advocates may find our data set useful.

Keywords: product recall, time-to-recall, managerial decision-making, data, automotive

1. Introduction

"Sitting here today, I cannot tell you why it took so long for a safety defect to be announced for this program, but I can tell you we will find out."

- Mary Barra, CEO of General Motors (GM), during a U.S. Congressional hearing about a GM recall related to an ignition switch defect, 2014.[1]

Safety defects in products are pervasive and can cause public harm. However, product manufacturers could be reluctant to recall potentially harmful products because recalls frequently result in significant negative publicity and lost sales. However, allowing a defective product to remain in the market can harm consumers (Cowley 2016; Mukherjee and Sinha 2018; Wowak et al. 2015), and lead to regulatory investigations (Ball et al. 2017; Jensen 2014), and class-action lawsuits (Janeway 2019; Maidenberg 2018). Thus, both manufacturers and regulators are under increasing pressure to be more transparent in their recall decisions. Unfortunately, researchers and practitioners understand little about a manufacturer's recall

[1] https://www.c-span.org/video/?c4548317/user-clip-statement-questions

decision-making process and the time taken to make this important decision. The primary reason for this gap in understanding is the absence of analyzable data.

We address this gap by providing a novel recall timing and process data set. We collect and clean unstructured data for 2,120 recalls initiated in the United States between 2009 and 2018 by 27 automobile manufacturers. Applying manual and computer-aided text analyses, we collect the chronology section of all recall reports submitted by automobile manufacturers to the federal government. We identify, for each recall, the defect awareness date, which is the date when the manufacturer first became aware of the defect. Subtracting the defect awareness date from the recall date allows us to compute an automobile manufacturer's time-to-recall (Eilert et al. 2017; Hora et al. 2011).

Each recall report submitted by automobile manufacturers to the National Highway Traffic Safety Administration (NHTSA), the automobile regulator in the United States, contains the textual chronology of all events leading up to the recall decision. We use this chronology to extract the date and the narrative for each event that occurs within the recall decision-making process. The event description reports what investigation the manufacturer performed in that step and which stakeholders—internal teams and groups, suppliers, dealerships, customers, the NHTSA, etc.—it involved. We break down the textual chronology into these distinct events. For each of the 2,120 recalls, our data set provides the number of steps the manufacturer took and the duration of the recall decision.

Many recalls involve supplier quality problems, which may affect timing and process. To identify whether a recall is related to a supplier, we use MarkLines' Who Supplies Whom, Standard & Poor's Capital IQ, and Bloomberg's Supply Chain (SPLC) databases to create a list of over 10,000 name variants of global automotive suppliers.[2] We then develop a software program that searches each recall's chronology for the presence of any variant of the suppliers' names in our list. These efforts classify 47% of automobile recalls in our data set as supplier recalls.

---

[2] "Name variants" include all versions of supplier names instead of just one standardized name. For example, "Aptiv Electric Systems Co., Ltd. Changchun Branch" and "Aptiv Electric Systems Co., Ltd. Guangzhou Branch" are name variants for "Apriv Electic Systems Co., Ltd." The list is thus exhaustive and inclusive of as many supplier name versions as possible. Section 3.3 describes how we use this list of name variants.

Following the theoretical lenses used by extant recall research (Haunschild and Rhee 2004; Liu et al. 2017), we briefly offer three theoretically motivated questions—one each from theories in crisis management, organizational learning, and information disclosure—that future, multidisciplinary research could answer using our data set. An important signal of the value of our data set also comes from the NHTSA. We held discussions with a senior leader at the NHTSA's Recall Management Division related to this data set. This discussion revealed that the NHTSA does not have these data in an analyzable form and that they would be interested in using our data set for its reports, such as the NHTSA's biennial reports to the U.S. Congress. This signal suggests that regulators, as well as researchers, practitioners, and other safety advocates may find our data set useful.

2. Literature Review

The preponderance of recall studies explores either the causes of recalls (Ball et al. 2018a; Bray et al. 2019; Haunschild and Rhee 2004; Shah et al. 2017; Steven and Britto 2016; Thirumalai and Sinha 2011; Wowak et al. 2021) or their consequences (Archer and Wesolowsky 1996; Mukherjee et al. 2021; Thirumalai and Sinha 2011). Few examine the recall decision, as we do herein. In fact, recall decision-making studies are the most recent and the most nascent stream of recall research. For example, in the context of consumer products, Hora et al. (2011) examined the determinants of the time-to-recall, measured as the days elapsed from the date of first sale of the recalled product (as defect awareness date was not available in their data source) to the date of its recall. They found that the time-to-recall is higher for preventive recalls (those that have non-zero reports of injury or death) versus for reactive recalls, for products with manufacturing defects versus design flaws, and for recalling firms that are more distant from the end-customer. Ball et al. (2018b) conducted a behavioral experiment involving medical device firms' managers as participants. Their results indicate that in deciding whether to recall a defective device, managers consider their understanding of the root cause of the defect and the physicians' ability to detect the defect before use the product on a patient customer. Çolak and Bray (2016) mined textual data on complaints received by the NHTSA to discover defect signals and relate these signals to the numbers of voluntary recalls and mandatory recalls in the following period. In the context of medical devices, Wowak et al. (2021) showed that female board representation expedites low-severity recalls and shortens the time-to-recall for devices with the most severe, life-threatening defects. In the context of automobile recalls, Eilert et al. (2017) found that defect severity increases an automobile manufacturer's time-to-recall, measured as the difference between the date when the NHTSA opened a formal investigation on the defective vehicles and the date when the manufacturer submitted its recall report to the NHTSA. Their measure thus does not consider voluntary recalls—i.e., recalls that do not have an associated NHTSA investigation. Further, their measure does not consider the time that a manufacturer is aware of the defect prior to the commencement of a formal NHTSA investigation.[3] As we describe next, our data set overcomes these limitations.

3.0. Recall Timing and Process Data Set Description

3.1. Data context

When an automobile manufacturer becomes aware of a product defect that may harm a consumer, and it decides to recall the defective products, it must furnish to the NHTSA what the Title 49 of the Code of Federal Regulations calls a "defect and noncompliance information report" (hereafter, "recall report" for brevity). This report is described in Part 573.6 (that is, 49 CFR § 573.6) of the law. Further, section (c) (6) of this report is required to contain "a chronology of all principal events that were the basis for the determination that the defect related to motor vehicle safety" (Legal Information Institute 2021:1). For example, the chronology section from a GM recall that is a part of our data set reads in part:

> *"On April 28, 2015, GM received notification from the steering supplier that there was an issue with the inner tie rod assembly machine. On May 1, 2015, GM opened a safety investigation into the 2014-2015 model year Chevrolet Cruze and 2014 Volt vehicles. In its investigation, the supplier was able to identify how the defective steering gear had improperly passed through the assembly process. During the investigation, GM analyzed VOQs, TREAD data, warranty data, and other field data, which concluded that the inner tie rod assembly error was isolated to the*

[3] We highlight two differences between our data set and what Eilert et al. (2017) used. First, because Eilert et al. (2017) considered only recalls that were preceded by a NHTSA investigation, their sample was limited to 381 investigations and 201 associated recalls initiated between 1999 and 2012. In contrast, we consider all recalls, regardless of whether precipitated by a NHTSA investigation, resulting in substantially broader and more complete coverage of 2,120 recalls. Second, by using the manufacturer defect awareness date from manufacturer-provided recall reports, and not the date when the NHTSA opened a formal investigation, we capture a complete picture of the time manufacturers take to make the recall decision. Notably then, in contrast to Eilert et al. (2017) who found Toyota to have the shortest mean time-to-recall of 213.91 days, we document that Toyota has one of the relatively longest time-to-recall of 429 days (see Table 3). Thus, we suggest that our data set may yield new insights by including recalls voluntarily initiated by manufacturers and considering self-disclosed awareness dates that may precede the initiation of an investigation.

> *population the supplier identified. On June 26, 2015, GM's Safety Field Action Decision Authority (SFADA) decided to conduct a safety recall."* [4]

While the above example mentions a supplier defect, our data set includes many other examples directly attributable to manufacturing or design problems. The next example from a BMW recall illustrates such a recall.

> *"BMW became aware of this matter through its quality control analyses and processes. On June 2, 2009, during internal driving at one of the BMW manufacturing plants, a failure of the brake disc occurred. Immediately, BMW initiated an analysis in order to determine the cause of the brake disc failure. It was determined that, due to problems with brake disc production tooling, cracks could form on the disc during manufacturing. Using production lot information contained on the brake disc, specific brake disc supply chain information was identified leading to the identification of potentially affected vehicles. On June 18, 2009, BMW decided to conduct a voluntary safety recall."*[5]

These examples demonstrate that the chronology section is rich in content. The chronologies vary in length and describe the events undertaken by automobile manufacturers in reaching their recall decision. These dates and descriptions may yield insights for research on recall decision-making. The NHTSA makes the recall report, including the chronology section, publicly available. However, these data have not yet been extensively used by academic researchers, the NHTSA, or, to our knowledge, the manufacturers. We attribute this lack of attention to the unstructured form of these data. The NHTSA does use parts of the recall report to provide values of variables such as the number of vehicles recalled. The NHTSA makes publicly available this recall data file[6], which prior recall studies have used (e.g., Eilert et al. 2017; Shah et al. 2017). Our data set complements the extant research that exploits these data provided in the recall data file. For example, we provide descriptive statistics of how the chronology data are associated with data from the recall data file, such as the number of vehicles recalled and whether the recall was voluntary or mandatory.

Before undertaking our data collection, we interviewed a senior representative at the NHTSA's Recall Management Division and confirmed that the NHTSA does not use the chronology text data because they are highly unstructured, excessively long, and very difficult to parse into individual events.

[4] https://static.nhtsa.gov/odi/rcl/2015/RCLRPT-15V442-4893.PDF
[5] https://static.nhtsa.gov/odi/rcl/2009/RCDNN-09V255-2663.pdf
[6] https://www-odi.nhtsa.dot.gov/downloads/folders/Recalls/FLAT_RCL.zip

This representative acknowledged that breaking down the chronology text into events and associated dates and descriptions would help them understand how manufacturers decide to recall, thus indicating the value of our data collection, cleaning, and coding for not only researchers but also the NHTSA. Indeed, following a series of highly publicized automobile recalls that triggered internal investigations, senior leadership at the Department of Transportation reprimanded the NHTSA for limited information sharing and inadequate data analysis (Office of Inspector General 2018). Our work may help in this sharing and analysis process for researchers and the NHTSA alike.

3.2. Data collection and cleaning

Our data range from 2009 to 2018 inclusive. We consulted the Wards Intelligence database for the 10 years of our timeframe to identify all 27 manufacturers that sold passenger cars and light-motor vehicles in the United States between 2009 and 2018. We also read the list of all recalling manufacturers in the NHTSA's recall data file to ensure we covered all relevant manufacturers. Each manufacturer initiated multiple recalls during the timeframe of our data set. We identified all 2,217 recalls from the NHTSA's recall data file in our time period. We removed 97 recalls that had either no chronology in any of the PDF files provided by the NHTSA or no dates provided in very tersely written chronology sections.[7] We thus provide data for 2,120 recalls (2,217 − 97), each identified by the NHTSA campaign number.[8]

We appended each NHTSA campaign number to a standard web link to obtain the web link from where we could download all PDF files[9] (see Figure A1 in the online appendix). The PDF file labeled "Defect Notice 573 Report" (the bottommost row in Figure A1)—with the file name prefix RCLRPT (RCL means recall and RPT means report)—is an example of the file of our interest. As we progressed from one recall to the other, we realized four characteristics of the raw data that demanded extra care in our data collection. First, over the years, the NHTSA has changed the label of the recall report and the

---

[7] The excluded 97 recalls are available upon request, but are not included in the data set as many of the key variables are missing.
[8] For example, 15V442000 (see https://www-odi.nhtsa.dot.gov/downloads/folders/Recalls/RCL.txt). The first two characters indicate the year in which the manufacturer initiated the recall (the "15" in 15V442000 indicates the year 2015), the third character is V, indicating a vehicle recall. The three digits from character 4 to character 6 comprise the sequence number that the NHTSA assigns to a recall in the focal year. The last three characters are always 000. For example, the first recall has the campaign number 09V001000, and the last campaign number is 18V935000.
[9] For example, see https://www.nhtsa.gov/recalls?nhtsaId=15V442. The web page contains a link to the multiple documents (PDF files) associated with the recall. For example, https://www.nhtsa.gov/recalls?nhtsaId=15V442 contains 13 associated documents. See Figure A1 in online appendix.

prefix given to the file name. For example, unlike the recall ID 15V442000 (Figure A1), the recall ID 09V001000 uses the label "Defect Notice (Part 573)" and the prefix RCDNN for the recall report (RC refers to recall and DNN means defect and noncompliance notice).[10] Second, manufacturers often submit amendments to the recall reports, where the amendment could be to the chronology section or any other. For example, Honda submitted three amendments to its reports for the recall 17V418000.[11] We thus had to read all the amended reports and identify the one with the most updated chronology section. Third, for several recalls, manufacturers provided the chronology section in a "miscellaneous" report (RMISC*.pdf) and referred to this file in the recall report. For example, for the recall 17V418000, Honda provided the chronology section in a separate PDF file.[12] As we note below, 10% of recalls in our data set had their chronology text in these miscellaneous files. Fourth, the PDF files of the reports of recalls initiated before 2015 contain images rather than selectable text. These four unique characteristics of the raw data meant that we had to download all 82,785 recall PDF files associated with the recalls in our data set to ensure we captured all recall chronologies. We used an optical character recognition program to convert the pre-2015 PDF files into text that we could select and copy for further processing.

We prioritized data integrity over expediency. In total, we carefully directed the day-to-days tasks of 12 research assistants (RAs) over four months to review the 82,785 recall PDF files and hand-collect the chronology data. The steps followed were to (1) identify the PDF file that includes the chronology section (file name beginning with RCLRPT, RCDNN, or RMISC), (2) locate the chronology section within that file (often, but not always, under heading 573.6 [c]), and (3) identify discrete events in the chronology text, and for each event, the date and the description. Each RA read the chronology section of each recall assigned to her/him and copied and pasted the date and description of each event into a spreadsheet. Next, we wrote a software program[13] to merge spreadsheets from each RA. We assigned another RA, over a three month period, the task of carefully re-reading each recall captured in this spreadsheet to verify the

[10] See https://www.nhtsa.gov/recalls?nhtsaId=09V001
[11] https://www.nhtsa.gov/recalls?nhtsaId=17V418
[12] https://static.nhtsa.gov/odi/rcl/2017/RMISC-17V418-5009.pdf
[13] We wrote the programs either in Python, R, or Stata, depending on the purpose of the software program.

values of all variables, including all dates, descriptions, events, and supplier information for each recall. The author team then reviewed, inspected, and approved all the data collection and cleaning and variable generation. Table 1 provides the name, data type, and description of each variable in our data set as described below.[14] The online appendix presents a numbered list of steps we followed to collect and clean the data. The list can help researchers extend our data set to recalls initiated after 2018.

3.3. Variables Descriptions.

NHTSA Campaign Number is a duplicate of the field named *CAMPNO* in the NHTSA's (1) recall data file (FLAT_RCL.txt), (2) quarterly report data file (FLAT_RCL_Qtrly_Rpts.txt), and (3) defect investigations data file (FLAT_INV.txt).[15] The NHTSA views a recall as a campaign and uses CAMPNO to uniquely identify a recall campaign. *NHTSA Campaign Number* can be used to merge our data set with the NHTSA's data files. This variable is a string of 9 characters. Examples include 09V001000 and 18V935000.

Manufacturer Name. Over the years, as the legal names of some automobile manufacturers have changed, so have the *Manufacturer Name* values in the NHTSA's recall data file (field named MFGNAME). We standardized the names for each of the 27 manufacturers in our data set. The *Manufacturer Name* in our data set is a string of 17 characters.

Manufacturer Country records a manufacturer's country of headquarter. We used Ward's Intelligence database to determine a manufacturer's country of headquarter.

File Name is the name of the PDF file from where we copied the chronology. The source PDF file can be accessed from the NHTSA website using the provided *File Name*.[16]

---

[14] Our data set is a Microsoft Excel 1997-2003 Workbook (XLS) file. Sheet 1—named Dictionary—is the same as Table 1. Sheet 2—named Data—provides the data, which includes 2,120 rows (one for each recall) and 110 columns. Sheet 3—named Complaint Data—includes additional consumer complaint data only for a subset of 21 recalls as explained in Section 4.0 and footnote 21.

[15] https://www-odi.nhtsa.dot.gov/downloads/folders/Recalls/FLAT_RCL.zip for data file, https://www-odi.nhtsa.dot.gov/downloads/folders/Recalls/FLAT_RCL_Qtrly_Rpts.zip for quarterly reports, and https://www-odi.nhtsa.dot.gov/downloads/folders/Investigations/FLAT_INV.zip for defect investigations.

[16] https://www.nhtsa.gov/recalls

Chronology contains the unedited/raw chronology text copied from the PDF file named in *File Name*[17] variable. *Chronology* ranges from 100 characters to 15,728 characters.

First Date and **Defect Awareness Date.** We read the *Chronology* of each recall and stored the values of the first date (*First Date*). We converted the *First Date* into *Defect Awareness Date* in three steps. First, for example, recall 09V042000's chronology text reads: "It was brought to Isuzu's attention from JBK on March 31, 2008 that an improper setting…" Therefore, we stored the values of *First Date* for 09V042000 as March 31, 2008. Of the 2,120 recalls, 1,239 had a clean *First Date* in a similar format to this example, which we simply copied to *Defect Awareness Date*. Second, some *First Date* values include date ranges. For instance, recall 09V001000's chronology section states that "In mid-June 2008, Maserati received information regarding a broken ball joint stud…". We converted the "mid-June 2008" period in *First Date* to "June 15, 2008" in *Defect Awareness Date*—that is, we replaced "mid" and "middle" with the 15th of each month. Similarly, some values of *First Date* included a month and a year but no day (e.g., "January 2009" for 09V032000). We used the 15th of the month as the date for such values, converting the *First Date* to *Defect Awareness Date* for an additional 528 recalls. Third, 353 recalls' *First Date* included only the words such as "early" (e.g., 09V358000) or "beginning" (e.g., 15V192000), which we treated as occurring on the first date of the period referenced. Similarly, *First Date* included the words "late" or "end" (e.g., early November, or late December), which we replaced with the last date of the mentioned month. Using these approaches, we converted the *First Date* to *Defect Awareness Date* for each of the 2,120 recalls. The variable *First Date* in our data set provides the raw values mentioned in the chronology text, whereas the variable *Defect Awareness Date* stores the converted values.

Recall Date is a duplicate of the field name RCDATE in the NHTSA's recall data file (FLAT_RCL.txt). It is the official date of recall initiation (Borah and Tellis 2016; Eilert et al. 2017; Gao et al. 2015).

[17] *File Name* is a 22-character string, which takes values such as RCDNN-09V001-1525.PDF, RCLRPT-13V383-1060.PDF, and RMISC-18V935-0432.PDF. That is, the *File Name* has three parts, separated by a hyphen. The first part indicates that NHTSA used the prefix RCDNN for years 2009 through 2014, and switched to the prefix RCLRPT from 2013 to 2018. Recalling manufacturers started report the chronology text in a separate miscellaneous file from 2014 onwards. Nearly 50% of recalls in our data set have chronology text in RCLRPT*.pdf, 40% in RCDNN*.pdf, and the remaining 10% in RMISC*.pdf. The second part is the six-character NHTSA recall campaign number, which matches with the first six characters of NHTSA *Campaign Number*. The last part is a four-digit integer that NHTSA uses for the specific file.

Time-To-Recall. We computed *Time-to-Recall* as the difference in days between the *Recall Date* and the *Defect Awareness Date.*

Recall Year is the year component of Recall Date. *Recall Year* allows us to graph the variation, by year, in the values of our key variables.

Supplier Mentioned. OM researchers have distinguished manufacturer recalls from supplier recalls—that is, whether the party responsible for the defect is the automobile manufacturer or its supplier (Shah et al. 2017). We created an indicator variable, *Supplier Mentioned*, which equals "Yes" if the *Chronology* mentions a supplier. We followed a two-step process to set the values of this variable. First, we manually read each *Chronology* and set the *Supplier Mentioned* to "Yes" if the text (1) mentioned any variant of the word "supplier" or (2) included the name of an organization that was not the recalling manufacturer. This step identified 978 recalls as supplier recalls.

Second, we realized that checking the presence of the word "supplier" or the name of an organization other than the recalling manufacturer has its limitations. Therefore, following prior OM research on recalls, we obtained automotive supplier names from three sources: the MarkLines' Who Supplies Whom (WSW) database (Bray et al. 2019), Standard & Poor's Capital IQ, and the Bloomberg supply chain (SPLC) database (Steven and Britto 2016). Capital IQ and Bloomberg report the legal names of the suppliers, and together provided us with 864 legal names of automotive suppliers for the 27 manufacturers in our data set. WSW provided 9,929 name variants for suppliers of the 27 manufacturers. Name variants mean any variation of the supplier's legal name concatenated with either the plant location or the component manufactured.[18] To be as inclusive as possible in our search for supplier recalls, we used all 9,929 name variants from WSW and all 864 legal names from Standard & Poor's Capital IQ and the Bloomberg SPLC databases. That is, we searched each of 10,793 (9,929 + 864) supplier name variants in the *Chronology* of each recall and found 94 recalls that indicated a potential coding mismatch from the manual step described above (that identified 978 supplier recalls). Reviewing each of these 94 exceptions

[18] For example, Anhui Zhongding Sealing Parts Co., Ltd.; Anqing TP Goetze Piston Ring Co., Ltd.; Aptiv Electric Systems Co., Ltd. Guangzhou Branch; and Aptiv Electric Systems Co., Ltd. Yantai Branch.

allowed us to identify 43 “false positives”—that is, recalls for which we marked *Supplier Mentioned* = “Yes” but whose *Chronology* neither had the “supplier” word nor included any of our 10,793 supplier name variants. We also found 51 “false negatives”—that is, recalls that we marked as *Supplier Mentioned* = “No” but whose *Chronology* included a name from our list of 10,793 name variants. At the end of this exercise, we classified 986 (47%) recalls as supplier recalls.

Voluntary or Mandatory. The NHTSA’s recall data file includes a field named INFLUENCED_BY, which takes one of three values: MFR, ODI, and OVSC. ODI refers to the Office of Defect Investigations and OVSC is the Office of Vehicle Safety Compliance. If a recall has INFLUENCED_BY equal to either ODI or OVSC, we set the value of *Voluntary or Mandatory* field in our data set to “Mandatory”, and if the value is MFR, we set it to “Voluntary”.

Recall Size. The NHTSA’s recall data file includes a field named POTAFF, which reports the number of vehicles potentially affected by the recall. Because a recall can affect multiple year-make-model combinations, we sum the value of POTAFF for all such combinations of a recall. The sum is the value of *Recall Size* field in our data set.

Events, Event Date, and **Event Description.** The law (49 CFR § 573.6) requires the recalling manufacturer to describe the principal events in the recall decision-making process. We thus counted the number of *Events* a manufacturer reported in the *Chronology* of its recall decision-making. We used the following two rules. First, if a sentence or a phrase mentions a date and includes action words suggesting the beginning of a new event[19], use the date as the event date and the associated verbatim text as the event description. Second, if the text includes action words suggesting a new event but does not accompany a date, leave the event date as blank and copy the relevant text against the event description. For event *n*, the event date is saved in *EnDt* variable, and the event description is saved in *EnDesc* variable. Table 2 provides one example of a chronology text that comprised three events. The number of *Events* in our full

[19] For example, “GM received notification...,” “GM opened a safety investigation...,” and “GM decided to conduct a safety recall” (see Table 2).

data set varies from 1 to 49. Consequently, our data set includes *E1Date* through *E49Date* and *E1Description* through *E49Description*.

3.4 Data summary

Table 3 summarizes—by manufacturer—the number of recalls, the average *Time-to-Recall*, the average number of events, and the percentage of supplier recalls. We note a substantial variation in the values of each variable. We illustrate this variation by the recalling manufacturer (Figures 1-2), by the recalling year (Figures 3-5), by supplier recalls (Figures 6-7), by recalling manufacturer's home country (Figures 8-10), by whether the recall was mandatory or voluntary (Figures 11-12), and by recall size (the number of vehicles recalled) (Figures 13-14).

We observe a few interesting patterns in the data. First, the average *Time-to-Recall* for a manufacturer varies across all the 27 manufacturers, ranging from a minimum of 124 days for Jaguar to a maximum of 733 days for Suzuki (Figure 1). We also observe significant variation within each manufacturer. For example, across the 198 recalls announced by GM across the timeframe of our data set, the *Time-to-Recall* varies from 5 days to 2,581 days. The Japanese manufacturers had some of the longest *Time-to-Recall*, as seen in Figure 1 (and further summarized in Figure 8). This finding contradicts our expectations because these manufacturers have a long-standing reputation for high quality and lean manufacturing practices (Womack and Jones 1997). In contrast, the mean *Time-to-Recall* for each of the traditional "Big Three" U.S. automobile manufacturers—Fiat Chrysler, Ford, and GM—are relatively short. This difference in *Time-to-Recall* counters the assumed customer-focused responsiveness of Japanese automobile manufacturers, and the sluggishness of their U.S. counterparts (Womack and Jones 1997).

Figure 2 presents the average number of events each manufacturer undertook to make the recall decision. Significant variation exists on this variable as well, with average values ranging from under three steps for Ducati, Lamborghini, and Ferrari to more than seven for Kia and Fiat Chrysler. Interestingly, there is not a consistent correlation between the number of steps that the manufacturer reports in its

Chronology and the *Time-to-Recall*. Fiat Chrysler, for example, takes more steps than almost all other manufacturers (Figure 2) in a comparably short decision-making time (Figure 1).

We next demonstrate how our variables change over time. Figures 3 and 4 respectively present the average *Time-to-Recall* and the average number of events for each manufacturer in the data set. While the *Time-to-Recall* does not seem to follow any discernable pattern, the number of events generally demonstrates an upward trajectory over time, peaking in 2016. Figure 5 indicates the proportion of supplier recalls. Perhaps not surprisingly, as supply chains have become more global over time, so has the proportion of investigations that involve a supplier (Bray et al. 2019). In other words, our data indicate that almost half of all automobile defect investigations involve suppliers. This percentage increased from 20% in 2009 to 52% in 2018, and peaked at 58% in 2016. Because the law requires the manufacturers to mention suppliers only when the recall is related to a supplier, we conclude that approximately half of all automobile recalls are due to issues attributable to suppliers. This finding is novel to the NHTSA, as we have learned from our discussions with senior NHTSA representatives.

Figures 6 and 7 explore whether the *Time-to-Recall* and the number of events vary by whether the *Chronology* mentioned a supplier. The graphs suggest that, on average, supplier recalls do not necessarily have a longer *Time-to-Recall* (Figure 6) but they involve more events (Figure 7). In addition, we prepare descriptive statistics by manufacturer home country, as depicted graphically in Figures 8, 9 and 10. These provide some tentative indication of potential cross-country differences. For example, the average time-to-recall was longest for recalls by Japanese manufacturers, whereas average time-to-recall was shortest for those issued by Swedish and U.K.-based manufacturers (Figure 8). Further, the number of events reported in the chronology was highest for recalls associated with Italy-based manufacturers (Figure 9), and a greater portion of recalls were associated with suppliers for recalls initiated by Korean manufacturers (Figure 10). However, because each country has few manufacturers, these potential differences should be interpreted with caution and may warrant further research.

Following prior research (Haunschild and Rhee 2004), we next differentiate the average *Time-to-Recall* and the average number of events by whether the recall is voluntary or mandatory (Figures 11-12).

When the NHTSA mandates a recall, the *Time-to-Recall* and the number of events are higher. This correlation indicates that the NHTSA may mandate a recall after the manufacturer has "dragged its feet" in making the decision voluntarily. This conjecture, motivated by our data set, is something that extant recall research has not yet explored because the data have been unavailable until now.

Finally, we explore whether the *Time-to-Recall* and the number of events vary by the size of the recall, measured as the number of vehicles affected by the defect (Figures 13-14). The higher the number of vehicles affected, the longer the investigations last and the higher the number of events in such investigations. The manufacturer thus appears to be less prompt in deciding to recall when the defect affects multiple vehicles. In aggregate, our Figures 1 through 14 present intriguing preliminary evidence that can drive questions for future research. Section 5 presents a few such questions.

4. Robustness Check

A key assumption in our data set is that the first date identified in the chronology reported by the manufacturer is the actual date when the manufacturer first became aware of the defect. In other words, *Time-to-Recall* is comparable across manufacturers only to the extent that manufacturers accurately report to the NHTSA the first date on which they became aware of the defect. We explore the validity of this assumption using the NHTSA complaints data (Bray et al. 2019; Çolak and Bray 2016). We note that the NHTSA does not directly associate individual complaints with a recall. We thus describe below the process by which we associated complaints with recalls.

Prior research has used the NHTSA complaints data file to examine supply chain proximity and component defects (Bray et al. 2019). Following this research, we downloaded on February 1, 2021, the NHTSA's complaints data file.[20] The file included 1,633,168 complaints. We first identified those complaints that matched any recall in our data set by matching firm, make, model, model-year, and component. Because recalls often affect multiple models and model-years, we made this comparison for all models and model-years impacted by each recall. We eliminated all complaints that did not match a recall

[20] https://static.nhtsa.gov/odi/ffdd/cmpl/FLAT_CMPL.zip

in our data set. Next, we eliminated matched complaints that had a complaint date (field named LDATE in the NHTSA's FLAT_CMPL.txt) later than the date of recall initiation (field named RCDATE in the NHTSA's FLAT_RCL.txt). While these complaints may directly relate to our recalls, they are not useful for validating our *Defect Awareness Date* if they were received after the recall was initiated. Similarly, we removed complaints that the NHTSA received before the manufacturing date of the vehicles affected by the recall (field named MFRDATE in the NHTSA's recall data file). We took this step because complaints received before the recalled vehicles were manufactured could not be associated with the recall. Lastly, we eliminated all complaints that the NHTSA received after the *Defect Awareness Date*. We did so because complaints received after our defect awareness dates indicate support for our data collection and cleaning, whereas complaints received before our defect awareness dates require closer scrutiny. After comparing all 1,633,168 complaints to the 2,120 recalls in our data set, we identified 109 recalls that had a complaint received before the *Defect Awareness Date*, and were not eliminated by any of the above screening criteria. However, merely because a complaint matches each of these criteria does not mean the complaint relates to the recall because many component categories used in the complaint and recall data, such as "engine," could indicate two very different problems. Thus, our final step was to carefully read the description of each of these 109 complaints (field named CDESCR in the NHTSA's FLAT_CMPL.txt) and determine whether the complaint is associated with the same defect as the one for which the manufacturer initiated the focal recall. To match complaint descriptions against recall descriptions, we leveraged Section 5 of the recall report which describes the defect related to the recall. Our bias in this comparison was to assume the complaint was related to the recall, and to exclude only those complaint-recall pairs that were unlikely to be related. Table 4 provides two examples of this comparison. Example 1 of Table 4 presents a complaint-recall pair that we determined was very likely to be connected to the same problem, and in which the complaint date pre-dated the *Defect Awareness Date*. Example 2 of Table 4 presents a similar comparison, but one in which the complaint-recall pair was unlikely to be connected to the same problem. The comparison conclusion paragraphs of Table 4 explain our justification for either associating, or not associating, the respective complaint with the recall, and indicate the logic we used in our complaint-recall matching process. From the 109 possible matches, our comparison yielded 21 complaints that were associated with the recall. For these 21 recalls, we provide additional data in our data set. Specifically, we include the NHTSA complaint identifier and complaint date from the NHTSA's complaints data file as additional data field in our data (fields named *CMPLID* and *LDATE* from FLAT_CMPL.txt).[21]

Our determination that just 21 recalls, or 1% of the 2,120 recalls, had a matched complaint that pre-dates the *Defect Awareness Date* indicates strong validity in our assumption that the *Chronology* mentions the correct defect awareness date. Further, to the extent that researchers have reason to expect inaccurate or incomplete disclosure, we provide the date of the earliest matched complaint to compute an alternate measure of *Time-to-Recall*.[22] However, even if all of the complaint-recall pairs that we deemed to not be associated were incorrectly categorized, there are at most 109 recalls, or 5% of the 2,120 recalls, that had a matched complaint that pre-dates the *Defect Awareness Date*.

5. Future Research Questions

We believe this data set has great promise in informing future recall research. We provide below three brief potential areas for future research.

Researchers have conceptualized a recall as an organizational crisis (Pearson and Clair 1998), examining how a firm's strategies in the aftermath of a recall help it recover from the ensuing crisis (Cleeren et al. 2013; Haunschild and Rhee 2004; Liu et al. 2017). While a recall decision date is the event when external stakeholders become aware of the crisis, the crisis begins when the firm becomes aware of the product defect that may require a recall. Examining firm behavior only after the firm decides to recall, as a majority of recall research does, compromises our knowledge in two ways. First, an exclusive focus on the post-recall phase does not offer insights into what actions the firm undertook before deciding to recall. The pre-recall actions may explain the difference between a preventive recall strategy and its

[21] Sheet 3 of the data set—named Complaint Data—includes three columns; NHTSA Campaign Number, Complaint ID and Complaint Date for these 21 recalls only.
[22] Considering the complaint date (and not the First Date) as the defect awareness date increases the average time-to-recall by 4.5 days. Further, this "revised" time-to-recall has 99.28% correlation with our primary time-to-recall.

reactive counterpart (Hora et al. 2011). Second, ignoring the pre-recall actions can produce biased insights. For example, researchers have documented that the announcement of a recall decision may lower the firm's stock price (Liu et al. 2017). The critical assumption here is that the recall decision date and time are exogenous. Our data set can help test this assumption.

In addition, since a product recall is a firm's acknowledgment of a quality failure in its product, recall researchers have cast a learning view of a recall. However, the empirical evidence has focused exclusively on the perceptible outcomes, ignoring the process that produces the outcomes (Haunschild and Rhee 2004; Kalaignanam et al. 2013; Thirumalai and Sinha 2011) because process data have, until now, been unavailable. As a result, we know little about how the recalling manufacturer develops knowledge about a product failure, and whether it involves a stakeholder, such as a supplier, in developing this knowledge. Although we know whether firms learn from recalls, we do not understand how they do so. This lack of knowledge characterizes not only automobile recalls but also other types of firm failures. We attribute this gap in our understanding to the absence of micro-process level data that could help researchers discover firm practices and processes in the aftermath of defect awareness. Our recall timing and process data set can help fill this gap.

One additional pertinent lens is information disclosure theory. The classic view of information disclosure focuses on the quantity of information (Galbraith 1974). However, too much information can exceed users' information processing capacity, leading to information overload and degrading the quality of user decisions. We thus propose the quantity and quality of the disclosed information as two relevant characteristics to be examined. Future research can text-mine the event descriptions we provide and measure simple yet insightful characteristics such as word count, sentiment, objectivity, and readability, all of which may lead to a deeper understanding of recall decision-making.

6. Conclusion

Recall research is a burgeoning field of study across multiple disciplines and has predominantly focused on recall consequences and causes. Few academic studies have explored the recall decision-making process because necessary data are sparse. Our collecting, cleaning, and coding of a decade of unstructured recall timing and process data help address this gap and hopefully stimulate numerous future recall studies.

References

Archer N, Wesolowsky G (1996) Consumer response to service and product quality: A study of motor vehicle owners. *Journal of Operations Management* 14(2):103-118.
Ball GP, Siemsen E, Shah R (2017) Do plant inspections predict future quality? The role of investigator experience. *Manufacturing & Service Operations Management* 19(4):534-550.
Ball GP, Shah R, Wowak KD (2018a) Product competition, managerial discretion, and manufacturing recalls in the US pharmaceutical industry. *Journal of Operations Management* 58-59:59-72.
Ball GP, Shah R, Donohue K (2018b) The decision to recall: A behavioral investigation in the medical device industry. *Journal of Operations Management* 62:1-15.
Borah, A, Tellis, GJ (2016) Halo (spillover) effects in social media: do product recalls of one brand hurt or help rival brands? *Journal of Marketing Research*, 53(2):143-160.
Bray, RL, Serpa JC, Çolak A (2019) Supply chain proximity and product quality. *Management Science* 65(9):4079-4099.
Cleeren K, Van Heerde HJ, Dekimpe MG (2013) Rising from the ashes: How brands and categories can overcome product-harm crises. *Journal of Marketing* 77(2):58-77.
Çolak A, Bray R (2016) Why do automakers initiate recalls? A structural econometric game. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2682645 [11 August 2021].
Cowley S (2016) From Tylenol to Fitbit: 10 notable product safety recalls. *The New York Times* 29 May. https://www.nytimes.com/2016/05/30/business/from-tylenol-to-fitbit-10-notable-product-safety-recalls.html [11 August 2021].
Eilert M, Jayachandran S, Kalaignanam K, Swartz TA (2017) Does it pay to recall your product early? An empirical investigation in the automobile industry. *Journal of Marketing* 81(3):111-129.
Gao H, Xie J, Wang Q, Wilbur KC (2015) Should ad spending increase or decrease before a recall announcement? The marketing–finance interface in product-harm crisis management. *Journal of Marketing*, 79(5):80-99.
Haunschild PR, Rhee M (2004) The role of volition in organizational learning: The case of automotive product recalls. *Management Science* 50(11):1545-1560.
Hora M, Bapuji H, Roth AV (2011) Safety hazard and time-to-recall: The role of recall strategy, product defect type, and supply chain player in the US toy industry. *Journal of Operations Management* 29(7):766-777.
Janeway K (2019) Samsung settles class-action lawsuit over 'exploding' washing machines. *Consumer Reports* 8 February. https://www.consumerreports.org/washing-machines/samsung-settles-washer-lawsuit/ [11 August 2021].
Jensen C (2014) Car seat maker Graco under investigation for delayed reporting of defect. *The New York Times* 1 December. https://www.nytimes.com/2014/12/02/business/car-seat-maker-graco-under-investigation-for-delayed-reporting-of-defect.html [11 August 2021].
Kalaignanam K, Kushwaha T, Eilert M (2013) The impact of product recalls on future product reliability and future accidents: Evidence from the automobile industry. *Journal of Marketing* 77(2):41-57.
Legal Information Institute (2021) https://www.law.cornell.edu/cfr/text/49/573.6 [February 14, 2021].
Liu Y, Shankar V, Yun W (2017) Crisis management strategies and the long-term effects of product recalls on firm value. *Journal of Marketing* 81(5):30-48.
Maidenberg, M (2018) Ford agrees to pay $299.1 million to resolve air-bag lawsuit. *The Wall Street Journal* 16 July. https://www.wsj.com/articles/ford-agrees-to-pay-299-1-million-to-resolve-air-bag-lawsuit-1531780925 [11 August 2021].

Mukherjee UK, Sinha KK (2018) Product recall decisions in medical device supply chains: a big data analytic approach to evaluating judgment bias. *Production and Operations Management* 27(10):1816-1833.

Mukherjee UK, Ball GP, Wowak KD, Natarajan KV, Miller JW (2021). Hiding in the Herd: The Product Recall Clustering Phenomenon. *Manufacturing & Service Operations Management*. Forthcoming.

Office of Inspector General (2018) NHTSA's management of light passenger vehicle recalls lacks adequate processes and oversight. *Report No. ST2018062,* 1-40.

Pearson CM, Clair JA (1998) Reframing crisis management. *Academy of Management Review* 23(1):59-76.

Shah R, Ball GP, Netessine S (2017) Plant operations and product recalls in the automotive industry: An empirical investigation. *Management Science* 63(8):2439-2459.

Steven AB, Britto RA (2016) Emerging market presence, inventory, and product recall linkages. *Journal of Operations Management* 46:55-68.

Thirumalai S, Sinha KK (2011) Product recalls in the medical device industry: An empirical exploration of the sources and financial consequences. *Management Science* 57(2):376-392.

Womack, JP, Jones DT (1997) Lean thinking—banish waste and create wealth in your corporation. *Journal of the Operational Research Society* 48(11):1148-1148.

Wowak AJ, Mannor, MJ, Wowak, KD (2015) Throwing caution to the wind: The effect of CEO stock option pay on the incidence of product safety problems. *Strategic Management Journal* 36(7):1082-1092.

Wowak, KD, Ball GP, Post C, Ketchen DJJ (2021) The Influence of Female Directors on Product Recall Decisions. *Manufacturing & Service Operations Management*. 23(4):895-913.

Table 1. Data Dictionary

Name	Data type	Description
NHTSA Campaign Number	Nine-character string	Unique alphanumeric ID that NHTSA uses to identify a recall. We source this variable from NHTSA's FLAT_RCL.txt data file.
Manufacturer Name	17-character string	Standardized name for the manufacturer brand.
Manufacturer Country	String	Name of the manufacturer's home country
File Name	24-character string	Name of the PDF file which contains the chronology text. Users can search the file name on the web and access the raw PDF file.
Chronology	15,728-character string	Text provided by the recalling manufacturer to NHTSA, containing a chronology of all principal events that were the basis for the determination that the defect related to motor vehicle safety
First Date	Date (MM/DD/YYYY)	First date mentioned in the *Chronology*
Defect Awareness Date	Date (MM/DD/YYYY)	The date when the manufacturer first became aware of the defect
Recall Date	Date (MM/DD/YYYY)	The date when the manufacturer notified NHTSA of its recall. We source this variable from NHTSA's FLAT_RCL.txt data file.
Time-to-Recall	Long integer	The number of days between the manufacturer's defect awareness date and the recall date
Recall Year	Short integer (YYYY)	The year component of *Recall Date*
Supplier Mentioned	Three-character string:Yes or No	Does the *Chronology* mention a supplier?
Voluntary or Mandatory	Nine-character string: Voluntary or Mandatory	Did the manufacturer initiate the recall? If yes, the recall is considered voluntary. If no, the recall is considered mandatory.
Recall Size	Long integer	The number of vehicles affected by the focal recall
Events	Short integer	The number of events mentioned in the *Chronology*
Event Date	String	Date when the event took place
Event Description	String	Description of the event

Table 2. Example of separating chronology into event dates and event descriptions

Chronology	**E1Dt** **E1Desc**	**E2Dt** **E2Desc**	**E3Dt** **E3Desc**
On April 28, 2015, GM received notification from the steering supplier that there was an issue with the inner tie rod assembly machine. On May 1, 2015, GM opened a safety investigation into the 2014-2015 model year Chevrolet Cruze and 2014 Chevrolet Volt vehicles. In its investigation, the supplier was able to identify how the defective steering gear had improperly passed through the assembly process. The supplier was also able to use torque and angle information stored as part of the production process to identify additional steering assemblies that may have been subject to the same error and therefore potentially contain the same defect. Those assemblies were then tracked by GM back to the individual vehicles produced by GM through vehicle identification numbers. During the investigation, GM analyzed VOQs, TREAD data, warranty data, and other field data, which concluded that the inner tie rod assembly error was isolated to the population the supplier identified. On June 26, 2015, GM's Safety Field Action Decision Authority (SFADA) decided to conduct a safety recall.	April 28, 2015 On April 28, 2015, GM received notification from the steering supplier that there was an issue with the inner tie rod assembly machine.	May 1, 2015 On May 1, 2015, GM opened a safety investigation into the 2014-2015 model year Chevrolet Cruze and 2014 Chevrolet Volt vehicles. In its investigation, the supplier was able to identify how the defective steering gear had improperly passed through the assembly process. The supplier was also able to use torque and angle information stored as part of the production process to identify additional steering assemblies that may have been subject to the same error and therefore potentially contain the same defect. Those assemblies were then tracked by GM back to the individual vehicles produced by GM through vehicle identification numbers. During the investigation, GM analyzed VOQs, TREAD data, warranty data, and other field data, which concluded that the inner tie rod assembly error was isolated to the population the supplier identified.	June 26, 2015 On June 26, 2015, GM's Safety Field Action Decision Authority (SFADA) decided to conduct a safety recall.

Table 3. Summary statistics

Manufacturer	Manufacturer's Home Country	Number of recalls (2009-2018)	Average time-to-recall (in days)	Average number of events	Percentage of supplier recalls
Aston Martin	UK	14	163	3.5	29%
Bentley	UK	5	236	4.6	0%
BMW	Germany	162	363	5.3	50%
Ducati	Italy	19	156	2.8	26%
Ferrari	Italy	14	452	2.9	50%
Fiat Chrysler	Italy	279	250	7.2	48%
Ford	USA	227	172	4.0	48%
General Motors	USA	198	160	5.2	40%
Honda	Japan	146	325	5.9	51%
Hyundai	Korea	72	300	3.9	47%
Isuzu	Japan	15	321	3.9	27%
Jaguar	UK	75	124	5.2	52%
Kia	Korea	45	291	7.4	67%
Lamborghini	Italy	9	234	2.6	11%
Maserati	Italy	22	177	6.0	59%
Mazda	Japan	61	558	5.4	41%
McLaren	UK	4	353	4.0	100%
Mercedes-Benz	Germany	138	245	4.5	45%
Mitsubishi	Japan	45	732	4.4	47%
Nissan	Japan	126	207	4.5	51%
Porsche	Germany	33	261	3.8	45%
Subaru	Japan	53	441	5.3	51%
Suzuki	Japan	35	733	6.9	40%
Tesla	USA	10	236	3.5	40%
Toyota	Japan	144	429	4.1	44%
Volkswagen	Germany	134	208	3.9	43%
Volvo	Sweden	35	139	5.4	40%
Total		2,120	282	5.1	47%

Table 4. Complaint to Recall Comparison: Robustness Check Examples

Example 1: Complaint **associated** with the same defect as the recall, pre-dating Defect Awareness Date

Component Information	
Complaint ID	845147
Complaint Date	06/13/2011
Model-years	2009 Mazda6
Component	Latches/Locks/Linkages: Doors: Latch
Complaint Description	I have a 2009 Mazda6 S Touring with 18,000 miles on it and the three screws that hold the drivers door latch assembly on came loose and fell out. These screws should of been locked tightened in place and it seems that all the latch screws on the car were not secured not to work loose.

Recall Information	
NHTSA Campaign Number	13V425000
Defect Awareness Date[a]	05/13/2013
Model-years	2009-2013 Mazda6
Component	Latches/Locks/Linkages: Doors: Latch
Recall Description[b]	In certain 2009-2013 model year Mazda6, the door latch mounting screws may become loose due to lack of torque during manufacturing and/or unevenness of the door shell in the latch mounting area. This situation may cause the screw's axial force to decrease, which in turn may cause the screw to loosen while the vehicle is in use. If all three door latch mounting screws become excessively loose the door latch mechanism would not engage and the door ajar warning light will illuminate. However, if the warning light goes unnoticed and the vehicle is driven, the door might open while the vehicle is in motion increasing the risk of accident or injury.

[a] *Defect Awareness Date* is our created variable as explained in Section 3.2.1 of this paper.
[b] *Recall Description* is what is provided in Section 5 of the defect and non-compliance reports (whereas we use Section 6 for Chronology data). Section 5 is titled "Description of Defect."

Comparison Conclusion: Both the complaint and recall discuss door latch screws that were not properly tightened. We concluded that the recall was related to the issue flagged in the customer complaint and associated this complaint with the recall.

Example 2: Complaint **not associated** with the same defect as the recall, pre-dating Defect Awareness Date

Component Information

Complaint ID	552698
Complaint Date	10/20/2005
Model-years	2005 BMW 745 LI
Component	Latches/Locks/Linkages: Doors: Latch
Complaint Description	The contact's finger was caught in the automatic door, causing injury to her finger. This was the passenger's side automatic door. The dealer has not currently been contacted. She felt this was a safety defect because it could have caused serious injury to one of her children.

Recall Information

NHTSA Campaign Number	12V504000
Defect Awareness Date[a]	05/31/2007
Model-years	2005-2007 BMW 7-series
Component	Latches/Locks/Linkages: Doors: Latch
Recall Description[b]	This recall involves the door latching system of 7-Series vehicles equipped with both the "Comfort Access" (CA) and "Soft Close Automatic" (SCA) options. The SCA system includes a control cam which is operated via an electric motor and is used to depress the door release lever which unlatches the door. In rare cases, due to the SCA software functionality, and the geometric design of the control cam, the cam may stop rotating at a position where the door release lever is depressed. If this were to occur, the door may not latch. This condition would be noticeable to the vehicle occupant. If the vehicle occupant then held the door closed in an attempt for it to latch, the cam would rotate for a few seconds and then stop (as part of the SCA software function). When the cam stopped rotating, it could be in a position where the door would appear to be latched closed, although it would only be closed due to the internal pressure of the SCA system and cam. An external influence, such as an irregular road surface, or inadvertent interior contact with the door, could lead to an unexpected opening of the door.

[a] *Defect Awareness Date* is our created variable as explained in Section 3.2.1 of this paper.
[b] *Recall Description* is what is provided in Section 5 of the defect and non-compliance reports (whereas we use Section 6 for Chronology data). Section 5 is titled "Description of Defect."

Comparison Conclusion: The complaint addresses an automatic door closing on a customer's finger. The recall addresses a door that appears to be closed but is not properly latched. Therefore, we concluded that the complaint was not related to the recall.

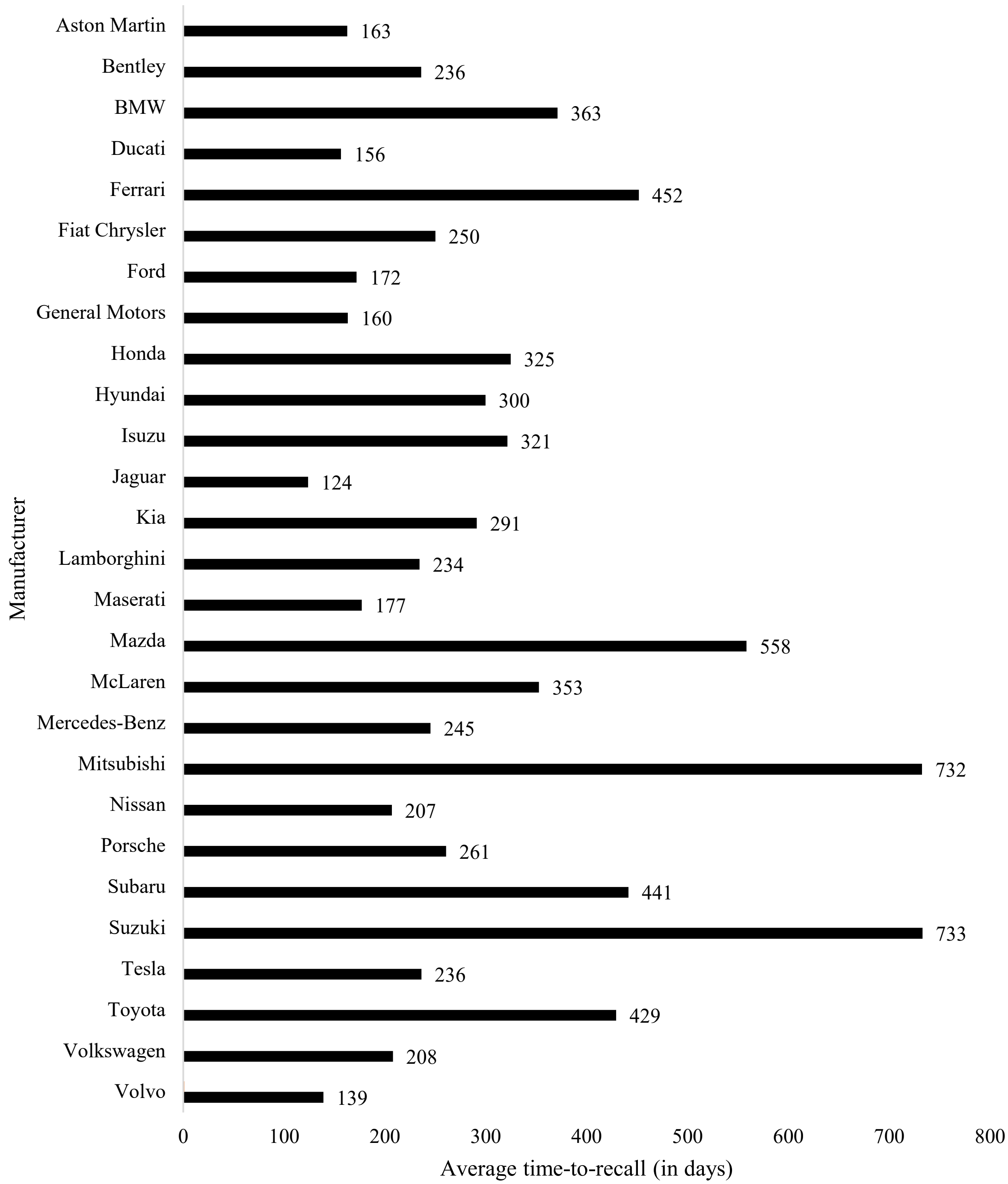

Figure 1. Average time-to-recall, by manufacturer

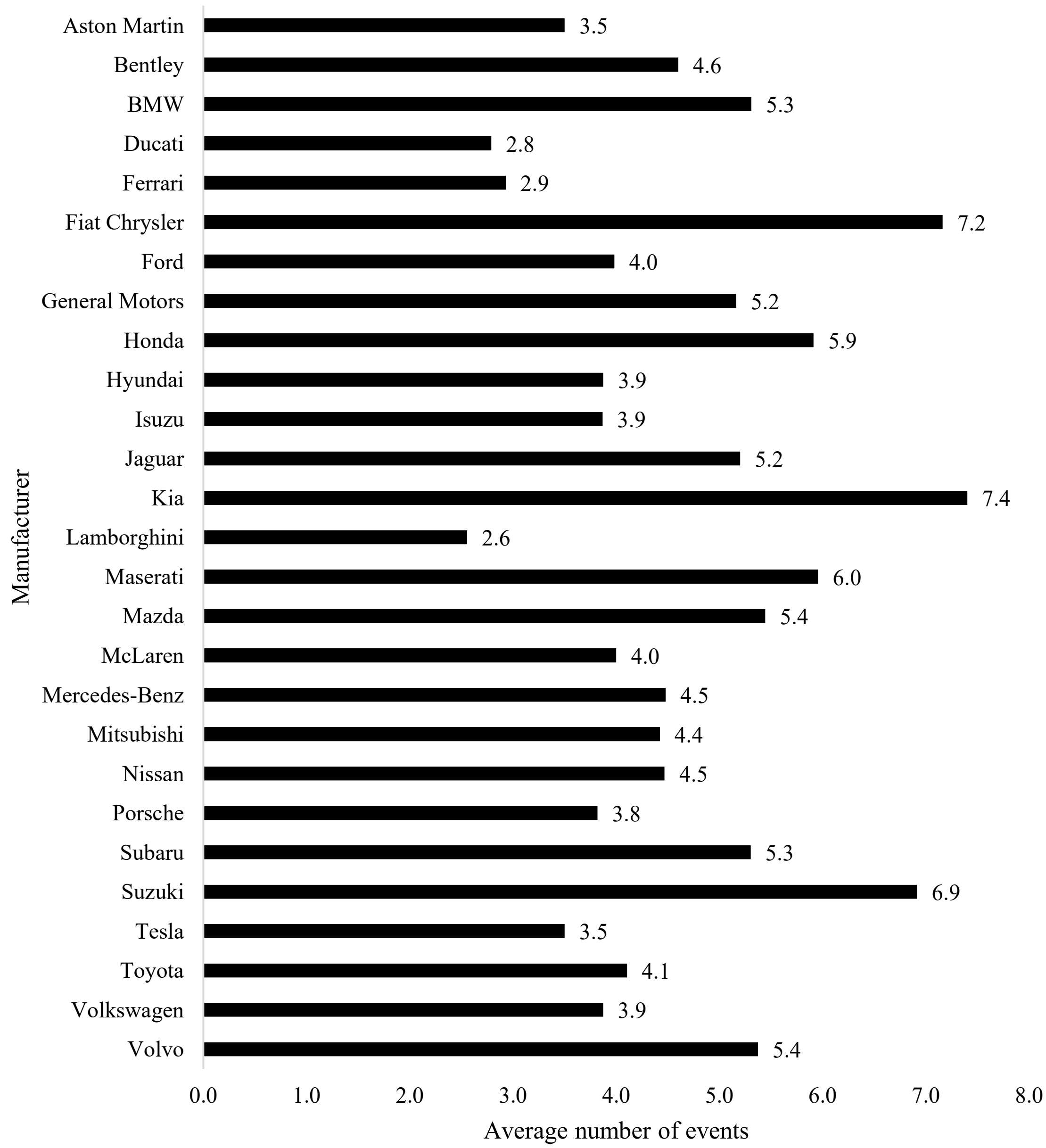

Figure 2. Average number of events, by manufacturer

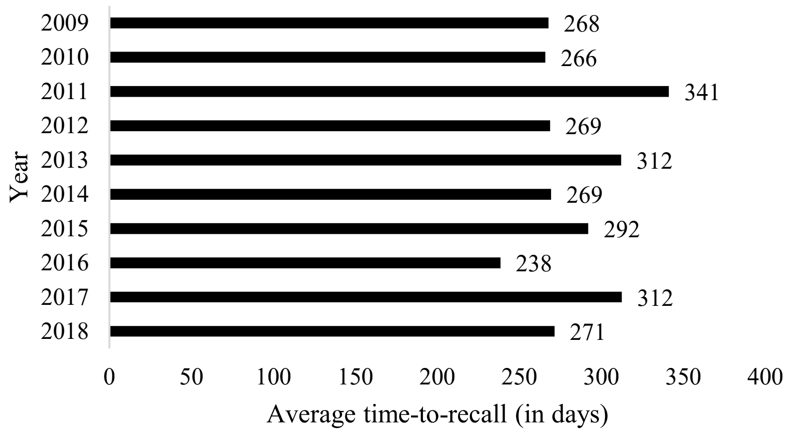

Figure 3. Average time-to-recall, by recall year

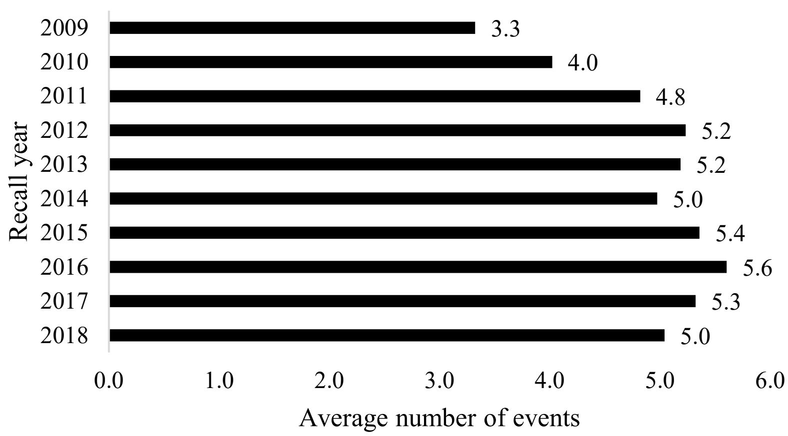

Figure 4. Average number of events, by recall year

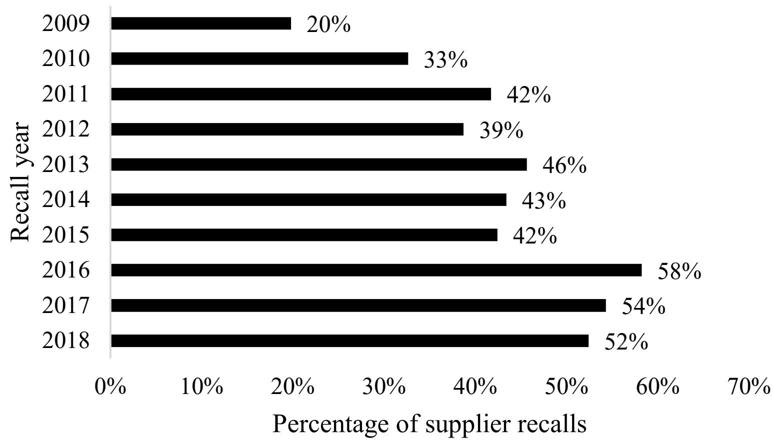

Figure 5. Percentage of supplier recalls, by recall year

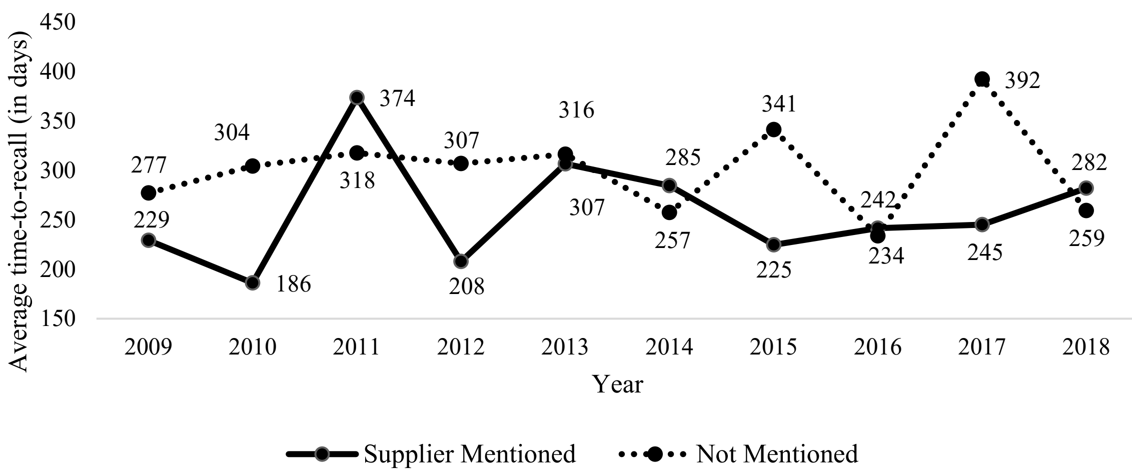

Figure 6. Average time-to-recall, by whether the *Chronology* mentioned a supplier

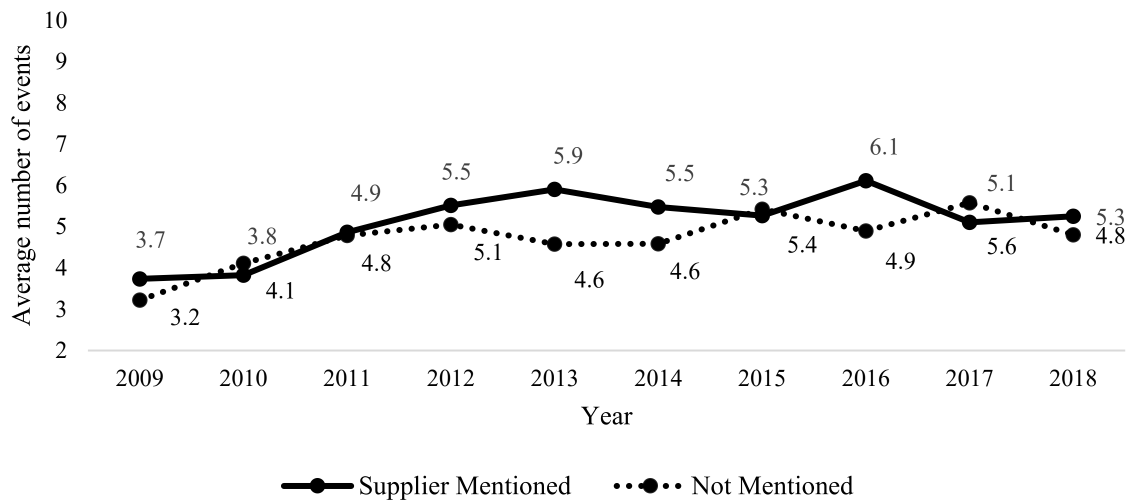

Figure 7. Average number of events, by whether the *Chronology* mentioned a supplier

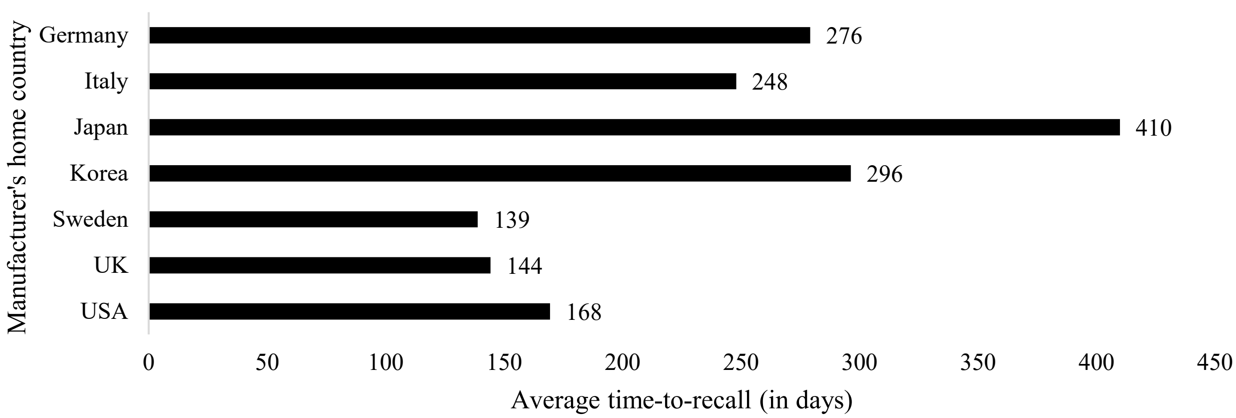

Figure 8. Average time-to-recall, by manufacturer's home country

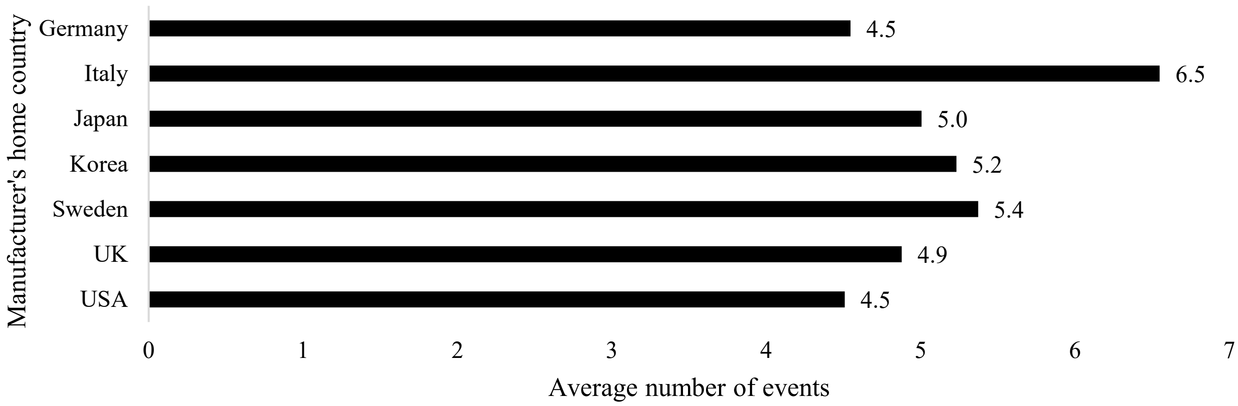

Figure 9. Average number of events, by manufacturer's home country

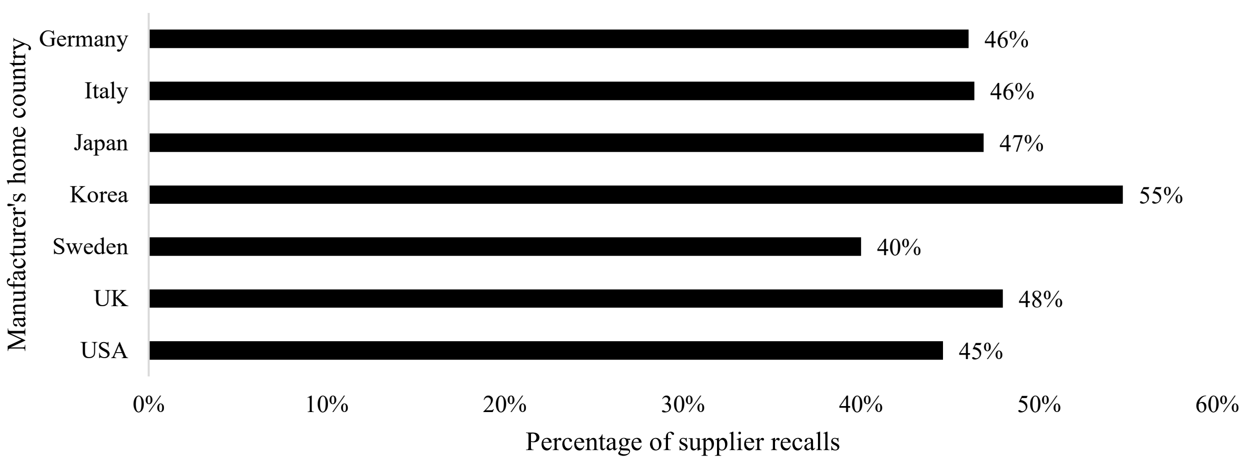

Figure 10. Percentage of supplier recalls, by manufacturer's home country

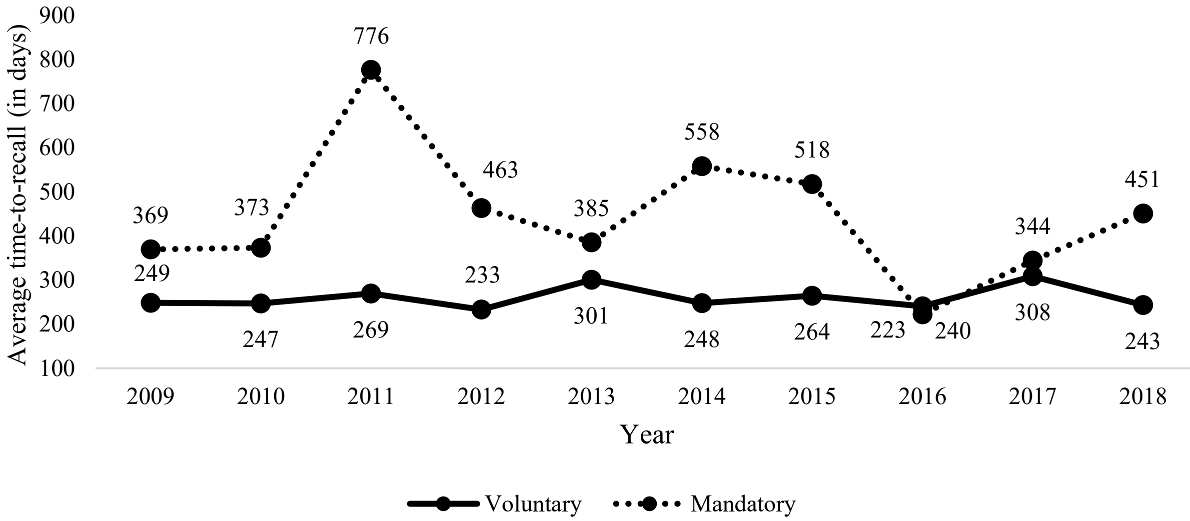

Figure 11. Average time-to-recall, by recall year and whether the recall is voluntary or mandatory

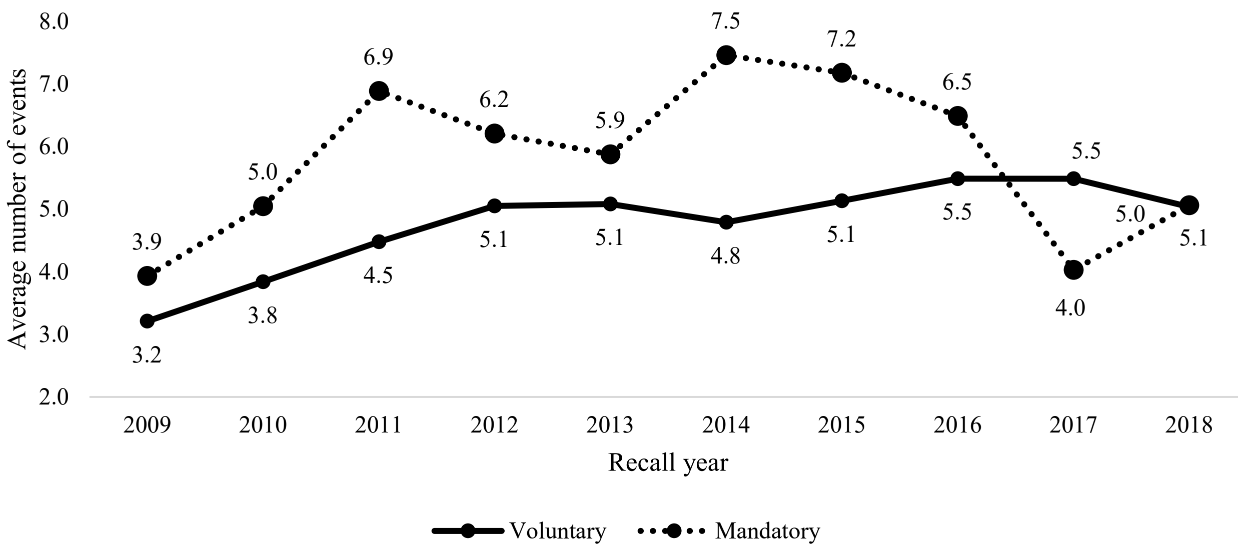

Figure 12: Average number of events, by recall year and whether the recall is voluntary or mandatory

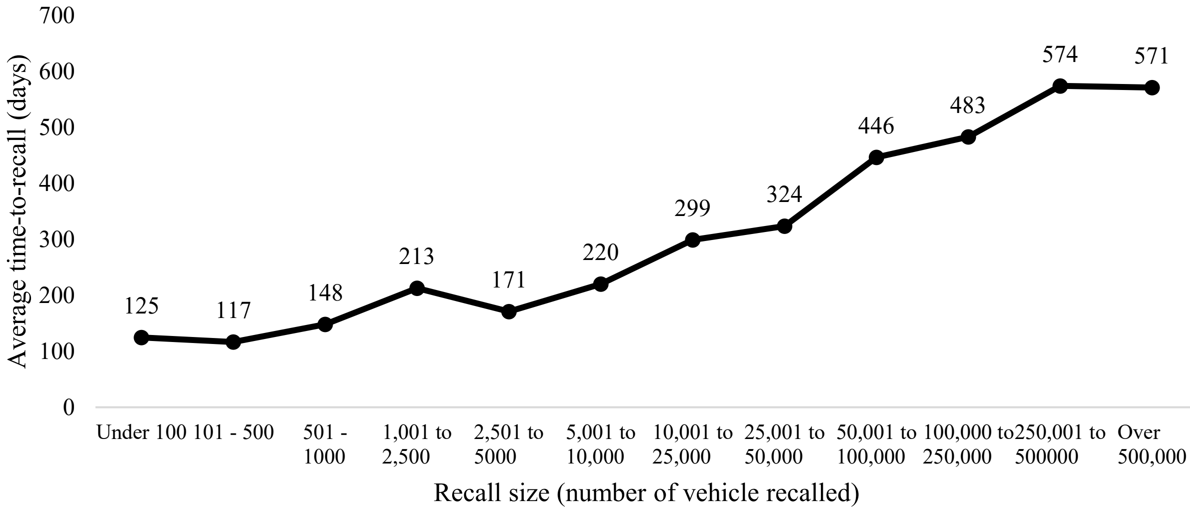

Figure 13: Average time-to-recall by recall size

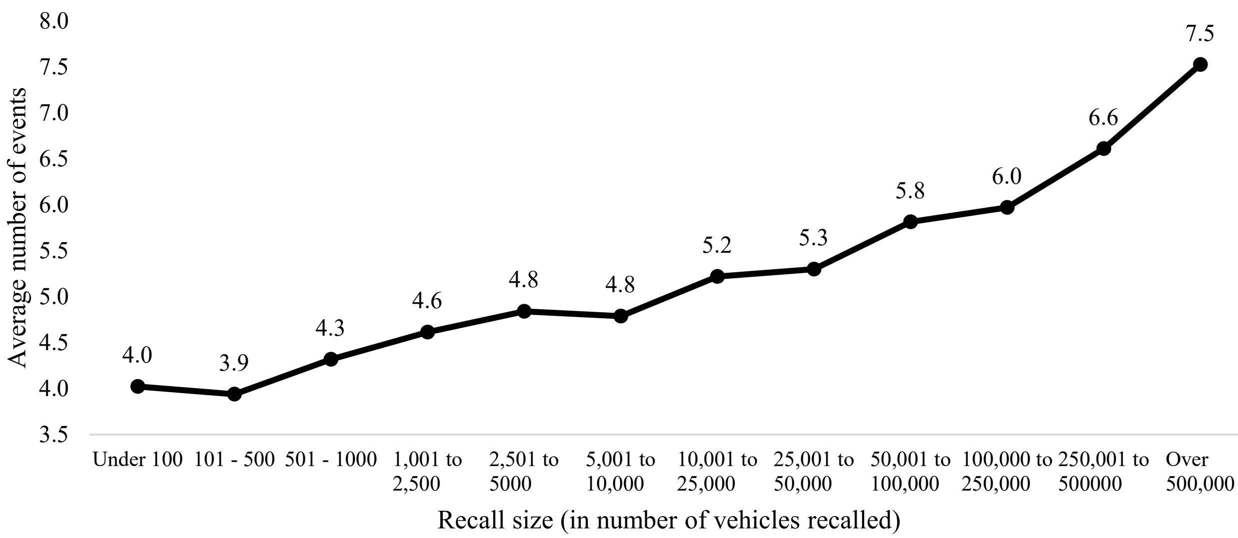

Figure 14: Average number of events by recall size

Online Appendix

July 10, 2015 NHTSA CAMPAIGN NUMBER: 15V442000

Steering Gear Not Tightened to Specification

Separation of the tie rod from the steering gear could result in the loss of steering, increasing the risk of a crash.

NHTSA Campaign Number: 15V442000

Manufacturer General Motors LLC

Components STEERING

Potential Number of Units Affected 10

Summary

General Motors LLC (GM) is recalling certain model year 2014-2015 Chevrolet Cruze vehicles manufactured November 8, 2013, to March 12, 2015, and 2015 Chevrolet Volt vehicles manufactured April 11, 2014. In the affected vehicles, the inner tie rod may not be tightened to specification allowing the tie rod to separate from the steering gear.

Remedy

GM will notify owners, and dealers will replace the steering gear assembly, free of charge. The recall began on July 28, 2015. Owners may contact Chevrolet customer service at 1-800-222-1020. GM's number for this recall is 15386.

Notes

Owners may also contact the National Highway Traffic Safety Administration Vehicle Safety Hotline at 1-888-327-4236 (TTY 1-800-424-9153), or go to www.safercar.gov.

3 Affected Products

13 Associated Documents

ISSUED Owner Notification Letter(Part 577)
RCONL-15V442-3707.pdf 157.448KB

Safety Bulletin - 15386 July 2015
RCSB-15V442-4478.pdf 103.275KB

Recall Acknowledgement
RCAK-15V442-5149.pdf 50.854KB

Recall Quarterly Report - 2015-Q4
RCLQRT-15V442-4823.PDF 28.937KB

Manufacturer Notices(to Dealers,etc)
RCMN-15V442-0430.pdf 2.661KB

Remedy Instructions and TSB
RCRIT-15V442-0095.pdf 30.615KB

Recall Quarterly Report #4, 2016-2
RCLQRT-15V442-5558.PDF 214.785KB

Recall Quarterly Report - 2016-Q1
RCLQRT-15V442-3315.PDF 214.701KB

Defect Notice 573 Report
RCLRPT-15V442-4893.PDF 29.18KB

Figure A1: Screenshot of web page for NHTSA Campaign Number 15V442: An Example

Steps for data collection and cleaning

Note: *Italicized* text indicates the name of a field in our data set.

1. Download and read the NHTSA's recall data file (FLAT_RCL.txt archived in https://www-odi.nhtsa.dot.gov/downloads/folders/Recalls/FLAT_RCL.zip) to select a specific set of NHTSA campaign numbers (field named CAMPNO, also known as recall identifier number; e.g., 21V001000) for which you want to collect data. This value appears in the *NHTSA Campaign Number* field in our data set.

2. For each recall, the value of MFR_NAME field in FLAT_RCL.txt provides the name of the manufacturer (see dictionary of FLAT_RCL.txt at https://www-odi.nhtsa.dot.gov/downloads/folders/Recalls/RCL.txt. A cleaned version of this name appears in the *Manufacturer Name* field in our data set.

3. Identify the manufacturer's home country name through internet sources such as the company website. This value appears as the *Manufacturer Country* field in our data set.

4. To obtain the recall report PDF file for each recall, visit https://www.nhtsa.gov/recalls?nhtsaId=<NHTSA campaign number>. For example, https://www.nhtsa.gov/recalls?nhtsaId=21V001000.

    a. Click the + icon on the web page. The web page will expand.

    b. Click the link named "Associated Documents." The web page will expand.

    c. Search on the web page for a file with name beginning with RCLRPT or file with description containing the term "573 Report." The name of this file is the value of the *File Name* field in our data set.

    d. Did the search yield a result (yes or no)?

        i. If no, the manufacturer did not provide chronology text. Exit the data collection process for this recall.
        ii. If yes, click on the search result to open the file in your web browser. For example, open https://static.nhtsa.gov/odi/rcl/2021/RCLRPT-21V001-1603.PDF. The file name (e.g., RCLRPT-21V001-1603.pdf) is the value of the *File Name* field in our data set.

5. In the recall report PDF file that you have opened, search "Chronology."
   Did the search yield a result (yes or no)?

    a. If no, then the manufacturer did not provide chronology text. Exit the data collection process for this recall and return to step 2 (that is, the next recall).
    b. If yes, did the manufacturer include the chronology text (yes or no)?
        i. If no, did the manufacturer refer to a "miscellaneous" file where it reported the chronology text (yes or no)?
            - If no, then the manufacturer did not provide chronology text. Exit the data collection process for this recall and return to step 2.

        - If yes, go back to step 4c and search for "Miscellaneous" document rather than "573 Report" and proceed with step 6.

    ii. If yes, proceed with step 6.

6. The text in this section is the value of the *Chronology* field in our data set. The first date mentioned in the chronology text is the value of *First Date* field. Set *Defect Awareness Date* equal to the *First Date*.

7. To determine the *Recall Date*, read the value of the field RCDATE in FLAT_RCL.txt.

8. Set *Time-to-Recall* equal to *Recall Date* minus *Defect Awareness Date*.

9. Obtain the year component of the *Recall Date*, and set *Recall Year* to this value.

10. To determine whether the recall is related to a supplier, check whether the chronology text mentions a supplier, by name or by role (i.e., the word supplier)? If yes, set *Supplier Mentioned* field to "Yes", otherwise to "No".

11. To determine whether the recall was voluntary or mandatory, check the value of the field INFLUENCED_BY in FLAT_RCL.txt. If the value is "ODI" or "OVSC", set *Voluntary Or Mandatory* field to "Mandatory". If the value is "MFR" set it to "Voluntary".

12. To determine the number of vehicles affected by the recall—that is, recall size—sum the value of the field POTAFF in FLAT_RCL.txt for all observations that have CAMPNO equal to the NHTSA campaign number (recall identifier) of the focal recall. (FLAT_RCL.txt has one observation for each year-make-model-component affected by the recall.) The sum is the value of the field *Recall Size* in our data set.

13. To determine the number of events, read the chronology text. Count the number of distinct events, which is the value of the *Events* field in our data set.

14. For each event, identify its date and a description (that is, the values in the *Event Date* and *Event Description* fields).

15. Validate the assumption that the first date in the chronology text is indeed the date when the manufacturer became aware of the defect (i..e, is *First Date* truly the *Defect Awareness Date*?). The alternative is that a vehicle owner might have submitted a complaint with the NHTSA about any of year-make-model-component combinations affected by the recall. The assumption is that a complaint to NHTSA becomes public on NHTSA's website very quickly, thus by default, a NHTSA complaint makes the manufacturer aware of the defect. This is because manufacturers monitor NHTSA's complaint website. This validation involves the following steps.

    a. Identify the year, make, model and component for each recall. These values are present in the NHTSA's recall data file (https://www-odi.nhtsa.dot.gov/downloads/folders/Recalls/FLAT_RCL.zip) fields YEARTXT, MAKETXT, MODELTXT and COMPNAME, respectively.

b. Identify all matching complaints for each recall by downloading and reading the NHTSA's complaints data file (FLAT_CMPL.txt archived in https://static.nhtsa.gov/odi/ffdd/cmpl/FLAT_CMPL.zip) and finding complaints that match as per the following:

    i. Value of YEARTXT in FLAT_CMPL.txt = Value of YEARTXT in FLAT_RCL.txt
    ii. Value of MAKETXT in FLAT_CMPL.txt = Value of MAKETXT in FLAT_RCL.txt
    iii. Value of MODELTXT in FLAT_CMPL.txt = Value of MODELTXT in FLAT_RCL.txt
    iv. Value of COMPDESC in FLAT_CMPL.txt = Value of COMPNAME in FLAT_RCL.txt

For each matched complaint

i. Is the value of LDATE in FLAT_CMPL.txt > the value of RCDATE in FLAT_RCL.txt? (yes or no)?

If no, continue with the next step.
If yes, ignore the observation. Such an observation relates to a complaint filed *after* the recall was initiated and is thus irrelevant to the validation.

ii. Is the value of LDATE in FLAT_CMPL.txt < the value of MFRDATE in FLAT_RCL.txt? (yes or no)?

If no, continue with the next step.
If yes, ignore the observation. Such an observation relates to a complaint filed *before* the manufacturer started manufacturing the recalled vehicle and is thus irrelevant to the validation.

iii. Is the value of LDATE in FLAT_CMPL.txt > the value of *First Date* that you found from the chronology text above (yes or no)?

If no, continue with the next step.
If yes, ignore the observation. Such an observation relates to a complaint filed *after* the manufacturer reported becoming aware of the defect and is thus irrelevant to the validation.

iv. Read "Description of the Defect" section in the 573 report PDF file and compare the text with the value of CDESCR field in FLAT_CMPL.txt for the complaint. Does the textual description of the recall defect report indicates the same problem as described by in the complaint (yes or no)?

If no, the value of *First Date* is the date when the manufacturer became aware of the defect.
If yes, use the value of LDATE for this observation as the *Defect Awareness Date* and recompute the *Time-to-Recall* as *Recall Date* minus *Defect Awareness Date*.