January 2013

Keeping Tabs on Financial Innovation: Product Identifiers in Consumer Financial Regulation

Patricia McCoy
Boston College Law School, patricia.mccoy@bc.edu

Daniel Carpenter

KEEPING TABS ON FINANCIAL INNOVATION: PRODUCT IDENTIFIERS IN CONSUMER FINANCIAL REGULATION

BY DANIEL CARPENTER* AND PATRICIA A. MCCOY+

For decades in the United States, policymakers and pundits touted financial innovation as an unquestioned good. Perhaps the most prominent proponent of this mindset in recent years was former Federal Reserve chairman Alan Greenspan, who championed Schumpeter's theory of creative destruction as making way for technological change.[1] Greenspan was so enamored of Schumpeterian economics that he used his time at the Fed to single-handedly pursue a conscious policy of deregulation in order to encourage financial innovations, ranging from subprime credit to mortgage-backed securities and credit default swaps.[2]

Of course, many financial innovations over the years have inured to the benefit of consumers and society. The checking account, which state banks pioneered in the late nineteenth century to evade the federal tax on state bank notes, is one, joined by the automated teller machine, the money market fund, and the debit card in more recent times.[3] As the 2008 financial crisis revealed, however, some financial

1. ALAN GREENSPAN, THE AGE OF TURBULENCE: ADVENTURES IN A NEW WORLD 250-52 (2007).

2. See KATHLEEN C. ENGEL & PATRICIA A. MCCOY, THE SUBPRIME VIRUS: RECKLESS CREDIT, REGULATORY FAILURE, AND NEXT STEPS 189-205 (2011). Greenspan was also enamored with the theories of Ayn Rand, id. at 189-90.

3. See, e.g., Henry N. Butler & Jonathan R. Macey, *The Myth of Competition in the*

innovations, if left to run rampant, can inflict serious negative externalities on the financial system and on society as a whole. Private-label mortgage-backed securities, collateralized debt obligations, and over-the-counter credit default swaps all came under scrutiny and new regulation in the wake of the financial crisis for exactly that type of harm.

The question for Congress and for policymakers, then, is how to harness the potential good from financial innovations while curbing their potential harm.[4] It is important for the government to operate with a deft touch to avoid stifling new financial products that are welfare-enhancing. At the same time, the government can no longer afford to hide its head in the sand: it *must* henceforth monitor financial innovations in view of the potentially grave harm to consumers and to the financial system as a whole that some of those innovations can pose.

Outside of the consumer finance arena, government regulators use a variety of regulatory techniques to reduce the potential harm from technological innovation. The Food and Drug Administration, for example, does not permit pharmaceuticals to be sold without exhaustive clinical trials and prior product approval.[5] The Consumer Product Safety Commission eschews prior product approval, but it does monitor consumer products for hazards and recalls products that violate mandatory safety standards or that are defective and pose a substantial risk of injury.[6] Meanwhile, state insurance regulators commonly

---

Dual Banking System, 73 CORNELL L. REV. 677, 681 (1988); MILTON FRIEDMAN & ANNA J. SCHWARTZ, A MONETARY HISTORY OF THE UNITED STATES, 1867-1960, at 18-19 (1963); BRAY HAMMOND, BANKS AND POLITICS IN AMERICA, FROM THE REVOLUTION TO THE CIVIL WAR 572-604, 728-34 (1957); Arthur E. Wilmarth, Jr., *The Expansion of State Bank Powers, the Federal Response, and the Case for Preserving the Dual Banking System,* 58 FORDHAM L. REV. 1133, 1153-54 (1990).

4. On regulation of financial innovations generally, *see, e.g.,* Zachary J. Gubler, *The Financial Innovation Process: Theory and Application,* 36 DEL. J. CORP. L. 55 (2011); Kathryn Judge, *Fragmentation Nodes: A Study in Financial Innovation, Complexity, and Systemic Risk,* 64 STAN. L. REV. 657 (2012); Robert C. Merton, *Financial Innovation and the Management and Regulation of Financial Institutions,* 19 J. BANKING & FIN. 461 (1995); Steven L. Schwarcz, *Regulating Complexity in Financial Markets,* 87 WASH. U.L. REV. 211 (2009); Charles K. Whitehead, *Reframing Financial Regulation,* 90 BOSTON U.L. REV. 1 (2010).

5. *See, e.g.,* DANIEL CARPENTER, REPUTATION AND POWER: ORGANIZATIONAL IMAGE AND PHARMACEUTICAL REGULATION AT THE FDA 21-22 (2010); Food & Drug Admin., *Development and Approval Process,* http://www.fda.gov/Drugs/ DevelopmentApprovalProcess/default.htm (last updated Oct. 15, 2012).

6. *See, e.g.,* CONSUMER PRODUCT SAFETY COMMISSION, 2011 ANNUAL REPORT TO THE PRESIDENT AND CONGRESS 1-21 (2012), *available at* http://www.cpsc.gov/

require insurance underwriters to file "forms" for new insurance products that they plan to market. For some types of insurance in some states, insurers must obtain prior regulatory approval before marketing new products. The more common trends today in insurance regulation, however, are file-and-use and use-and-file systems.[7]

Applying pre-approvals and recalls to financial innovations is not straightforward. For clinical trials to be effective, they would have to be conducted over the entire business cycle, resulting in protracted delays. This is one reason why insurance regulation has increasingly migrated away from pre-approval requirements. Similarly, in federal regulation of consumer finance, when products have been banned, it has been through statutory enactment or full notice-and-comment rulemaking (and even the latter is highly controversial), not through recalls.

Mindful of these problems and this history, in this Article we propose a far less intrusive, market-compatible method for monitoring financial innovations without blocking them: data collection and analysis using unique product identifiers. A central hypothesis of this essay is that the beneficial and harmful effects of financial innovations could be better understood if regulators and third parties could harness product identifiers to identify product-specific risks. Furthermore, product identifiers need to be available to the public as well as to the government in order to enlist the considerable resources and insights of independent, outside researchers to improve the monitoring of consumer and systemic financial risk.

Our proposal builds on a number of key reforms mandated by Congress in the Dodd-Frank Wall Street Reform and Consumer Protection Act[8] in 2010. In the systemic risk provisions of Title I of Dodd-Frank, for example, Congress created the Financial Stability Oversight Council (FSOC) and charged it with identifying and monitoring risks to the financial stability of the United States.[9] To discharge its monitoring functions, FSOC, working through the new Office of Financial Research, can require financial companies to submit

---

/PageFiles/123357/2011rpt.pdf.

    7.   *See, e.g.,* 2 NEW APPLEMAN ON INSURANCE LAW § 10.1 (2009).

    8.   Pub. L. No. 111-203, 124 Stat. 1376 (2010) (codified in scattered sections of the U.S. Code).

    9.   12 U.S.C. §§ 5322(a)(1)(A), (b) (2012).

periodic and other reports to assess the extent to which a financial activity poses a threat to the financial stability of the United States.[10] Similarly, in Dodd-Frank, Congress instructed the newly created Consumer Financial Protection Bureau to "monitor for risks to consumers in the offering or provision of consumer financial products or services, including developments in markets for such products or services."[11]

The monitoring that Congress contemplated in Dodd-Frank cannot be adequately performed without the ability to identify distinct financial products and isolate their effects. Fortunately, creating product identifiers increases monitoring abilities substantially at a relatively low cost. Once reasonable fixed costs have been incurred, databases with product identifiers are relatively cheap to maintain, and, if properly deployed and presented to citizens in a user-friendly manner, they can assist regulatory policy in at least three ways: (1) by improving regulatory information and policy so that actions focus on the products with the largest risk profiles, both in terms of risk to consumers and to the financial system as a whole, (2) by inducing financial services providers to monitor their products' risks more carefully, and (3) by improving consumer information so as to reduce information asymmetries between producers and consumers (and between regulators and consumers), thereby allowing consumers to choose products that better fit their needs and ability to bear risk. We examine these possibilities and discuss features of risk databases that permit prediction of product-specific risk profiles in the financial services sector, even in cases where product innovation may change the available set of consumer alternatives on a regular and expected basis.

The Article proceeds as follows. In Part I, we canvass the growth of publicly searchable federal databases on consumer product safety.[12] Part II examines the properties of these existing systems and their limitations.[13] Many of these limitations are well recognized among observers and participants in the regulatory regime but have yet to be remedied or addressed due to political, fiscal, or agenda constraints. In

---

10. 12 U.S.C. § 5322(d)(3)(A) (2012).

11. 12 U.S.C. § 5512(c)(1) (2012); *see also* 12 U.S.C. §§ 5511(c)(3), 5512(c)(4)(B)(ii) (2012).

12. *See infra* Part I.

13. *See infra* Part II.

Part III, we explain how product identifiers significantly improve the analytical power of monitoring in several respects.[14] These identifiers do so by facilitating linkage of multiple databases to permit analysis of the interaction of one type of financial product with another, by reducing the costs of that linkage, and by reducing error in the identification and classification of new financial products. Part IV proposes a new system of product identifiers in the regulation of consumer finance to improve the governance and monitoring of consumer financial innovations.[15] Finally, Part V provides a brief conclusion.[16]

I. THE GROWTH IN PUBLICLY SEARCHABLE FEDERAL DATABASES ON CONSUMER PRODUCT SAFETY

A variety of regulatory policies attempt to monitor risks associated with consumer products. These range from the consumer product safety apparatus in the United States to state and federal automobile safety regimes, to risk regulation for insurance and consumer finance and pre-approval for pharmaceutical products. The academic rationales for these policies are varied, though few of the policies enacted in regulation are or were enacted for the ex post rationalizations that scholars give them.[17] Many are enacted in the name of "consumer protection" and "safety," and arguments about information asymmetries between producer and consumer usually figure into the enactment and design of many such institutions.

A growing number of federal regulatory agencies govern markets in which product innovation constantly alters the set of objects and firms regulated and in which risks to consumers vary heavily by product in their incidence, harm and risk. In recent years, a number of these regimes have begun to host large information systems dedicated to reporting, tracking and aggregating consumer risk experiences. At some level, such systems have characterized markets in advanced industrialized nations for decades if not centuries. Better business bureaus and public health agencies have long kept complaint files on

---

14. *See infra* Part III.
15. *See infra* Part IV.
16. *See infra* Part V.
17. *See generally* CARPENTER, *supra* note 5, at 35-45, 73-117, 228-97.

different businesses.[18] Yet the rise of the Internet and electronic and statistical technologies has fundamentally transformed the possibilities of these databases. In particular, electronic databases may permit statistical prediction (and perhaps causal analysis) of product-specific risks with some precision, given the large sample sizes contained in these databases.

In a limited number of these areas, agencies have established databases, some of which are open to the public, that allow reporting of product-specific attributes and risks. The most prominent of these databases are ones spearheaded by the Consumer Product Safety Commission, the National Highway Traffic Safety Administration, the Consumer Financial Protection Bureau, and the Food and Drug Administration. Users' ability to analyze risk posed by specific products to consumers by discrete product attributes varies considerably, depending on the particular database.

A.   The Consumer Product Safety Commission

The most prominent effort in the United States to create a consumer product safety database has occurred at the Consumer Product Safety Commission (CPSC), which established the Consumer Product Safety Information Database in 2010.[19] The Commission created the database pursuant to a rulemaking authorized by Section 212 of the Consumer Product Safety Improvement Act of 2008, which directs the Commission to establish and maintain a database on the safety of consumer products, and other products and substances regulated by the Commission, that is publicly available, searchable, and accessible through the Commission's website.[20] In May 2010, the Commission released a proposed rule with the outlines of a database in mind. Particular issues, namely the permeability of the database to manipulation by one company at the expense of another, split the

---

18.   Laura Nader, *Disputing Without the Force of Law*, 88 YALE L.J. 998, 1003 (1979).

19.   CPSC, Publicly Available Consumer Product Safety Information Database: *Final Rule*, 75 Fed. Reg. 76832 (Dec. 9, 2010) (codified at 16 C.F.R. § 1102 (2013)). In a similar vein, the Department of Health and Human Services (HHS) created a Household Products Database that allows readers to search online for safety information on household products by brand name and manufacturer for health effects and ingredients. *Household Products Database*, HHS, http://householdproducts.nlm.nih.gov/index.htm (last updated Jan., 2013).

20.   Consumer Product Safety Improvement Act of 2008, Pub. L. No. 110-314, § 212(a), 122 Stat. 3048 (codified at 15 U.S.C. § 2055a(a)(1) (2012)).

2013] *PRODUCT IDENTIFIERS* 201

Commission on a three-to-two vote in November 2010, with majority Democrats casting the necessary votes to approve the final rule and minority Republicans voting against it.[21]

The raw components of the Consumer Product Safety Database have existed for decades in the form of consumer product incident reports. The Commission maintains a database of these reports.[22] Complaints come to the database from several sources, including the National Electronic Injury Surveillance System (NEISS), which aggregates data from hospital emergency rooms on product-related injuries, death certificates, and the Injury or Potential Injury Incident (IPII) database, which includes data from what are known as Medical Examiners and Coroners Program (MECAP) reports. Once aggregated, these reports were always publicly accessible in the minimal sense that citizens could file a Freedom of Information Act (FOIA) request to obtain them. Yet the costs of access for any one citizen are prohibitive, and the FOIA process has long been known to move slowly, especially as it is overburdened by lawyers using it for the purpose of discovery. What is more, even if querying the database were cheaper, aggregation of the reports would be prohibitively costly, which is one reason why many public interest research groups and third parties like Consumer Reports have been unable to do it systematically.[23]

The CPSC's database addresses these problems by providing publicly available and immediately searchable web-based data on reports of harm and recall notices involving consumer products and

---

21.   Andrew Martin, *Partisan Rift Mires Product Safety Database Plan*, N.Y. TIMES, Nov. 24, 2010, at B1; Andrew Martin, *In Split Vote, Safety Panel Approves Database for Consumer Complaints*, N.Y. Times, Nov. 25, 2010, at B4; *see* CPSC, Publicly Available Consumer Product Safety Information Database: *Final Rule*, 75 Fed. Reg. 76832 (Dec. 9, 2010) (codified at 16 C.F.R. § 1102 (2013)). The National Association of Manufacturers sided with the Republican Commissioners—Nancy Nord and former Commissioner Anne Northrup—on this issue, while consumer safety advocates in public interest research groups tended to side with the Democratic Commissioners, including chair Inez Tenenbaum.

22.   "For several decades, the Commission has gathered and maintained a database of consumer complaints known as consumer product incident reports involving a description of incidents related to the use of consumer products that fall within the scope of the Commission's jurisdiction." CPSC, Publicly Available Consumer Product Safety Information Database: *Proposed Rule*, 75 Fed Reg. 29156 (May 24, 2010) (codified at 16 C.F.R. § 1102 (2013)).

23.   Memorandum from Mary Kelsey James & Ming Zhu, CPSC staff, to the CPSC & Todd A. Stevenson (Nov. 17, 2010), http://www.cpsc.gov/PageFiles/90803/dbfollowup.pdf (regarding the "Publicly Available Consumer Product Safety Information Database Proposed Final Rule").

other products and substances regulated by the CPSC. The breadth and depth of this database far exceeds data currently available in any other country. The database allows members of the public to search a publicly available CPSC website for incident reports and recalls by a variety of product-specific variables, including consumer product type, brand and model, and manufacturer.[24]

Electronic incident reporting and electronic search are two of the central architectural components of the Commission's database. Now that the online database is completed, it is much easier for citizens to file complaints about products.[25] In addition, it is now possible for any user with an Internet connection to search the database for a particular product.[26]

B.   The National Highway Traffic Safety Administration

Part of the inspiration for the Consumer Product Safety Commission's database was the assemblage of vehicle accident and recall reports compiled and maintained by the National Highway Traffic Safety Administration (NHTSA). For several decades, the Administration has maintained data on accidents and recalls. These data are sufficiently fine-grained as to permit analysis of the effects of recalls.[27] NHTSA maintains these data in a large, publicly searchable database of vehicle accidents and recalls, with detailed information on all individual recalls since 1966, housed at the Department of Transportation's National Center for Statistics and Analysis (NCSA).[28] The data include a variety of product-specific variables, including "model year, beginning and ending dates of manufacturing, potential number of units affected, potential number of units defective, recall initiator, the number of units corrected, hazard category (only up to

---

24.   *See* SAFERPRODUCTS.GOV, http://saferproducts.gov/Default.aspx (last visited May 25, 2013); *see also* CPSC, REPORT TO CONGRESS PURSUANT TO SECTION 212 OF THE CONSUMER PRODUCT SAFETY IMPROVEMENT ACT OF 2008: IMPLEMENTATION OF A SEARCHABLE CONSUMER PRODUCT SAFETY INCIDENT DATABASE (SAFERPRODUCTS.GOV) (2009), *available at* http://www.cpsc.gov//PageFiles/127269/cpsia212.pdf.

25.   *See File a Report*, CPSC, https://www.saferproducts.gov/CPSRMSPublic/ Incidents/ReportIncident.aspx (last visited May 25, 2013).

26.   *See* SAFERPRODUCTS.GOV, *supra* note 24.

27.   Yong Kyun Bae & Hugo Benitez-Silva, *Do Vehicle Recalls Reduce the Number of Accidents? The Case of the U.S. Car Market*, 30 J. POL'Y ANALYSIS & MGMT. 821, 823, 827-28 (2011) [hereinafter *U.S. Car Market*].

28.   *See* SAFERCAR.GOV, http://www.safercar.gov/ (last visited Oct. 20, 2013).

2001), summaries of defects, possible consequences, and the correction required to eliminate the defects."[29]

The number of vehicle recalls is large and has jumped appreciably in the last two decades; "after a record of over 30 million cars recalled in 2004, in the last few years it has consistently reached between 15 and 17 million, and in 2010 alone 20 million cars were recalled."[30] NHTSA obtains data on accidents from Police Accident Reports (PARs) collected through the General Estimates System (GES), which dates from 1988. The GES is produced from a nationally representative sample, culled from approximately six million police-reported crashes annually.[31] For these accidents, the GES data include information on the severity of injuries for those involved, as well as information on vehicle models and vehicle-year models.[32]

In two interesting and innovative papers, Bae and Benitez-Silva adduce panel-data observational evidence that vehicle recalls reduce both the number and severity of accidents.[33] On severity of crashes, their estimates suggest that "the probability of resulting in an incapacitating injury goes down from 7.97% to 3.73% during the year after the recall" for recalled cars but not for non-recalled cars, and that for recalled models compared to non-recalled models, "the probability of resulting in fatal injury also goes down (0.0054 vs. 0.0016)."[34] The results are undoubtedly driven by a small number of events, but if correct, they provide a rationale for information safety databases. First, they suggest that the product identifiers in the current system of vehicle safety are working to produce benefits by providing a platform for the quick removal of defective cars from the roads. There is also evidence suggestive for the hypothesis that, for recalled cars that remain on the road, drivers may drive more carefully for those that remain on the road.

---

29. *U.S. Car Market, supra* note 27, at 828.

30.   Yong Kyun Bae & Hugo-Benitez Silva, *The Effects of Automobile Recalls on the Severity of Accidents*, 51 ECON. INQUIRY 1232, 1232 (2013) [hereinafter *Automobile Recalls*].

31. *See* *National Automotive Sampling System (NASS)*, NHTSA, http://www.nhtsa.gov/NASS (last visited May 25, 2013) [hereinafter *NASS*].

32.   For description of these data, we have relied heavily upon *U.S. Car Market, supra* note 27, at 833-34, and *Automobile Recalls, supra* note 29, at 1232. *See also Safety Recall Compendium*, NHTSA, DEFECTS AND RECALL INFO. ANALYSIS DIV. (2001), http://www-odi.nhtsa.dot.gov/recalls/documents/recompendium.pdf.; *NASS, supra* note 31.

33. *U.S. Car Market, supra* note 27; *Automobile Recalls, supra* note 30.

34. *Automobile Recalls, supra* note 30, at 1247.

It is clear that this effect would not be possible but for product identification—through recalls—and is understood, at least implicitly if not explicitly, by consumers who are drivers.

The results of Bae and Benitez-Silva similarly point to an important limitation of information systems in the realm of transportation policy and auto safety. Each car in the United States is identifiable by a Vehicle Identification Number (VIN). If recalls tracked the VINs of the cars that were returned during a recall, then policymakers would be able to monitor the fraction of recalled vehicles that have been returned and repaired. Furthermore, insurance companies would be able to incorporate the information into their pricing behavior, rewarding drivers who returned their recalled cars for inspection and remedy. Due to the opposition of vehicle manufacturers in the 1970s, however, a proposal for attaching VINs to recalls was rejected by the Administration. Bae and Benitez-Silva (2010, 2011) see the availability of VIN data for recalls as a welfare-improving and relatively cheap policy measure.[35]

C.   *The Consumer Financial Protection Bureau*

Since its inception in July 2011, the Consumer Financial Protection Bureau (CFPB) has developed two publicly searchable databases offering product-specific information.   The Credit Card Agreement Database, launched in 2012, allows the public to search the CFPB's website for credit card agreements, either by issuer or by specific language in the agreement.[36] The public can also download all of those agreements for further analysis. One drawback of the database is that credit card issuers contribute agreements to the dataset strictly on a voluntary basis. Accordingly, it does not encompass the universe of credit card agreements and is likely to be subject to reporting bias. Furthermore, credit card issuers are permitted by law in many respects to amend their agreements on an ongoing basis and in ways that are

---

35.   *See, e.g., id.* at 1234.

36.   *Credit Card Agreement Database*, CFPB, http://www.consumerfinance.gov/credit-cards/agreements/ (last visited May 23, 2013); Marla Blow, *Launching the College Credit Card Agreement Database*, CFPB BLOG (Nov. 1, 2012), http://www.consumerfinance.gov/blog/launching-the-college-credit-card-agreement-database/.

tailored to individual consumers.[37] This chameleon-like feature of credit card contracts means that the product features of a given customer's credit card are ever-changing. The credit card agreements in the database cannot and do not capture this dynamic element.

Subsequently, in 2013, the CFPB opened its electronically searchable Consumer Complaint Database to the public, containng data at inception on more than 90,000 complaints about credit cards, mortgages, student loans, bank accounts, services, and other consumer loans.[38] Among other things, the database allows users to download, sort, and visualize complaint data. However, the dataset has a number of limitations. Chief among those is that consumer finance products are only sorted by relatively crude product features such as fixed versus adjustable rates, without the additional, granular data that are needed to analyze the true product features at issue in any given complaint. Similarly, the data come solely from unverified consumer complaints. While those complaints are a potentially rich source of data, they are prone to error and bias in reporting.

D.   *The Department of the Treasury*

In Section 1483 of the Dodd-Frank Wall Street Reform and Consumer Protection Act,[39] Congress required the Department of the Treasury to make loan-level data on mortgage modification requests under the Making Home Affordable (MHA) program publicly available within sixty days after the deadline for reporting. The Treasury has done so by posting downloadable periodic datasets to its website. Each dataset contains loan-leval mortgage information for each trial and permanent modification applications evaluated under MHA's Home Affordable Modification Program (HAMP), including those that are

---

37. *See, e.g.,* Oren Bar-Gill & Elizabeth Warren, *Making Credit Safer,* 157 U. PENN. L. REV. 1, 10 (2008). While the Credit Card Accountability Responsibility and Disclosure Act of 2009 or CARD Act places some restrictions on such changes, many other types of changes are still permitted. *See, e.g., CARD Act Factsheet,* CFPB, *available at* http://www.consumerfinance.gov/credit-cards/credit-card-act/feb2011-factsheet/ (last visited May 26, 2013).

38. *Consumer Complaint Database,* CFPB, http://www.consumerfinance.gov/ complaintdatabase/ (last visited May 23, 2013); Scott Pluta, *Releasing Complaint Data about Credit Cards, Mortgages, Student Loans, Bank Accounts, Services, and Other Consumer Loans,* CFPB BLOG (March 28, 2013), http://www.consumerfinance.gov/ blog/category/credit-cards/.

39. 12 U.S.C. § 5219b (2012).

approved, denied, and cancelled. The data fields contain product-specific information, including the product type, term, interest rate, and monthly payment information, both pre- and post-modification.[40]

E.   The Food and Drug Administration's MedWatch System

Perhaps the most advanced system of post-market safety and consumer experience is the MedWatch system established by the U.S. Food and Drug Administration (FDA).[41] Like other agencies, the FDA has been collecting Adverse Event Reports (AERs) from manufacturers of medical products and physicians since the 1960s,[42] but the inauguration of an electronic and publicly accessible database in 1993 created a much richer, more accurate, and more comprehensive system of information. For each drug for each month, the MedWatch system aggregates all adverse event reports and includes information about the date of incident, patient gender and age, and the severity of the incident, among other variables.

AERs are not searchable publicly without a Freedom of Information Act (FOIA) request. Members of the public, however, can search MedWatch on the FDA's website for safety alerts by product. In addition, the FDA has begun an electronic notification system, based upon MedWatch, that transmits updates from the MedWatch system to physicians and patients, including by smartphone and by Twitter.[43]

AERs lack sufficient detail for many inferences, and there is evidence of substantial reporting bias. Yet regulatory officials and drug safety specialists maintain that, combined with epidemiologic studies, these reports can serve as data for hypothesis formation. The

---

40.   *See Making Home Affordable File Summary,* U.S. DEP'T OF THE TREAS. (Jan. 31, 2011),   available   at   http://www.treasury.gov/initiatives/financial-stability/reports/Documents/MHA%20Data%20File%20Summary_new.pdf ; *Making Home Affordable Appendix C, Data File Data Dictionary v3.0,* U.S. DEP'T OF THE TREAS., at C-8 (Dec.   3,   2012),   available   at   http://www.treasury.gov/initiatives/financial-stability/reports/Documents/MHA%20Data%20File%20Appendix%20C%20v3.0%20FINAL.pdf.

41.   *MedWatch: The FDA Safety Information and Adverse Event Reporting Program,* FDA, www.fda.gov/Safety/MedWatch/default.htm (last updated Sept. 10, 2013).

42.   *See* CARPENTER, *supra* note 5, at 591-92; ARTHUR A. DAEMMRICH, PHARMACOPOLITICS: DRUG REGULATION IN THE UNITED STATES AND GERMANY 116-50 (2004).

43.   *Subscribe to MedWatch Alerts,* FDA, http://www.fda.gov/Safety/MedWatch/ucm228488.htm (last updated Jan. 6, 2011).

hypotheses then can be fed into randomized trials or more tightly controlled epidemiological studies to produce more internally valid estimates of product risk.[44] For example, these data have been used in the analysis of pharmaceutical risk by Mary K. Olson,[45] among others.

What differentiates the MedWatch system from other product information systems is that in the United States, each drug is uniquely identified by its New Drug Application (NDA) number. This is so even for generic versions of a molecule that long ago lost its patent protection. Hence, more precise attributions of product risk can be made at the micro level (by physicians, pharmacists, drug companies and patients themselves), and macro-level inferences can be made about the relative observed risks of drugs. The unique identifier is a function of the FDA's gatekeeping power over the American pharmaceutical market (and by extension from this highly profitable market, much of the global pharmaceutical marketplace). In order to be sold in interstate commerce in the United States, each new drug must pass through the new drug application process, at the beginning of which it is assigned a unique NDA number that serves as an identifier in regulatory and pharmacoepidemiological work. There is a large literature on the effects of product recalls and warnings that is beyond summary here, but it is sufficient for now to note that all drug product withdrawals and labeling decisions correspond to NDA numbers.

There is reason to believe that efforts like these will continue, in part because policymakers in the United States have embarked upon a transparency and disclosure initiative affecting numerous levels of government,[46] and in part because other nations' regulators have also begun to call for greater disclosure of risks. In just one example of that trend, six federal agencies with vastly different jurisdictions joined forces in 2004 to create an online "one stop shop" permitting the public to search for U.S. recalls of foods, medicines, cosmetics, motor vehicles, car seats, environmental products, boats, and other consumer

---

44. *See* Timothy Brewer & Graham A. Colditz, *Postmarketing Surveillance and Adverse Drug Reactions: Current Perspectives and Future Needs*, 281 J. AM. MEDICAL ASS'N 824, 828 (1999); Robert Temple, *Meta-analysis and Epidemiologic Studies in Drug Development and Postmarketing Surveillance*, 281 J. AM. MEDICAL ASS'N 841, 841 (1999).

45. Mary K. Olson, *The Risk We Bear: The Effects of Review Speed and Industry User Fees on New Drug Safety*, 27 J. HEALTH ECON. 175 (2008).

46. *See, e.g.*, ABOUTDATA.GOV, http://www.data.gov/about (last visited May 26, 2013).

products government-wide.[47] Congress has also expressly mandated the creation of other publicly searchable databases in the consumer safety area.[48] The robustness of many of these systems is limited, however, by the lack of granular data allowing researchers to pinpoint the effect of a product's exact features and to quantify joint effects of the use of multiple products on consumer welfare.

II. CONSUMER PRODUCT RISK AND PRODUCT IDENTIFICATION

Despite the rise of these information systems, the kinds of predictive databases most commonly used by risk analysts are ill-equipped to address some of the most important sources of consumer risk. In particular, the ability to analyze the interaction between someone's use of one type of product with their use of another type is in its infancy. There are examples of multiple-product databases in retail finance (*e.g.*, the study of multiple credit card instruments in Gross and Souleles[49] and studies of multiple products using credit reporting agency data[50]), yet outside of medicine these examples are rare.

---

47. *See Your Online Resource for Recalls,* WWW.RECALLS.GOV, http://www.recalls.gov (last visited May 23, 2013); News Release, CPSC, Government Website Now Provides Instantaneous Information on Recalls (Nov. 16, 2004), http://www.cpsc.gov/en/Newsroom/News-Releases/2005/Government-Website-Now-Provides-Instantaneous-Information-on-Recalls-/. The six agencies are the CPSC (consumer products), FDA (food, medicines, and cosmetics), the U.S. Department of Agriculture (meat and poultry), NHTSA (motor vehicles and car seats), the Environmental Protection Agency (pesticides and vehicle emissions), and the U.S. Coast Guard (boats and boating safety).

48. *See, e.g.,* 12 U.S.C. § 2803 (2012) (requiring public disclosure of home mortgage data under the Home Mortgage Disclosure Act); 21 U.S.C. § 2013 (2012) (requiring creation of a searchable database on the effects of methamphetamine use and treatment options); 21 U.S.C. § 2103 (2012) (directing the FDA to create a searchable database of recalled human food and of recalled pet food).

49. David B. Gross & Nicholas S. Souleles, *Do Liquidity Constraints and Interest Rates Matter for Consumer Behavior? Evidence from Credit Card Data,* 117 QUARTERLY J. ECON. 149 (2002).

50. *See, e.g.,* Kenneth P. Brevoort & Cheryl R. Cooper, *Foreclosure's Wake: The Credit Experiences of Individuals following Foreclosure* (Fed. Reserve Bd. Fin. & Econ. Discussion Series, Working Paper No. 2010-59, 2010) (analyzing effect of foreclosure on subsequent delinquencies affecting other types of consumer loans); Ronel Elul, et al., *What "Triggers" Mortgage Default?,* 100 AM. ECON. REV. 490 (2010) (exploring joint propensity to default on mortgages and credit cards). In addition, the forthcoming National Mortgage Database will allow federal officials and possibly outside researchers to study, among other things, emerging trends in the joint use of different consumer finance products, including home mortgages, auto loans, and student loans. *See* Press Release, CFPB, Federal Housing Finance Agency and Consumer Financial Protection Bureau to partner on development of National Mortgage Database (Nov. 1, 2012), http://www.consumerfinance.gov/

Without the ability to track the use of different types of products by individual consumers, regulators, policymakers, and analysts are unable to understand or measure the following types of risks:

    •   *Superadditive risk specific to individual consumers:* Superadditive risk arises when the joint distribution of consumer risk differs materially from the aggregated product-specific distributions of risks.[51] Consumers may happen to use more than one product at a time and the joint distribution of risk from some bundles of products may not be predictable from the marginal distributions of the bundled products themselves. An obvious example comes in contraindicated medications such as the clinical recommendations against lithium use with angiotensin-converting enzyme (ACE) inhibitors, the proscription of monoamine oxidase inhibitors (MAOIs) with dextromethorphan, or the added risk of aspirin for pain management to patients taking warfarin or other blood thinners. The suffocation risk from a certain kind of blanket or toy may be raised when the infant is placed in a certain kind of crib that limits motion or creates higher probabilities of limbs getting caught in the bars or corners. Many of these product interactions are well known, but some of them are dependent on consumer type, and there are likely many such reactions that remain unknown. Superadditive risk is rarely if ever studied by economists in consumer finance.

    •   *Superadditive risk across consumers:* Superadditive risk can also arise where one consumer's use of a dangerous product is positively correlated with harm to other consumers. This can occur in the distribution phase, where the growing market dominance of a risky product crowds out the availability of a safer product. For instance, during the housing bubble that precipitated the financial crisis of 2008, the low initial monthly payments on relatively risky interest-only and option adjustable-rate mortgages cut into the market share of fixed-rate loans.[52] This can also occur in the use phase, where one person's use of

---

pressreleases/federal-housing-finance-agency-and-consumer-financial-protection-bureau-to-partner-on-development-of-national-mortgage-database/.

    51.   *See* James S. Koopman, *Interaction Between Discrete Causes*, 113 AM. J. EPIDEMIOLOGY 716 (1981); Tyler J. VanderWeele & James Robins, *Biologic Interactions and Their Identification* (COBRA Preprint Series, Paper No. 12, 2006), *available at* http://biostats.bepress.com/cobra.

    52.   *See, e.g.*, *Consumer Protections in Financial Services: Past Problems, Future Solutions: Hearing Before the S. Comm. on Banking, Housing, and Urban Affairs*, 111th Cong. 3-11 (2009) (statement of Patricia A. McCoy, George J. and Helen M. England Professor of Law, U. Conn. Sch. of L.), *available at* http://papers.ssrn.com/sol3/

a risky consumer product can have adverse spillover effects increasing the likelihood of harm to other consumers from use of the same or similar products. For example, when the sharp rise in delinquencies from subprime and Alt-A mortgages pushed down U.S. home prices significantly starting in the first quarter of 2007, other homeowners who had been current on their mortgages became delinquent due to job loss from the recession, combined with being underwater on their mortgages.[53] In the financial arena, this type of correlated risk can give rise to concern about the stability of the financial system.

• *Marginal propensity toward use of additional products:* Accumulation of experienced risk with one set of products may lead to utilization of others. A consumer who has accumulated debt on one credit card may attempt to transfer the debt to another set of debt instruments (including credit cards with different terms or a consumer loan). A patient on atypical antipsychotics or certain forms of antidepressant may gain so much weight as to induce metabolic syndrome, requiring the prescription of oral antidiabetics. There is in general little information about the joint distribution of risk with these products.

• *Joint distribution of consumer products:* Certain products may be marketed together in bundles, or the geographic availability of one product may be differentially associated with another (payday loans and certain kinds of mortgage instruments).

• *Bespoke products:* In certain areas, the features of consumer products are increasingly being tailored to the individual consumer, which can affect the associated risk. Advances in DNA analysis, for example, are leading to individualized drugs tailored to specific patients' genetic make-ups. While this movement is in its infancy, it is expected to gain steam, bringing with it the growing potential for unknown risks. This trend is considerably further along in the retail consumer finance sector, where advances in credit scoring models and risk-based pricing—and in other, sometimes illegal, ways to segment consumers—have spurred the individually tailored loan features,

---

papers.cfm?abstract_id=1367977.
    53.   *See, e.g.,* Patricia A. McCoy, *The Home Mortgage Foreclosure Crisis: Lessons Learned* 21 (Working Paper, 2013), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2254672.

affecting price terms and other product features as well.[54] This individualization can reduce or increase the risk profile of the specific instruments involved.

• *Dynamic nature of consumer product features:* In an extension of the bespoke trend, financial product providers may draft their contracts to allow for revisions to those contracts on an ongoing basis. This is particularly common in the area of open-end consumer credit, such as credit cards and home equity lines of credit (HELOCs), where the credit line may be open and used for decades, over a gamut of changing competitive conditions, business cycles and life circumstances of the user. For these products, a product's initial characteristics can and do morph over time and so do the associated consumer risks. Moreover, in this era of "big data," servicers often use their observations of consumer use and performance to tailor these modifications to individual consumers, with some becoming more likely to receive certain types of unilateral amendments than others.[55] A similar bespoke trend can be seen in the differential marketing of "add-on" products, such as credit insurance, or refinancing offers, according to a consumer's individual credit profile.[56]

Despite their increasing and potentially widespread use of risk and regulatory databases, it may be difficult to satisfy the conditions necessary for researcher to be able to study these important types of risk. A number of difficulties may arise in the utilization of consumer risk databases (CRDs). For example, consumer risk databases may inefficiently reveal risk information, due to the lack of product identifiers or other data. Similarly, CRDs may be subject to selection or sorting bias because certain consumer types may be more likely to utilize certain products. The resulting distribution of observed risk may result more from consumers than from the products themselves. Product identifiers can help identify—and rule out—suspected sources of that risk.

---

54. *See, e.g.,* Kathleen C. Engel & Patricia A. McCoy, *A Tale of Three Markets: The Law and Economics of Predatory Lending*, 80 TEX. L. REV. 1255, 1278-79 (2002).

55. *See, e.g.,* Bar-Gill & Warren, *supra* note 37, at 10. *Cf.* Daniel M. Schwarcz, *Reevaluating Standardized Insurance Policies*, 78 U. CHI. L. REV. 1263, 1277-1317 (2011) (analyzing differences in personal lines insurance policies).

56. *See* Tom Baker & Peter Siegelman, *Protecting Consumers from Add-On Insurance Products: New Lessons for Insurance Regulation from Behavioral Economics* (U. PA. Inst. for Law & Econ., Research Paper No. 13-1, 2013), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2199569.

Product identifiers can also aid in the measurement of the other types of risk just discussed. They can allow the estimation of joint marketing strategies and superadditive risk—both for individual consumers and for the financial system as a whole—by permitting the linkage of databases to identify the concurrent use of discrete financial products. With the use of panel data, these same linkages can permit researchers to measure the marginal propensity toward the use of additional products by detecting the sequential use of different financial products. Product identifiers can also help analyze the trend toward individually customized financial products with changing features and the consumer welfare implications of that trend.

III. CONSUMER PRODUCT DATABASES AS NETWORKED INFORMATION SYSTEMS

The databases described earlier have a number of components in common. They aggregate information from reports by consumers, providers, or third parties. They aggregate these reports by appropriately chosen categories and then make the results of aggregation (in the sense of lists or statistical summaries) available to end users and in some cases to the public. Regulators sometimes also premise warning systems and even product recalls and withdrawals on these databases; the information extends beyond analytic utility and quite clearly informs decisions that impose costs upon firms and consumers.

For purposes of representing relationships among databases, and among consumers and various products (including bundles of products owned or used simultaneously), this Section describe a model of consumer risk databases.[57] We start by examining the most common databases in financial regulation, which are single-product databases, and noting their limitations. Next, we discuss how to overcome those limitations through linking databases and the associated problem of linkage's high potential cost. Finally, we explain how the introduction

---

57. This model extends and, to some extent, generalizes the work of those in informatics and computer science who have considered these systems in other policy contexts. *See, e.g.,* Mark D. Flood, *Embracing Change: Financial Informatics and Risk Analytics,* 9 QUANTITATIVE FIN. 243 (2009); Ting Yu, et al., *PRUNES: An Efficient and Complete Strategy for Automated Trust Negotiation over the Internet, in* PROCEEDINGS OF THE 7TH ACM CONFERENCE ON COMPUTER AND COMMUNICATIONS SECURITY (2000).

of product identifiers can reduce that cost substantially while also minimizing errors in the identification and classification of different financial products.

## A.   Single-Product Databases and Their Limitations

Ideally, a consumer risk dataset would observe, for any given point in time, the risk outcome for every individual consumer for every product that he or she used. We can refer to this ideal dataset as the *aggregate risk profile*. Normally, however, reality falls short of this ideal. For any regulated sector, few if any data sources encompass an exhaustive representation of the aggregate risk profile. Instead, there are various data sources that most often are limited to a single product. Some of these data sources may be publicly available, either for free or for a price; others are not available to public users apart from the government; and still others are out of reach of both the public and the government, absent a subpoena or similar compelled process. Examples of the latter might include product health provider data on patients within a given plan who are prescribed a certain drug or bank-based data on credit card accounts. Analysts in investment or regulatory applications will often combine these databases, either examining a larger consumer population than they could have with just one of the databases, or linking consumer characteristics, market characteristics, or product characteristics from multiple data sources.

Take, for instance, a dataset such as a consumer complaint database that represents the risk experience of a given subpopulation, say the population of consumers some of whom who have defaulted on some kind of loan or mortgage, or who have some kind of credit card and may have filed for bankruptcy, or the population of individuals who have been involved in car accidents. A key feature of this dataset is that it possesses considerable information about individual citizens or consumers, but no information (unless implicit and extractable through learning) about the products being used. The only situation in which this dataset is sufficient to comprise the aggregate risk profile is when individuals are using one product and one product only. Of, course comparative risk analysis across products is impossible in this case.

B.   Linking Datasets and Their Costs

Given the number of data sources to be linked and the number of analysts (including third party groups, consumers themselves, and possibly regulators or investors) wishing to link them, establishing a system to link them requires a myriad of linkages to connect each analyst to each data source.

Figure 1. The Multiplicity of Linkages Among Analysts and Data Sources.[58]

Data Sources   Analysts
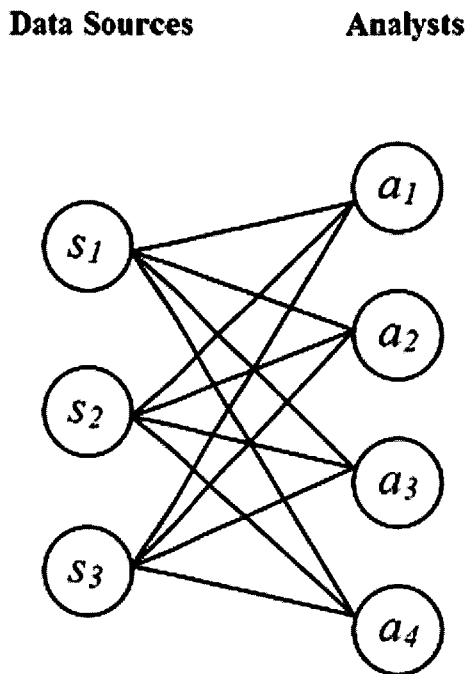
This system entails considerable expense, which can be broken down into two major costs. One cost involves the expense of designing the system to link each analyst to each data source. The other cost entails the expense of actually accomplishing each of those multiple linkages. The important point here is that the total costs do not increase additively (*i.e.*, by the number of data sources plus the number of

---

58.   Flood, *supra* note 57, at 246.

analysts). Rather, the total cost increases multiplicatively (*i.e.*, by the number of data sources *times* the number of analysts).[59] For example, Figure 1 assumes that four analysts are each trying to link three data sources. If c denotes a unit of cost, the total cost of those linkages will be $4c \times 3c = 12c$, not $4c + 3c = 7c$. Consequently, with each additional analyst and/or data source that is linked, the total cost will rise geometrically.

Figure 2. Example of the Effect of the Number of Data Sources and Analysts on Cost of Linkage[60]
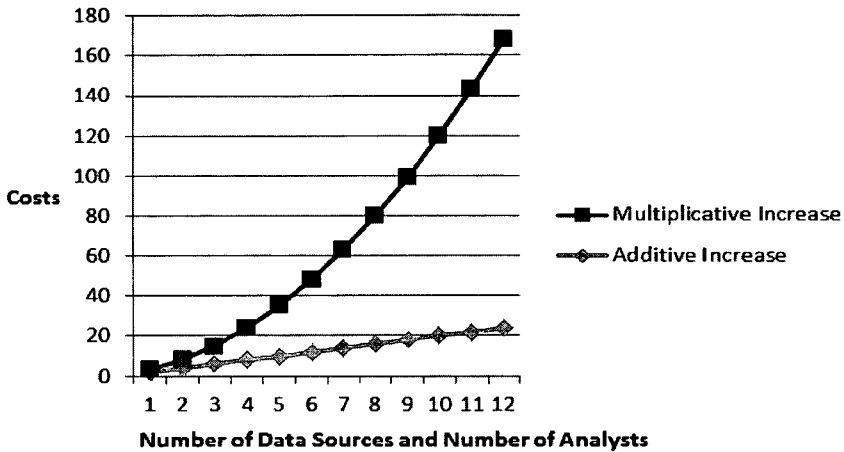
C.   *Facilitating Linkage Through a Product Identifier*

In consumer finance, as well as other areas of product safety, an important aim of risk analysis is to predict the joint distribution of risk, conditioned upon the consumer's utilization of more than one product at a time. As discussed above, databases that allow for this type of risk estimation are rare. Furthermore, linking multiple databases to permit

---

59. *See* Flood, *supra* note 57. As a matter of strategic institutional design, the question of how many analysts there are and what contracts govern them is examined in Sean Gailmard & John W. Patty, *Stovepiping*, 25 J. THEORETICAL POL. 388 (2013) and John W. Patty, *The Politics of Biased Information*, 71 J. POL. 385 (2009).

60. For purposes of illustration, this example assumes that the number of data sources and the number of analysts are the same.

estimation of joint distribution of risk is extremely costly and error-prone. The task, then, is how to harness the analytical power of linked databases while reducing the costs of linkage and errors in classification. Product identifiers that attach a unique code to every different product can accomplish both of these objectives.

1. Reducing the Costs of Linkage

For purposes of risk analysis, product identifiers offer two related advantages. The first is that a product identifier can help reduce costs by providing a common intermediate link—otherwise known as a *numeraire typology*[61]—for linking various data sources and analytics (think of multiple banking regulators linking static loan origination data, credit bureau data, and neighborhood demographic data for use in multivariate regressions).

Figure 3. The Effect of Introduction of a Unique Product Identifier in Streamlining Linkages.[62]

Unique Product Identifier

---
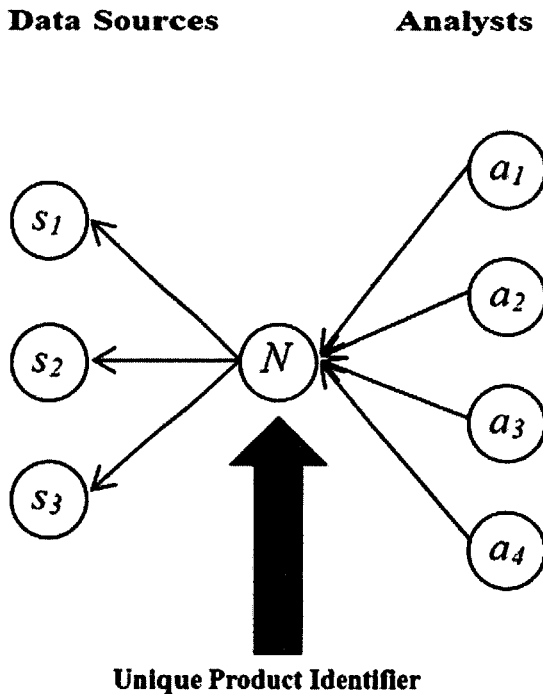
61. *See, e.g.,* Flood, *supra* note 57, at 246.
62. Flood, *supra* note 57, at 246.

Doing this reduces the rise in costs from adding analysts or datasets from a geometric increase to a much more affordable additive increase (see Figure 2).[63] Comparing Figure 1 to Figure 3, for instance, by introducing a product identifier, the number of linkages needed drops from 12 to 7. Thus, adding an identifier is a powerful way to reduce the cost of linking databases in most instances.

As an example of the power of product identifiers in this regard, let us imagine a pharmacoepidemiologist trying to examine whether one mental health drug is associated with greater liver toxicity than another, or is associated with more seizures than another. And let us further imagine a world without product identifiers (no NDA numbers), but with manufacturer-specific databases on the experience of users with their products. In order to assemble a comparative database, independent researchers would, in short order, experience costs that increase exponentially with the number of products examined. If a set of ten third-party psychiatrists (for example, different offices within the FDA, or various universities, or international health agencies) wanted to move from examining the post-market experience of five different psychotropic drugs to fifteen, the linkage costs to society would increase one-hundred fold. With a product identifier, however, the costs would increase only tenfold. If data assembly required repeated querying of data sources, as is common, these cost differentials would only magnify. At some basic level, there would be vast duplication of effort unless the various analysts could coordinate upon a product identifier, which would require the solution of a collective action problem and the institutional problem of assigning the identifier.

2. Error Reduction

A skeptical reader might object at this point that in the twenty-first century, linkages like these can be made approximate even without a product identifier. But a product identifier has a second powerful

---

63.   For a full exposition of the underlying mathematical model, *see* Daniel Carpenter, *Product Identifiers in Consumer Regulation: Notes Toward a Theory, with Applications to Retail Financial Epidemiology* (Working Paper, Nov. 15, 2011).

advantage over linkages performed without an identifier, which is substantial reduction in the propensity toward errors.

One could describe the set of products taken by their common trade name or chemical name and, even without an NDA identifier, use textual strings to identify the various drugs. Experience with the MEDWATCH system suggests that, even in the relatively well-honed case of FDA adverse event reports, many assumptions must be made to identify drugs whose NDA number has not been entered into the system, leaving millions of AERs unutilized. The same problems would presumably affect any epidemiological database premised upon a health plan, where physicians filing reports on adverse consequences of pharmaceutical utilization may be subject to the same error.

The possibility that a set of consumers (or consumer experiences) is lumped into the wrong bin is a special case of classification error. But it can also constitute partition error. In partition error, analysts attribute product risks from one class of products to a subclass that is appreciably different (inclusion error), and/or fail to include a subclass into a larger class in which the subclass properly belongs (exclusion error).

Imagine a set of mortgages or credit cards that differ in their mortgage terms (fixed-rate versus adjustable-rate), their pre-payment penalties, the declension of their rates (whether they have up-front "bonus" rates, "teaser interest rates," or conditional prizes attached), their amortization schedules, and the terms and transparency of their disclosure. A set of credit cards sold by the same bank would qualify as separable products once these characteristics are taken into account and so would mortgages, consumer loans, and other debt and credit instruments.

At present there is no universal system assigning a unique product identifier that can be used to distinguish retail financial products from one another. This is so even when one focuses upon a market for particular kinds of consumer financial products, say excluding the mortgage market from analysis and focusing purely on the market for credit cards. Analysts such as consumer financial regulators, university scholars, and consumer advocates may observe a sample of loan defaults, personal bankruptcies, or other risk-associated events. Yet these databases do not generally contain data that allow analysts to differentiate among products. Even in cases where

proprietary datasets are used, or where analysts have regional databases of consumer loans, the data often exclude important product characteristics differentiating one contract from another, and almost always exclude data on the other kinds of loans and credit/debt instruments held by consumers.

This opens the possibility that in trying to make even observational inferences about the interaction of products with certain consumers, analysts are committing implicit or explicit partition error. Analysts might make inferences about the properties of a class of subprime mortgages, for instance, when in fact the risk is heterogeneous depending on whether the subprime mortgage comes with a teaser interest rate or not. In order to properly classify consumer financial products, and in order to properly reclassify them to reflect new information and new analysis, analysts must have a product identifier that permits a clear assignment of the product into a category. A product identifier can reduce partition error by enhancing the process of classifying products according to certain critical characteristics, characteristics that may be associated with (and even cause) consumer risk. Without a unique product identifier, the costs of this classification (and the costs and even possibility of reclassification) can become prohibitive, as an exhaustive search through text, or perhaps a blind reliance upon the text, would be required. Hence partition error will lead to classification error, which will generate attributional error (invalid or inefficient inferences).

The lack of a product identifier can make this problem worse in ways that again quickly rise in proportion to the complexity of the problem being addressed.[64] This will happen if misclassification increases more quickly than successful classification. If we assume that the risk of misclassification weakly increases along with the possible number of product characteristics available, then this suggests genuine trouble for a system without a reliable product identifier.[65] When the

---

64. This follows Marina Meilă, *Comparing Clusterings—An Information-Based Distance*, 98 J. MULTIVARIATE ANALYSIS 873, 875 (2007), originally, and Justin Grimmer & Gary King, *Quantitative Discovery from Qualitative Information: A General-Purpose Document Clustering Methodology* (July 13, 2009) (unpublished manuscript) (on file with the Dept. of Gov't and Inst. for Quantitative Soc. Sci., Harvard Univ.).

65. By "weakly increasing" we mean that the risk never decreases, and increases strictly at least once, while the number of product characteristics rises. For a full exposition of the underlying mathematical model, *see* Carpenter, *supra* note 63.

number of product characteristics that should be taken into account grows, and when the analyst is limited to crude categorizations among these products, then the number of possible errors grows even further, and does so exponentially. A carefully designed product identifier reduces this chance of error.

IV. POSSIBILITIES FOR CONSUMER FINANCE: UNIQUE PRODUCT IDENTIFIERS FOR RETAIL FINANCIAL PRODUCTS

Product identifiers might be applied to a newly regulated world that has never before seen them: consumer financial products, especially mortgages, credit cards, consumer loans such as auto loans, and other retail financial products. Since 2011, these products have come under the jurisdiction of the new Consumer Financial Protection Bureau (CFPB). While the CFPB is still assembling its data sources, what is quite clear is that for the time being, there is no product identifier either across types of financial products or within markets (e.g., within credit cards) to assist the agency.

Yet the prospect of a unique product identifier for consumer financial products (which are essentially variable forms of contract) should not be dismissed. While the Bureau does not have pre-market approval power over new financial products, a financial product identifier can still be synthesized using existing statutory authorities by analyzing consumer credit contracts. Based on that analysis, the Bureau could perform the following tasks:

- identify each product and assign it a unique alphanumeric identifier;
- compile and analyze previous consumer experience and safety data relevant to the safety and use of the loan product, including (a) past experience with similar products or the proposed product; and (b) reference to relevant studies on how average consumers as well as subpopulations may be predicted to use the product;
- link the identifier to post-origination data on loan performance, workouts, and termination,

and analyze the performance of all products,
including new products, going forward. (The
Bureau should collect these data if it does not
have them already); and

- for a subset of new products that are not
  generally recognized as low risk to consumers
  and/or the financial system, conduct "roll-out"
  experiments in the laboratory or field to assess
  the safety of the product.

A.   *Creation of a System to Generate Product Identifiers*

To create product identifiers, a minimally invasive system,
based on the Bureau's existing statutory authority, to collect form
contracts for consumer financial products under the CFPB's jurisdiction
would be required institutionally. These contracts would be uploaded
into a database to allow the Bureau to construct product identifiers
based on the contract terms.

This system would involve four different processes. First, the
Bureau would require all consumer financial product providers under its
jurisdiction to submit the underlying form contracts for each of the
providers' different products within a prescribed timeframe. As new
products are developed, providers would submit the form contracts for
those products to the Bureau on a quarterly basis. Providers would also
submit amendments to existing consumer finance contracts (e.g., in the
context of credit cards) quarterly.

Each of these submissions would contain a proposed label, the
product contract,[66] and a full description of the product. The product
description should include a schedule of the full range of parameters for
numeric terms such as principal, interest rates and fees, the product term
and amortization schedule, interest rate indices, margins and caps,
prepayment penalties, any equity-sharing arrangement, and other terms
that the Bureau deemed relevant. The Bureau could mandate this
reporting under its statutory monitoring powers, which give the CFPB

---

66.   The contractual terms for some consumer financial products involve two or more
contractual documents. Residential mortgage loans, for instance, always involve both a note
and a mortgage (or deed of trust). The submission would encompass all of the documents
comprising the contract between the provider and the customer.

"authority to gather information from time to time regarding the organization, business conduct, markets, and activities of covered persons and service providers."[67]   Alternatively, FSOC and the Office of Financial Research could require such reporting using their statutory power to mandate financial companies to submit other reports so FSOC can evaluate threats that financial activities pose to the financial stability of the United States.[68]

Second, for the various consumer financial products embodied by the contracts, a team of analysts would render a categorization consistent with finding the most refined mesh available for a set of products that would leave sufficient sample sizes for statistical cross-product comparisons.

Third, the Bureau would report each product identifier back to the provider.[69]   Each provider would be required to assign the correct identifier prospectively to all of its individual loan files and report that identifier in all loan-level data reported to the Bureau. In addition, the Bureau would work with developers of data standards and with the Federal Housing Finance Agency, the government-sponsored entities Fannie Mae and Freddie Mac, the Federal Housing Administration (FHA), and proprietary data providers to adopt a product identifier field for both static and dynamic consumer financial product datasets.

Finally, the Bureau would provide on its website a publicly available and searchable database with the form contracts, the associated product identifiers, the providers' names, and the components of each product identifier for the purpose of public research.   This database would build on the Bureau's existing credit card agreement database, except that submissions would be mandatory and the database would feature the addition of unique product identifiers.

---

67.   12 U.S.C § 2807 (2012).   The Bureau also has power to mandate reporting of information and data under its supervisory powers, under the Home Mortgage Disclosure Act, and under a mortgage default and foreclosure database (to be created jointly with the Department of Housing and Urban Development). *See* 12 U.S.C. §§ 2801 et seq. (2012).

68.   12 U.S.C. § 5322(d)(3)(A) (2012).

69.   This function could be automated, for example, with email notifications directing providers to a Bureau website containing each submission and the associated product identifier.

B.   *Construction of the Identifier*

Given a contract collection system like the one described, a unique identifier could be established using the following procedure. First, have a group of experts (perhaps in consultation with the CFPB's Consumer Advisory Board[70] and Academic Research Council,[71] joined by consumers themselves) create a set of product bins that establish significant differences across product types (APR and simple interest paid, payment and amortization schedules, closing costs and other fees where applicable, and penalties and triggers) while permitting sufficient aggregation of products within each bin to permit statistical analysis.[72]

Next, for a given product field (credit cards and mortgages), the agency would represent each consumer financial product by two data sets: (1) the text of the contract and (2) symbolic data representing individual contract terms. Call this dual data set the *document-symbol pair* (DSP).   Following this, the agency would classify individual document-symbol pairs into the bins established by experts,[73] thus establishing an alphanumeric product type identifier. Each contract, along with all summary data provided with the contract, would be made publicly available in searchable format on the Bureau's website. Once a product identifier was established with different databases, the Bureau would be in a position to perform several analyses that, in   our estimation, no financial regulator can now perform: (1) whether a particular financial product carries greater observed risk to consumers than other, similar products in the same market, and (2) whether a particular combination of financial products places consumers in greater (possibly superadditive) risk than do others.

C.   *Linkage of Identifier to Post-Loan Performance Data*

Once each loan has an identifier, then the identifier can be used in post-loan performance and risk data sets collected by public agencies

---

70.   12 U.S.C. § 5493 (2012).

71.   *See Advisory Groups*, CFPB, http://www.consumerfinance.gov/advisory-groups/ (last visited May 29, 2013).

72.   This would probably be time-consuming, but could be performed through "mechanical Turking" so as to exploit the wisdom and labor of many decentralized agents.

73.   This could be accomplished using a Sammon multidimensional scaling algorithm, following Justin Grimmer & Gary King, *General Purpose Computer-Assisted Clustering and Conceptualization*, 108 PROC. NAT'L ACAD. SCIENCES 2643 (2011).

(such as Home Mortgage Disclosure Act data) and private firms (such as DataQuick and CoreLogic data). Indeed, once an identifier is created, private firms will have strong incentives to better organize their internal data and their risk analytics by using the identifier. This raises, of course, the question of why such an identifier does not already exist. The answer lies in well-known problems of collective action and coordination. If an identifier were created, all firms would benefit but they would benefit whether or not they paid the up-front costs of cooperating with others to create a common standard for the identifier. Hence each firm has an incentive to free-ride off of the efforts of others. The coordination problem arises because unique identifiers usually need a centralized infrastructure—a central repository where the identifiers can be kept and where duplicates are avoided—and while this central agent need not be a government agency, it should be a government agency if the creation and enforcement of the identifier requires legal authority and the ability to monitor whether loan originators are complying with the standards of the identifier.

D.   *Roll-Out Experiments with New Financial Products*

While the architecture for experiments with new products is not yet constructed, one could imagine a process whereby certain firms are incented to randomly assign an inducement or advertisement to one kind of loan versus another. The "treatment" for such an experiment would be the differential inducement provided to consumers to take one kind of loan form (a new financial product) versus another similar loan, and then the performance of these loans along with various risk indicators (financial stress, delinquency, underwater status, foreclosure, bankruptcy) could be tracked. Various studies have occurred in the retail financial field, and various ideas about so-called field experiments with new financial products have been proposed, yet data monitoring for such studies would probably be greatly facilitated by the creation of an identifier which would permit comparison of the products used by research subjects.[74]

---

74.   *See, e.g.,* Philomena M. Bacon & Peter G. Moffatt, *Mortgage Choice as a Natural Field Experiment on Choice Under Risk,* 44 J. MONEY, CREDIT & BANKING 1401 (2012); Stephanie Moulton, et al., *Field Experiments on the Impact of Financial Planning Interventions for Recent Homebuyers* (Ctr. for Fin. Sec., Univ. of Wis.-Madison, Working Paper 2011-CFS.5, 2011).

V. CONCLUSION

The benefits of a unique product identifier for consumer financial products would, we believe, far outweigh the costs, which would largely be experienced upfront and would wane over time as adaptation to the identifier would be experienced across numerous markets. The time is ripe for action by the federal government, including by the CFPB, that will create such an identifier and erect an infrastructure in which it is used and maintained.