World Privacy Forum

The Scoring of America: How Secret Consumer Scores Threaten Your Privacy and Your Future

By Pam Dixon and Robert Gellman
April 2, 2014

Brief Summary of Report

This report highlights the unexpected problems that arise from new types of predictive consumer scoring, which this report terms *consumer scoring*. Largely unregulated either by the Fair Credit Reporting Act or the Equal Credit Opportunity Act, new consumer scores use thousands of pieces of information about consumers' pasts to predict how they will behave in the future. Issues of secrecy, fairness of underlying factors, use of consumer information such as race and ethnicity in predictive scores, accuracy, and the uptake in both use and ubiquity of these scores are key areas of focus.

The report includes a roster of the types of consumer data used in predictive consumer scores today, as well as a roster of the consumer scores such as health risk scores, consumer prominence scores, identity and fraud scores, summarized credit statistics, among others. The report reviews the history of the credit score – which was secret for decades until legislation mandated consumer access -- and urges close examination of new consumer scores for fairness and transparency in their factors, methods, and accessibility to consumers.

About the Authors

Pam Dixon is the founder and Executive Director of the World Privacy Forum. She is the author of eight books, hundreds of articles, and numerous privacy studies, including her landmark Medical Identity Theft study and One Way Mirror Society study. She has testified before Congress on consumer privacy issues as well as before federal agencies. Robert Gellman Robert Gellman is a privacy and information policy consultant in Washington DC. (www.bobgellman.com.) He has written extensively on health, de-identification, Fair Information Practices, and other privacy topics. Dixon and Gellman's writing collaborations include a reference book on privacy *Online Privacy: A Reference Handbook,* as well as numerous and well-regarded privacy-focused research, articles, and policy analysis.

About the World Privacy Forum

The World Privacy Forum is a non-profit public interest research and consumer education group focused on the research and analysis of privacy-related issues. The Forum was founded in 2003 and has published significant privacy research and policy studies in the area of health, online and technical, privacy, self-regulation, financial, identity, and data brokers among other many areas. www.worldprivacyforum.org.

Contents

THE SCORING OF AMERICA: **6**

HOW SECRET CONSUMER SCORES THREATEN YOUR PRIVACY AND YOUR FUTURE **6**

INTRODUCTION 6

PART I: SUMMARY AND BACKGROUND **7**

WHAT IS A CONSUMER SCORE? 8

WHO HAS A SCORE? 8

GAPS IN CONSUMER PRIVACY RIGHTS AND PROTECTIONS AROUND CONSUMER SCORING (AND WHY EXISTING LAWS DON' ALWAYS APPLY) 9

Key Issue: Score Secrecy *11*

Key Issue: Score Accuracy *12*

Key Issue: Identity Theft and Consumer Scoring *13*

Key Issue: Unfairness and Discrimination *13*

Key Issue: Sensitive Health and Lifestyle Information and Consumer Scoring *14*

Key Issue: Consent and Use of Consumer Data in Predictive Scores *15*

SCORES THEN: A HANDFUL OF FACTORS. ……SCORES NOW: THOUSANDS OF FACTORS 15

SCORING METHODS AND MODELS ARE OPAQUE 18

EXAMPLES AND NUMBERS OF CONSUMER SCORES 19

USES OF CONSUMER SCORES, REGULATION, AND MODERN ELIGIBILITY 19

DEJA VU: WHY THE HISTORY OF THE CREDIT SCORE IS IMPORTANT 21

SUMMARY OF FINDINGS AND RECOMMENDATIONS 23

Findings: *23*

Recommendations: *24*

ADVICE FOR CONSUMERS: 26

PART II. CONSUMER SCORES: WHAT GOES INTO THEM, AND HOW THEY ARE MADE. **27**

DEFINING CONSUMER SCORES MORE DEEPLY 27

MAKING A CONSUMER SCORE, STEP 1: WHAT CONSUMER INFORMATION GETS PUT INTO A CONSUMER SCORE? **30**

TRADITIONAL SCORE INGREDIENTS: CREDIT SCORES 30

MODERN SCORE-MAKING INGREDIENTS: RAW CONSUMER DATA IN THE DIGITAL AGE 32

CONSUMER DATA AVAILABLE FOR PURCHASE AND USE IN ANALYTICS 33

Demographic Information: *33*

Contact Information: *34*

Vehicles: *34*

Lifestyle, Interests and Activities data (including medical): *34*

Financial and Economic – Property and Assets data: *36*

Financial and Credit data: *37*

SCORING MODELS: HOW THE CONSUMER SCORES ARE MADE, PART TWO **38**

ALGORITHM, INC. 38

LAYERING SCORING MODELS 40

HOW MANY VARIABLES? 41

POLICY QUESTIONS 41

PART III. THE CONSUMER SCORES **42**

CATEGORY: FINANCIAL AND RISK SCORES 43
ChoiceScore *43*
Median Equivalency Score (Summarized credit statistics) *44*
Risk IQ Score *45*
Consumer Profitability Score *45*
Job Security Score *47*
Consumer Prominence Indicator Score *47*
Discretionary Spending Index Score *48*
Invitation to Apply Score *48*
Charitable Donor Score *49*
Household Segmentation Scoring Systems (Personicx, Mosaic, etc.) *50*
Collection and Recovery Scores *50*
Churn Scores *51*
CATEGORY: FRAUD SCORES 52
FICO Falcon Fraud Manager *53*
Other Fraud Scores *54*
CATEGORY: CUSTOM SCORES 55
The Emergence of Custom Scores and the Pregnancy Predictor Score Example *55*
CATEGORY: REGULATED CREDIT AND FINANCIAL SCORES 57
FICO Score *58*
Vantage Score *58*
Beacon Score *59*
Small Business Intelliscore *59*
Tenant Scores *60*
CATEGORY: IDENTITY AND AUTHENTICATION SCORES 60
ID Analytics ID Score *60*
Insurance Scores *61*
Category -- Health Scores *61*
Affordable Care Act Individual Health Risk Score *62*
FICO Medication Adherence Score (MAS) *63*
Frailty Scores: General *65*
CMS Frailty Adjustment Score *66*
Hopkins Frailty Score *66*
Other Health Scores *66*
Personal Health Scores: WebMD, others *67*
Resource Utilization Group Scores *68*
SF-36 Form *69*
Complexity Scores *69*
CATEGORY – SMART GRID AND ENERGY SCORES 70
Peer-to-Peer Energy People Meter Score (EPM) *70*
CATEGORY - SOCIAL SCORING 72
Klout Score *73*
Employment Success Score *75*
TAX RETURN SCORES 75
CATEGORY – LAW ENFORCEMENT SCORES, INCLUDING POLICE, TRANSPORTATION, SAFETY, AND OTHER 76
Automated Targeting System Score *76*
Richard Berk Algorithm *77*
Youth Delinquency Scores *77*
Predictive Anti-Fraud Scores: US Postal Service Office of Inspector General *77*

CATEGORY -- ENVIRONMENTAL SCORES 78
EP Health Risk Score *78*
AIQ Green *79*
CATEGORY - OTHER CONSUMER SCORES 79
Potential Scores *80*
Non-included Scores: *80*

PART IV: THE HISTORY OF SCORING: HOW THE CREDIT SCORE AND CONSUMER SCORES BEGAN, AND WHY IT IS RELEVANT TODAY 80
THE BEGINNINGS OF CONSUMER SCORING 80
CREDIT SCORING BECOMES ENTRENCHED 81
HOW THE FORMERLY SECRET CREDIT SCORE BECAME AVAILABLE TO THE PUBLIC 82
ONGOING DISCLOSURE CHALLENGES AND OTHER ISSUES WITH CONSUMER CREDIT SCORES 83

CONCLUSION 84

ABOUT THIS REPORT AND CREDITS 86

APPENDIX A 87
TIMELINE: HIGHLIGHTS IN SCORING 87

APPENDIX B 89
SCORE TAXONOMY 89

WORLD PRIVACY FORUM

April 2, 2014 | Pam Dixon and Robert Gellman

The Scoring of America:
How secret consumer scores threaten your privacy and your future

Introduction

To score is human. Ranking individuals by grades and other performance numbers is as old as human society. Consumer scores — numbers given to individuals to describe or predict their characteristics, habits, or predilections — are a modern day numeric shorthand that ranks, separates, sifts, and otherwise categorizes individuals and also predicts their potential future actions.

Consumer scores abound today. Credit scores based on credit files receive much public attention, but many more types of consumer scores exist. They are used widely to predict behaviors like, spending, health, fraud, profitability, and much more. These scores rely on petabytes of information coming from newly available data streams. The information can be derived from many data sources and can contain financial, demographic, ethnic, racial, health, social, and other data.

The Consumer Profitability Score, Individual Health Risk Score, Summarized Credit Statistics that score a neighborhood for financial risk, fraud scores, and many others seek to predict how consumers will behave based on their past behavior and characteristics.

Predictive scores bring varying benefits and drawbacks. Scores can be correct, or they can be wrong or misleading. Consumer scores – created by either the government or the private sector – threaten privacy, fairness, and due process because scores, particularly opaque scores with unknown ingredients or factors, can too easily evade the rules established to protect consumers.

The most salient feature of modern consumer scores is the scores are typically secret in some way. The existence of the score itself, its uses, the underlying factors, data sources, or even the score range may be hidden. Consumer scores with secret factors, secret

sources, and secret algorithms can be obnoxious, unaccountable, untrustworthy, and unauditable. Secret scores can be wrong, but no one may be able to find out that they are wrong or what the truth is. Secret scores can hide discrimination, unfairness, and bias. Trade secrets have a place, but secrecy that hides racism, denies due process, undermines privacy rights, or prevents justice does not belong anywhere.

Broader transparency for consumer scores with limited secrecy may offer a middle ground. Knowing the elements but not necessarily the weights of a scoring system provides a partial degree of openness and reassurance. Knowing that there is a scoring system and how and when it is used helps. Knowing the source and reliability of the information used to make a score helps. Being able to challenge a score and correct the data on which it is based helps. Knowing that some types of information will not be used for scoring helps. Knowing that data collected for one purpose will not be used for another or in violation of law helps. Knowing that the person running the scoring system is accountable in a meaningful way helps.

The history of the credit score provides a useful model for the new batch of predictive consumer scores. Developed in the 1950s, the credit score became part of consumer credit granting. The credit score was largely secret to the consumers that it scored and affected until 2000, when a long and well-documented history of unfair uses and abuses finally culminated in the credit score being made available to consumers. Eventually, public pressure caused the credit score's use and even its underlying factors to become public. The use of factors such as race, gender, and religion were prohibited and this was spelled out in detail in law.

No similar protections exist for most consumer scores today. Consumer scores are today where credit scores were in the 1950s. Data brokers, merchants, government entities, and others can create or use a consumer score without notice to consumers. For various reasons laws governing credit scores do not typically extend protection to the new consumer scores. We need rules that will make consumer scores fair, accountable, accurate, transparent, and non-discriminatory.

This report discusses and explores consumer scores, what goes into them and how they are made, how they are used, the regulations in place that control some but not most new consumer scores, and how scores affect broader privacy and fairness issues. The discussion of findings and recommendations points toward solutions and reforms that are needed.

Part I: Summary and Background

As the numbers of predictive consumer scores increase and their usage expands, Americans face a future that may be shaped in significant ways by consumer scores. By itself, consumer scoring is not necessarily good or bad. Scoring orders consumers along a

mathematically defined scale. However, scoring has the prospect of being used to affect individuals in significant ways that may not always be fair or even legal.

If a predictive score unknown to a consumer determines how that consumer is treated, the results may not be acceptable to the American public. The quality and relevance of the data used, the transparency of the methodology of how the score was created, plus the reasonableness of the application of the consumer score are the major factors that determine the fairness of any scoring activity. These issues should be the central focus of any policy debate about consumer scoring. These issues also suggest the elements of best practices that should apply to consumer scoring.

What is a Consumer Score?

With this report, the World Privacy Forum introduces a term: *consumer scores*. Consumer scores – the ones we discuss in this report – are built using predictive modeling. Predictive modeling uses copious amounts of information fed through analytical methods to predict the future, based on past information.

Predictive consumer scores are important because they affect the lives, privacy, and wellbeing of individuals. Many people know about credit scores, but few know about the broader range of new consumer scores. Consumer scores are already abundant and are in active use. Consumer scores are not just an online phenomenon. Consumer scores are found in a wide array of "offline" arenas, including businesses, health care providers, financial institutions, law enforcement, retail stores, federal and state government, and many other locations. Some social consumer scores may have online applications, but mostly, consumer scores are not solely focused on just online activities. And unlike credit scores, consumer scores remain largely secret and unregulated.

The World Privacy Forum defines a consumer score as follows:

> A consumer score that describes an individual or sometimes a group of individuals (like a household), and predicts a consumer's behavior, habit, or predilection. Consumer scores use information about consumer characteristics, past behaviors, and other attributes in statistical models that produce a numeric score, a range of scores, or a yes/no. Consumer scores rate, rank, or segment consumers. Businesses and governments use scores to make decisions about individual consumers and groups of consumers. The consequences can range from innocuous to important. Businesses and others use consumer scores for everything from predicting fraud to predicting the health care costs of an individual to eligibility decisions to almost anything.

Who has a Score?

Consumer scoring is already more widespread than most people realize. Many hundreds of consumer scores exist, perhaps thousands. How many Americans have them? Almost

all do. Minors are less likely to be scored than adults, although they, too can have or influence some consumer scores. For example, household scores often reflect interests and activities of minors.[1]

Among American adults, each individual with a credit or debit card or a bank account is likely to be the subject of one or more scores.[2] Many individuals signed up under the Affordable Care Act have a score.[3] Individuals who buy airline tickets have a score.[4] Individuals who make non-cash purchases at large retail stores likely have a score.[5]

Scores such as the medication adherence score, the health risk score, the consumer profitability score, the job security score, collection and recovery scores, frailty scores, energy people meter scores, modeled credit scores, youth delinquency score, fraud scores, casino gaming propensity score, and brand name medicine propensity scores are among the consumer scores that score, rank, describe, and predict the actions of consumers.

In short, almost every American over the age of 18 has at least one score, and most adult Americans have many scores. An individual could easily be the subject of dozens or even hundreds of secret consumer scores. We can safely predict that there will be many more consumer scores in the future. Fed by the masses of consumer data now available, consumer scoring is quickly becoming a form of shorthand to make sense of a sea of information.

Gaps in Consumer Privacy Rights and Protections around Consumer Scoring (And why existing laws don't always apply)

This report's analysis is that many new consumer scores exist, and many of these new scores do not appear to fall under the narrow protections offered by the Fair Credit Reporting Act[6] or the Equal Credit Opportunity Act[7] for a variety of reasons. Scores built from factors outside a formal credit bureau file, scores designed to predict the behavior of groups of people instead of individuals, and new scores in emerging and unregulated areas may all fall outside of existing protections.

For example:

---

[1] A good example of this would be an *aggregate credit score*, which scores neighborhoods versus individual consumers.
[2] Likely scores for an adult with a credit or debit card would be real-time or near real-time fraud and/or identity scores.
[3] This is the ACA Health Risk Score. Specific scores are discussed in detail in Part III of the report.
[4] This score is generated by the federal Transportation Security Agency. This score is discussed in Part III of the report.
[5] See the Custom Scores section of Part III.
[6] 15 U.S.C. § 1681 et seq.
[7] 15 U.S.C. 1691 et seq.

- Energy consumption scores, churn scores, and identity scores are not likely to fall under the FCRA and other laws as currently written. This is because those scores do not meet the layers of qualifications that would bring them under the FCRA.

- Scores that identify the approximate credit capacity of neighborhoods instead of individuals also appear to be unregulated. This is because the FCRA applies to individuals, not neighborhoods. Formal credit scores may only be used in certain circumstances, for example, for extending a firm offer of credit or insurance. Credit scores cannot be used for general marketing purposes, but aggregate credit statistics tied to a neighborhood do not appear to be subject to the same restrictions for the reasons mentioned. Lead generation is not the same thing as a formal offer of credit under the FCRA.

- Risk scores -- like health risk scores -- that use broad demographic information and aggregate financial statistics about consumers to assess financial or other risks (credit bureau files are not typically used) also don't appear to fall under the layers of requirements that would bring them under current regulation.

- The Equal Credit Opportunity Act requires credit scoring systems to not use race, sex, marital status, religion, or national origin as factors comprising the credit score. But this law applies only to what is today a narrowly defined credit scoring system. Other scores which fall outside of the narrow definitions – like identity, fraud, churn, and other predictive scores can incorporate factors that would in other situations be considered prohibited factors to use.

As a result, consumers may have scant rights to find out what their non-FCRA consumer scores are, how the scores apply to them and with what impact, what information goes into a score, or how fair, valid, or accurate the score is. Even if the input to a score is accurate, consumers do not know or have any way to know what information derived from their lifestyle, health status, and/or demographic patterns is used to infer patterns of behavior and make decisions that affect their lives.

Further, consumers can have difficulty exercising basic Fair Information Principles for many if not most new consumer scores.[8] Fair Information Principles form the base for most global privacy law today, including some US privacy laws. However, those who create unregulated scores have no legal obligation to provide Fair Information Practices or due process to consumers.

---

[8] In this report we refer to Fair Information Practices as a baseline and standard by which to judge consumer scoring. FIPS are an established set of eight principles guiding privacy. The U.S. has ratified the FIPs twice since the 1970s. The FIPS include the principles of collection limitation, data quality, purpose specification, use limitation, security safeguards, openness, individual participation, and accountability. See Robert Gellman, *Fair Information Practices: A Basic History*. <http://bobgellman.com/rg-docs/rg-FIPShistory.pdf>. A brief introduction is here <http://www.worldprivacyforum.org/2008/01/report-a-brief-introduction-to-fair-information-practices/>.

These significant gaps in consumer protections mean that consumer scores may include or use discriminatory factors in their composition, or uncorrected or otherwise inaccurate information could be included. Scores developed to characterize individuals or predict their behaviors need to provide fairness and due process. The credit score is already subject to some regulation, but that is not to say that consumers would not benefit from better rules for credit scores.

There is a great need to examine the effects and fairness for all consumer scores now in use. Intriguing possibilities exist that a certain stratification of consumer experience based on opportunities offered to each consumer could become commonplace. Victims of identity theft, for example, may consistently receive different and less desirable marketing treatment than individuals with clean credit scores, even if most other demographic factors are similar.

Disparate treatment, even in the area of marketing opportunities granted to consumers, raises many questions, questions that the general field of risk-based pricing has raised. Oddly, direct marketing lists and activities have the potential to strike deeply into the lives of individuals in quirky ways that can have an impact on consumer lifestyle. Much remains to be learned about the impact of consumer scoring in the direct marketing arena, as well as eligibility issues and edge-eligibility issues like scores for identity and authentication.

Some of the specific issue areas around gaps in protection and information fairness in scoring including the following.

Key Issue: Score Secrecy

There are good reasons why credit scores are not secret anymore, nor are the foundational factors that comprise the score. By law, consumers have the right to see credit scores now. This report finds that with the exception of the credit score and a handful of other consumer scores, at this time, secrecy is the hallmark of many consumer scores.

The factors that go into most scores are usually secret, the models used are usually secret, and in many cases, the score itself is also secret. This report's analysis is that consumer scores that are risk scores bear many similarities to scores regulated under the FCRA. Yet industry treats these risk scores as falling outside the FCRA so that consumers have none of the rights guaranteed by the FCRA.

Consumers have no formal rights to find out what their non-FCRA consumer scores are, or how these scores affect their lives. Victims of identity theft and other individuals may have errors or omissions affecting their scores, but they do not necessarily have a right to see or correct the scores. Even if information is accurate, consumers do not know or have any way to know how companies use information derived from their lifestyle, health status, and/or demographic patterns to infer patterns of behavior and make decisions that affect their lives. Unseen scores can affect consumers' marketplace experiences and much more.

Key Issue: Score Accuracy

Because consumers do not have the right to correct or control what personal information goes into a consumer score as an attribute or factor, the accuracy of the scores is suspect. Consumers also do not have the right to see the scoring models used to make the score, nor do they typically have information about the model validity. Because of the lack of transparency, consumers cannot be assured of the reliability, fairness, or legality of scoring models. Inaccurate, incomplete, and illegal factors may be used today to make decisions about consumers without any oversight or redress.

Credit scores offer a useful model here. Credit scores are based on credit reports. Credit report accuracy has been the subject of substantial, meaningful scrutiny over decades. The CFPB, in its 2012 study on credit reports, noted the significant problems with inaccuracy that occur:

> Given the volume of data handled, the challenges of matching tradelines to the correct consumer files, and the number and variety of furnishers, inaccuracies in some credit files inevitably occur. Inaccuracies in credit files and credit reports can occur where information that does not belong to a consumer is attached to his or her file, where information belonging to a consumer is omitted from the file, or where there are factual inaccuracies in trade line or other information in the consumer's file. Some of these inaccuracies can be attributed to matching challenges in assigning a trade line to a consumer's file. Other causes of inaccuracies include data and data entry errors, NCRA system or process inaccuracies, furnisher system or process inaccuracies, identity fraud, or time lags.[9]

In a ten-year, Congressionally-mandated study published in 2013, the FTC found that overall "one in five consumers had an error on at least one of their three credit reports."[10] The FTC found that these credit report errors did impact the credit score. The FTC found that, specifically:

- "Slightly more than one in 10 consumers saw a change in their credit score after the CRAs modified errors on their credit report; and;
- Approximately one in 20 consumers had a maximum score change of more than 25 points and only one in 250 consumers had a maximum score change of more than 100 points."[11]

---

[9] Consumer Financial Protection Bureau, *Key Dimensions and Processes in the U.S. Credit Reporting System* at 23 (Dec. 2012).
[10] Federal Trade Commission, *Section 319 of the Fair and Accurate Credit Transactions Act of 2003: Fifth Interim Federal Trade Commission Report to Congress Concerning the Accuracy of Information in Credit Reports*, <http://www.ftc.gov/reports/section-319-fair-accurate-credit-transactions-act-2003-fifth-interim-federal-trade>.
[11] Id.

Errors in credit scores abound, and credit scores are based on credit reports, which also are subject to significant errors. If a transparent score with few factors has these kinds of errors, what about consumer scores? Consumer scoring relies on dozens, hundreds, or thousands of data elements that have no standards for accuracy, timeliness, or completeness. The quality of data matters: errors in data used to make a score create a score that is not predictive. With thousands of factors, error rates and false readings become a big issue.

Key Issue: Identity Theft and Consumer Scoring

Victims of identity theft – both financial and medical forms of the crime -- may have significant and stubbornly ongoing errors or omissions affecting their scores. ID theft victims can be seriously affected by identity scoring because their identity scores and fraud scores may be incorrect as a direct result of criminal activity. This can cause a range of problems from being denied services to being tagged as a potential fraudster. Yet even this vulnerable group has no right to see or correct many consumer scores.

Key Issue: Unfairness and Discrimination

One of the fundamental policy issues regarding scoring activities is the question of what characteristics it is appropriate to use in scoring consumers. In the world of home loans, ECOA has answered that question. But in the world of direct marketing, this area is nearly without boundaries. In a prescient early critique of scoring policy, Columbia University professor Noel Capon wrote in 1982:

> Since prediction is the sole criterion for acceptability, any individual characteristic that can be scored, other than obviously illegal characteristics, has potential for inclusion in a credit scoring system.[12]

As a bewildering plethora of new databases of consumer information become available, these databases may be scored in various ways by being run through one or more scoring models. More databases of consumer information fundamentally can mean more potential scores, and more potential characteristics to score.

The Equal Credit Opportunity Act protects consumers from invidious discrimination in formal credit granting situations. Notably, the ECOA requires that credit scoring systems may not use race, sex, marital status, religion, or national origin as factors comprising the score. The law allows creditors to use age, but it requires that seniors be treated equally.[13]

---

[12] Noel Capon. *Credit scoring systems: a critical analysis*, 46 Journal of Marketing 82-91(1982).
[13] For more information, see Federal Trade Commission, *How Credit Scores Affect the Price of Credit and Insurance*, <http://www.consumer.ftc.gov/articles/0152-how-credit-scores-affect-price-credit-and-insurance>.

But in the modern consumer scores, marital status – a protected factor under ECOA – is commonly used as a consumer score factor. Consumer scores may also contain underlying factors of race, sex, and religion without disclosure to consumers. In some cases, health factors may also be included in scores, for example, if a person smokes, or has a chronic illness. (See the section on Factors below for an example of a score that incorporates smoking and ethnicity).

As discussed in Part II of this report, a single score is often created from the admixture of more than 600 to 1,000 to even 8,000 individual factors or data streams. These factors can include race, religion, age, gender, household income, zip code, presence of medical conditions, zip code + 4, transactional purchase information from retailers, and hundreds more data points about individual consumers.[14] Therefore, one individual score can have the potential to contain hidden factors that range from bland – like mail order buyer of sports goods -- to quite sensitive – like ethnicity.

A score designed to assess or assign consumer value to a business could easily include factors that would be entirely unacceptable or that, in the context of either the Equal Credit Opportunity Act (ECOA) or the Fair Credit Reporting Act, would be flatly illegal. If ECOA factors are present in consumer scores, in most cases it would be difficult or impossible for consumers to find out if the scoring system or its factors were secret.

While carefully directed and controlled use of credit scoring and credit automation has reduced some discriminatory practices, new consumer scoring that uses elements that correlate with prohibited factors such as race can reintroduce discrimination and hide the effects behind a secret or proprietary screen that falls entirely outside of current consumer protection regulations. This is not acceptable.

Key Issue: Sensitive Health and Lifestyle Information and Consumer Scoring

Health scores already exist. This category of score deserves special attention and scrutiny. Some health scores are used in the HIPAA context, some are used outside the HIPAA context. The health scores used outside the HIPAA context are of most concern. Actuaries already use some new consumer scores to underwrite risk, for example, the Brand Name Medicine Propensity Score from a health category and the Underbanked Indicator from the financial category.[15] Scores can contain health information as hidden information within the score, and used for health purposes, or used for non-health related purposes such as marketing, or risk scoring. Many consumers with chronic health conditions would object strenuously to having their financial risk be determined by their health status. While health risk may be very predictive in a score, is it fair to use without consumer knowledge?

---

[14] Part II contains a substantive list of scoring factors.

[15] For more details about this issue, see the Health Score discussion in Part III of the report. See also Tim Hill, *Predictive Modeling Topic*, Future of Preferred Underwriting, Society of Actuaries, August 25-27, 2013.

Just because a score contains information about a consumer's health status, it does not mean the score will be subject to the federal health privacy rule (HIPAA). In fact, much of health information available for commercial use outside of the healthcare environment falls outside the scope of HIPAA. HIPAA, for example, provides no consumer privacy rights over health data held by list and data brokers.

Health information often leaks outside of HIPAA protections when it is revealed by consumers through surveys, website registrations, and other online activities. After a consumer reveals his or her health information to a non-HIPAA third party, that information is considered out of HIPAA's bounds. It is in this way that consumers' most sensitive health information can wind up used as fodder for a consumer score, with unknown consequences.

Consumer scores that use health or other sensitive information such as sexual orientation as factors need close examination for fairness, and consumers need rights over whether their health information is used in predictive scores, whether for marketing or any other purpose.

Key Issue: Consent and Use of Consumer Data in Predictive Scores

If a consumer fills out a registration on a health-related web site or a consumer warranty card that accompanied a purchase, the consumer did not give ***informed consent*** that the information can be used in downstream consumer scoring in ways that affect the consumer's marketplace opportunities. A buried statement in an unread privacy policy that "we may share your information for marketing purposes with third parties" is not informed consent to allow unfettered use information for predictive scoring.

Does making a purchase with a credit or debit card at a retailer grant consent for use of a consumer's purchases and other information to be used in a score? Part II of this report contains a detailed discussion of what kinds of information go into consumer scores. Many individuals would be quite surprised to learn just how the details of their lives are fodder for scores they may never see or have access to – and did not knowingly consent to. The issue of consent becomes increasingly important for scores that affect any kind of eligibility, such as jobs, credit, insurance, identity verification, or other significant opportunities.

Scores Then: A Handful of Factors. ……Scores Now: Thousands of Factors

The research for this report found that consumer scores may rely on hundreds or thousands of pieces of consumer information coming from many different data sources. This report identifies a large roster of raw consumer data that includes demographic information like age, race, gender, ethnicity, and home address as well as religion, mobile phone number, online and offline purchase history, health conditions like Alzheimer's, diabetes, and multiple sclerosis, as well as intimate financial details such as net worth,

card holder information, low or high-end credit scores, money market funds, ages of children, and a great deal more.

Statistical scoring methods rely on the increasing availability of large amounts of new source data from social media, the web in general, and elsewhere. The input for consumer scores can include information that is mostly unobjectionable or public. But, as discussed, consumer scores also can incorporate highly sensitive information that in other contexts could be used in a prejudicial, unfair, or unethical way in making decisions about consumers. Some data, such as social media data, can be unobjectionable in one context, but inappropriate as a factor, for example, in credit decisioning models.

An example, and a fairly common one, is of a predictive model that a major US health insurer worked with an analytics company to create. The idea was to determine whether or not publicly available consumer data could enhance the quality and effectiveness of their predictive risk models. They tested approximately 1,500 factors at the household level and found that the consumer information that showed the most value in predicting individual level risk included:

- Age of the Individual
- Gender
- Frequency of purchase of general apparel
- Total amount from inpatient claims
- Consumer prominence indicator
- Primetime television usage
- Smoking
- Propensity to buy general merchandise
- Ethnicity
- Geography – district and region
- Mail order buyer - female apparel
- Mail order buyer - sports goods[16]

Those unfamiliar with predictive models can find it surprising to learn that information about purchasing sporting goods can become a part of a predictive score for a health insurer. But the factors used in this example are not surprising factors to find in a modern predictive consumer score model. This is actually a fairly short list compared to some models with thousands of factors.

The raw source material for the factors fed into consumer scores comes from sources such as:

- Retailers and merchants via Cooperative Databases and Transactional data sales & customer lists.
- Financial sector non-credit information (PayDay loan, etc.)

[16] Satish Garla, Albert Hopping, Rick Monaco,& Sarah Rittman,*What Do Your Consumer Habits Say About Your Health? Using Third-Party Data to Predict Individual Health Risk and Costs. Proceedings*, SAS Global Forum 2013. <http://support.sas.com/resources/papers/proceedings13/170-2013.pdf>.

- Commercial data brokers
- MultiChannel direct response
- Survey data, especially online
- Catalog/phone order/Online order
- Warranty card registrations
- Internet sweepstakes
- Kiosks
- Social media interactions
- Loyalty card data (retailers)
- Public record information
- Web site interactions, including specialty or knowledge-based web sites
- Lifestyle information: Fitness, health, wellness centers, etc.
- Non-profit organizations' member or donor lists
- Subscriptions (online or offline content)

(Part II of the report contains a more complete list.)

Traditionally, much of this data came from data brokers or mailing list sellers. That is still the case, but now many new data streams are now available. So-called big data (large data sets) is one source. Other new data, particularly mobile and social data streams, comes via application programming interfaces (API).

Data sets that used to be too large for all but the largest of companies to handle computationally can now be replicated and massaged by smaller firms and dedicated analytics teams within companies. Small analytics companies now compete with large data brokers to offer predictive analytics as well as data. One company states they use 300 billion data attributes in compiling their predictive scores, compiled from 8,000 data files.[17] This is no longer an extraordinary feat, it is competitive and to be expected in a world with large data flows.

Analytics tools will continue to come down in price, just as consumer data has become a commodity item. Widespread and inexpensive data and analytics have the potential to allow broader use of predictive analytics. Consumer scores may proliferate, especially in the absence of any need for accuracy, fairness, or transparency. Consumer scoring may expand just because it is cheap and fashionable. Merchants themselves may have little ability to judge the accuracy of consumer scoring.

In short, given abundant data and more data tools, factors used to create consumer scores could continue to increase. With each new unverified factor comes the risk of extra errors or unfairness due to sensitive or prejudicial or irrelevant factors.

[17] See Part III, Churn and Fraud scores for further discussion.

Scoring Methods and Models are Opaque

To create a consumer score, the score modeler feeds raw information (factors about consumers) into an algorithm designed to trawl through reams of data to detect consumer behavior patterns and to eventually sift consumers into a ranking by their scores. Each score generally has a name and predictive or descriptive function.

Credit scores are the best-known example of this. With credit scores, information culled from a consumer's credit bureau file becomes the raw input into a formal credit scoring model. **Credit** scores are built on **credit file** data. There is a nexus between the score and the data. The data is intrinsic to the score. The Fair Credit Reporting Act lays out, in concert with the Equal Credit Opportunity Act, a variety of responsibilities and restrictions in the uses of credit report data to use in credit scores. It is a balanced approach, and the Fair Credit Reporting Act remains a strong privacy law that enables Fair Information Practices for consumers.

Today, though, scoring models are easily built from data that is **extrinsic** to the final score. No nexus may exist between the **input** to a score and the **output**. In the financial scoring area, companies can now build financial scores from social media, demographic, geographic, retail purchase history, and other non-traditional information that may not be included in the formal credit file. In the health arena, analysts can now build health risk scores from mere wisps of demographic data, without any actual patient records.

In this is a new world of scoring, where analysts use factors extrinsic to the purpose of the score to build scores, that a person has red hair can be used as a factor. And the more factors, the better. Instead of using 30 factors, why not 3,000?

The use of credit information for pricing insurance risk is an example of this.[18] Statisticians and actuaries predict the cost of providing car or homeowners insurance using selections of credit report factors about a driver or homeowner. These insurance scores reportedly have a predictive capability. Yet there is no overt reason why credit worthiness correlates with the risk involved with driving safety. There is much controversy about the use of a statistical correlation that does not appear to be causal. Some states restrict the use of credit information for insurance, but the practice remains common.

Many of the consumer scores discussed in Part III are new classes of scores. When scores have hundreds and thousands of factors, it stands to reason that a causal link becomes much more tenuous. The more factors, the less casual the link may be. Risks associated with models are discussed in Part II of the report.

---

[18] See generally, Federal Trade Commission, *How Credit Scores Affect the Price of Credit and Insurance*, <http://www.consumer.ftc.gov/articles/0152-how-credit-scores-affect-price-credit-and-insurance>.

Examples and Numbers of Consumer Scores

Consumer scoring is growing. In 2007, the research for this report uncovered less than 25 scores. In 2014, the research uncovered hundreds of scores, with the strong likelihood that thousands of custom scores exist beyond our ability to confirm them.

Here are some examples of consumer scores:

> **Consumer profitability scores** predict, identify, and target marketing prospects in households likely to be profitable and pay debt.
>
> **The Job Security Score** claims to predict future income and capacity to pay.
>
> **Churn scores** seek to predict when a customer will move his or her business or account to another merchant (e.g., bank, cell phone, cable TV, etc.)
>
> **The Affordable Care Act (ACA) health risk score** creates a relative measure of predicted health care costs for a particular enrollee. In effect, it is a proxy score for how sick a person is.
>
> **The Medication Adherence Score** predicts if you are likely to take your medication according to your doctor's orders.
>
> **Brand Name Medicine Propensity Score** – will you be purchasing generics or brand name medications?
>
> **Fraud Scores** indicate that a consumer may be masquerading as another, or that some other mischief is afoot. These scores are used everywhere from the Post Office at point of sale to retailers at point of sale to behind-the-scenes credit card transactions. This is a very widely used score, and a number of companies compete in the fraud score area.

Part III of the report discusses these and other specific scores in detail.

Uses of Consumer Scores, Regulation, and Modern Eligibility

After a consumer is scored, ranked, described, or classified, companies, governments, private enterprises, health care entities, and others including law enforcement, can then use the resulting score to make decisions about an individual or group.[19] This is why scores impact consumers every day.

---

[19] We note in passing that consumer scores are typically created by analytics companies or professionals, and then the company or individual can either sell the score or sell their abilities to create custom scores for third parties. This is neither good nor bad, it is simply the basic business model, and it is quite old, stretching back many decades now. See discussion in Part IV.

Scores are gaining footholds as part of routine business processes for an expanding number of purposes for everything from marketing to assessing a person's identity to predicting a person's likelihood to commit fraud, and more.[20] The consumer score acts as a form of predictive evaluation to measure, predict, and generally facilitate making a decision about things such as an individual's:

- Credit worthiness,
- Popularity,
- Reputation,
- Wealth,
- Propensity to purchase something or default on a loan,
- Measure health,
- Measure/predict likelihood to commit fraud,
- Measure/predict identity
- Measure/predict energy consumption
- Job success probability
- Etc.

In the traditional credit score realm, some have argued that scoring is a foundational activity in the credit market, as well as a wholly positive factor. Others have said that the sub-prime meltdown of the late 2000's was fueled by overreliance on scoring products.[21] Some of the reasons for using credit scores have the potential to be helpful directly to consumers. Better and faster credit decisions help consumers, for example.

In new consumer scoring, some have argued that the scores are mainly just for marketing and are largely beneficial.[22] There can be potential benefits for consumers, For example, consumer risk scores that prevent fraud are helpful up to a point. But any potential benefits are real only if the scoring models are correct and non-discriminatory, the data is timely, and the scores are something that consumers want. Credit score regulation provides transparency and imposes some limits on use and construction. That offers some assurance to consumers. But when other consumer scores enter the marketplace without transparency or the limits that apply to credit scoring,[23] consumer benefits are much more uncertain, and unfairness is more likely.

---

[20] A trend in the data business is that consumer data itself has become a commodity due to the ease with which much consumer data can be acquired. Predictive analytics are becoming the key drivers of the data business. Instead of just lists of consumer information, a predictive score is a "value add" to data offerings.
[21] Federal Reserve Board Chairman Alan Greenspan, remarks at the annual convention of the American bankers association, (October 7, 2002), page 4. The extent to which the same credit scoring technologies touted by Chairman Greenspan may have been responsible for the mortgage meltdown and financial crisis that started in 2008 is beyond the scope of this report.
[22] FTC – Alternate Scoring Workshop, March 19, 2014. <http://www.ftc.gov/news-events/press-releases/2014/03/ftc-announces-agenda-panelists-alternative-scoring-seminar>.
[23] We do not mean to suggest that consumer scores have flaws and lack a full range of consumer protections, only that some limits and rights exist.

There is continuum of concern regarding consumer scores. Some scores are used for straightforward marketing purposes. These scores may be of less concern (however the fairness of factors and secrecy and validity are still a concern). Of greater concern are the consumer scores that are used for what we call "modern eligibility." This includes identity verification and fraud assessment scores, as well as credit decisioning scores and scores that are used to predict job success or decide between job applicants. These scores are especially worrisome because errors in these scores could lead to significant deleterious consumer impacts.

Whether a consumer receives a coupon for a free soda is not a big deal. In comparison, whether a consumer can complete a transaction is of significant consequence. Any score used for eligibility – like being approved for credit or a job -- becomes important. The most casual social scores meant just to measure social reach have on occasion been used as a criterion for judging applicant hiring qualifications, so all scores need to be explored and assessed.

Some scores –for example, aggregate credit scores not subject to the FCRA – can determine a neighborhood's general credit score or range. Opportunities for individuals living in that neighborhood will be affected in ways that they cannot anticipate and in ways that bear no relationship to their personal situation. Forms of redlining – the practice of turning someone down for a loan or insurance because they live in an area deemed to be high risk – is a threat in these situations.

By all appearances, consumer scoring has sped beyond the old constraints that were imagined in a largely analog era, and real consumer harms can be the result.

Deja Vu: Why the History of the Credit Score is Important

History is repeating itself with consumer scoring. Before secret predictive consumer score issues, there were secret credit score issues. Credit scores had many of the same problems: secrecy, unfairness, inaccuracy, and opacity. Part IV of the report contains a detailed history of the credit score, including how that score became public and how consumers got important rights regarding credit scores and reports.

In brief here, credit scores were unknown to most consumers through the 50s, 60s, 70s, and 80s. Trickles of a score that was not disclosed to consumers but that could be used to deny a person credit began to leak out slowly to some policymakers, particularly around the time ECOA passed. In May 1990, the Federal Trade Commission failed to protect consumers when it wrote commentary indicating that risk scores (credit scores) did not have to be made available to consumers. But when scoring began to be used for mortgage lending in the mid 90s,[24] many consumers finally began hearing about a "credit score,"

[24] In 1995 Freddie Mac and Fannie Mae endorsed the use of credit scores as part of the mortgage underwriting process. This had a substantial impact on the use of credit scores in the mortgage loan industry. See for example Kenneth Harney, *The Nation's Housing Lenders might rely more on credit scores*, The Patriot Ledger, July 21 1995.

many of them for the first time, and mostly when they were being turned down for a loan.[25] A slow roar over the secrecy and opacity of the credit score began to build.

By the late 90s, the secrecy of credit scores, the underlying methodology or factors that went into the score, and the scoring range became a full-blown policy issue. Beginning in 2000, a rapid-fire series of events – particularly the passage of legislation in California that required disclosure of credit scores to consumers – eventually ended credit score secrecy. Now, credit scores **must** be disclosed to consumers, and the context, range, and key factors are now known.[26] This is an example of how brave State privacy legislation serves as a model for state and federal policy makers. In this case, the US "laboratory of democracy" took state legislation and turned it into a federal rule that protects consumers everywhere.

Credit scores are no longer secret anywhere in the United States, and this was and still is the right policy decision. Why are other scores used for important decisions about consumers still secret? Why do score factors and numeric ranges remain secret, when the risk of the data comprising the score of a factor used in modern eligibility practices such as identity verification or fraud identification is very high?

Consumer scores stand today where credit scores stood in the 1950s: in the shadows. While there are some happy exceptions to this, such as most social scores and a few other consumer scores, most consumer scores are not available for consumers to see. As a result, consumers have little to no ability to learn when their lives are affected in a major or minor way by a consumer score that they never heard about. Credit scores are not perfect and still present some issues, but we have learned much from the credit score.[27] What we have learned most of all is that there should be no secret consumer scores and no secret scoring factors. If a score is being used in any meaningful way in a consumer's life, he or she needs to know about it and have some choices regarding that score.

---

[25] See for example, comments of Peter L. McCorkell, Senior Counsel to Wells Fargo, to the Federal Trade Commission, August 16, 2004 in response to FACT Act Scores Study.

[26] As of December 2004, the Fair Credit Reporting Act as modified by the Fair and Accurate Credit Transactions Act, or FACTA, ended score secrecy formally, and required consumer reporting agencies to provide consumers with more extensive credit score information, upon request. Also made available to the public was the context of the score (its numeric range), the date the score was created, some of the key factors that adversely affected the score, and some other items.

[27] Historically, some known consumer issues with the credit score include the following:

- Credit scores reflect inaccuracies in the credit reports they are based on, and credit reports have repeatedly been found to contain errors.
- Victims of ID theft can experience changed credit scores.
- Consumers who experience major life events such as medical events or divorce can pay a long price in the scoring world.
- The FTC has brought cases around "mission creep" in the use of credit score outside of its regulated uses. (Credit scores may only be used for firm offers of credit or insurance, not for general marketing use.)

Summary of Findings and Recommendations

Key Findings:

Consumer scores are expanding in type, number, and use because of the growth of predictive analytics and the ready access to hundreds and thousands of factors as raw material. Just as credit scores were secret for decades until state and federal legislation mandated that consumers could see their credit scores, today consumer scores are largely secret.

While new scores multiply, consumers remain in the dark about many of their consumer scores and about the information included in scores they typically don't have the rights to see, correct, or opt out of. A primary concern is how these scores affect individuals and meaningful opportunities available to them. Another area of concern is the factors used in new consumer scores, which may include readily commercially available information about race, ethnicity, religion, gender, marital status, and consumer-reported health information. This report's other key findings are:

- Unregulated consumer scores – as well as regulated credit scores – are both abundant and increasing in use today.

- The information used in consumer scores comes from a large variety of sources. Some scores use thousands of factors or consumer attributes.

- Many consumer scores, the ranges of the scores, and the factors used in them are secret.

- A consumer score may, without any public notice, rely on an underlying factor or attribute that has discriminatory implications (e.g., race or gender) or that most consumers consider sensitive (e.g., health or financial).

- Consumer scores in use today affect a consumer's marketplace opportunities. Some of these opportunities are major (e.g., financial, employment, health,), some are minor (e.g., receiving a coupon, spam, or junk mail), and many are in between. Consumers are adversely affected by scores that are kept secret, and consumers are adversely affected when they do not have rights to correct scores.

- Consumer scores are found in a wide array of "offline" arenas, including businesses, health care providers, financial institutions, law enforcement, retail stores, federal and state government, and many other locations. Some of the more social consumer scores may be online, but mostly consumer scores are not solely focused on just online activities.

- Consumers usually have no way to know what the scores predict or how the scores are used.

- Consumers typically have no notice or knowledge about the data sources used in scores predicting their behavior or characterizing them. Consumers typically have no rights over the data about themselves, and consumers usually have little to no ability to control use of the data.

- Consumers typically do not have the right to opt out of being the subject of a consumer score or to prevent use of a consumer score.

- Except where the Fair Credit Reporting Act applies to a consumer score, most consumer scores are not subject to any regulation for privacy, fairness, or due process. A lack of transparency makes it difficult or impossible to determine if creation or use of the scores violates a law that prohibits discrimination.

- Consumers who are victims of identity theft can have their credit or consumer scores affected thereby and may have little recourse even though errors may have major consequences for their ability to function in the economic marketplace can be major. Other consumers can also have their lives affected by the use of consumer scores to determine eligibility for important opportunities in the marketplace. Some consequences may be less significant.

- Consumers have remedies under state and federal law with respect to correcting and seeing their credit reports, but not necessarily with respect to the many records that contribute to consumer scores. Secret consumer scores do not provide consumers with correction rights of underlying information.

Key Recommendations:

Consumer scoring is not inherently evil. When properly used, consumer scoring offers benefits to users of the scores and, in some cases, to consumers as well. Some uses are neutral with respect to consumers. Consumer scores can also be used in ways that are unfair or discriminatory. The goal of these recommendations is to protect the benefits of consumer scoring, guarantee consumer rights, and prevent consumer harms.

- No secret consumer scores. No secret factors in consumer scores. Anyone who develops or uses a consumer score must make the score name, its purpose, its scale, and the interpretation of the meaning of the scale public. All factors used in a consumer score must also be public, along with the nature and source of all information used in the score.

- The creator of a consumer score should state the purpose, composition, and uses of a consumer in a public way that makes the creator subject to Section 5 of the Federal Trade Commission Act. Section 5 prohibits unfair or deceptive trade practices, and the FTC can take legal action against those who engage in unfair or deceptive activities.

- Any consumer who is the subject of a consumer score should have the right to see his or her score and to ask for a correction of the score and of the information used in the score.

- There are so many consumer scores in existence that consumers should have access to their scores at no cost in the same way that the law mandates credit reports be available at no cost, as mandated by Congress. Otherwise, if a consumer had to pay only one dollar for each meaningful score, a family could easily spend hundreds or thousands of dollars to see the scores of all family members.

- Those who create or use consumer scores must be able to show that the scores are not and cannot be used in a way that supports invidious discrimination prohibited by law.

- Those who create or use scores may only use information collected by fair and lawful means. Information used in consumer scores must be appropriately accurate, complete, and timely for the purpose.

- Anyone using a consumer score in a way that adversely affects an individual's employment, credit, insurance, or any significant marketplace opportunity must affirmatively inform the individual about the score, how it is used, how to learn more about the score, and how to exercise any rights that the individual has.

- A consumer score creator has a legitimate interest in the confidentiality of some aspects of its methodology. However, that interest does not outweigh requirements to comply with legal standards or with the need to protect consumer privacy and due process interests. All relevant interests must be balanced in ways that are fair to users and subjects of consumer scoring.

- The FTC should continue to examine consumer scores and most especially should collect and make public more facts about consumer scoring. The FTC should establish (or require the scoring industry to establish) a mandatory public registry of consumer scores because secret consumer scoring is inherently an unfair and deceptive trade practice that harms consumers.

- The FTC should investigate the use of health information in consumer scoring and issue a report with appropriate legislative recommendations.

- The FTC should investigate the use of statistical scoring methods and expand public debate on the proprietary and legality of these methods as applied to consumers.

- The Consumer Financial Protection Bureau should examine use of consumer scoring for any eligibility (including identity verification and authentication) purpose or any financial purpose. CFPB should cast a particular eye on risk

scoring that evades or appears to evade the restrictions of the FCRA and on the use and misuse of fraud scores. If existing lines allow unfair or discriminatory scoring without effective consumer rights, the CFPB should change the FCRA regulations or propose new legislation.

- The CFPB should investigate the selling of consumer scores to consumers and determine if the scores sold are in actual use, if the representations to consumers are accurate, and if the sales should be regulated so that consumers do not spend money buying worthless scores or scores that they have no opportunity to change in a timely or meaningful way.

- Because good predictions require good data, the CFPB and FTC should examine the quality of data factors used in scores developed for financial decisioning and other decisioning, including fraud and identity scores. In particular, the use of observational social media data as factors in decisioning or predictive products should be specifically examined.

- The use of consumer scores by any level of government, and especially by any agency using scores for a law enforcement purpose, should only occur after complete public disclosure, appropriate hearings, and robust public debate. A government does not have a commercial interest in scoring methodology, and it cannot use any consumer score that is not fully transparent or that does not include a full range of Fair Information Practices. Government should not use any commercial consumer score that is not fully transparent and that does not provide consumers with a full range of Fair Information Practices.

- Victims of identity theft may be at particular risk for harm because of inaccurate consumer scores. This is a deeply under-researched area. The FTC should study this aspect of consumer scoring and try to identify others who may be victimized by inaccurate consumer scoring.

Advice For Consumers:

- Consumers can take several steps to help reduce some but not all of the data flows regarding scores. If a consumer opts out of pre-screened offers of credit and insurance (Opt Out Prescreen)[28], this can help reduce the overall volume of credit information circulating about them. Opting out of affiliate information sharing (as allowed) under the Graham Leach Bliley opt out[29] can also help reduce information flows. For scores regulated under the Fair Credit Reporting Act, consumers can get one free credit report each year from the major bureaus at

[28] <https://www.optoutprescreen.com>.*World Privacy Forum Top Ten Opt Outs*, <http://www.worldprivacyforum.org/2013/08/consumer-tips-top-ten-opt-outs/>.
[29] <http://www.fdic.gov/consumers/privacy/privacychoices/index.html#yourright> and *World Privacy Forum Top Ten Opt Outs* <http://www.worldprivacyforum.org/2013/08/consumer-tips-top-ten-opt-outs/>.

www.annualcreditreport.com. For more information about consumer rights under the FCRA, ECOA, and other laws, see FTC's consumer resources[30] and CFPB's consumer resources.[31]

Part II. Consumer Scores: What Goes Into Them, and How They Are Made

Part II discusses how consumer scores are constructed in more depth. This part includes three main segments: 1) a technical discussion of how scores are made (which may be skipped for those not interested in the technical details of predictive analytics); 2) a list of data sources used in scores, and policy questions; and a more detailed definition of consumer scores.

Defining Consumer Scores More Deeply

We repeat here the definition we introduced earlier in this report:

> Consumer scores are scores that describe an individual or sometimes a group of individuals (like a household), and have a demonstrated ability to predict one or more consumer behaviors or outcomes. Consumer scores use information about consumer characteristics, behaviors, and other attributes in different amounts and combinations in statistical models that produce a numeric score, a range of scores, or a yes/no. Consumer scores are used to rate, rank, segment, and make decisions and predictions about individual consumers and groups of consumers. Those decisions can range from innocuous to important.

Generally, there are three elements in scoring, scores, factors, and models (or algorithms). The score is a metric, often but not always a number (e.g., categorical), that measures some quality of an individual (or group) or a transaction. A score is often used to determine a course of action regarding an individual or a transaction. *Consumer* scores are a class of scores used to make a determination about a consumer or a transaction related to or affecting the consumer either directly or indirectly.[32] Scores can be

[30] <http://www.consumer.ftc.gov/topics/credit-and-loans> and <http://www.consumer.ftc.gov/articles/0347-your-equal-credit-opportunity-rights>.
[31] <http://www.consumerfinance.gov/fair-lending/> and <http://www.consumerfinance.gov/askcfpb/search?selected_facets=category_exact:credit-reporting>.
[32] In the scoring literature, consumer score as a term does not appear frequently, but its occurrences generally concur with the definition the World Privacy Forum has used in this report. See for example Dan Meder, *Blended scores are better scores*, 109 Business Credit 48-49 (2007).

generated by a variety of means, including fully automated algorithms[33] and hybrid models.

Factors can be thought of as pieces of information that describe or relate in some way to the consumer, or to consumer behavior that the creator of the model feels or has determined is important to increase the predictive power of the model. Predictive mathematical models that generate scores can use many inputs, including but not limited to heuristics, demographics, transaction behavior, user history, comparative and/or existing profiles and results from other scores. Other factors can be payment history, number of late payments, and length of credit are all factors about an individual's transactions related to credit that may be used in assigning a credit score. Age, income, race, geographic location, education level, and patterns of behavior (for example, how many times a person has returned merchandise to a particular store, or how many times an individual has bounced a check) can be relevant data, depending on the goals of the score.

The model or algorithm takes the personal factors associated with an individual or a class that the individual belongs to (e.g., household or neighborhood), measures the factors against the model, and assigns a score to that individual. A scoring model can be simple or complex. It can use vast quantities of personal data, or just a little. A big data approach to consumer scoring typically requires analysis of large amounts of data to forecast future behavior, outcomes, or qualities.[34]

Scores, and the models and factors that produce the scores, can be controversial. A score is only as predictive or as fair as the score model and the factors used in that model. One score can use a factor that cannot or is not used in another score. In some situations, for example, home mortgages, score models cannot use factors that would discriminate, such as age or race. But other score models, such as auto insurance, can use some factors prohibited in home insurance. Many consumer scores are completely unrestricted in the choice of factors.

Another controversy that comes up frequently in generating predictive models is the problem of over-fitting. Over-fitting arises when an algorithm is trained to perform very well on an existing set of data, but has been tailored so well to that data set that it can behave erratically or incorrectly outside of the specific scenario it has trained for.

---

[33] Algorithms are procedures for solving a problem, often a mathematical problem, in a series of finite steps. In scoring models, algorithms also refer generally to the processes by which data are analyzed within a predictive model in order to generate a score. See Fair Isaac Corporation Overview, *Decisions Made Simple*, Glossary of EDM, <http://www.fairisaac.com/NR/rdonlyres/A609962F-371C-4FCA-BB2A-85DE8FD936F5/0/FairIsaacCorpOverview06.pdf>.

[34] This definition was culled from a wide variety of literature sources and from author interviews with experts in the scoring field. See bibliography for source list, see especially D.J. Hand & W.E. Henley, *Statistical Classification Methods in Consumer Credit Scoring: A Review*, 160 Journal of the Royal Statistical Society,523-541 (1997).See also Allen N. Berger, W. Scott Frame, *Small Business Credit Scoring and Credit Availability*, 45 Journal of Small Business Management (2007). See also Fair Isaac Corporation Overview, *Decisions Made Simple*, Glossary of EDM. <http://www.fairisaac.com/NR/rdonlyres/A609962F-371C-4FCA-BB2A-85DE8FD936F5/0/FairIsaacCorpOverview06.pdf>.

To understand overfitting, consider an algorithm training a robot to go to the kitchen and bring its master a drink. If trained specifically in its master's house, the robot may in time learn how to get the master's drink from the kitchen without fail. However, it does so by creating a set of rules specific to the master's house. For example, go past the pink chair 10 feet, turn right at the blue doorway two feet from the refrigerator, and then lift the arm exactly 20 inches to open the door and retrieve the diet soda on the third shelf. The problem with this scenario is that it doesn't generalize well across all housing,all masters, and all drinks.

What the developer of the algorithm wants is a model that works in all circumstances. That is a much harder task, and use of shortcuts risks overfitting. Overfitting is a constant danger for people who create algorithms, and models must be constantly tested in a variety of settings to ensure that they are not overfitting a specific scenario. Consequently, scores that have not been broadly tested may be inaccurate for some uses.

All consumer scores use some kind of consumer information – characteristics, behaviors, transactions, and/or attributes[35] – that describe an individual or sometimes a group of individuals and have demonstrative ability to predict some kind of behavior. Huge quantities of data may be collected and organized for this purpose. Each type of consumer score uses different factors, or consumer information, in different amounts and combinations.

Some scoring models, especially in the testing phases, can use as many as 1,000 variables or more to create a single score. Some use only one or two factors. But most use many factors. In the past few years, scoring has matured rapidly as predictive analytics teams and expertise have become part and parcel of how companies seek to monetize and understand their customer base, among other activities.

In this discussion, it is important to state what a consumer score is not. [36] For the purposes of this report, a consumer score does not measure a consumer's skill or abilities. For example, an SAT score is not consumer score in and of itself. Similarly, a health score that uses clinical health factors in a clinical setting solely for health diagnosis or treatment by a health care provider is not a consumer score. [37] However, if an SAT score or a health score is used for a different purpose as part of a predictive consumer score, then those scores will become factors of a consumer score.

---

[35] Characteristics are, for example, the questions asked on the credit application. Characteristics can also be performance categories of the credit bureau report. Attributes are, for example, the answers given to questions on the application, or entries inn the credit bureau report. Education is a characteristic, college degree or highest level of education achieved is the attribute.

[36] Some media reports have called alternative non-credit scores "e-scores." This term is actually a proprietary name for a product from a company called eBureau. <http://www.ebureau.com>. Even if the term were available for general use, the term e-score is too narrow and limiting for the broad range of health, energy, social and other consumer scores now in use.

[37] We assume without investigation that scores used for health care diagnostic or treatment purposes are fully transparent.

Also, if a score uses health factors for decisions outside of a strictly clinical setting, or if a score predicts a health outcome using extrinsic, non-clinical factors, then those scores are consumer scores. Health scores are discussed more in the discussion of individual consumer scores in Part III.

An appendix to this report includes a score taxonomy that describes consumer scores and score types, along with a decision tree to assist in determining when a score is a consumer score or not.

Making a Consumer Score, Step 1: What Consumer Information Gets Put Into a Consumer Score?

The underlying factors that go into a consumer score are important indicators of the fairness, accuracy, and non-discrimination of the score. If the factors selected to create a score are inaccurate, unfair, or discriminatory, then the score itself will be susceptible to the same biases.

Traditional Score Ingredients: Credit Scores

Much is known about what goes into credit scores and about credit score models. As with all scores, a key to a good credit score is the quality (e.g., accuracy, currency, and completeness) of the factors used in the scoring model. This is scoring 101.

David T. Kresge, formerly senior vice president of analytic services at Dun & Bradstreet noted the "data hunger" of scoring models in congressional testimony:

> "One of the keys to the implementation of a knowledge-based decision system is to incorporate and make effective use of the widest possible range of information. The data should be gathered from all available sources and it should be as wide-ranging as possible. Experience clearly demonstrates that good credit decisions cannot be based on just a small number of factors."[38]

Examples of some factors that go into consumer scores in the United Kingdom include:

- Time at present address
- Home status (owner, tenant, other)
- Credit card information
- Type of bank account
- Telephone (yes, no)
- Age

[38] Prepared testimony of David T. Kresge before the House Committee on Small Business, Subcommittee on Government Programs and Oversight (July 17, 1998).

- County Court judgments (number)
- Purpose of loan
- Type of occupation (coded)
- Marital status (married, divorced, single, widow, other)
- Time with bank (years)
- Time with employer (years) [39]

In the U.S., commonly used predictive variables for traditional financial scoring include:

- Payment history
- Public record and collection items
- Delinquencies
- Prior credit performance
- Outstanding debts
- Relationship between total balances of credit and total limit
- Age of oldest trade line
- Pursuit of new credit (applications to obtain additional credit)
- Time at present address
- Time with current employer
- Type of residence
- Occupation.[40]

From these two examples, we see that characteristics included in a financial score model vary from country to country. They may vary from state to state, depending whether laws restrict the use of some characteristics or variables. The data available in different countries may differ, and that may explain in part the construction of the model. It is not unusual for *missing information* to be an actual characteristic and included as a factor in the scoring model. This is a much-debated area of scoring.[41]

Until 2000, the factors that went into credit scores were not public.[42] However, these factors are now known. Fair Isaac reveals that for its FICO score,

- 35% is based on borrower's history payment history
- 30% is based on how much a borrower had drawn on available credit (amounts owed)

---

[39] These factors are culled from Table 1, Characteristics typical of certain credit scoring domains, D.J. Hand, *Statistical Classification Methods in Consumer Credit Scoring: A Review*, 160 Journal of the Royal Statistical Society, 527 (1997).

[40] These factors are culled from two sources. See Margaret Howard, *Shifting risk and fixing blame: The vexing problem of credit card obligations in bankruptcy*, The American Bankruptcy Law Journal, Winder 2001, 76 Am. Bankr. L.J. 63. See also *Scoring and Modeling*, FDIC – Risk Management Examination Manual for Credit Card Activities, Chapter VIII. Division of Supervision and Consumer Protection, March 2007. <http://www.fdic.gov/regulations/examinations/credit_card/ch8.pdf>.

[41] See, e.g., Hand, DJ &Henley WE, *Can reject inference ever work*? 5 IMA J Maths Appl Bus Ind 5-55 (1993).

[42] See the heading in this report "A Brief History of Consumer Scores" for a discussion of how credit score factors became public.

- 15 % is based on the length of the credit history
- 10% is based on the types of credit used
- 10 % is based on new credit[43]

We know the factors and their weights, but we do not know FICO turns a particular consumers payment history into a number that becomes a score.

Modern Score-Making Ingredients: Raw Consumer Data in the Digital Age

The carefully selected scoring factors and the much-debated weights for regulated credit scores as discussed above are like a landscaped garden in a well-tended public park compared to the untamed jungles of the data factors available and used in the new consumer scores. The new kinds of consumer scores use a much wider array of data sources, to the point that the new data sources make traditional credit scores look under-sourced by comparison. Whether 500 factors result in a better algorithm than 5 factors is unknown, and the answer may vary from score to score. More may be better sometimes, but not all the time.

Data for consumer scores can come from many sources, including data broker lists, retail purchases, social scores, census tract data, purchasing patterns, health conditions, ethnicity, book purchasing patterns, exercise patterns, and many other factors. Data used may be individual to a consumer or modeled (e.g., all consumers in a census tract). As described above, the effectiveness of a model depends on its ability to predict accurately from a variety of real world datasets and designated factors. Understanding what a model is trying to predict, what data is used for testing, and how the elements mesh to achieve a result are important to assessing the value, impact, and potential pitfalls of the scoring model. In some cases, it may be that a model is equally effective with less information, negating the need for collection and storage of vast quantities of information and data that could have privacy implications. More data may not be better or necessary.

Below is a list of the most common elements of consumer data available and in circulation today. Most consumers would be stunned to learn the number of data elements available in the commercial marketplace. Not every consumer score and not every data broker file for sale includes each item on this list. Different scores use combinations of different elements and plug those into differing score algorithms and models.

This list includes independent data sets with both structured and unstructured data. This list is sourced in part from 2013 Government Accountability Report on information resellers. Other information came from a WPF review and analysis of data broker data cards viewed through NextMark [44]over the course of a year (primarily 2013), and also from WPF review and analysis of reliable data broker web sites that list data sources. For

[43] MyFico, "Credit Education," <http://www.myfico.com/CreditEducation/>.
[44] Nextmark List Finder <http://lists.nextmark.com>.

example, the Acxiom *About the Data* portal[45] lists many categories of information collected and used for consumer marketing.

The data sets available for purchase today listed here – along with others we did not identify – can create multiple layers of predictive analysis of how consumer behavior, finance, demographics, geography, and the other factors listed here interact. That does not necessarily mean that the results are better.

Consumer Data Available for Purchase and Use in Analytics

The range of consumer data available for use in data analytics is broad and deep. The categories listed here is not exhaustive, but it offers an idea of the range of consumer information that goes into consumer scores.

Demographic Information:

- Age
- Age range
- Date of birth
- Education
- Exact date of birth
- Gender
- Marital status
- Home ownership
- Own or rent
- Estimated income
- Exact income
- Ethnicity
- Presence of children
- Number of children
- Age range of children
- Age of children
- Gender of children
- Language preference
- Religion
- Occupation - category of occupation
- Examples: Beauty (cosmetologists, barbers, manicurists) civil servants, clergy, clerical/office workers, doctors/physicians/surgeons, executives/administrators, farming/agriculture, health services, middle management, nurses, professional/technical, retail service, retired, sales, marketing, self-employed, skilled/trade/machine operator/laborer, teacher/educator.
- Occupation - title of occupation
- Military history

[45] Acxiom About the Data Portal <https://aboutthedata.com>.

- Veteran in household
- Voter party
- Professional certificates (teacher, etc.)
- Education level reached or median education

Contact Information:

- Full name
- Email address
- City
- State
- ZIP
- ZIP + 4
- Home Address
- Land-line phone
- Social IDs / social media handles and aliases
- Mobile phone number
- Carrier
- Device type
- Email address

Vehicles:

- Vehicle make, model and year
- VIN
- Estimated vehicle value
- Vehicle lifestyle indicator
- Model and brand affinity
- Used vehicle preference indicator

Lifestyle, Interests and Activities data (including medical):

- Antiques
- Apparel (women, men & child)
- Art
- Average direct mail purchase amounts
- Museums
- Audio books
- Auto parts, auto accessories
- Beauty and cosmetics
- Bible purchaser
- Bird owner
- Books
- Book purchases - plus types. (Mystery, romance, religious, etc.)

- Book clubs
- Career
- Career improvement
- Cat owner
- Charitable giving indicators:
- Charitable donor by type of donation (religious, health, social justice)
- Charitable donor by ethnicity or religion (Jewish donors, Christian donors, Hispanic donors)
- Charitable donor by financial status (wealthy donors)
- Children or teen interests
- Fashion and clothing (Multiple: sports, high fashion, shoes, accessories, etc.)
- Collectibles
- Collector
- Christian families
- Computer games
- Computers
- Consumer electronics (Many categories, including electronic fitness devices)
- Dieting and weight loss
- Telecommunications and mobile
- Dog owner
- Investing
- DVD purchasers
- Electronics - home, computing, office, other
- Empty nester
- Expectant parents
- Frequent mail order buyer
- Frequency of purchase indicator
- Getting married
- Getting divorced
- Gun ownership
- Health and beauty
- Health and medical: for example, Allergies, Alzheimer's disease, angina, arthritis/rheumatism, asthma, back pain, cancer, clinical depression, diabetes, emphysema, erectile dysfunction, epilepsy, frequent heartburn, gum problems, hearing difficulty, high blood pressure, high cholesterol, irritable bowel syndrome, lactose intolerant, ulcer, menopause, migraines/frequent headaches, multiple sclerosis, osteoporosis, Parkinson's disease, prostate problems, psoriasis/eczema, sinusitis/sinuses.
- High-end appliances
- Home improvement
- Household consumer expenditures — many categories.
- Jewelry
- Magazine subscriptions
- Mail order buyer
- Mobile location data (some analytics companies)
- Movies - attendance / collector

- Musical instruments
- Music
- New mover
- New parent
- Online and continuing education
- Online purchasing - many categories
- Parenting
- Pets - other
- Plus size clothing purchase
- Political affiliation
- Recent home buyer
- Recent mortgage borrower
- Retail purchasing - many categories.
- Science-related
- Sexual orientation
- Social media sites likely to be used by an individual or household, heavy or light users
- Spa
- Sports interests: (large category, these are examples)
- Birdwatching
- Equestrian
- Exercise and fitness
- Gardening
- Golf
- Fishing
- Outdoor interests - hiking, camping, climbing
- Swimming, diving, snorkeling
- Spectator Sports
- Stamps/coins
- Yoga
- Television, Cable, Satellite/Dish
- Travel: Vacations, domestic and/or international
- Purchase of international hotel or air flights
- Frequent flyer
- Types of purchases indicator
- Veteran in household
- Vitamins
- Volunteering

Financial and Economic – Property and Assets data:

- Estimated income
- Estimated household income
- Home value
- Length of residence
- Payment data: 30, 60, 90-day mortgage lates

- Purchase date
- Purchase price
- Purchase amount
- Most recent interest rate type
- Most recent loan type code
- Sales transaction code
- Most recent lender code
- Purchase lender code
- Most recent lender name
- Purchase lender name
- Fuel source
- Loan to value
- Purchase interest rate type
- Most recent interest rate
- Purchase interest rate
- Pool or spa
- Home - year built
- Air conditioning
- Boat ownership
- Plane ownership
- Motorcycle ownership
- Commercial assets or business ownership

Financial and Credit data:

- Bankruptcy
- Beacon score
- Credit score - actual
- Certificates of deposit/ money market funds
- Estimated household income ranges
- Income producing assets indicator
- Estimated net worth ranges
- IRAs
- Life insurance
- Low-end credit scores
- Mutual funds/annuities
- Summarized credit score or modeled credit score by neighborhood
- Payday loan purchaser
- Number of credit lines
- Tax liens
- Card data:
- Card holder - single card holder
- Range of new credit
- Debit or credit card present in household
- Card holder - brand (Discover, Visa, Mastercard, etc.)
- Card holder - type (Gas, bank, premium, luxury, prepaid, etc.)

- Frequent credit card user
- New retail card holders
- Underbanked or "thin file"
- Stocks or bonds
- Average online purchase
- Average offline purchase

In addition, a business may use enterprise data (historic data from its own customer files) to create proprietary or custom scores for its own use.

Scoring Models: How the Consumer Scores are Made, Part Two

Just as underlying factors going into a score should be fair and accurate, the algorithms that analyze the information should be of a high quality, should generate an accurate prediction, and should be validated against real-world data. As models are ultimately judged on their ability to make useful predictions from data, understanding how they actually perform and against what data sets is key. Without constant validation, scores might have no actual predictive value in reality. A bad or ineffective model ultimately means that the score does not offer accurate predictions. Bad scores based on a faulty or overfit model can still affect the treatment of individual consumers, the most important being eligibility and health care availability decisions.

Score creators have a good reason to get their models right. The marketplace is likely to weed out bad models, although it may take considerable time before this happens. The effects on individuals of poorly predictive consumer scores are uncertain. It would be useful for model makers to disclose their assumptions, predictive accuracy and model limitations. If a model is inherently a bad predictor of something important, then model users and model data subjects will want to know. Is it the data, the assumptions, overfitting, or other issues? A good faith, robust, public dialogue here could be helpful to all parties.

Algorithm, Inc.

Consumer scores generated by sophisticated mathematical models that detect patterns in information are often predictive, involve one or more algorithms, and rely on factors that describe individuals in some way.[46] The historical databases and raw consumer information that supply information to a score model can be both wide and large. Credit

[46] In a formal statistical model process, for example, data characteristics and attributes that describe a consumer are compared with a scoring table, or scorecard, and can be awarded points according to where they fall within the table. The points can be tallied to arrive at the overall score. Whether a high score means low or high risk depends on the model's construction. For more information about scoring models, see the discussion of this topic in this report. See also D.J. Hand, *Statistical Classification Methods in Consumer Credit Scoring: A Review*, Journal of the Royal Statistical Society, 523-541. (1997).

scoring databases used to build score models, for example, may contain records of well over 100,000 individuals, and the model may measure over 100 factors, or variables. Behavioral scoring databases that store transactional data on for example, retail or other purchases, repayment, or other activities can be even larger.[47]

Score models have advanced rapidly from their inception to the present time. The first widely used scoring model dates back to 1941, A paper published by Durand proved that discriminant analysis could produce reliable predictions of how individuals would repay credit extended to them. For detecting fraud, some of the older methods (pre-neural networks) of detecting fraud used lists of risky transactions and thresholds. It is worth noting that today's *score modeling* is an established field of study[48] as well as a competitive business.

In the world of scoring, the model used to generate the score is as important as the score itself. Two different score models, using identical factors, will almost always come out with a different score or metric for the same individual. A small error in the original calculations or inputs can magnify errors in the outcome.

The Federal Deposit Insurance Corporation, discussing credit score models, describes score models succinctly:

> Scoring models summarize available, relevant information about consumers and reduce the information into a set of ordered categories (scores) that foretell an outcome. A consumer's score is a numerical snapshot of his or her estimated risk profile at that point in time. Scoring models can offer a fast, cost-efficient, and objective way to make sound lending decisions based on bank and/or industry experience. But, as with any modeling approach, scores are simplifications of complex real-world phenomena and, at best, only approximate risk.[49]

The type of model a company uses to score consumers depends on the information analyzed, the purpose of the score, how much data is available, and a complex maze of other issues.[50] Ultimately, it is about how well the score model predicts what the company wants it to predict. Currently, widely used scoring models include discriminant analysis, linear regression, logistic regression, and decision trees or recursive partitioning, among other classification techniques.[51] Credit scoring is a well-established methodology at this point, and models have been fine-tuned for decades. The most typical approach in

---

[47] Id. at 526.
[48] See for example the Journal of Operational Research Society, Oxford, England; The Royal Statistical Society, UK; Applied Statistics (Journal).
[49] *Scoring and Modeling*, FDIC – Division of Supervision and Consumer Protection, March 2007. <http://www.fdic.gov/regulations/examinations/credit_card/ch8.pdf>.
[50] See Taylor James, *Predictive Analytics: Making Little Decisions with Big Data*, Information Management. (September 26, 2012).
[51] D.J. Hand, *Statistical Classification Methods in Consumer Credit Scoring: A Review*, 160 Journal of the Royal Statistical Society, 524-531-35 (1997).

a credit model is a logistic regression model.[52] Some consumer scores, such as fraud scores, employ completely different models, such as probabilistic or neural networks models. Banks experiment with neural network models,[53] and the credit card services industry uses neural networks for fraud detection.

Newer modeling techniques include the use of genetic algorithms, and an antibody approach. Ted Crooks, a leading model developer, described how antibody systems currently in development work:

> The next step after neural networks, the things we're working on these days are systems that use antibody approaches. It [an antibody system] looks at hundreds of millions of possible combinations of transactions, and recognizes those that individual fraudsters or fraud rings have found popular lately.[54]

Layering scoring models

Models can be layered on top of one another as well, and hybrid models that blend together results from various models are common[55]. Consumer information is often fed into not just one, but a variety of scoring models, compared with a variety of test or control data, and then the "best" or most accurate model is then validated and chosen for final deployment in the business setting. Because varying models can change the quality and meaning of the final score so much, which scoring model is chosen makes a difference. The design is a complex balancing:

> "Each model may employ a different subset of observations, consider different variables, make different assumptions about the relationship among the variables, and use different design concepts."[56]

When just one model is chosen, some information diversity is inevitably lost. To combat this, some experiment with combining two separate scores together to produce a *blended score* or a *combined score*. For example, a blending of a credit bureau score plus an application credit score would be a blended score. The idea is that the two scores together make a stronger predictive whole value.[57]

---

[52] LC Thomas, RW Oliver, DJ Hand, *A Survey of Issues in Consumer Credit Modeling Research*, 56 The Journal of the Operational Research Society (2005).
[53] Id. at 536.
[54] Interview, American Public Media, Marketplace Money, *Define Suspicious Activity*, (March 30, 2007).
[55] The model that won the Netflix prize consisted of blending 107 results. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.142.9009>.
[56] H. Zhu, P.A. Beling, & G.A. Overstreet, 52 Journal of the Operational Research Society 974-980 (2001) (Special Issue: Credit Scoring and Data Mining).
[57] *Scoring and Modeling*, FDIC – Division of Supervision and Consumer Protection, March 2007. <http://www.fdic.gov/regulations/examinations/credit_card/ch8.pdf>. "While most scores and models are generally established as distinct devices, a movement to integrate models and scores across an account's life cycle has become evident."

Not only can a consumer characteristic such as payment history be used in a scoring model, a score can be used as a characteristic, too. As such, it is not unusual to learn of combined uses of consumer scores in scoring models, and even entire score networks.[58] So what goes into a consumer score can be characteristics about individuals, or can be a score that has already been created about the consumer, or both.

How many variables?

Some see the use of large numbers of variables in modern models as a positive.

> In general, the ability to effectively use many different variables increases the strength of predictive models as it incorporates all available customer data that may be predictive as compared to traditional systems that are unable to scale since each variable must be tested by hand. This is particularly relevant in the insurance industry where there is a vast amount of customer information and a high correlation between the data and predictive outcomes.[59]

The type of variables used in a consumer scoring model have been controversial over the years. One of the complaints with consumer scores is that the scores derive from a black box of big data without a lot of thoughtful selection. One analyst noted that in the past, the factors fed into a model were in some way tied to the model. Credit file factors would go into credit scoring models. But in today's world, numerous unrelated factors may go into a model. For example, one financial company uses information about whether a person types in all caps or all lowercase letters as a predictive factor for loan repayment.[60]

Some models produce scores that give explicit estimates, others are numerical scales that reflect increasing levels of risk.[61] Scoring models can score and rank individual consumers, the characteristics of portfolios of loans, [62] or can even score neighborhoods. "Neighborhood scoring" has led to aggregate credit scores and other proxies for consumer credit classifications that are discussed elsewhere in this report.

Policy Questions

---

[58] H. Zhu, P.A. Beling, & G.A. Overstreet, 52 Journal of the Operational Research Society 974-980 (2001) (Special Issue: Credit Scoring and Data Mining).
[59] MMA Deploys KXEN's Infinite Insight to Boost Up-Sell, Cross-Sell and Customer Retention. ENP Newswire, 30 May 2013.
[60] Kenneth Cukier, Co-author, "Big Data: A Revolution That Will Transform How We Live, Work and Think, Council on Foreign Relations Federal News Service Media Call Subject: Big Data (May 9, 2013). The company Cukier mentioned on the call was *Just Finance*, according to the transcript.
[61] John Copas, *The Effectiveness of Risk Scores: The Logit Rank Plot*, 48Applied Statistics 166 (1999).
[62] The Basel Capital Accord, which regulates banks' lending from 2007 forward, has pushed portfolio scoring forward in the financial sector. See LC Thomas, RW Oliver, & DJ Hand, *A Survey of Issues in Consumer Credit Modeling Research*, 56 Journal of the Operational Research Society (2005).

A key policy issue for consumer scoring models is who gets to see the underlying information fed into the model. Another important issue is whether the factors are discriminatory or prejudicial in any way. The use of health factors in a non-medical consumer score and the use of other sensitive factors in a scoring model are also problematic.

Another issue focuses on the type of model used for the analysis, and if the model was appropriately and fully validated and kept up to date. Credit and consumer score models must work, be accurate, and be updated regularly. If they are not, even small deviations can lead to inaccuracies. An inaccuracy in a credit report can make a consumer the target of predatory lenders.

Another issue arises with the use of a consumer score as an underlying factor in a new or different consumer score. Error upon error can accumulate in a way that even transparency will not enable a highly-educated consumer to untangle. Again, prejudicial, inappropriate, or unfair factors could be in the mix, but a consumer wouldn't know it.

When a predictive model assigns a value or a range to a consumer, the model used to create that value must be transparent, accurate, reliable, and kept up to date. The numeric range should be well-quantified, and the results validated.

Part III. The Consumer Scores

This section describes and documents major consumer scores and categories. We found the scores included here through a lengthy and diligent literature search. We also conducted interviews with scoring experts and others knowledgeable about scores. The list here is not comprehensive because much information about consumer scoring is not public.

In 2007, when the initial stages of pre-research for this report began, many fewer consumer scores existed and the documentation was sparse. The first iteration of this report was only a handful of pages long. Now seven years later, an entire business sector has grown around predictive analytics and consumer scoring. The growth in consumer scoring has been rapid, suggesting that we are at the beginning of a significant growth period, probably closer to the beginning of the predictive analysis bell curve than at the middle.

Category: Financial and Risk Scores

Consumer scores that measure risk are a large category comprising several major subtypes. These scores are not usually subject to FCRA rules, but the law's application depends on the structure, use, and factors of each score.[63]

The most typical risk scores measure forms of consumer financial risk using non-credit factors or measure consumers for various types of fraud. In this report, we include authentication and identity scores in this category, as these scores ultimately seek to reduce or mitigate risk.

ChoiceScore

Experian produces **ChoiceScore**, a type of financial risk score.[64] Experian creates the score from consumer demographic, behavioral, and geo-demographic information. This score is used to segment consumers, as described here by a reseller of the data:

> ChoiceScore by Experian UnderBanked and Emerging Consumers
>
> ChoiceScore helps marketers identify and effectively target under-banked and emerging consumers. Using the most comprehensive array of non-credit data available from Experian. A financial risk score (indicating the potential risk of future nonpayment) provides marketers with an additional tool for more precise targeting.[65] The data card also indicated that the ChoiceScore could be used to suppress some consumers from getting information. [66]

Experian's web site indicates that the ChoiceScore is not likely accessible by consumers. The score appears to be available for non-FCRA uses.[67] The score's factors are not published, so it is difficult to know what kind of underlying data is included in the score. It is also difficult if not impossible to determine what businesses are buying or using the score.

Modern data analytics have made child's play of unearthing people who are in various credit score brackets without revealing the actual credit score. Congress acted to protect

---

[63] Sending an advertisement to a consumer is not the same thing as sending a formal, pre-approved offer of credit as described in the FCRA. This risk score category includes risk scores that may well be used to generate leads, but the advertisements themselves are not formal pre-approved offers of credit. This difference was discussed at length at the FTC Alternative Credit Scoring Workshop (March 19, 2014). <http://www.ftc.gov/news-events/events-calendar/2014/03/spring-privacy-series-alternative-scoring-products>.

[64] Experian ChoiceScore, <http://www.experian.com/marketing-services/data-digest-choicescore.html>.

[65] CHOICESCORE BY EXPERIAN UNDER BANKED AND EMERGING CONSUMERS, <http://datacardhub.adrearubin.com/market?page=research/datacard&id=268601>.

[66] <http://datacardhub.adrearubin.com/market?page=research/datacard&id=268601>.

[67] According to the data broker's data card, two entities purchased this data: Achievecard, and Figi's Incorporated. Figi's Incorporated appears to be a food gift retailer. <http://www.fbsgifts.com/about.html#figis>.

the use of specific credit score information with good reason. From a consumer perspective, the same underlying principles still need to be at work: fairness, accuracy, transparency, and some reasonable limits on use.

Median Equivalency Score (Summarized credit statistics)

Experian's **Median Equivalency Score** "assesses the potential risk for seriously derogatory behavior."[68] Experian's material states: "The scores range from 360 to 840 (high score equals low risk) to accommodate the industry standard use of credit scores."[69]

Summarized credit scores or statistics use geography to designate a credit statistic tied to a neighborhood.

> "Summarized Credit Statistics are calculated by aggregating the available consumer credit data within a ZIP+4 geographic area. They do not communicate any individual consumer credit histories, but rather depict the consumer credit activity in a finite neighborhood."[70]

This type of aggregation is typically done by analysis of census data overlaid with copious amounts of non-credit data, like consumer spending data. The summarized credit scores can then be used for marketing purposes broader than what the FCRA allows, because the FCRA applies to an individual's credit scores – not to this type of a neighborhood score.

An exemplar of the use of this summarized credit statistic is Experian's Summarized Credit Statistics mailing list. [71] Experian described the list as follows:

> "Summarized Credit Statistics data is derived from Experian's national consumer credit file and provides consumer credit activity in a neighborhood. The information is calculated by aggregating the available consumer credit data in each ZIP+4™. Choose from more than 300 variables providing valuable information pertaining to tradeline status and specific types of tradelines.
>
> Experian's Median Equivalency Score™ is a Zip+4 level score that helps you identify areas that may be more or less likely to have future derogatory credit activity. The score is statistically derived using payment information, utilization, mortgage, retail and other tradeline information aggregated at the ZIP+4 level."[72]

---

[68] <http://www.experian.com/assets/marketing-services/product-sheets/summarized-credit-stat.pdf>.
[69] Id.
[70] Id.
[71] Experian ConsumerView - Summarized Credit Statistics Mailing List. NextMark List ID #93574. <http://lists.nextmark.com/market?page=order/online/datacard&id=93574>.
[72] Id.

Often, scores of this type are used for lead generation. (Our analysis finds that lead generation is not the same thing as a pre-approved offer of credit. Risk scores might be used to generate leads, but the advertisements themselves are not formal pre-approved offers of credit.)[73]

Additional Experian summarized credit statistic scores for ZIP + 4 neighborhoods include the:

- **National Risk Score**
- **National Equivalency Score**
- **National Bankruptcy Score**

The National Equivalency score is available to a consumer by either requesting an Experian file, or through a service like CreditSesame.[74]

Risk IQ Score

The **Risk IQ Score** from AnalyticsIQ [75] uses summarized credit statistics to predict the "likely credit risk of individuals in a small geographic area."[76] In this sense, it is similar to other risk scores that use summarized credit data that are applied to neighborhoods. The score is built from 100-plus sources, it uses 1,500 factors, and it is updated quarterly.

Most importantly, the Risk IQ score does not apparently use ECOA factors. It states directly in its materials "no protected class demographics are used in the model."[77] This is a welcome statement in a risk model which does not likely fall under the FCRA. (As mentioned frequently in this report, the FCRA applies to individuals, not to neighborhood groupings.)[78]

Consumer Profitability Score

---

[73] See e.g., <http://lists.nextmark.com/market;jsessionid=480F77AC04EECF4E3BA8F18CB50CDFBF?page=order/online/datacard&id=93574> ("Applications: Target candidates for invitations to apply for credit; Use as a predictive variable for acquisition and cross-sell models; Identify loyal prospects in ideal neighborhoods for publishing and continuity programs; Locate neighborhoods with recent and/or heavy credit purchase activity; activity may indicate families in new housing developments and neighborhoods undergoing revitalization where households have diverse product and service needs. Suggested users: Car dealerships/auto marketers; Catalogers and continuity clubs; Insurance providers; Investment planners; Tax services; Travel companies." [Edited for space only. Copy of data card as of publication date available.]

[74] <http://www.creditsesame.com>.

[75] <http://analytics-iq.com>.

[76] <http://analytics-iq.com/download/RiskIQ.pdf>.

[77] Id.

[78] See supra 63 regarding difference between lead generation and formal offers of credit.

This score is designed to predict, identify, and target marketing prospects in households likely to be profitable and pay debt. Experian does not offer this score as a score subject to FCRA limits. Instead of being sourced from credit data, Experian sources this score from its proprietary ConsumerView database, which includes information about 235 million consumers and 117 million households from hundreds of data sources. The score is "rescored" or updated monthly.[79]

While marketers may like these scores for identifying profitable consumers, these kinds of scores can potentially serve as unregulated proxies for credit risk. This score ranges from 1 − 13. Those households with a 1 are described as "High profitability, high likelihood to perform." Those at 13 are designated as "Low profitability, unlikely to perform."

It is important to understand that the **Consumer View Profitability Score** applies to a household. Household scores do not fall under the FCRA's protections, because the FCRA applies to individuals, not households. Households at the lower end of the spectrum here are most affected by the absence of the rights that the FCRA gives to individual consumers. This score does not appear to use traditional credit file factors such as credit scores do. A credit score is limited in how it can be used for marketing. Companies who use a credit score must make a firm offer of credit or insurance. That is not the case with a risk score.

Experian describes the score:

> "The ConsumerView Profitability Score combines a robust scoring model that offers high levels of refinement for selecting top profitability prospects with the best consumer database, ConsumerView, to deliver greater precision in predicting, identifying and targeting prospects at the household level.
>
> The ConsumerView Profitability Score offers 13 levels with three high-profitability levels. This provides clients with additional precision in selecting the best prospects that will respond and comply with the terms of their Invitation- to-Apply lending, credit or continuity program offers."[80]

One of the ways Experian uses the consumer profitability score is in bundled targeted to different sectors. For example, Experian has a Healthcare bundle. This bundle is for healthcare organizations and marketers and includes the ConsumerView Profitability score, along with other Experian scores and a range of consumer demographic data. Here is Experian marketing pitch to healthcare companies:

> "ConsumerView Healthcare: Healthcare organizations and marketers can leverage information about consumer's lifestyles, interests and activities to help them distinguish relationships between various demographic and socioeconomic

[79] Experian Profitability Score (March 2014) <http://www.experian.com/marketing-services/profitability-score.html>.
[80] Id.

groups. This strengthens analyses, bolsters health risk assessments and makes consumer outreach initiatives more effective by identify those who are likely to be responsive to similar interventions, educational programs and communication initiatives (compliance, continuation of prescribed treatments, etc.) Individual characteristics such as age, marital status, education and occupation are provided along with other household insights including income, information about other members of the household, their dwelling type and length of residence among others. Furthermore, Mosaic USA lifestyle segmentation and a compact set of household expenditure propensities give healthcare marketers comprehensive visibility across a variety of dimensions." [81]

Job Security Score

Scorelogix's Job Security Score is an income risk-based score. The company describes the score as follows:

> "Scorelogix is the inventor of the Job Security Score or JSS. The JSS is the industry's first income-risk based credit score and the only score that predicts borrowers' ability to pay by factoring their income stability. The JSS dramatically improves banks' ability to reduce credit losses and marketers' ability to reduce mailing costs."[82]

Scorelogix allows consumers to see their score, and the company includes an analytic report with the score. The score range is from 1- 1000, and the report includes a prediction of unemployment risk, a consumer's relative ranking compared to others, key positive and negative risk factors influencing the score, and some additional information.[83]

Consumer Prominence Indicator Score

Acxiom offers a **Consumer Prominence Indicator Score** that "quantifies the size of a specific consumer's economic footprint, indicating the historical consumer purchasing and relative amount of marketing activity surrounding that individual." The service appears to be aimed at marketers.[84]

---

[81] The Experian ConsumerView Healthcare (March 2014) <http://www.experian.com/small-business/listdetails/consumerview-healthcare.html>.
[82] Scorelogix Home Page (March 2014), <http://www.scorelogix.com>.
[83] JSS landing page, (March 2014), <http://www.scorelogix.com/jss_landing_page02.asp?product_id=1>.
[84]See <http://lists.nextmark.com/market;jsessionid=0EB27D0D2B66AB1A48F0DA8EA9E66BBC?page=order/online/datacard&id=131838> Title: Acxiom – InfoBase Consumer List Mailing List. Consumer Prominence Indicator noted in Selects. See also: *It's All About the Data*, Slide presentation, Sandy Hurst, Acxiom Sales Team Leader, p. 11. <http://webapps.franserv.com/spf11/downloads/8AcxiomData%20Services.pdf>. See also Tom Hill, *The Future of Preferred Underwriting*, Society of Actuaries, Predictive Modeling Topics.

Discretionary Spending Index Score

Equifax offers a **Discretionary Spending Index** (DSI). The DSI is a household scoring system based on the discretionary spending power of consumers. This service appears to be aimed at marketers.

> "Discretionary Spending Index (DSI) is a continuous household-based score of 1 to 1000 that ranks households by likely spending capacity and spending behaviors. It enables marketers to rank customers and prospects by estimated spending power.

DSI can be used alone or incorporated into models where consumer spending is a factor. Marketers could use DSI to enhance account management, identify cross-sell opportunities, and provide appropriate offers."[85]

Invitation to Apply Score

These scores fall into the lead generation space, and as such, generally measure how likely a person is to respond to an offer. These generally do not fall under FCRA regulation. Several companies offer scores that fall generally in this category.

Fair Isaac at one time offered a marketing score (Qualify score) that allowed financial service marketers operating invitation-to-apply campaigns the ability to focus solicitations more narrowly on those more likely to respond. The score no longer appears to be in use. It is mentioned here as an historic exemplar.[86]

Experian offers a product called **Veriscore** that helps to identify customer value potential.[87]

> "Assess the lifetime value of existing and prospective customers.

(August, 2013), <http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=8&ved=0CFsQFjAH&url=http%3A%2F%2Fwww.soa.org%2Ffiles%2Fpd%2F2013-il-pref-under-sem-pred-mod-hill.pdf&ei=9IIvU8KoNs3yqwGTjYCwBA&usg=AFQjCNFbGU5SNzReUTqy3zQMGYxIh50g8Q&bvm=bv.62922401,d.aWM>.

[85] Equifax IXI Services (March 2014). <http://www.ixicorp.com/products-and-services/customer-targeting-and-scoring/discretionary-spending-index-dsi/>.

[86] Fair Isaac Qualify marketing score, <http://www.businesswire.com/news/home/20040621005329/en/Fair-Isaac-Qualify-Score-Helps-Businesses-Reduce#.Uy-K_FyaHxg>. This score appears to be defunct now, we mention it for historic purposes only. The score did not rely on credit data so the FCRA's requirements for pre-screened offers of credit would not apply. For more information on prescreening, see the FTC publication at <http://www.consumer.ftc.gov/articles/0148-prescreened-credit-and-insurance-offers>.

[87] Experian-VeriScore (March 2014), <https://www.experian.com/marketing-services/customer-value-marketing.html?cat1=>.

> VeriScore[SM] predicts response and lifetime value of new customers generated from alternate media sources such as call centers and registration forms. It evaluates their potential for fulfillment, cross-sell, up-sell and optimizes your database. Perfect for industries such as catalog, financial, fundraising, media, retail and telecommunications. Our models can be applied during merge/purge processing to produce a more targeted list without adding time to your production cycle. By allowing you to extract the best prospects from every list, VeriScore helps you find prospects most likely to respond and become loyal customers."

InfoGroup Targeting Solutions, as part of its ITS Consumer Data services, creates a predictive response ranking from 1 to 9 based on consumer transactions and other information.

> "Proprietary predictive response values ranking from 1-9 are based on *known transactions* taking into account recency of purchase, frequency of purchase activity, and dollars spent (RFM) within each market. The higher the RFM score indicates multiple purchases, frequent number of orders tied to dollars spent.[88]

Charitable Donor Score

Donor scores seek to classify and rank those who donate to charities. Donor scoring can be done internally within an organization, or donor scoring can be outsourced. Generally, donor research collects large amounts of information on potential donors from numerous sources including public securities filings, public ethics disclosure forms, probated wills, and other sources, including from non-profits' existing donor lists.

SMR Research offers a **Donor Score**. The donor score is meant to identify the best prospects for large charitable donations.

> "This Score predicts which U.S. households will make the largest contributions to charitable causes. The higher the score, the larger the donation usually will be."[89]

Blackbaud Sphere, offers analytic modeling to non-profits to assist with identifying a variety of donors."[90] The company has robust analytics capacity and provides several models under its Target Analytics ProspectPoint modeling services. For example,

> "The Major Gift model ranks and scores supporters and determines which of these individuals are most likely to make a major gift. It identifies not only which individuals have the capacity to make a major donation, based on overall wealth, income levels, and hidden assets, but also the propensity to give to the organization in significant amounts, as demonstrated by their profile and past behavior. The model is far more accurate than utilizing either capacity or

---

[88] See <http://www.infogrouplistservices.com/b2c-data-solutions/its-consumer-data>.
[89] <http://www.smrresearch.com/Charitable_Donors_Score_&_Lists.pdf>.
[90]<http://internet.blackbaud.com/site/c.duIXLgOXJrIaE/b.8646093/k.99CB/Blackbaud_Sphere.htm>.

> propensity alone and significantly reduces the risk that an organization will waste time and money investing in a nonproductive prospect. "[91]

DonorTrends offers a **DonorScore**.[92] This model works with nonprofits' internal databases to predict donor response:

> "Our scientific DonorScores system assigns a value from 0 to 1,000 to each donor in your database. This value predicts the future actions each donor is likely to take. This enables you to target your donors more effectively to increase revenue and decrease cost.[93]

Household Segmentation Scoring Systems (Personicx, Mosaic, etc.)

Several companies offer targeting systems that match lifestyles, demographics, and spending habits with neighborhoods and households. These classifications are typically based on predictive analytics using varying models.

One of these companies is Claritas, which has a product called **PRIZM** that divides consumers into 66 segments with catchy names, such as Blue Blood Estates, Young Digerati, Gray Power, and Old Milltowns.[94]

A similar system is **Personicx** by Acxiom. Personicx places each US household into one of 70 segments based on that household's specific consumer and demographic characteristics.[95] These targeting systems qualify as consumer scoring systems even though they do not use a score because of the categorization of household by label.

Experian offers a similar product called **Mosaic**.[96]

Collection and Recovery Scores

The use of collection and recovery scores likely falls under the FCRA, but it is unclear whether these categories of scores are actually exposed to consumers.[97]

---

[91] Blackbaud (March 2014), <https://www.blackbaud.com/files/resources/downloads/01.14.ANLY_%20ProspectPoint.datasheet.pdf>. See also <https://www.blackbaud.com/ModelingExplorer>.
[92] <http://donortrends.com>.
[93] <http://donortrends.com/performance-reporting/maap-report/> See also <http://donortrends.com/donorscores/>.
[94] Claritas – MyBestSegments Segments Explorer, <http://www.claritas.com/MyBestSegments/Default.jsp?ID=30>..
[95] <http://acxiom.com/personicx/>. See Also *My Cluster*, <https://isapps.acxiom.com/personicx/personicx.aspx>.
[96] Experian. <http://www.experian.com/marketing-services/consumer-segmentation.html>.
[97] Consumer Finance Protection Bureau, Fair Debt Collection Practices Act and the Dodd-Frank Act (Bulletin 2013-08). There has been much discussion of the impact of debt collection on credit scores. This is beyond the immediate scope of this report. For more information, see <http://files.consumerfinance.gov/f/201307_cfpb_bulletin_collections-consumer-credit.pdf>.

FICO offers scores for delinquent accounts that predict whether the account is likely to pay, and if so, how much will likely be paid over a given time period.[98] The company calls these scores simply **FICO Collection Scores**.

> "Adding more analytics for more precise decisions. FICO® Collection Scores are rapidly deployable analytics that typically boost collection performance by 15–20%. They include early-stage scores for cycle 1 and cycle 2 that rank-order accounts by their probability of rolling, as well as a late-stage score that ranks accounts by expected collection amount. FICO custom analytics include a wide range of predictive modeling (behavior, propensity, strategic default, attrition, etc.), decision modeling and optimization techniques."[99]

Experian has a product called **PriorityScore** for Collections.[100] It is designed to score and segment debt collection accounts.

Churn Scores

Churn scores seek to predict when a customer will move his or her business or account to another merchant (e.g., bank, cell phone, cable TV, etc.). These scores are abundant today. Any company that has historic customer sales data, can use the data to help build its own churn score.[101] Many businesses create churn scores in-house or have an outside analytics company crunch the numbers for them.

Churn scores are interesting in many ways due to the unpredictability of customers.[102] Churn scores are nevertheless well understood, to the point there is at least one patent application for a method of calculating churn.[103] Examples of companies that often have churn scores for customers are wireless telecommunications companies and cable providers, among many other business sectors.

Versium is one analytics company that creates churn scores for businesses.[104] Its churn scores are often custom scores made for specific businesses, and based on custom

---

[98] <http://www.fico.com/en/products/fico-collection-score/>.
[99] FICO, Insights White Papers, Five Imeratives in a Shifting Collections Landscape. Feb. 2013, No. 66. <http://www.fico.com/en/products/fico-collection-score/>.
[100] <http://www.experian.com/consumer-information/debt-collections-strategies.html>.
[101] Emily Parkhurst, *What's Your Churn Score? Big Data Startup raises $2.5M* (Sept. 10, 2013) Puget Sound Business Journal, <http://www.bizjournals.com/seattle/blog/techflash/2013/09/whats-your-churn-score-big-data.html?page=all>.
[102] See, for example, Findable Consulting Blog, *3 Steps to Building Customer Churn Scores*, <http://blog.findable.me/post/52410161427/3-steps-to-building-customer-churn-risk-scores>. See also a discussion of creating the analytics for churn scores at <http://www.analyticbridge.com/forum/topics/issues-with-predicting-churn?commentId=2004291%3AComment%3A72409>.
[103] Eilam, Barak, Lubowich, Yuval, & Lam, Hila, *Method and Apparatus for Predicting Customer Churn*, US Patent Application 20090292583, <http://www.freepatentsonline.com/y2009/0292583.html>.
[104] Versium. <http://versium.com/predictive-

enterprise data.[105] Versium is an exemplar of the trend of smaller analytics companies competing in a space that used to be reserved for much larger companies. Versium uses 300 billion attributes from over 8,000 compiled lists[106] covering up to, for example, 410 million unique emails, 240 million records of demographic data, 90 million mobile phone numbers, 120 million land line numbers, and 1.6 billion records of address history trail, among others. Some parts of this raw consumer data is going into score models for the churn score.

Another company with a churn score is Analytics IQ's **ChurnIQ** score.[107]

Again, many churn scores exist. Due to the sheer number of businesses and analytics companies creating churn scores, the value ranges for the score may vary widely, as can the underlying factors used and the update schedule for the score. It is unlikely that many businesses make churn scores visible to their customers.

Category: Fraud Scores

Fraud scores are an important type of consumer risk score, and they comprise a major category of consumer scores. Fraud scores are prevalent. A large number of fraud scores are available today covering many risk types, with some of these scores in wide deployment. Those who received a phone call from their credit card company after making an unusually high credit card purchase have experienced a predictive fraud score in action. Fraud scores fall outside of the Fair Credit Reporting Act in almost every case. While there may be a fraud score that falls under the FCRA, the research undertaken for this report did not uncover such a score.

As a result, a consumer's fraud score — and most consumers will have multiple fraud scores — will not normally be available for viewing, or for correction or dispute. It is easy to understand the benefit of fraud scores. Companies that build predictive fraud scores have statistics showing how much reduction in fraud the score creates. In some cases, the amount of fraud reduction can be substantial, above 50 percent. There are downsides, though. False positives are highly problematic for the consumers saddled with them. This is particularly challenging for victims of identity theft − either financial, medical − who may have damaged or erroneous fraud scores imputed to them.

Fraud scores are part of the fabric of many businesses today, in particular retailers and financial sector businesses. A consumer with a fraud score that indicates high risk can have difficulty transaction business routinely. Declined credit purchases, declined loans, and declined financial or in some cases health services are among the most common impacts. False positives for fraud scores can vary significantly depending on the scoring

---

scoring/?PHPSESSID=844c277cc5f173b37a1bd0255c33caec>.

[105] WPF Interview with Versium (March 2014).

[106] Versium Data Sheet. <http://versium.com/wp-content/uploads/2014/01/versium.com-data-sheet.pdf> and <http://versium.com/lifedata/>.

[107] Analytics IQ. <http://analytics-iq.com/products/loyaltychurn/>.

model used and the factors fed into the model. Consumers subject to a false positive may find it difficult to clear their records because of the lack of transparency and formal rights.

Data brokers and analytics companies that sell fraud scores are secretive about the score factors and models. Little is known about them among the general public in comparison, to, for example, consumer credit scoring. It is not possible for consumers to approach FICO and request their Falcon Fraud score in the same way it is for FICO customers to request their FICO credit score. Consumers have no rights because the FCRA does not apply here.

This report discussed the problems with opacity of scores, particularly those that have noticeable impact on consumers. At one time, the credit score was secret due to concerns that consumers would game the credit system if they knew their scores. However, as stories grew of the abuse of credit scores, lawmakers took progressively stronger action to ensure consumers could see their credit scores, including structural protections such as constraining use of the scores and providing consumers a right of access. Fraud scores play a significant role in consumers' lives. We need a discussion about the fairness, accuracy, underlying factors used in the scores, and about the non-transparency of the scores.

FICO Falcon Fraud Manager

FICO sells an anti-fraud product that creates a near-real-time fraud score for consumers called **Falcon Fraud Manager**[108], which it describes as follows:

> "Accurately detect fraud on payment card authorization and electronic payment transactions in a fraction of a second. Identify suspicious account holder,cardholder and, optionally, device behavior patterns, generating a score indicating likelihood of fraud."[109]

Falcon stands for Fuzzy Adaptive Logic Control/Decision Network. The Falcon score relies on a neural network, and FICO claims a high score accuracy.[110] Accuracy rates vary, depending on the product and its usage. Fraud scores can hover around an 85 percent or higher accuracy rate. FICO states its false positive rate is around 4 percent.

Of course, claims of accuracy are common for consumer scoring products, but independent verification of the claims are rare. Generally, little is known about the values or ranges for fraud scores. Credit card purchase behavior, device behavior, and other known consumer demographics are likely candidates making up the fraud score at any given moment for an individual.

---

[108] FICO provided WPF an extensive in-person background on this product.
[109] <http://www.fico.com/en/products/fico-falcon-fraud-manager/>.
[110] See for example, FICO page on Falcon Fraud Manager, multiple white papers available. <http://www.fico.com/en/products/fico-falcon-fraud-manager/>. See also Fraud Score Accuracy, <http://versium.com/fraud-score-accuracy/> and FICO.

Other Fraud Scores

Corelogic has a range of **LoanSafe** products that check and score loan applications for fraud, among other mortgage fraud prevention services.[111]

CoreLogic has another product called **ThirdParty Scorecard**. It assesses an agent's loan quality, assigning a risk score to each agent that can be compared against internal, local market and industry performance standards.[112] These scores can indirectly affect consumers, who have no way of know if their brokers or agents are viewed as trustworthy within the industry. A consumer using a broker who has a poor risk score may not be able to obtain a loan or may pay a higher price without knowing the real reason.

Interthinx has a range of mortgage fraud and verification products such as **FraudGuard** and **SafeCheck**.[113]

VISA Risk Manager / Visa Advanced Authorization– cardholder real-time risk scoring**.**[114]

ReD PRISM (Proactive Fraud Risk Management) (neural network). ReD PRISM® is a transaction monitoring and risk management tool for card issuers, merchant acquirers. This tool generates scores and reason codes for transactions.

> "Patented neural network, pattern-recognition software, a fraud detection model and an Active Cardholder History Database that typically holds 30 days of cardholder activity.
>
> The engine generates a score and reason codes for each transaction processed. Scores can be generated in real-time, to be part of an authorization system, and/or in near-real-time for post authorization analysis."[115]

MasterCard fraud scoring solution, a collection of products under the umbrella of **Expert Monitoring Solutions**.[116]

Versium has a fraud score that has reported high accuracy levels of 85% with a false positive rate of 4%.[117]

---

[111] CoreLogic Mortgage Fraud Solutions, <http://www.corelogic.com/solutions/mortgage-fraud-solutions.aspx>.
[112] <http://www.corelogic.com/products/data-repository-solutions.aspx#container-ProductDetails>.See also, *Fifth Third Mortgage Correspondent Seller Guide* (Sept. 28. 2012), <https://www.53.com/files/doc/cl/seller-guide/Seller%20Guides%2010.05.12.pdf>.
[113] <http://www.interthinx.com/solutions/mortgage-fraud-verification-services?pi_ad_id=36143850429&gclid=CMyljPiEvL0CFcqUfgod8xUAPQ>.
[114] <http://www.visa.com/visariskproducts/> and <http://visa.ca/merchant/security/layers-of-security/advanced-authorization.jsp>.
[115] ReDPrism <http://www.redworldwide.com/fraud-prevention/red-prism/>.
[116] <https://www.mastercard.com/us/company/en/whatwedo/products.html>.

There are too many fraud scores to be comprehensive here. A sampling of other types of fraud scores include:

- **Medicare Advantage** is risk scoring for fraud
- **FICO Insurance Fraud Manager**[118]
- **LexisNexis FraudPoint** (applicant fraud prevention)[119]
- **Volusion** credit card fraud score[120]
- **Kount Score** (Prevent fraudulent web purchases)[121]

Category: Custom Scores

Custom scores are those scores uniquely calibrated for a particular business, or that use a particular set of proprietary customer data, or both. The scores can use a combination of internal customer information from the business and information from data brokers and other external sources. Custom scores can be about almost anything to do with customer patterns.

Thanks to the sophistication of today's data brokers, availability of large streams of data, and accurate data analytics, custom scores are in wide use now. Their use may well increase over time as retailers and other businesses seek to increase understanding, segmentation, and targeting of new consumers and existing customer bases. Custom scores are typically closely held, and only limited information is little available on the public record.

The Emergence of Custom Scores and the Pregnancy Predictor Score Example

This report reviews only a portion of available consumer scores. A significant part of consumer scoring is entirely hidden from public view because of the emergence of custom scores. Custom scores can assess almost anything to do with customer patterns. If a business can leverage its own customer data with a custom score, it may have a unique asset.

Perhaps the most famous custom score thus far that became public is Target's **Pregnancy Predictor Score**. This score came to light due to the reporting of Charles Duhigg, who wrote a 2012 article for the New York Times, *How Companies Learn Your Secrets*. In

---

[117] <http://versium.com/fraudscore/>. See also <http://versium.com/resources-roc-curve/>.
[118] <http://www.prnewswire.com/news-releases/fico-unleashes-new-analytics-for-fighting-americas-700-billion-healthcare-fraud-waste-and-abuse-problem-172324041.html>. For general information about health fraud, see <http://jama.jamanetwork.com/article.aspx?articleid=191726>.
[119] <https://www.lexisnexis.com/risk/solutions/fraudpoint-fraud-prevention.aspx>.
[120] <http://onlinebusiness.volusion.com/articles/volusion-launches-fraud-score/>.
[121] <http://www.kount.com/products/complete/kount-fraud-score>.

this article, Duhigg described in detail the Target pregnancy score and how Target developed it.[122] Target's deep databanks of past customer behaviors across its broad customer base was the fodder for the score. Target used predictive analytic models to draw conclusions from the data.

Duhigg described how Target acquired such a robust customer database. Target − at least at the time Duhigg published his article describing the practice − assigned its customers a Guest ID whenever possible, which effectively linked purchases and activities over time. Duhigg also described what is now known to be a common practice among large retailers especially of adding outside data from data brokers to existing customer data. Duhigg describes it this way:

> "Also linked to your Guest ID is demographic information like your age, whether you are married and have kids, which part of town you live in, how long it takes you to drive to the store, your estimated salary, whether you've moved recently, what credit cards you carry in your wallet and what Web sites you visit. Target can buy data about your ethnicity, job history, the magazines you read, if you've ever declared bankruptcy or got divorced, the year you bought (or lost) your house, where you went to college, what kinds of topics you talk about online, whether you prefer certain brands of coffee, paper towels, cereal or applesauce, your political leanings, reading habits, charitable giving and the number of cars you own."[123]

This process, called data appending, reverse data append, or data enhancement, is well-documented as a common practice. For example, a landmark California case, Pineda v. Williams Sonoma, [124] revealed reverse data append activity that is usually non-available to the public eye. The complaint states:

> "Plaintiff visited one of defendant's California stores and selected an item for purchase. She then went to the cashier to pay for the item with her credit card. The cashier asked plaintiff for her ZIP code and, believing she was required to provide the requested information to complete the transaction, plaintiff provided it. The cashier entered plaintiff's ZIP code into the electronic cash register and then completed the transaction. At the end of the transaction, defendant had plaintiff's credit card number, name, and ZIP code recorded in its database.
>
> Defendant subsequently used customized computer software to perform reverse searches from databases that contain millions of names, e-mail addresses, telephone numbers, and street addresses, and that are indexed in a manner resembling a reverse telephone book. The software matched plaintiff's name and

---

[122] Charles Duhigg, *How companies learn your secrets*, New York Times (Feb. 16, 2012) ,. <http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?_r=1&pagewanted=all>.
[123] Id.
[124] See, e.g., *Pineda v. Williams Sonoma Stores, Inc*., No. S178241 (CA Sup. Ct, 2012), <http://caselaw.findlaw.com/ca-supreme-court/1555490.html>.

> ZIP code with plaintiff's previously undisclosed address, giving defendant the information, which it now maintains in its own database. Defendant uses its database to market products to customers and may also sell the information it has compiled to other businesses."[125]

This description was one of the first full explanations of the inner workings of data append to the public.

Using its customer data, Target's predictive formulas combed through customer patterns and identified approximately 25 products that, if purchased together in a certain period of time, suggest a likelihood of pregnancy. Predictive analytics allowed Target to assign a pregnancy score to shoppers who matched the criterion. It then used the scores to send ads to women whom Target predicted might be pregnant. The result reported by Duhigg was of a father storming into Target angrily with a handful of baby-related advertisements from the retailer stating that his daughter was not pregnant. Later, he discovered that his high-school aged daughter – unbeknownst to him – was indeed pregnant.

We know because of Duhigg's article about just one Target custom score. We do not know how many other customer scores Target uses. We do not know what other retailers use custom scores. If trends are any indication, the rapid upswing in analytics companies offering custom scores and custom analytics as well as access to massive, personally identifiable, consumer data sets hints that use of custom scores may at some point overtake off-the-shelf prefabricated scores. Custom scoring makes it even harder for consumers to learn how merchants use their data.

As stated throughout this report, some scores are less troublesome to consumers. But many may be problematic to a greater or lesser degree, and consumers do not know what they do not know about consumer scoring in general and about custom scoring in particular. We all deserve to live in a world where we have the opportunity to know how we are being sorted and sifted and to have transparency and fairness in any ranking process.

Category: Regulated Credit and Financial Scores

Traditional credit scores are regulated. In our taxonomy, a credit score is a type of consumer score. There are hundreds of credit scores available on the market today. Any credit grantor can develop and use its own general or specifically focused scoring system. Although consumers have rights to see credit scores, they may not know about the existence or use of those systems, especially for scores developed internally by a credit grantor and not purchased from external vendors.

Any businesses can develop a credit scoring system and can sell consumers the ability to see their scores, whether or not anyone actually uses the credit scores to make real world decisions about consumers. A household that wants to see multiple credit scores for

[125] Id.

household members could spend a considerable sum buying scores. Consumers may not know which scores are actually used and which are not. This has the potential to be lucrative for the business but potentially expensive for consumers, and this is the reason consumers need to be cautious about which credit scores they are purchasing.

The use of credit scores outside the immediate credit granting process is hard to assess. There is one report that an electric utility in Texas wanted to use credit scores to set electricity prices to some residential users. The Texas Office of Public Utility Counsel objected and the plan apparently never went into effect.[126] See the discussion about insurance scores.

FICO Score

A well-known credit score is the FICO Score.[127] There are numerous credit scores, but FICO appears to lead in terms of sales.[128] Fair Isaac states:

> "The FICO® Score is the most widely used credit score in North America. Lenders purchased more than 10 billion FICO Scores in 2013, and 90 percent of all U.S. consumer lending decisions use the FICO Score. The 25 largest credit card issuers, the 25 largest auto lenders and tens of thousands of other businesses rely on the FICO Score for consumer credit risk analysis and federal regulatory compliance."

The FICO score falls squarely under the Fair Credit Reporting Act.

FICO also has an **Expansion Score** that draws on alternative credit data such as bank account records, payday loan payment records, and installment purchase plans, to produce a credit score that may not be based solely on the consents of a credit report.[129] Other companies also offer similar credit scores for people with "thin" credit reports.[130]

Vantage Score

The three national credit bureaus have jointly developed the VANTAGE score, which is also a popular credit score.[131] The Vantage score is regulated under the Fair Credit Reporting Act.

---

[126] See <http://www.liheap.ncat.org/news/Sept04/Texas.htm>. Title: Texas OPC Says TXU Credit Scoring Puts Poor at Risk.
[127] <http://www.myfico.com>.
[128] Press Release, FICO, *New Version of the Industry standard FICO Score Will Be Available Beginning This Summer*, <http://www.fico.com/en/about-us/newsroom/news-releases/new-version-industry-standard-fico-score-will-available-beginning-summer/>.
[129]<http://www.fairisaac.com/Fairisaac/Solutions/FICO+Expansion+Score/Expansion+Score+Overview/FICO+Expansion+Score.htm>.
[130] <http://biz.yahoo.com/brn/060908/19082.html>.
[131] <http://www.vantagescore.com>.

Beacon Score

A variant on the general credit score is Equifax's **BEACON** service, which seeks to predict the likelihood that a new or existing account will become delinquent within 24 months.[132] The BEACON score is regulated under the FCRA.

When low Beacon scores are used to market to consumers

It is possible that an individual's low Beacon score can be used indirectly for marketing purposes through circumstances the authors of the FCRA did not predict. For example, consider an individual whose Beacon scores was too low to qualify for a cell phone contract. That individual could end up on a data broker list of "Cell Phone Turndowns." WPF first recorded this list in 2007,[133] and has confirmed the existence of a similar list in 2014.

The 2014 Cell Phone Turndowns list read in part:

> "This file is comprised of individuals who attempted to set up a contract with a cell phone provider but did not meet the required Beacon score requirements. These consumers are ready and eager to receive offers and opportunities in the following categories: secured and sub-prime credit, Internet, legal and financial service, health insurance offers, home equity loans, money making opportunities, and pre-approved credit with a catalog purchase."[134]

Even if a score begins life as subject to FCRA limits, it is easy for industry to track a consumer's activities and extract those that reflect a score at a particular level. This is what happened with the cellphone turn down list. Those consumers clearly have low credit scores. The resulting list of those denied phone would not fall under any regulation, and it could be used as a proxy for a regulated credit score.

Small Business Intelliscore

The focus of this report is on scores for individuals. There are also numerous scores for businesses available for purchase. It is worth mentioning at least one exemplar of this type of score. Experian offers a product called Small Business Intelliscore.[135] It uses commercial and consumer credit data to generate a risk score. This is an example of a scoring system that, for small businesses, fuzzes the line between consumer and commercial activities.

---

[132] <https://www.eport.equifax.com/eport/eport_beacon.htm>.
[133] *Cell Phone Turndowns*, DirectListFinder2.0, NextMark ID 188161, <http://listfinder.directmag.com>, (June 16, 2007). PDF copy of the list available from WPF.
[134] *Cell Phone Turndowns*, NextMark ID188161, <http://lists.nextmark.com/market?page=order/online/datacard&id=188161> (March 13, 2014).
[135] <http://www.experianbizinsight.com/data_enhancement/intelliscore.shtml>.

Tenant Scores

Numerous tenant scores are available. These scores give a history of evictions, and use credit bureau and other data.[136] CoreLogic is one of the companies offering a tenant score, theirs is called **CoreLogic SafeRent**.[137] Tenant scores fall under FCRA regulation.[138]

Category: Identity and Authentication Scores

Identity and authentication scores are in widespread use. These are a form of modern eligibility scores, although current regulations do not view these scores as regulated eligibility scores.

Consumers encounter these scores, for example, when logging on to an online patient portal hospital system or to a financial institution for online banking. Some forms of ID or authentication scoring is invisible to the consumer. Entire businesses specialize in authentication. These kinds of scores are a subcategory of consumer risk scores. It is unusual for a consumer to find or see their ID or authentication score, but there are exceptions. The authentication of online users is a reasonable activity and a protection for both the consumer and the business. The problem arises when an innocent consumer – perhaps the victim of identity theft – loses access to accounts or is denied service and cannot readily learn the reason or correct the problem.

ID Analytics ID Score

ID Analytics is one of several companies that offer ID scores. Its **ID Score** "calculates the risk associated with an identity, allowing businesses to focus Identity Risk Management efforts on identities with the highest likelihood of fraud without alienating legitimate customers."[139] ID scores are in widespread use, and can be troublesome to consumers when something is amiss. For example, if a person fails an ID score, it is an indication that they may either be an identity thief, or be an identity theft victim. At times, a low ID score will prevent a person from purchasing items like a cell phone, and can be troublesome in other verification situations.

ID Analytics is one of the few companies that allow consumers to view their ID score at no cost to the consumer.[140] This has been in place for about six years. Consumers cannot

---

[136] See for example <http://myrental.com/reports/tenant-score/>.
[137] <http://myrental.com/aboutus/>.
[138] See *FTC Warns Data Brokers That Provide Tenant Rental Histories They May Be Subject to Fair Credit Reporting Act* (April 3, 2013), <http://www.ftc.gov/news-events/press-releases/2013/04/ftc-warns-data-brokers-provide-tenant-rental-histories-they-may>. See also Privacy Rights Clearinghouse on Renter's Rights for more on Tenant Screening and the FCRA. <https://www.privacyrights.org/renters-guide-to-privacy-what-to-know>.
[139] <http://www.idanalytics.com/solutions/score.html>.
[140] <http://myidscore.2advanced.com>.

learn the factors that go into the score nor dispute the score, but it is nevertheless a step in the right direction that consumers can see their ID score.

Fair Isaac has an ID product called **FICO Identity Resolution Manager**.[141]

Insurance Scores

Insurance scores, sometimes called credit-based insurance scores, fall under the FCRA. The scores use credit scores and credit information to analyze prices for automobile insurance and homeowners insurance. Insurance scores differ from credit scores, and it appears that insurance companies may have their own algorithms. More information about the use of credit scoring in insurance is available from the National Association of Insurance Commissioners[142] and from Consumer Reports.[143]

ChoiceTrust, a **ChoicePoint** company, offers to sell home and auto insurance scores to individuals.[144] **FICO** also has an insurance score product. A growing number of insurance carriers use custom scores that have been developed to meet that company's specific underwriting criteria."[145]

It is unknown whether or how insurance companies use the insurance industry property claim databases (generally referred to as CLUE or Comprehensive Loss Underwriting Exchange) for scoring.[146] Consumers can request a copy of their CLUE reports under the FCRA.[147]

Category -- Health Scores

Initial research for scoring done for this report in 2007 found few health scores. In 2014, research uncovered significant and high-impact consumer health scores in use. Health scores are now in full circulation with little consumer awareness. The same questions raised above about transparency, secrecy, factors, and use are relevant here. Other

---

[141] <http://www.fico.com/en/products/fico-identity-resolution-manager/>.
[142] <http://www.naic.org/cipr_topics/topic_credit_based_insurance_score.htm>.
[143] <http://www.consumerreports.org/cro/personal-finance/car-insurance-8-06/overview/0608_car-insurance_ov.htm>. See also <http://www.insurancescored.com>.
[144] <http://www.choicetrust.com/servlet/com.kx.cs.servlets.CsServlet?channel=welcome&subchannel=insscore#>.
[145] <https://choicetrust-solutions.custhelp.com/cgi-bin/choicetrust_solutions.cfg/php/enduser/std_adp.php?p_faqid=617&p_created=1050502344&p_sid=fYAN1Nri&p_accessibility=0&p_lva=&p_sp=cF9zcmNoPTEmcF9zb3J0X2J5PSZwX2dyaWRzb3J0PSZwX3Jvd19jbnQ9MTEmcF9wcm9kcz0mcF9jYXRzPTE3NiZwX3B2PSZwX2N2PTEuMTc2JnBfcGFnZT0x&p_li=&p_topview=1>.
[146] More information about CLUE is available from the Privacy Rights Clearinghouse <http://www.privacyrights.org/fs/fs26-CLUE.htm>.
[147] <https://personalreports.lexisnexis.com/fact_act_claims_bundle/landing.jsp>.

questions come into play as well. For example: can employers purchase health scores? Are health scores ever shared with debt collectors?

New health scoring systems that fall in the category of consumer scores will be developed and used in the near future. It is also possible to foresee the development of family and neighborhood health scores based either a combination of traditional medical histories, genetic data, census data, data broker lists, environmental data, or histories of actual health treatments that may fall outside of HIPAA.

Health records held by health care providers or insurers are subject to the federal health privacy rules known as HIPAA.[148] While these records are available for many non-consensual uses, the information in the records should not normally be available to data brokers and score creators. However, the HIPAA rules do not cover health information held by gyms, websites, banks, credit card companies. many health researchers, cosmetic medicine services, transit companies, fitness clubs, home testing laboratories, massage therapists, nutritional counselors, alternative medicine practitioners, disease advocacy groups, or marketers of non-prescription health products and foods. This vast class of largely unregulated health information is available as input to a health scoring algorithm. Further, consumers routinely disclose health information to companies that promise to provide coupons. Consumers rarely understand that companies can collect personal information that they can later sell.

Personal health records (PHR) maintained by companies outside HIPAA protections may also become a source of unregulated health information for scoring.[149] Information disclosed through web searches or Internet browsing also typically remains unregulated by HIPAA, and all of the information can be fodder for scores.

Affordable Care Act Individual Health Risk Score

Each individual in a health plan subject to risk adjustment under the Affordable Care Act (ACA) will be assigned a health risk score. This is a new score, and it is an important score especially because it is part of a federal program. In establishing the rules for health risk scores for individuals, the Health and Human Services Department effectively created a score that ultimately measures how sick a person is. The stated goal of the risk score is to create a relative measure of predicted health care costs for a particular enrollee. The scores are supposed to be phased out over the next four years.

The rules for the individual health risk score became official in March 2012, when the Department of Health and Human Services issued a final rule on reinsurance and risk adjustment under the Affordable Care Act.[150] The overall purpose of risk adjustment is to mitigate the impact of potential adverse selection and stabilize premiums in the individual

---

[148] Health Insurance Portability and Accountability Act, 45 C.F.R. Parts 160, 162, & 164.
[149] See World Privacy Forum, Personal Health Records: Why Many PHRs Threaten Privacy (2008), <http://www.worldprivacyforum.org/2008/02/blog-legal-and-policy-analysis-personal-health-records-why-many-phrs-threaten-privacy/>.
[150] <https://www.federalregister.gov/articles/2012/03/23/2012-6594/patient-protection-and-affordable-care-act-standards-related-to-reinsurance-risk-corridors-and-risk>.

and small group markets under the Affordable Insurance Exchanges that are part of the Affordable Care Act.

A key element of the risk adjustment is the calculation of a plan's average actuarial risk so that the plan's average risk can be compared to other plans. The scores will be important because they will determine whether a plan pays or receives funds through the premium adjustment system. A plan might have an incentive to assign its insured a higher score. The use or disclosure of that score for another purpose could harm an individual. Even disclosure of an honest score could be harmful. This is a new area and a new score, and there is much uncertainty about the use or misuse of the score.

A plan's average risk is based on the risk score of each enrollee in that plan. An individual's health risk score will be a measure of how much that individual is likely to cost the health plan. The risk score measures likely health costs and is, in a very general way, a proxy for how sick an individual is. How expensive an individual will be to insure is important to insurers and employers, and the score can easily be misused.

The HHS rule took some care to protect the privacy and security of an individual's risk score, including limits on the disclosure of identifiable elements when individual risk scores are passed on by a plan for use in State risk adjustment programs. Nevertheless, each individual in plans subject to risk adjustment will have his or her own health risk score.

The regulation is silent about individuals seeing their health risk score. If an insurer has a risk score for an individual, then it appears that it would be *Protected Health Information* as defined in the privacy rules issued under the rules of the Health Insurance Portability and Accountability Act (HIPAA). If that conclusion is correct, the score should be available to individuals under standard HIPAA rules. It is possible to foresee that an employer or lender or someone else with power over an individual might coerce the individual into obtaining his or her score and disclosing it.

FICO Medication Adherence Score (MAS)

Launched on June 23, 2011 by analytics firm Fair Isaac Corp., this score identifies a patient's propensity to adhere to a medication prescription plan during the next 12 months. It is a predictive score designed to let pharmacies and insurers know when or if a patient is at risk and needs a medication reminder.[151] The score pulls from public data and from patients' prescription histories when available. The score ranges from 1-500, with a score above 400 indicating that a patient is likely to take medications as prescribed. Patients who score 200 or below may get a reminder, as a low score predicts non-adherence. The company created the scoring algorithm from a randomized sample set of

---

[151] Jeremy M. Simon, *New medical FICO score sparks controversy, questions*, Creditcards.com (July 28, 2011), <http://www.creditcards.com/credit-card-news/fico-score-medication-adherence-1270.php>. See also Tara Parker Pope, *Keeping score on how you take you Medicine* (June 20, 2011). NYT Blog Well, <http://well.blogs.nytimes.com/2011/06/20/keeping-score-on-how-you-take-your-medicine/?_php=true&_type=blogs&_r=0>.

Pharmacy Benefit Manager patients numbering in the millions. (More on what a PBM is here).[152]

By the end of 2011, FICO scored 2 to 3 million patients, with an additional 10 million expected during 2012. New numbers were not available for later years.

Factors in the score include:

- Employment
- Homeownership
- Living situations
- Age
- Gender
- Family size
- Asset information (ex., likelihood of car ownership).

FICO states its analytics identify medication adherence risk by "using data from a range of publicly available third-party data sources. Because the FICO **Medication Adherence Score** requires minimal information from the patient, and no prescription claims or sensitive health information, the score can be generated for members of any patient population."[153] FICO states that with just a name and home address it can "pull the remainder of the necessary information from publicly available sources."[154] As of 2014 FICO also states "The Medication Adherence Score will use a patient's prescription claims history when available and pull on other publicly available third-party data sources when no other information is present."[155]

These differences in what factors are used to create the score make a difference. If FICO calculates the score without any health information obtained from a covered entity regulated under the HIPAA federal health privacy rules,[156] then the information is not regulated as health information under HIPAA. This illustrates a limitation of the HIPAA privacy rules that allow information about patients to be used and disclosed, bought and sold, by data brokers and others without application of any health privacy rules.

If, however, FICO calculates a score about an individual based on any information – even just the individual's name – obtained from a business associate under HIPAA rules, then it would appear that FICO uses protected health information under HIPAA and it would have to be a business associate of the provider or insurer that disclosed the information to FICO. This would bring this score under the HIPAA regulations.

---

[152] For more on PBMs and HIPAA see Gellman and Dixon, *A Patient's Guide to HIPAA*, Basic Rights. World Privacy Forum, Sept. 2013. <http://www.worldprivacyforum.org/2013/09/hipaaguide40/>.
[153] FICO, Press release, *New FICO Analytics Predict Likelihood of Patient Adherence to Prescription Medication*,(June 23, 2011). <http://www.fico.com/en/about-us/newsroom/news-releases/new-fico-analytics-predict-likelihood-of-patient-adherence-to-prescription-medication/>.
[154] <http://www.fico.com/en/products/fico-medication-adherence-score/>.
[155] Id.
[156] 45 C.F.R. Parts 160, 162, 164.

Because of the two different scenarios that are possible here, it is impossible to tell from the outside just what FICO does with the score. It seems possible that FICO and HIPAA-covered entities could potentially organize the sharing of information to evade HIPAA's requirements. For example, FICO could take its entire list of individual scores, give that list to a HIPAA covered entity and allow the covered entity to select information about patients of the covered entity. That would result in no sharing of HIPAA-protected health information with FICO.

Possible customers for FICO's Medical Adherence Score are pharmaceutical manufacturers. Unlike health plans, drug manufacturers do not have any direct way of learning who is taking the drugs that they manufacturer. They need the assistance of intermediaries, like pharmacies and pharmacy benefit managers, to send prescription reminders. HIPAA allows these reminders.

Drug manufacturers fund many if not most prescription reminder programs. To send a reminder, the manufacturer pays a HIPAA-covered entity – most likely a pharmacy or pharmacy benefit manager – to contact the patient lawfully. The manufacturer must pay for the cost of the notice to the patient and provide an incentive to the intermediary. The full cost might be a few dollars per notice. If the manufacturer can identify those patients who are likely to refill prescriptions anyway, it can tell the intermediaries to send reminders only to those who have a low adherence score. The effect is to pay less to FICO and avoid paying a larger amount for a notice. We do not know if this reflects how the scores are actually used.

FICO states that patients can ask their health care providers if they have a score. For patients with a MAS score, FICO directs them to ask their health care providers about their opt out policies. Under HIPAA, patients should be able to request this score, as it should be Protected Health Information and subject to HIPAA transparency rules if a HIPAA covered entity maintains the score. However, an opt-out from a third party health score is uncharted territory for HIPAA. WPF's Patient's Guide to HIPAA has a section detailing how to request health records under HIPAA.[157] It is not always an easy or a simple process, and it can require a great deal of persistence just to find the right provider who has the information. It is not clear to us how providers might treat a MAS score request, and it is unknown if any would honor a request for an opt out. In short, it is somewhat disingenuous for FICO to direct patients to the HIPAA process when it is FICO that maintains a patient's MAS score.

Frailty Scores: General

Frailty Scores usually apply to the elderly. A good bit of research has been conducted using this score as a measure. As a result, a frailty score has become much more important in recent years. Research found that some frailty scores could predict mortality

[157] Gellman and Dixon , *A Patients Guide to HIPAA*, Right to Inspect and Copy Your Record, FAQs 18-24. <http://www.worldprivacyforum.org/2013/09/hipaaguidepart2/>.

within one year.[158] Separate research indicated some frailty scores can usefully predict the likelihood of patient post-operative surgical complications or readmission to a hospital. While the scores can predict care needs, the scores can also be used to simply project costs, and this raises questions about possible misuse in non-health scores or marketing activities. Unless a HIPAA-covered entity calculates a frailty score using health records, the score is not likely covered by the HIPAA health privacy rules.

CMS Frailty Adjustment Score

The Centers for Medicaid developed a frailty score in the late 1990s. In 2004, after refinement, the CMS frailty measure was extended to more Medicare managed care organizations.[159]CMS is a HIPAA-covered entity so the score should be subject to the HIPAA health privacy rules. After CMS developed its score, several other models of frailty scores developed.

Hopkins Frailty Score

Johns Hopkins University developed the **Hopkins Frailty Score**. Designed for use before surgery, the score would be calculated by a health care provider and would be subject to HIPAA. This predictive score in its original form has low factors compared to other scores and a small range. The factors are highly predictive, however, and this score is in widespread use.[160]

It is unknown how many patients are assigned frailty scores, and it is unknown how many patients ever request their scores. Conceivably, a score held by a health care provider should be covered under HIPAA and patients should be able to request their score if one is there.

The concern with any predictive score, particularly a frailty score, is that it can escape into the hands of third parties where it can be used outside of the original intent of the score. The frailty score can be highly predictive, and therefore its use needs to be carefully guarded.

Other Health Scores

---

[158]Dave Levitan. *Frailty Score Predicts 1-Year Mortality But Not Procedural Complications in TAVR* (Sep. 17, 2012), <http://www.tctmd.com/show.aspx?id=113395>.

[159] John Kautterand Gregory C. Pope, *CMS Frailty Adjustment Model*. 26 M.S. Health Care Financing Review (2004-2005). <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/HealthCareFinancingReview/downloads/04-05winterpg1.pdf>.

[160] See Johns Hopkins Medicine, L*evel of Frailty Predicts Surgical Outcomes in Older Patients, Johns Hopkins Researchers Find* (May 12, 2010), <http://www.hopkinsmedicine.org/news/media/releases/level_of_frailty_predicts_surgical_outcomes_in_older_patients_johns_hopkins_researchers_find>.

A medical data breach revealed in 2011 that a company called Accretive collected detailed and sensitive health information about hospital patients in Minnesota via contract with those hospitals and then used that data to develop scores. The information included:

- Patient's full name
- Gender
- Number of dependents
- Date of birth
- Social Security number
- Clinic and doctor
- A numeric score to predict the "complexity" of the patient
- A numeric score to predict the probability of an inpatient hospital stay
- The dollar amount "allowed" to the provider
- Whether the patient is in "frail condition"
- Number of "chronic conditions" the patient has

Patients had no knowledge of the use of the information for scoring:

> "Upon information and belief, the hospitals' patient admission and medical authorization forms do not identify Accretive by name or disclose the scope and breadth of information that is shared with it. Upon information and belief, patients are not aware that Accretive is developing analytical scores to rate the complexity of their medical condition, the likelihood they will be admitted to a hospital, their "frailty," or the likelihood that they will be able to pay for services, among other things.[161]"

The Minnesota Attorney General found that company promoted its health data activities to investors as:

- Risk scoring of patients
- Has an "intense focus" on "reducing avoidable hospital admissions"
- Identifies the "sickest and most impactable patients" for "proactive management"
- Identifies "real-time interventions with significant revenue or cost impact"[162]

The full details of Accretive's activities are beyond the scope of this report because they involved matters beyond scoring. However, when the lawsuit ended, Accretive agreed to a ban from doing business in Minnesota for a period of time.[163]

Personal Health Scores: WebMD, others

---

[161] Complaint, *State of Minnesota vs. Accetive Health, Inc* (USDC Minn 2012).
[162] Minnesota Attorney General, Press Release, Attorney General Swanson Sues Accretive Health For Patient Privacy Violations (Jan. 19, 2012), <http://www.ag.state.mn.us/Consumer/PressRelease/120119AccretiveHealth.asp>.
[163] Star Tribute.com, *Accretive is banned from Minnesota* (July 31, 2012), <http://www.startribune.com/lifestyle/health/164313776.html>.

A personal health score is a growing category of scores that, at the moment, are relatively casual and aimed primarily at enhancing a consumer's self-understanding. These scores do not carry the same underwriting weight as for example, a large sample-set based, formal statistical score would. The scores generally appear to be largely educational in nature, and voluntary.

Under the Affordable Care Act, wellness programs and health improvement are priorities. It is no surprise, then, that new health self-scoring activities for patient self- monitoring are coming online. These scores, by their nature, are typically generated by an online survey taken by the patient, with the resulting scores available to the patient. Although many variations exist of these sorts of more casual health scores, at this point most of the scores do not appear to be tied to benefit costs.

WebMD is a good exemplar of this kind of "personal" health score.[164]

One Health Score is another exemplar.[165] This score allows a consumer to rate their physical activity and its benefits. This score has a range of 1-100, with a score of 60 and above indicating that the person being scored has a basic level of physical activity. Scores of 90 and over are generally attained by professional athletes.

These scores are likely not subject to HIPAA protections. If the scores derive from information supplied by a consumer, then they are not protected health information under HIPAA unless a HIPAA covered entity calculated the score. A commercial website offering services to consumers is normally not a HIPAA covered entity. Even if the website maintains health records for the consumer and with the consent of the consumer, use of the records is subject to the privacy policy and terms of service of the website. Most consumers would likely not understand that health information held on their behalf by commercial websites has no legal privacy protections. How the use of these health scores will evolve and whether they will "escape" into the hands of marketers and data brokers is not known.

Resource Utilization Group Scores

Medicaid uses resource utilization groups (RUG) to classify residents in nursing homes based on the relative resources that an individual is likely to use. Medicare pays for Part A skilled nursing facility stays based on a prospective payment system that categorizes each resident into a payment group (RUG) depending upon his or her care and resource needs. Skilled nursing facilities determine a RUG based on 108 items on an assessment of the resident known as the Minimum Data Set (MDS).[166] Calculated by a HIPAA covered entity (the nursing home), the score remains subject to HIPAA privacy rules.

---

[164] WebMD, *Health Manager, Personal Health Score* <http://www.webmd.com/health-manager>.
[165] One Health Score, <http://www.onehealthscore.com/faq>.
[166] See HHS Inspector General, *Review of Nursing Facility Staffing Requirements at Beverly Healthcare of Reading* (2005), <http://www.oig.hhs.gov/oas/reports/region3/30400214.htm>.

SF-36 Form

A standard health industry/research form is the SF-36 (the SF-12 is a shorter version). The SF-36 is a multi-purpose health survey that produces an 8-scale profile of functional health and well-being scores as well as physical and mental health summary measures and a health utility index. The SF-36 is used for surveys of general and specific populations, to compare the relative burden of diseases, and to differentiate the health benefits produced by different treatments. The data subject completes the survey.[167] Whether HIPAA protections apply to the information on the survey depends on who collects the information. If a HIPAA covered entity (health care provider or insurer) collects the survey, the information falls under HIPAA. If a patient completes the survey for a health researcher, it may or may not fall under HIPAA. Many researchers are not covered entities under HIPAA, and many fall under no privacy regulations at all.

This type of instrument could produce a consumer score depending on who uses it and the purpose. If used for general actuarial purposes by a health plan, the results would probably not qualify as a consumer score. If used by a medical device merchant to target likely wheelchair purchasers, it would qualify as a consumer score. If used in a research project, the SF-36 would not result in a consumer score.

Complexity Scores

Complexity scores are beginning to spring up for various patient types and situations (See Frailty score). Grants have been set aside for the development of new complexity scores, for example, work to create a Complexity Score to Identify Hospitalized Patients at High Risk for Preventable Adverse Drug Events was funded in 2013.[168] It is likely that complexity scores will be developed for many patients' situations. A complexity score used for treatment fall under HIPAA and does not qualify as a consumer score. A complexity score used for marketing or to set rates may be a consumer score.

An exemplar complexity score is the Aristotle Complexity Score. This score was developed over the course of five years by the Aristotle Institute. Used in its original context, this score is **not** a consumer score because it is for diagnostic use.

> A group of 50 internationally accepted experts has been working for more than five years on a new method to evaluate the quality of care in Congenital Heart Surgery (CHS) that is called Aristotle. Senior, experienced congenital heart surgeons considered the possible risk factors for each procedure and assigned scores based on potential for mortality, potential for morbidity, and anticipated surgical difficulty.

---

[167] See <http://www.sf-36.org/tools/sf36.shtml>.

[168] See <http://www.ashpfoundation.org/MainMenuCategories/PracticeTools/Drug-Therapy-Management-Complexity-Score-Index> and <http://www.ashpfoundation.org/PR2013ComplexityScore>.

> The Aristotle system, electronically available, has been introduced by both the European Association for Cardio-Thoracic Surgery (EACTS) and Society of Thoracic Surgeons (STS) as an original method to compare the performance of Congenital Heart Surgery (CHS) centers. Pediatric cardiologists have joined the project and are currently developing a complexity score for interventional cardiology procedures.[169]

The Aristotle score, allows precise scoring of the complexity for 145 CHS procedures.[170] Again, as with the complexity score, if used in a clinical setting, these scores should fall under HIPAA and should be viewable by patients. Also as with frailty scores, complexity scores could be subject to abuse if layered into scores outside the health care context.

Category – Smart Grid and Energy Scores

The Internet of Things and the Smart Grid stands to generate a significant number of consumer scores in the very near future. At least one examplar score is already in use, with many others set to come into use soon.

Peer-to-Peer Energy People Meter Score (EPM)

This score measures a residential customer's energy consumption and seeks to engage the customers to evaluate their own energy consumption patterns. Consumer scores arising from Smart Grid[171] or Internet of Things[172] usage is an emerging field. These scores are of great interest due to the approaching tsunami of information that connected devices in and out of the home, including cars, will provide. The **EPM score** is a proprietary score from Trove Data. The company has a range of analytics in the area of energy, not all of which qualify as consumer scores. The Energy People Meter score is of specific interest here. Trove Data describes it as follows:

> "Trying to engage your customers? How can you score their actions as a customer? TROVE's Peer-to-Peer Scoring application and proprietary Energy People Meter (EPM) captures the attention of the customer and inspires them to evaluate their energy usage. By providing individual recommendations for each customer, this tool allows utilities to increase customer service and engage customers in new ways."[173]

---

[169] <http://aristotleinstitute.org/aboutScore.asp>.
[170] The Aristotle score: a complexity-adjusted method to evaluate surgical results. <https://www.ncbi.nlm.nih.gov/pubmed/15144988>.
[171] See Comments of EFF and CDT to the Department of Energy on Implementing the FIPS in the Smart Grid,(Nov. 1, 2010), <https://www.eff.org/files/DOE%20Comments-Nov1.pdf>.
[172] See Comments of EPIC to the FTC on the Privacy and Security Implications of the Internet of Things, (June 1, 2013), <https://epic.org/privacy/ftc/EPIC-FTC-IoT-Cmts.pdf.>
[173] Trove Data Utility Applications, <http://www.trovedata.com/solutions/utility-applications/>.

Trove Data lists its data sources used to compile its scores in a FAQ section. Its list includes meter data, satellite imagery, and appliance data. Its FAQ states:

"We aggregate and fuse thousands of data of attributes from hundreds of sources including the following:

- Meter data
- Demographic data
- GIS
- Distributed energy data
- Satellite imagery
- Financial data
- Wireless/Mobile data
- Appliance data
- Social behavior data
- Other disparate data sources that can provide deep insight into energy behavior patterns."[174]

While it is likely that other companies provide competing analytics scores in this area, Trove Data is an intriguing case study as a data aggregator because it was among the first to work with Smart Grid information.[175] A benefit of the Trove EPM score is in its environmental potential, allowing energy companies to fine-tune power supply and usage.

The initial EPM score is not exposed to the consumer, however, the company de-identifies consumer data and aggregates the information. The company has taken numerous steps to remove personally identifiable information from the data the utilities companies receive.[176] **Versium** has an energy score it is planning to introduce as well.

The issue of Smart Grid and other Internet of Things data flows and how to handle the privacy implications of these devices and the data they shed over the long term is complex. The Federal Trade Commission held a conference on this topic, which generated robust conversations on how privacy protections might be configured in Smart Grid scenarios.[177] FTC Chairman Edith Ramirez recommended privacy by design, simplified consumer choice, and transparency as important privacy starting points.[178] This is an area of consumer scoring where much can be done right now to be proactive in

---

[174] Trove Data FAQ, *What data sources does Trove use?* (March 15, 2014), <http://www.trovedata.com/solutions/faq/>.
[175] See an early article about the company, Katie Fehrenbacher, *GridGlo mines big data for real time energy apps*, GigaOm (May 10, 2011), <http://gigaom.com/2011/05/10/gridglo-mines-big-data-for-real-time-energy-apps/>.
[176] Interview, Trove Data, (March 2014).
[177] Federal Trade Commission, *The Internet of Things: Privacy and Security in a Connected World* (Nov. 19, 2013) (Conference, webcast, transcripts, and public comments). <http://www.ftc.gov/news-events/events-calendar/2013/11/internet-things-privacy-security-connected-world>.
[178] At 9-10, <http://www.ftc.gov/sites/default/files/documents/public_events/internet-things-privacy-security-connected-world/final_transcript.pdf>.

terms of mitigating any potential consumer downsides so as to reap the benefits of the data analysis.

Category - Social Scoring

Much has been made of social scoring over the past few years. Entire companies, web sites, books, college courses focus intensively social scoring as social scores themselves continue to proliferate. Social scores seek to measure influence, or Social Networking Potential (SNP) The scores derive from a specific type of algorithm called a "social algorithm."[179] Michael Schafer writes in his book *Return on Influence*:

> "This trend of social scoring is creating new classes of haves and have-notes, social media elites and losers, frenzied attempts to crash the upper class, and deepening resentments."[180]

Some social scores enjoy a transparency that most of the scores mentioned in this report do not. **Klout**,[181] **Booshaka**,[182] **Kred**,[183] **PeerIndex**,[184] **PROskore**,[185] **SocialIQ**,[186] **Tweet Grader,**[187] and **Twitalyzer**[188] are among those providing social media analytics and metrics directly to the public. Among these, Klout is currently the most important and is in mainstream use. People can readily see their own − and others' − Klout scores, which has created a complex dynamic. People can opt out of having a Klout score if they dare, but in some professions, not having a Klout score would be a professional liability.

Social scoring systems used by traditional data brokers are more difficult to find, evaluate, and quantify. Nevertheless, some information is available. Acxiom collects household social media predictors such as "Social media sites likely to be used by an individual or household, heavy or light user, whether they engage in public social media activities such as signing on to fan pages or posting or viewing YouTube videos."[189] However, the metrics for this collection, noted in a 2013 GAO report, do not appear on Acxiom's About the Data portal, so it is not possible to readily understand how Acxiom assesses social data. Analytics firm Versium offers a social influencer scoring, it uses its own observational scores.[190]

---

[179] See for example <http://en.wikipedia.org/wiki/Viral_advertising#References>.
[180] Mark W. Schaefer. Return on Influence: The revolutionary power of Klout, social scoring, and influence marketing at Introduction(2012).
[181] http://klout.com/home
[182] http://www.booshaka.com/product/overview
[183] http://www.kred.com
[184] http://www.peerindex.com
[185] http://proskore.com
[186] http://socialiq.com
[187] http://www.tweetgrader.webs.com
[188] http://twitalyzer.com/5/index.asp
[189] Government Accountability Office, *Information Resellers, Consumer Privacy Framework Needs to Reflect Changes in Technology and the Marketplace*, at 52 (Table 2) (2013), <http://www.gao.gov/assets/660/658151.pdf>.
[190] <http://versium.com>.

It is unlikely that data brokers would ignore the important and relevant category of social scoring, but whether and how they use specific scores like Klout as factors in their own score calculations is an unknown. Data brokers that create their own custom, non-public social scores may be a larger category than is generally known, but it may take time for this information to come to public light, if it ever does.

Klout Score

The Klout score is the best-known social-media based score. Anyone who has a Twitter account usually also has a Klout score. People can see their Klout score and can opt out of the score if they wish.[191] Large brands use the score to give free or discounted products and services to high scorers, called Klout Influencers. Common examples are reduced rates on a hotel room or free upgrades on flights. The score is fairly well-understood, and although the Klout algorithm is secret, the score itself is transparent. Mark Schaefer described Klout concisely:

> "Klout compiles more than 100 different factors across dozens of social media platforms, pumps billions of pieces of data through it algorithms, and creates a personalized assessment of influence that ranges from 1 to 100. The world average is about 19. Someone with a score of 30 shows expertise, wheras a score of 50 or more means leadership and expert status. And a perfect 100? That is reserved for one person alone: Justin Bieber."[192]

The Klout score has ignited some privacy controversies. One early privacy snafu[193] occurred in 2011 when Klout was found to be inadvertently scoring minors. The company stopped the practice immediately and apologized. Other questions remain. What are the privacy consequences when people are socially scored and others can readily see those scores? It is the fundamental premise of this report that scoring is a modern decision-making methodology and shorthand that, relying on complex algorithms, is going to become important in modern life due to its prevalence and use. Given that the Klout score can be seen and has an opt out, some of the transparency and choice concerns that plague many of the other scores may diminish, but the secrecy of the Klout score's composition and opacity of use remain prime concerns.

An unusual factor in the Klout score is that those who have a score and have not opted out have a public score that anyone can see. Just to make a point, credit scores can be seen by the individual, but are not available to the public without constraint. The public nature of social scoring is something new, and it creates intriguing social effects that are absent in other types of scoring. This is at the heart of the complexity of what comprises

---

[191] <http://klout.com/corp/optout>.
[192] Mark W. Schaefer. *Return on Influence: The revolutionary power of Klout, social scoring, and influence marketing* at chapter 7 (2012).
[193] At one point, Klout was scoring minors. In a blog post, the CEO apologized and Klout discontinued the practice. <http://blog.klout.com/2011/11/we-value-your-privacy/>.

modern privacy. Anyone who really wants to can usually achieve a high Klout score, or at least a passable score, as tips for accomplishing this abound. In fact, on eBay, one can outright purchase Twitter followers to ostensibly enhance the score.[194]

High scorers receive benefits − and these benefits can be financial or professional. Individuals with high Klout scores can be and have been treated preferentially by companies seeking to capatalize on their perceived clout. A high Klout score could help those seeking jobs in public relations, marketing, or other fields where Klout familiarity and understanding is helpful.[195]

Mark Schaefer tells the story of Sam Fiorella, a top marketing executive who lost a job interview because his Klout score of 45/100 was deemed too low. When Fiorella followed up with the company after a post-interview period of silence, he was told by the company that his online influence was "not sufficient for the job requirements."[196] Fiorella's Twitter profile now does not disclose a Klout score, and he has written a book about social media. His analysis is that the social scores are fading in importance to some degree as more understanding about how brand influence in social media marketing works.[197] The use of any score in an eligibility decisions is tricky, and if the Klout score were to be used in this way with any consistency it would become the subject of much debate.

It may not be clear, however, how those with low or no Klout scores will fare in job or other marketplaces. An individual may never realize that he or she did not receive an interview, job, discount, premium, coupon, or opportunity due to a low score. It is hard to hear the dog that doesn't bark. An individual denied credit based on a credit report is entitled to know the reason. An individual denied *something* based on a Klout score or the absence of a Klout score has no similar entitlement.

Beyond eligibility, the use of social scores for performance review or for public labeling is also a challenge. One sales conference posted a list of its speakers' Klout scores in a top-40 roster.[198] The good news about the Klout score is that it is public. But perhaps because it is so public, available to all, new kinds of challenges are created. For example,

---

[194] Search term "Twitter followers" typed on eBay led to 35 items. One item advertised "4,000 REAL Twitter followers!." Other items included books and other materials on how to increase Twitter followers. <http://www.ebay.com/sch/i.html?_trksid=p2050601.m570.l1313.TR0.TRC0.H0.Xtwitter+followers&_nkw=twitter+followers&_sacat=0&_from=R40>.Search conducted March 15, 2014.

[195] For example, a social media marketing position. One such position was advertised on CareerBuilder, and listed among the job requirements: "Familiarity with the key tools involved in social marketing, including measurement devices such as Radian 6, and influence tools, such as Klout.",<http://www.careerbuilder.com/JobSeeker/Jobs/JobDetails.aspx?source=jrp&APath=2.21.0.0.0&job_did=J3G0ZF6F6FQ17RCN03N&sc_cmp1=js_jrp_viewjobdesc&IPath=ILKV0A>. Retrieved March 15, 2014.

[196] Schaefer, *Return on Influence*, at Chapter 1.

[197] Email exchange, March 2014. See also Fiorella's book in which he discusses this issue at length, Pearson Publishing: *Influence Marketing: How to Create, Manage, and Measure Brand Influencers in Social Media Marketing* (2013).

[198] InsideSales.com, *Top 40 Klout Scores at the 2014 Sales Acceleration Summit* (March 12, 2014), <http://www.insidesales.com/insider/social-selling-2/sales-acceleration-summit-top-40-klout-scores/>.

opting out is not a realistic option for people in certain occupations. Having a low score could be professionally detrimental, just as having a high score could be beneficial.[199] A change in Klout's algorithm could have a major impact on an individual's status or marketplace opportunity without any notice or way for the individual to find out. Even if Klout scores lose their influence, individuals will be affected in various ways during the transition.

Another significant question is whether data brokers use or combine Klout scores and other publicly available scores in individual profiles of consumers. That answer is shrouded in obscurity, but it stands to reason that publicly available scores can be found and used by data brokers if they choose. If so, the Klout score may have more influence than even the company itself realizes, because the score could be used in algorithms that determine consumer placement and rank on lists for a wide range of consumer offers and non-offers. Klout scores could influence prices individuals pay through differential pricing algorithms.[200] Data broker scores have too little transparency for any reasonable fact-finding in this area. It will be important to discover how data brokers are using social scores in their product ranges, and if these are publicly available scores, or custom social scores. The absence of transparency in this area is troublesome in the short and long term.

Employment Success Score

Researchers from a trio of universities figured out a way in 2012 to analyze a person's Facebook page to create a score predicting their job success.[201] Employers have not used this score in any formal way or to scale. If a score like this was developed for more widespread use, it would likely fall under the Fair Credit Reporting Act. (It would also be non-compliant with Facebook's Terms of Use.)

Ta Return Scores

The Internal Revenue Service scores tax returns using several different scoring systems. Computer program calculate a numeric score for each tax return. The Discriminant Function System (DIF) score rates the potential for change as a result of an audit, based on past IRS experience with similar returns. IRS also has an Unreported Income DIF (UIDIF) score that rates a return for the potential of unreported income. RS personnel uses the scores to select for audit and to identifying the items on these returns that are the

---

[199] Intriguingly, this applies not just to individuals, but also groups, businesses, schools, and other entities. For example, see Molly Greenberg, *You'll Never Guess Which DC Area College Has the Best Klout Score* (March 12, 2014). <http://inthecapital.streetwise.co/2014/03/12/best-college-klout-scores/>.

[200] See generally, Robert Gellman, *Differential Pricing and Privacy: Good, Bad, or Otherwise?* (2014), <http://www.concurringopinions.com/archives/2014/03/differential-pricing-and-privacy-good-bad-or-otherwise.html>.

[201] Eve Tahmincioglu, *Facebook profiles predict job success* (Feb 22, 2012), Today.com, <http://www.today.com/money/facebook-profiles-predict-job-success-1C8368043?franchiseSlug=todaymoneymain>.

best candidates for review.[202] Given the internal use of these scores and the mildly restrictive law[203] on the privacy of IRS records, the IRS score may be of lesser immediate concern.

Category – Law Enforcement Scores, including Police, Transportation, Safety, and other

The government creates and uses risk scores for its work.[204] A number of government scores exist around transportation, safety, anti-terrorism, and other law-enforcement related activities. This information, however, is very difficult to find in the public domain.

However, a substantive 2013 Rand report on predictive policing shed important light on a wide range of police use of predictive scoring techniques to determine which individuals would be most likely at high risk to offend in the future. It is an important baseline study on this emerging issue.[205]

Automated Targeting System Score

The screening of airline passengers by the Department of Homeland Security has been a subject of ongoing controversy for several years. The details of the system are still not fully publicly known, but what is known is that the program collects data about passengers and links the data with other sources of information to establish a risk score for each passenger. The Transportation Security Administration uses the scores to screen passengers.

The Privacy Impact Assessment for the program states:

> "ATS provides equitable treatment for all individuals in developing any individual's risk assessment score because ATS uses the same risk assessment process for any individual using a defined targeting methodology for a given time period at any specific port of entry."[206]

---

[202] <http://www.irs.gov/newsroom/article/0,,id=151888,00.html>.
[203] 26 U.S.C. § 6103. While this law has many loopholes, it should keep IRS records out of the hands of data brokers.
[204] See Martin Bosworth, *FBI Uses Data Brokers, Risk Scores to Hunt Terrorists*, (July 11, 2007), <https://www.consumeraffairs.com/news04/2007/07/fbi_risk_scores.html> and Ellen Nakashima, *FBI Plans Initiative to Profile Terrorists*,Washington Post (July 11, 2007), <http://www.washingtonpost.com/wp-dyn/content/article/2007/07/10/AR2007071001871.html>.
[205] Walter L. Perry, Brian McInnis, Carter C. Price, Susan C. Smith, & John S. Hollywood. *Predictive Policing: The role of crime forecasting in law enforcement operations* (2013) (Rand Corporation),<https://www.ncjrs.gov/pdffiles1/nij/grants/243830.pdf>.
[206] Department of Homeland Security,DHS *Privacy Impact Assessment for ATS* (Aug. 3, 2007), <http://www.dhs.gov/xlibrary/assets/privacy/privacy_pia_cbp_ats_updated_fr.pdf>. See also Department of Homeland Security, *PIA Update*, (January, 2014), <https://www.dhs.gov/sites/default/files/publications/privacy-pia-cbp-atsupdate-01312014.pdf>.

More information on passenger screening is available from the Electronic Privacy Information Center[207] and the Identity Project.[208] A DHS report reviews legal and other problems with the Automated Targeting System.[209] The DHS scores must be shared with airlines, but whether the scores leak into the commercial marketplace is uncertain.

Richard Berk Algorithm

Criminologist Richard Berk developed a predictive model to identify murderers. Kim Zetter of Wired wrote, "To create the software, researchers assembled a dataset of more than 60,000 crimes, including homicides, then wrote an algorithm to find the people behind the crimes who were more likely to commit murder when paroled or put on probation."[210] Pennsylvania, Maryland, and Washington D.C. use the software.

Youth Delinquency Scores

The Foundation for Information Policy Research in the United Kingdom completed a report identifying the growth in children's databases and assessing the data protection and privacy implications.[211] The report describes structured assessment tools for the youth justice system in England and Wales that create profiles of young offenders by examining the factors that may have brought each youth into contact with the criminal justice system. The assessments are scored for adverse factors, and the score is used to predict the likelihood of re-offending.[212] Whether any comparable U.S. scoring systems exist is unknown. What is known is that the UK scores are subject to the UK data protection law. Similar scores created by U.S. states would not necessarily fall under any privacy regulation.

Predictive Anti-Fraud Scores: US Postal Service Office of Inspector General

The US Postal Service Office of Inspector General has a predictive analytics team that uses predictive fraud scores to address point-of-sale fraud issues. As described, the Postal Service has a customized fraud detection tool. "Using more than 30 indicators to search a wide variety of data, the fraud detection tool flags and ranks instances of suspicious

---

[207] http://www.epic.org/privacy/airtravel.
[208] <http://papersplease.org/wp/>.
[209] <http://www.dhs.gov/xlibrary/assets/privacy/privacy-secure-flight-122006.pdf> and DHS Privacy Impacy Assessment ATS System, <http://www.dhs.gov/xlibrary/assets/privacy/privacy_pia_cbp_ats_updated_fr.pdf>.
[210] Kim Zetter. *Precog Software Predicts Crime*, Wired (Jan. 2013), <http://www.wired.com/2013/01/precog-software-predicts-crime/>.
[211] Foundation for Information Policy Research, *Children's Databases – Safety and Privacy* (2006), <http://www.fipr.org/childrens_databases.pdf>.
[212] Id. at 49-50.

activity, allowing investigators in the Postal Service's Office of the Inspector General to decide which leads to pursue."[213]

Category -- Environmental Scores

EPA Health Risk Score

The EPA uses substantial predictive analytics tools, and has a **Human Toxicity Risk Score** that can be computed in aggregate, by neighborhood/per square mile.

In a groundbreaking series of articles in 2005, the Associated Press used the EPA data[214] to map the air quality risk scores for every neighborhood in the U.S. The AP mapped the EPA toxicity risk scores to socio-economic and racial factors for each neighborhood from the 2000 Census to determine the makeup of who was breathing the dirtiest air in America. The headlines across the country read, in some variation, that minorities suffer most from industrial pollution.[215]

The results established important understandings about neighborhoods and toxicity, and the resulting snapshot of where and how factory pollution was impacting neighborhoods and people were deservedly much-discussed. These results are examples of beneficial uses of scores and what today would be called large datasets or "big data."

It is helpful that the EPA has a set of meaningful best practice guidelines for analyzing its data. The EPA has a Risk Characterization Handbook. It discusses EPA's use of risk characterizations in some detail. The EPA analysis of risk analysis is valuable here:

> "Risk characterizations should clearly highlight both the confidence and the uncertainty associated with the risk assessment. For example, numerical risk estimates should always be accompanied by descriptive information carefully

---

[213] See US Postal Service, RFP, *USPS OIG Countermeasures and Performance Evaluations (CAPE) Data mining, predictive analytics, and Data Management Services* (Oct. 24, 2013), <http://postalmag.com/datamining.pdf>. See also: Shawn Hessinger – *Data Mining for Fraud at the US Postal Service*. Oct. 21, 2011. <http://www.allanalytics.com/author.asp?section_id=1412&doc_id=234817>.

[214] See <http://www.epa.gov/risk/health-risk.htm>. From the AP article: "The scores aren't meant to measure the actual risks of getting sick or the actual exposure to toxic chemicals. Instead, they are designed to help screen for polluted areas that may need additional study of potential health problems, EPA said."

[215] David Pace, *More Blacks Live With Pollution*, Associated Press (Dec. 13, 2005), <http://onlinenews.ap.org/work/pollution/wrap.py?story=./linked_story/part1.html>. See also http://www.nbcnews.com/id/10452037/ns/us_news-environment/t/minorities-suffer-most-industrial-pollution/> The EPA uses toxic chemical air releases reported by factories to calculate risk for each square kilometer of the United States. The scores can be used to compare risks from long-term exposure to factory pollution from one area to another.
The scores are based on:
_The amount of toxic pollution released by each factory.
_The path the pollution takes as it spreads through the air.
_The level of danger to humans posed by each different chemical released.
_The number of males and females of different ages who live in the exposure paths.

> selected to ensure an objective and balanced characterization of risk in risk assessment reports and regulatory documents."[216]

Further, the EPA created excellent documentation on how the analysis of its own data is to be used.[217] The documentation is for its own researchers, and is of note here because of its quality.

It stated, in part:

> "The methods used for the analysis (including all models used, all data upon which the assessment is based, and all assumptions that have a significant impact upon the results) are to be documented and easily located in the report. This documentation is to include a discussion of the degree to which the data used are representative of the population under study. Also, this documentation is to include the names of the models and software used to generate the analysis. Sufficient information is to be provided to allow the results of the analysis to be independently reproduced."[218]

These recommendation could be readily applied to consumer scores and would increase fairness and transparency for many of the scores.

AIQ Green

IQ Analytics's **AIQ Green** scoring tool "identifies prospects with a high propensity to show interest in environmentally friendly products."[219] This is a marketing score.

Category - Other Consumer Scores

The borders of consumer scoring are not fully clear. We view our definition as a work in progress. It may be too broad or too limited. Individuals may be affected in some way by patterns of usage yet to develop or in ways that are hidden from public view. If so, then the definition for consumer scores may need to change. These are some other scores that we found but did not otherwise include here.

- **Brand Name Medicine Propensity Score**
- **Rx Online Search Propensity**
- **Casino Gaming Propensity Score**

---

[216] U.S. Environmental Protection Agency, *Risk Characterization Handbook. Prepared for the U.S. Environmental Protection Agency by EPA's Science Policy Council* At A5 (2000), <http://www.epa.gov/spc/pdfs/rchandbk.pdf>.
[217] U.S. Environmental Protection Agency, *Policy for Use of Probabilistic Analysis in Risk Assessment*, (May 15, 1997), <http://www.epa.gov/spc/pdfs/probpol.pdf>.
[218] Id at 2.
[219] <http://analytics-iq.com/download/AIQ_Green_PDF.pdf>.

- **Economic Stability Indicator Financial**
- **Prescriptions by Mail Propensity**
- **Underbanked Indicator**

Potential Scores

eBay feedback score: This score is public, standard-less, and subjective, but eBay publishes the scores, and vendors and purchasers use the scores to make decisions about each other. At the least, it fall at the lower end of the spectrum of concern. eBay has its own rules for monitoring the scores, with opportunities for appeal. Crowd-source scores like eBay's are likely to grow in popularity, and they may be worthy of more detailed analysis in the future as a separate class of scores.

SenderScore: This score comes from a database of email sender reputation maintained by a commercial company.[220] According to the sponsoring website, the score derives from a proprietary Return Path algorithm, and represents a domain's overall performance against metrics important to both ISPs and recipients of email. This score represents the overall health of an email system as it appears to receiving systems.

Non-included Scores:

A host of proprietary Business-to-Business and business-focused scores are available. We acknowledge the considerable number of these scores, but exclude them on the basis of our definition of consumer scores.

Part IV: The History of Scoring: How the Credit Score and Consumer Scores Began, and Why it Is Relevant Today

The credit score is the progenitor of all consumer scores. The scoring story begins in 1941 with credit scoring, and continues today with the broadening of scoring to encompass consumer scoring in finance, insurance, health, and more.[221]

The beginnings of consumer scoring

Today's broad array of activities in consumer scoring grew from credit scoring activities, which date back to 1941. It was then that a mathematician named David Durand used

[220] See <https://www.senderscore.org>. *SenderScore – Search for Email Sender Reputation* (July 21, 2007).
[221] This discussion of the development of scoring is not intended to be comprehensive, but rather a look at the highlights of how scores have developed.

discriminant analysis to produce a scoring system to help predict risk in the granting of credit. His 1941 publication in the National Bureau of Economic Research is widely viewed as the first known account of using an analytical model to score credit risk.[222] One company in particular, Fair Isaac and Company, saw the business potential of scoring and developed scoring products for sale to business beginning in 1950. Fair Isaac was early to monetize the field, and as such the company is deeply intertwined with the popularization of scoring in business settings.

Credit scoring becomes entrenched

Over the course of the next few decades, mathematicians tweaked and refined models, tried new ones, compared models and combined scores, and in general pushed the entire body of research forward dramatically. These mathematical advances mirrored rapid advances in computing. The combination of computing and scoring allowed for increasingly rapid deployment of scoring in the credit environment. Credit scoring received a formal nod when the Equal Credit Opportunity Act (ECOA) cited credit scoring systems in one of its amendments.[223] By1979, William Fair of Fair Isaac estimated that between 20 and 30 percent of all consumer credit decisions were made by credit scoring.[224] By the 80s and 90s, scores had been adopted widely in the U.S. and were spreading across the world, particularly in mature economies such as the U.S. and Europe. In 2006, a press release noted that FICO scores were used by U.S. lenders to make decisions about more than 75 percent of mortgage loan originations. [225] During the creation, spread and use of the credit score from about 1941 to 2000, the score was largely secret to consumers. A decision by the FTC in the 1990s sealed the secrecy of the

---

[222] David Durand, *Risk Elements in Consumer Installment Financing, Study #8* (1941), National Bureau of Economic Research.. Interestingly, a 1942 letter reporting on 1941 research activities written by National Bureau of Economic Research board member C. Reinold Noyes reveals that the bureau did not understand the profound implications of what they had published: "Undoubtedly the most important event to record in this report on the National Bureau of Economic Research for the year 1941 is the appearance of Simon Kuznets' National Income and its Composition, 1919-38…" Durand's work received a small, passing mention in the letter. *Report by our Representative on the Board of Directors of the National Bureau of Economic Research*, 32American Economic Review 519-521 (1942) (Supplement, Papers and Proceedings of the 54th Annual Meeting of the American Economic Association).

[223] 12 CFR § 202.6(b)(2)(ii). "To qualify as an empirically derived, demonstrably and statistically sound, credit scoring system, the system must be:

(i) Based on data that are derived from an empirical comparison of sample groups or the population of creditworthy and noncreditworthy applicants who applied for credit within a reasonable preceding period of time;

(ii) Developed for the purpose of evaluating the creditworthiness of applicants with respect to the legitimate business interests of the creditor utilizing the system (including, but not limited to, minimizing bad debt losses and operating expenses in accordance with the creditor's business judgment);

(iii) Developed and validated using accepted statistical principles and methodology; and

(iv) Periodically revalidated by the use of appropriate statistical principles and methodology and adjusted as necessary to maintain predictive ability."

[224] Credit Card Redlining 1979, p. 183-184. As quoted in Noel Capon, *Credit scoring systems: a critical analysis*, 46 Journal of Marketing 82-91 (1982) .

[225] Fair Issac, Press Release, *Fair, Isaac 'Demystifies' FICO Scores with list of Score Factors, Web-Based Explanation Service* (June 8, 2000) ("FICO scores are used by U.S. lenders to make billions of credit decisions each year, including more than 75 percent of mortgage loan originations.").

credit score, but was later reversed. The history of how the credit score became public is an important precedent for current eligibility-related scores that are currently not available to consumers.

How the formerly secret credit score became available to the public

Scores were unknown to most consumers through the 50s, 60s, 70s, and 80s. Trickles of a score that was that could be used to deny a person credit but which was not revealed to consumers began to leak out slowly to some policymakers, particularly around the time ECOA passed. But scores had not entered the minds of most people.

In May 1990, the Federal Trade Commission wrote commentary indicating that risk scores (credit scores) did *not* have to be made available to consumers. But when scoring began to be used for mortgage lending in the mid 90s,[226] many consumers finally began hearing about a credit score, most for the first time, and most when they were being turned down for a loan.[227] A slow roar over the secrecy and opacity of the credit score began to build.

By the late 90s, the secrecy of credit scores and the fact that people could not see the underlying methodology or factors that went into the score or the range of the score to determine how the number should be interpreted was a full-blown policy issue. Beginning in 2000, events pushing toward increased credit score disclosure began to escalate, culminating in a rapid-fire series of events that eventually dismantled credit score secrecy and non-disclosure.

It is fair to say that a good deal of the escalation of events began when E-Loan, an Internet lender, took the extraordinary step of making credit scores public in February 2000 via a web site.[228] The scores were free, and the word got out quickly to consumers. In one month, the site attracted more than 25,000 customers, and a lot of attention.[229] The web site was shut down after six weeks; Fair Isaac, at that time, had a rule prohibiting the disclosure of FICO scores to consumers unless they were turned down for a loan.[230] But although the site was been shut down, consumer appetite for their scores had been whetted. This incident was a tipping point due to how it popularized the score issue among consumers.

---

[226] In 1995 Freddie Mac and Fannie Mae endorsed the use of credit scores as part of the mortgage underwriting process. This had a substantial impact on the use of credit scores in the mortgage loan industry. See for example Kenneth Harney, *The Nation's Housing Lenders might rely more on credit scores*, The Patriot Ledger (July 21 1995).

[227] See for example, comments of Peter L. McCorkell, Senior Counsel to Wells Fargo, to the Federal Trade Commission, August 16, 2004 in response to FACT Act Scores Study.

[228] A good first-hand account of the E-Trade web site incident may be found in an E-Loan press release: E-LOAN, Inc., *A full credit score disclosure pioneer, calls for national legislation; New credit score disclosure law is a giant step forward for California consumers, but consumers everywhere else in America remain in the dark* (June 27, 2001).

[229] Brian Angell, *A Score to Settle: Consumer demand is high for credit scores. What's the holdup?* American Banker-Bond Buyer(August 2001).

[230] Id.

That same month, on February 22, 2000, California senator Liz Figueroa introduced SB 1607 which would give Californians access to their credit scores. Specifically, the bill required lenders to give customers a copy of the credit score obtained to solicit a loan or accept a loan application.[231] Bowing to the growing pressure, Fair Isaac began to release some information about the factors that were used in its credit scoring model, FICO, in June 2000,[232] but they did not release the actual score at that time. One of the arguments they made was that too much disclosure would allow manipulation of the score.[233]

Governor Grey Davis signed the credit score disclosure legislation in September 2000, and the law took effect July 1, 2001.[234] An uncomfortable situation then arose for federal lawmakers: Californians were the only ones who had access to their credit score. It was a classic recipe for national legislation on credit score disclosure.

In 2002, the FTC reversed its 1990 decision and concluded that consumers should be able to see their credit scores. As of December 2004, the Fair Credit Reporting Act as modified by the Fair and Accurate Credit Transactions Act, or FACTA, ended score secrecy formally, and required consumer reporting agencies to provide consumers with more extensive credit score information, upon request.[235] Also made available to the public was the context of the score (its numeric range), the date the score was created, some of the key factors that adversely affected the score.[236] The Federal Trade Commission is required by FACTA to study various aspects of credit scoring, insurance scoring, disparities, modeling, and more.[237] Much still remains unknown about scoring models, even those that fall under FACTA such as credit scoring models. The formulas, which are important in verifying many aspects of the scoring model, are still secret.

Ongoing disclosure challenges and other issues with consumer credit scores

During the FACTA process, a growing trend was captured via the public comment process, that is, that the use of credit scores was greatly expanding to other areas of

---

[231] See <http://info.sen.ca.gov/pub/99-00/bill/sen/sb_1601-1650/sb_1607_bill_20000222_introduced.html> for the bill as introduced, and <http://info.sen.ca.gov/pub/99-00/bill/sen/sb_1601-1650/sb_1607_bill_20000930_chaptered.html> for the final version of the bill as chaptered.
[232] *Fair, Isaac 'Demystifies' FICO Scores with list of Score Factors, Web-Based Explanation Service*, PR Newswire, June 8, 2000
[233] Fair, Isaac, Freddie to offer credit info to consumers, Credit Risk Management Report, June 12, 2000. Vol. 10, No. 11. Regarding arguments for not releasing the score to consumers, see Bonnie Sinnock, *Fair Isaac site offers credit score details*, American Banker –Bond Buyer Association, November 6, 2000.
[234] C.A.R.'s Credit Scoring Bill Signed into law by Governor Gray Davis; Landmark Legislation Gives Consumers Access to Credit Scores. PR Newswire, October 2, 2000.
[235] 15 U.S.C. § 1681 g(f).
[236] For a guide to consumer rights granted under FACTA, see *Appendix F: Summary of Consumer Rights Under the Fair Credit Reporting Act*, FTC, November 19, 2004. <http://www.ftc.gov/opa/2004/11/facta.shtm>.
[237] The Federal Trade Commission has already released some reports, which may be found at <http://www.ftc.gov/>.

business. One area of concern was the use of credit scores to determine homeowner and auto insurance rates. Some individuals who had good driving records, for example, all of a sudden, upon renewal, were receiving much higher insurance rates due to a weak credit score. This practice has been the subject of much discussion, study, consternation, and some lawmaking, with varying results. During this general period of time that FACTA was being debated, the crime of identity theft began to become known and understood. New laws regarding the setting of insurance rates by a credit score impacted by identity fraud have been percolating through the states now as a result.

Disclosure of credit scores in now a non-issue. But while credit scores have been made public, it is not so with all other consumer scores. Consumers who want to see their identity score, their Z score, or many other scores cannot. Consumers who inquire about scores, or even the existence of a possible score, are not always told whether or not a score is being used. Similar if not identical arguments are used today to keep some consumer scores secret as were used to keep the credit score secret. While the credit score and its use, has been regulated by FACTA and now also by Basel II, this is not so for the broadening range of consumer scores that are increasingly attaching themselves to consumers.

The heightened availability and almost complete lack of oversight and regulation of the newer consumer scores combined with almost complete opacity regarding consumer scores' (minus credit and some forms of insurance scores) models, factors, ranges, validation, bias, sample size, and so forth has created a swath of non-disclosure and secrecy that consumers are at this point largely unaware of.

Conclusion

Because consumers cannot see most of the new consumer scores, cannot know the factors underlying many of the scores, there is no real application of Fair Information Principles to many of the new and unregulated consumer scores. Consumers who do not know about the existence or use of consumers scores cannot have any say in who used the scores, or how. Scores affect the lives of consumers, but only with reform will consumers receive rights to protect their interests.

The data business is changing and is becoming much more sophisticated. Consumer scores are a significant way that this is happening. Consumer scoring has substantial potential to become a major policy issue as scores with unknown factors and unknown uses and unknown validity and unknown legal constraints move into broader use.

Secrecy, fairness of the factors, accuracy of the models, and the use of sensitive information are some of the key issues that must be addressed. It is exquisitely unlikely that self-regulation will solve all of the dilemmas consumer scoring introduces. However, we already have at least a partial model for what would constitute fair regulation from the history of the credit score. The protections consumers receive with respect to credit

scores need to be expanded to all consumer scoring, and the rules for credit scores may warrant some reexamination as well.

About this Report and Credits

Authors: Pam Dixon, Executive Director, World Privacy Forum and Robert Gellman, Privacy Consultant.

Editing, proofing, and reading: Tim Sparapani, Linda Ackerman, Nathan Good, Blake Hamilton.

Cover illustration: John Emerson of Backspace.

Publication date: April 2, 2014.

Published at: www.worldprivacyforum.org

Version: 1.2

URL for report: http://www.worldprivacyforum.org/category/report-the-scoring-of-america/
Future version and updates will be located at the above URL.

Background interviews for this report were conducted from a period of 2007 to 2014. The World Privacy Forum thanks the US Federal Trade Commission and the many companies, experts, analysts, and others we interviewed for this report.

Appendix A

Timeline: Highlights in Scoring

1941 David Durand publishes first account of the use of discriminant analysis to produce a scoring system for the use in granting credit. [238]

1950 Glueck, S. and Glueck, E.T. Unraveling Juvenile Delinquency, Cambridge: Harvard University Press. Develop a point scoring method.

1950 Stanford University researchers Bill Fair and Earl Isaac set up a new business in a San Rafael garage.[239]

1963 Myers and Gorgy (compared discriminant analysis and regression analysis)

1971 Orgler used regression analysis to create a behavioral score to evaluate outstanding loans based on past performance. (Orgler, Yair E. *A Credit Scoring Model for Commercial Loans*, Journal of Money, Credit and Banking, 2 November 1970, 435-45.)

1977, 1978 Eisenbeis presented assessment of the use of discriminant analyss in business, finance, and economics in general. [240]

1980 Wiginton one of first published accounts of logistic regression applied to credit scoring.

1981 Grablowsky and Talley (1981) compared linear discriminant analysis and probit analysis by using data from a large U.S. Midwestern retail chain. [241]

1984 Breiman et al publish on recursive partitioning, or decision trees.

1991 Safavian and Landgrebe – survey of recursive partitioning in scoring, including artificial intelligence.

1992 Blackwell and Sykes described the use of behavioral scoring to determine credit limits.

1993 Leonard described an expert system for detecting fraudulent use of credit cards.

---

[238] D.J. Hand, *Statistical Classification Methods in Consumer Credit Scoring: A Review*,160 Journal of the Royal Statistical Society 532 (1997).
[239] Edmund Sanders, *California Firm that developed FICO credit scores is still sailing*, Orange County Register (September 26 1997).
[240] D.J. Hand, *Statistical Classification Methods in Consumer Credit Scoring: A Review*,160 Journal of the Royal Statistical Society 532 (1997).
[241] Id.

1994 Rosenberg and Gleit-- applications of neural networks to corporate credit decisions and fraud detection.

1995 Henley described a fraud score card built by a linear regression analysis model.

1995 FAIR ISAAC introduces its first model of a Small Business Credit Score.[242]

1996 Henley and Hand develop an adaptive metric "nearest neighbor" method for credit scoring.

2000 In February California senator Liz Figueroa introduces legislation to allow consumers to see their credit scores. The bill is signed into law in September.

2001 July 1: Californians have the legal right to view their credit scores.

2003 The FACT Act is enacted December 4, consumers nationwide given the legal right to view credit scores, score range, and some additional rights.

2004 FTC requests public comments on the use of credit scores in setting insurance rates.

2007 Antibody scoring models in development, Theodore Crooks.

2008 Klout is born, and social scoring becomes a reality.

2011 FICO launches Medical Adherence Score, one of the first major medical consumer scores. The score does not have to rely on medical files for its predictions.

2012 Charles Duhigg breaks the existence of theTarget *pregnancy predictor score* in a New York Times feature article; raises awareness of predictive analytics and use of masses of factors in scoring algorithms.

2013 HHS creates The Health Risk Score for individuals using the Affordable Care Act (Obamacare) program.

2014 FTC holds alternative scoring models conference, first high-level attention to non-FCRA scores.

2018 The target date for phasing out the Health Risk Score. (Planned).

---

[242] Allen N. Berger, W. Scott Frame. *Small Business Credit Scoring and Credit Availability*, 45Journal of Small Business Management (2007).

Appendix B

Score Taxonomy

In minds of consumers, there is just one score, the credit score. But the credit score is just one final outcropping of a layered and complex taxonomy of scoring. This taxonomy can assist consumers in seeing the full range and depth of scoring activities that exist, and may impact them.

This taxonomy is also important in understanding the scores this report focuses on. This report is focused on *consumer scores* that are used for *consumer purposes*. Or to use the taxonomic language, consumer scores derived from formal predictive models and used for consumer-related purposes, that is, used in a way that impacts a non-clinical (non-medical) decision about a consumer or a group of consumers.

I. Predictive Statistical Models

II. Formal Scoring Models

III. Consumer Scoring Model*s*

IV. Consumer Scoring Model Type (application, behavioral, or combined)

V. Consumer Scoring Function: the broad function of the score card, as follows:

Propensity score cards: will the consumer, for example, defaul, what is the propensity of a certain result. Credit scoring is a propensity scoring function. Health Scoring is a propensity function if it falls under the full taxonomy preceding this point.

Response score cards: will the consumer respond to a direct marketing offer

Usage score cards: will the consumer use the credit (or other) product if given the product

Attrition score cards: will the consumer continue with the lender, especially if there is some special offer available for an introductory period only.

Customer profit scoring score cards: estimates the total profitability of the customer to the lender

Product profit score cards: seeks to estimate the profit the lender makes on this product from the customer[243]

VI. Source of the Score Model and score (Generic, custom, or vendor supplied score)

VII. The Specific *Type* of Score (fraud, credit, etc.) Here, the term credit refers to the broad type of score.

VIII. Application of Score (what purpose is the score used for)
Consumer-related: test: does the score impact a decision about an individual consumer or a group of consumers?
Research-related: (esp. Health research)[244] test: is the score used to primarily to understand or explain a process or a disease and never used to make a decision about an individual consumer beyond a clinical medical decision? (If a financial or risk decision is taken, then the score becomes a consumer score, not just a clinical score.)

IX. Actual Scores (This includes all specific scores resulting from the taxonomy, Z score, Falcon score, FICO score, etc.) *Note: this report is focused on Consumer-related scores, or scores that are used for consumer purposes. If at any point a pure research-related score is used in a consumer score model as a predictive factor and the resulting final score is used for consumer purposes, the final score would be considered a blended consumer score and would be included in the consumer category.* See Taxonomy step VII.

[243] The discussion of the function of score cards in this segment is derived closely from LC Thomas, RW Oliver, DJ Hand, *A Survey of Issues in Consumer Credit Modeling Research*, 56 Journal of the Operational Research Society(2005).
[244] There are a number of health-related scores in particular that are originally created solely for research purposes, particularly public health research. Scores of this type are not considered in this report. But if the research score is later combined with a consumer score and is used for consumer purposes, then that score would be a blended consumer score and would be considered in the report. Because of how scoring models operate, it is possible that some pure research-related health scores later became part of some consumer-related scores. Any characteristic, such as a health score, can be input into a consumer scoring model. In this way, a pure research score can *contribute to* a consumer-related score. It would be nearly impossible to determine how many health scores originally created for research purposes have or are being used in this way. Scoring models generally do not reveal formulae to this level of specificity.