Chapter 5
Targeting by Numbers. The Uses of Statistics for Monitoring French Welfare Benefit Recipients

Vincent Dubois, Morgane Paris, and Pierre-Edouard Weill

Abstract The targeting of welfare recipients in order to control their situation and their entitlement has increased from the beginning of the 2000s onwards. Data mining has recently been included in the set of techniques used for this purpose, in addition to traditional bureaucratic checks of documents, home inspection visits, and to data crossing. Imported from the private sector, this statistical tool is part and parcel of a "risk management" policy of the family branch of French social security. It has been promoted as the cornerstone of recipients' monitoring since 2010. Analyzing the use of this method enables us to show the new relationships between statistical instruments, legal norms, and performance indicators which define the administration of the Poor in the neomanagerial era. Thanks to statistical correlations, this tool identifies welfare recipients' features significantly associated with the highest level of risks of irregularities. Then, scoring algorithms enable local managers to target high-risk populations over which in-depth checks are performed. This has led to positive financial results, but also to an increasing focus of surveillance on the most disadvantaged households. Based on interviews with executives of the National Family Benefits Fund (Caisse nationale des allocations familiales - CNAF) and with local managers, ethnographic observation of street-level bureaucrats' daily work and quantitative analysis of national and local data, our contribution is twofold: on the use of statistical modeling in welfare policies implementation; on the compounding of control in the contemporary government of the poor.

Keywords Statistics · Data mining · Welfare benefits · Lower class · Social control

---

V. Dubois · M. Paris
Université de Strasbourg – SAGE, Strasbourg, France
e-mail: vincent.dubois@misha.fr; morgane.paris@misha.fr

P.-E. Weill (✉)
Lab-LEX (EA 7480), Université de Bretagne Occidentale, Brest, France
e-mail: pierreedouard.weill@univ-brest.fr

L. Barrault-Stella, P.-E. Weill (eds.), *Creating Target Publics for Welfare Policies*, Logic, Argumentation & Reasoning 17,
https://doi.org/10.1007/978-3-319-89596-3_5

5.1 Introduction

In France, until the mid-nineties, monitoring welfare benefit recipients and combatting fraud were issues that seldom arose in public debate and were not part of the strategic direction for social policies. However, they became key issues in the subsequent period, in France as in most European countries, and were ubiquitous both in the media and the political arena, from heads of departmental councils asserting the need for tighter controls of welfare benefit (e. g. *Revenu de solidarité active (RSA)*),[1] to the highest government levels (a particularly striking example being President Sarkozy's speech in Bordeaux on 15 November 2011[2]). At the same time, welfare institutions adopted an increasingly systematic monitoring policy with stronger organization, greater legal/human resources and new technical tools.

This article focuses on data mining, which has become a key tool in the implementation of this monitoring policy. The generic term "data mining" refers to a set of statistical methods used to process large volumes of data and to develop models that, in turn, make it possible to systematize observations on the same type of data. Unlike other methods based on hypothetico-deductive reasoning (a hypothesis that is tested empirically then validated or discarded), data mining proceeds by induction: it starts from the exploratory observation of data, identifies regularities and correlations between variables, then develops predictive models that will be used to process large-scale data. For the case in point, the system is used to process information from the files of some 12.8 million recipients of the Departemental Family Benefits Funds (*Caisses d'allocations familiales – CAF*).

That a welfare institution uses this technology for monitoring purposes is indicative of three current trends in public policies. First, it is an example of technology transfer from the private to the public sector where data mining is used to model consumer behaviour, improve marketing-related targeting, reach potential donors in fund-raising campaigns, and in particular manage customer-related risks, such as actuarial calculations of rates in the baking and insurance sector and the identification of prospective defaulters in service activities such as mobile telephony (Lazarus 2012). Data mining is increasingly used in the public sector, particularly in the United States (Federal Agency Data Mining Reporting Act of 2007). It is used to identify criminals, tax and customs fraudsters and other potential offenders. In France, data mining has especially been promoted as an instrument to improve targeting in the fight against fraud by the National Family Benefit Fund (*Caisse natio-*

---

[1] *Revenu de solidarité active (RSA)* is a French welfare benefit designed to give recipients a basic income, whether or not they are capable of work. In 2016 there were 2.5 million recipients. Claimants must look for employment or define and undertake a vocational project that will improve their financial situation to remain eligible. See Wacquant (2010) on the relationship between "activation" policies and controls of underprivileged populations in France, from an international comparison perspective.

[2] Online video: https://www.youtube.com/watch?v=aXPhEPRTyBs

Methodology

This research, following up on a CNAF-funded study on the same subject in the early 2000s (Dubois 2003), was carried out at three levels:

1. national: the domestic political and institutional sphere in which monitoring policies are either directly developed or indirectly influenced (the CNAF, relevant government ministries, the National Anti-Fraud Unit (*Direction nationale de lute contre la fraude – DNLF*), the National Court of Auditors (*Cour des comptes*), etc.). This was examined on the basis of interviews conducted within the institutions involved (n = 13) and the collection of internal documents.
2. local: the implementation of monitoring policies at this level was examined on the basis of local studies conducted in two CAF centres: one in a highly disadvantaged urban area and the other in a more favourable socio-economic environment, with a city centre of over 100,000 inhabitants, several mid-sized towns, and rural areas. In both instances, interviews were conducted with staff members (n = 22: inspectors, accountants, legal department staff, senior managers); additional interviews with other institutions (general councils (*Conseils généraux*)); observations of home inspection visits, fraud commission meetings, and other meetings related to our study object; collection of internal documents and statistics.
3. the population under study was observed during spot checks and studied through local and national statistical data, respectively supplied by the two CAF centres and the CNAF statistics directorate (*Direction des statistiques, des études et de la recherche - DSER*).
4. Combining qualitative and quantitative methods, our investigation started by considering CNAF statistical analyses and re-processing secondary data produced by the same institution. This multi-level perspective enabled us to study the targeting practices used from the highest levels of government down to the "street-level bureaucrats" (Lipsky 1980), which in this case are the inspectors in direct contact with the welfare beneficiaries.

nale des allocations familiales – CNAF)[3] since 2010, customs services and the National Employment Agency (*Pôle Emploi*) since 2013, and the tax authorities since 2014. The use of data mining also shows a second trend in public policies: the implementation of risk management by predictive modelling. The rise of predictive policing in the United States is a good example of this trend. Its objective is to

[3] The National Family Allowance Fund (*Caisse nationale des allocations familiales – CNAF*) manage the family branch of the French social security system and supervises Departemental Family Benefit Funds (*Caisse des allocations familiales – CAF*).

anticipate crimes before they are committed by estimating the likelihood of re-offending on the basis of past actions. It is also used as a decision-making tool for setting bail or prison sentences according to the risk of reoffending (Harcourt 2005). With regard to welfare benefit recipient monitoring, the aim of data mining is to predict the risk of irregularity – a prediction which then triggers an investigation into the recipient's situation. In this context, data mining is in line with the much wider trend of the growing use of big data in public policy since it relies on the collection and processing of large amounts of data that can be used by public or private organizations in their quest for information on individuals (Harcourt 2014; Ollion 2015). The family branch of the French social security system holds a lot of detailed information on nearly 13 million beneficiary households, which equates to nearly 30 million individuals (including spouses and children). A substantial amount of data is held for each person, not only on benefits received, but also on marital and professional status, income, age and schooling of child/children, housing situation, etc. Finally, and we place a particular focus on this point, data mining is being used for a targeting policy that should allow public policies to be more systematically and efficiently adjusted to the diversity of situations, in this case through the identification of regularity risks in welfare beneficiary files.

This article will examine these three major trends via a case study on the uses of data mining in the monitoring policies of the French social security system's family branch: (i) technology transfer from the private to the public sector, (ii) risk management using predictive models based on handling large volumes of data and (iii) the implementation of monitoring through targeting population. Its objective is to show how the risk control management model helps to bring closer a policy direction that supports tighter recipient controls and a statistical tool that enhances efficiency but inevitably leads to the targeting of the most vulnerable population groups.

To find out why data mining is used to detect irregularities in files, one must look at how it was transferred from the private to the public sector by the welfare institutions, in this case the CNAF. Data mining was promoted from 2010 onwards as the key element of a systematized and rationalized risk management policy using increased beneficiary monitoring (1). The refocusing of this control policy on data mining was accompanied by changes in the relationship between institutional stakeholders at the national and local level, and also helped to change monitoring practices and increase their effectiveness (2). These changes had an impact on the everyday lives of people on welfare benefits recipients, particularly for those who rely on benefits the most and are subject to stepped-up control (3).

5.2 A New Statistical Tool for Monitoring

A statistical technique can be said to be successful when "it coincides with new ways of perceiving and organizing social relationships" and "a network of innovators finds a sufficient number of allies to circulate it" (Desrosières 2014: 81). For instance, although the family branch of the CNAF was the forerunner in this matter and had been using data mining for the large-scale processing of beneficiary files since 2011, it was the growing requirement to ensure accurate benefit payments and reduce unjustified spending that led to the concurrent intellectual, institutional and technical creation of a new global risk management strategy. The conditions in which this tool was transferred meant that, in its early phases at least, it was preferentially defined and used to monitor social benefit recipients, in particular for the detection of undue payments and fraud. Undue payments are due to lateness, (in) voluntary errors/omissions in beneficiary declarations or faulty data processing by the institution. Benefit fraud is committed by people who deliberately make a false declaration in order to receive money for which they are not entitled.

5.2.1 A Risk Management Tool

The use of data mining fits into a broader context of the emergence of social fraud, and to a larger extent of welfare benefit monitoring in open debate and government policies. In this respect, it is a technical solution to a political problem. Initially, this drive to combat social fraud stemmed from more or less directly injunctions on the CNAF, formulated by successive governments and various institutions. The government was able to become more directly involved in the organization and accounts of the social security system following the 1996 reform that mainly gave rise to the signing of Target and Management Agreements or COGs (*Conventions d'objectifs et de gestion*). These agreements laid down multi-annual commitments for institutions, in particular regarding anti-fraud tools and techniques. The 2001–2004 COG provided for the establishment of a general risk database that would include beneficiary declarations, and also recommended the analysis of risks and their securitization procedures to establish control standards. Likewise, there was a specific action plan covering prevention and improvements in undue payment recovery that was based on analyses identifying the causes and prospects of recovery. This more analytical approach integrated the new emphasis on targeting by reducing random or concern-based checks and using statistical analysis to better define targets and increase efficiency. The subsequent agreements reflected the constant incentive to use new technical tools that would identify undue payments and to monitor their outcomes. These external incentives to reinforce monitoring tools also came, more indirectly, from the National Anti-Fraud Unit (*Délégation nationale de lutte contre la fraude – DNLF*) – an inter-ministerial structure created in 2008 that helps to promote the use of technology such as data mining and to make the results visible. The Court of Auditors (*Cour des comptes*) has played a very direct role in this process, occasionally in its annual or thematic reports and continuously in its role of accounts certification since 2006. Monitoring effectiveness became a criterion in account validation and the auditors of the Court made recommendations notably for risk management, which played a decisive role in the efforts made in this domain and was a direct incentive for the use of data mining.

External criticisms of the supposed lack of fraud detection and prevention activities meant that the family branch was pressured into reorganizing and reinforcing their activities in this respect. The old divide between internal and external monitoring (monitoring of staff operations and that of incoming beneficiary data, respectively) was replaced by the notion of risk management, mainly through the creation of the Risks Management department in 2005. This reorganization and new outlook tended to cause an increase in beneficiary monitoring in the institution, with more and more CAF employees participating in this external risk management from a financial and accounting perspective. Accounts officers who were previously unfamiliar with this type of monitoring became major operators. At the same time, the generic term of risk management, that as such concerns all risk management staff, is virtually synonymous with welfare beneficiary monitoring, as its activity is essentially focused on external risks that are defined by the reliability of incoming data, i.e. information provided by the beneficiaries. Initially a relatively low-level sectorial expertise linked to welfare benefits, recipient monitoring became a cross-cutting concern which plays a large part in the institution's new management modes.

There is a direct affinity between organizational thinking in terms of risk management and a probabilistic and predictive tool such as data mining. Unlike coherence checks (a comparison of information from various sources), this tool is not designed to identify anomalies, but to assess the probability (risk level) of an anomaly occurring. This predictive statistic – upstream from but still supporting the administrative logic of compliance checks – serves as a management logic primarily geared towards financial rigour.

This tool does not, however, determine its potential uses by itself. It is a technological tool that on the face of it seems to be neutral in that it can be used for a variety of equally important operations and purposes. Targets of data mining models can be files with a financial risk, but also people who eligible for benefits. It is imaginable that the *Caisse nationale* could inversely use the data mining tool to identify potential recipients. In this way, data mining would not only help to combat fraud but also to diminish the non-claiming of benefits by eligible claimants – these two uses are not mutually exclusive. However, in the same way as risk management policy is widely confused with beneficiary monitoring, data mining was primarily designed to detect undue – including fraud-related – payments.

5.2.2 The Institutional Appropriation of Data Mining

The second condition for success identified by Alain Desrosières is the tool's perimeter of use which in this case is a factor that has also oriented its uses towards the detection of undue payments and fraud. The National Employment Agency (*Pôle emploi*) and the sickness branch of social security also use the tool in the same way, which gives the supposedly "neutral" technique of data mining the profile of a welfare state monitoring tool.

To some extent, this direction was determined by the origins of the use of data mining. Prior to being promoted as the national tool for monitoring policies, data mining was first piloted at the local level. This was done through the initiative of an accounting officer who researched new methods because even though the anti-fraud tools at his disposal were limited he could be liable in cases of undetected fraud. And this is how he discovered data mining, which the Gironde Family Benefits Fund (*CAF*) had been testing since 2004 in conjunction with the national risk management policy. The first step was to build up a national database of 3000 fraudulent applications as a benchmark used to draw up the first models tested by the Gironde CAF. This profiling was refined in 2006–2007 when the database was expanded. From 2009, data collection based on the random sampling of beneficiary files was rolled out nationwide. Each year the system selects 10,500 beneficiary files that are systematically checked by certified monitoring officers using both desk and home monitoring.

In keeping with what had been carried out at the local level then extended to several departments, this specific operation called *Cible 021*, "was like a survey data investigation on risks and their changes" (Collinet 2013: 130), making it possible to develop then adapt models based on correlations between the most statistically predictive variables of undue payment. These various models, which will be discussed in more detail later, make it possible to give a risk score to beneficiary files, and on this basis to select the cases to be monitored. It is therefore a pre-monitoring instrument that can identify the files with a risk level justifying an actual investigation. The low cost of implementation improves the efficiency of monitoring operations. In 2010, some CAF centres voluntarily piloted the instrument to trigger monitoring, and it was rolled out across the entire network in 2011. In addition, a smaller sample of 7000 beneficiaries was used to calculate the financial risks not covered by the monitoring system (Residual Risk Indicators).

The way in which data mining has been used by the *Caisse nationale*, *Pole Emploi* and other social security branches, is in line with the changing functions that have been assigned to and demanded by social security institutions. These institutions have highly integrated management logics against a background of diminishing resources. Data and statistical instruments are no longer used solely for the socialization of risks to which individuals are exposed (sickness, ageing, unemployment, accidents, family events) in what François Ewald (1986) referred to as the "insurance society" logic. Nor is it a question of identifying high-risk populations who, because of their situation or potentially dangerous behaviour that puts either themselves or society at risk, generate prevention actions from specialized institutions, as analysed by Robert Castel (1981). The risks data mining assesses in its risk management mode, are those that missing, incomplete or false beneficiary data – much of which comes from the recipients themselves – present to the smooth legal and financial running of the paying institution. In both cases, statistics are indeed a "tool for governing" (Desrosières 2014); they primarily support a form of risk management synonymous with expenditure control.

5.3 A Control Mechanism Realigned Around Data Mining

Validated by conclusive financial results, data mining now appears as the new cornerstone of monitoring in a context of increased sharing of computerized data and the targeting of home monitoring for so-called "high-risk" welfare beneficiaries, who are often also the most vulnerable. The use of data mining is a part of "social work's electronic turn" in Western Europe since the beginning of the 2000s, promoting, among other things, the nationwide collection of local data for government at a distance (Parton 2008). The major role it now plays in the family branch monitoring policy has given more power to the central level, which in turn has limited flexibility and initiative of the departmental management team and their field officers.

5.3.1 A New Cornerstone of Monitoring

Data mining has become an essential element of national monitoring and anti-fraud policies because it was piloted beforehand and since has been presented as a reason for the constant progression in the detection of undue payments – and to a lesser extent non-payment and underpayments – since the late 2000s. CNAF progress reports highlight increasing returns on targeted home checks since the use of data mining was generalized. The proportion of spot checks that have resulted in accruals has considerably increased from 17% undue payments and 14% non-payment and underpayment in 2009 to 44% and 30%, respectively, in 2015. There is also a significant change in the amounts detected, from 122 to 225 million euros in undue payments, and 40 to 59 million euros in non-payment and underpayments over this period. This positive correlation is thought to be due to the ability of the statistical model to identify risk factors of irregularities with a high financial impact, even where the predictive models for the amounts of undue payments/non-payment and underpayments are not yet developed. This being true, the use of models to target the type of situations presenting the greatest financial risks can only become more prevalent.

If data mining is seen as the new cornerstone of monitoring policies it is also because its programmed development, justified by its financial results, is connected to two other notable evolutions in this policy. First, increased monitoring effectiveness cannot be separated from a constant increase and improvement in the quality of incoming data. This is based on a range of innovations in the storage, consultation and sharing of personal data between social security branches, tax authorities and private partners, notably in the banking sector. The exchange of computerized data and, more broadly, of on-file verifications, has been boosted by the implementation of a full range of identifiers, files, and cross-checking between files, enabling wide-scale monitoring from a distance that is low-cost in terms of staff resources. Legally authorized since 1998 but only in use since 2008, a national database (*Référentiel* *national des bénéficiaires*) identifies each social benefit recipient by a single register number (*Numéro d'identification au repertoire - NIR*) and groups all CAF files together, thus making it possible to identify multiple entries and to link information from other databases. A national directory (*Répertoire national commun de la protection sociale - RNCPS*) makes it easier to access personal data held by other welfare institutions. Set up in 2009, it is accessed through a common computer portal called *Espace des organismes partenaires de la protection sociale (EOPPS)* used by social welfare partner organizations. The sharing of this type of computer interface helps to strengthen pre-existing institutional links (Baudot 2011).

At the same time, the number of home checks has decreased. However, this does not mean that these processes are any less important to the monitoring process, but rather that they are now focused on the riskiest, and therefore most complex, files identified by the data mining algorithms – the primary means of triggering a monitoring process. The identification of a potential, undefined risk reinforces the need for home checks, but these increase the workload for certified officers, whether for upstream investigations or reporting purposes. It goes without saying that the significant and continuous decrease of spot checks, from 280,000 in 2009 to 166,000 in 2015, is because more home checks are being triggered by data mining. Several trends overlap: the increased use of data mining for triggering home checks, specialization in difficult cases, longer assessment time for more targeted and exhaustive investigations, and improved performance in detecting irregularities.

5.3.2 More Centralized Supervision in Localized Targeting

Geared towards the improvement of monitoring effectiveness and an increase in the detection of undue payments and fraud, data mining has helped to limit the autonomy of local CAF centres and lessen the differences that used to exist between their practices. Information requests focused on individual social welfare beneficiaries, based on knowledge gained from local benefit management, was previously one of the only options that local CAF centres had for shaping their control policies. Data mining has progressively become the primary tool in selecting files to be monitored. In the beginning, some departments were chosen to participate in the wide-scale pilot scheme (such as Seine-Saint-Denis) before it was rolled-out nationally in 2011. Prior to this date, less than a quarter of spot checks were decided at the national level; in 2016, nearly two thirds of spot checks were triggered through a nationwide data mining-based targeting process. This increase in the proportion of data mining-related controls – which seem however to have reached a peak and evened out since 2015 – coincides with a drastic drop in the number of controls triggered by advisory staff (only 20% in 2016 compared to 51% in 2011), and less significantly, with a drop in the number of local targets or activity/resource checks (RACs).

Breakdown of spot checks concluded in year N according to the monitoring target

	2011 (%)	2012 (%)	2013 (%)	2014 (%)	2015 (%)	2016 (%)
Data mining	23	40	53	60	63	63
Local targets	8	5	3	5	6	6
Advisory staff	51	42	30	24	20	20
External flagging	3	3	3	3	4	4
Activity/resource checks	4	3	3	1	1	1
Other	11	7	8	7	6	8
Total	100	100	100	100	100	100

This evolution is in keeping with the national monitoring plans, which recommend always giving data mining priority over other means of triggering checks, including monitoring requests made by advisory staff, that are now limited in number. The introduction of data mining not only reduces the number of targets identified by CAF centre directors, but also field officer initiatives. These officers still play a part in combatting fraud, but they are now expected to turn their attention to the identification of residual cases that are not detected by the statistical tool. The technical solution does not replace officers' skills and know-how but it regulates their action and directs it toward the most complex cases.

The rising power of an automated risk-detection system has altered the nature of spot checks. Data mining and risk scores have given comptrollers a sort of black box device that is shrouded in mystery and that has some undesirable effects such as repeated checks performed year after year on the same welfare beneficiary, whose profile has been flagged, often wrongly, by the statistical tool. The objective of spot checks has been redefined, and the role of officer as councillor is being phased out in favour of officer as monitor. Through training sessions and also more indirectly through the generalization of fraud objectives assigned to them (that are also determined through data mining), comptrollers are encouraged to base their meeting with the social welfare beneficiary around the verification of benefits to which they are entitled and even on the detection of a deliberate anomaly, i.e. fraud. Since the cases verified are those with the highest score on the risk scale, and since the data mining targeting mode is designed to detect fraud, the comptrollers tend to presume the files they are auditing are fraudulent and to conduct their investigation accordingly.

CNAF usefully reinforces its guidance of local targeting practices by monitoring the integrated data used to provide beneficiary scores. The selection of files and situations to be monitored is updated every month through a more efficient survey of changes in beneficiary situations. This is made possible through the development of data sharing with the above-mentioned partner institutions. National network administrators also use the statistical models when they set local goals for CAF centres in terms of monitoring and fraud objectives, with a view to harmonizing the coverage of risk management nationwide. Therefore, the local CAF centre directors have limited flexibility, in principle, for the monthly adjustment of the amount of checks to be performed in accordance with the annual goals set at the national level.

These objectives are always difficult to obtain, especially in high-risk departments. More specifically, the objective is to align local CAF centres practices with those promoted at the national level. To achieve this, the 2013–2017 COG assigned a central role to the data mining process that had been implemented over previous years. At the organizational level, the goal is to limit local disparities and align practices nationwide, as evidenced by the development of indicators to track differences between CAF centres in terms of the return on monitoring, warnings and financial penalties. The indicator policy previously implemented was further strengthened, notably by the tracking of detected frauds showing the proportion of repeat offences, and especially the residual risks indicator which could be said to close the cycle of risk management logic.

Data mining has become the key instrument in the family branch's monitoring policy through the reconfiguration of roles within its central administration. The skillset of accounting officers has become broader and the data mining tool has made it possible to gear the verification of incoming data to the detection of fraud. This shows that the use of data mining has an effect on the economy of the monitoring policy.

5.4 Towards Increased Monitoring of the Most Vulnerable Populations

The application of data mining is, in theory, based on identifying the risk factors of undue welfare payments and not of at-risk populations. Furthermore, risk prediction models focus on the characteristics of the beneficiary's situation rather than on benefits received or how their files are managed. However, the multiplication of certain risk factors for vulnerable or isolated households results in them being mechanically flagged as prime monitoring targets. In this way, the progressive generalization of data mining in this domain leads to the underprivileged being overexposed to checks.

5.4.1 Risk-Prediction Models that Correspond to Welfare Beneficiary Characteristics

The identification and assessment of risks of undue payment studied here follows a well-defined protocol. This section will provide an overview of the statistical techniques and empirical material used by the CNAF statistics department to better understand, without revealing confidential data, the various forms of predictive risk modelling on which recipient scores are based. This will show how the models take into account certain distinctive situational characteristics of welfare recipient households.

The first challenge for statisticians was in choosing the characteristics or behaviours that presented a particular risk from the profusion of data available on welfare beneficiaries and their households. Two methods of statistical modelling were available: decision trees and logistic regression. The *de facto* use of the second method seems relatively common in the field of risk prediction (Peretti-Watel 2005). Moreover, it was in keeping with the initial experiment conducted by the Bordeaux CAF and it provided CNAF statisticians with much more empirical material. It is important to point out that all data on welfare beneficiaries surveyed was accessed on a single reference date – April 2009 – which made it easier to create a homogeneous database but led to observations based on attributes and behaviours situated at a given point in time, but which can evolve over time. However, exploiting this database made it possible to measure (until the database is renewed) the risks of undue payments made over a period of more than 6 months among more than 1500 possible variables – which represents a great diversity. These variables are infinitely multipliable due to the recoding carried out. For instance the "contact" variable is based on a subtraction of the number of beneficiary calls and visits to the CAF centre.

Various logistic regression models were then designed to identify the risks of undue payments. These are possible variations on the risks associated with recipients and their households based on their personal/social situations rather than the benefits they receive, even if these benefits are considered, *a posteriori*, to be relevant to the targeting process for future controls. There are five regression models. The first four models correspond to four types of risks linked to household resources (1); household composition according to civil status (2); housing status (3); professional status of adults in the household (4). The fifth model, called the "global" model, gather the characteristics most significantly associated with undue payments. Each model has its own specificities, beginning with the larger or smaller proportion of the sample they make it possible to define. In the model related to professional status, for instance, the jobless status of the spouse is particularly associated with a risk of undue payment. In the model taking into account the various household situations listed, the risk is twice as high when the recipient is single or a single parent than when the spouse has a regular income. Risk is also greatly increased when children over 18 years old live at home. As for the housing situation, for the few recipients who declare a spouse who is either a student or a pensioner, increase by 2.6 the risk of undue payment (particularly for housing-related benefits), compared to beneficiaries who declare no spouse. As for the resources-related risk modelling, these are shown to be higher when beneficiary income is low, particularly when the housing affordability ratio is over 35%. Finally, once again, the global model flags situations of unemployment, dependent child/children over 18 and low incomes. Added to these beneficiary household characteristics are file management variables (for instance a non-certification via the NIR), or variables related to ways in which welfare beneficiary-CAF contact (the difference between the number of calls and trips to the reception desk) and other form of payment than by bank transfer. Although the global model processes the greater part of the sample, risk-type modelling applies to smaller groups. This implies a more rapid drop in the significance of the correlations made when explanatory factors are given for undue payments. This is why these factors are found in smaller numbers than in the global model, which tends to reinforce the specific nature of the results. A new version of the global model is currently being developed for wider use from 2016. It is necessary to briefly set out the uses of these models that were developed from a sample of welfare beneficiaries to assign an undue payment risk score to all beneficiaries of the reference population. In fact, the creation of a teaching database on the data mining models made it possible to model the risk of undue payment flagged by a change of information related to housing, family situation, income or job status. The global model is also used. The results of three successive local experimental campaigns led the statistics department to focus on a specific way of combining previous logistic regression models to assign scores to all welfare beneficiaries nationwide. The score variable is generated from a set of five scores calculated by modelling per risk type: it is equal to the maximum reached by the scores in each risk subtype, or by the score within the so-called global model. *In fine*, a beneficiary with a score of 0.3 would have a 30% risk of being found to have received undue payment when monitored, which justifies the beneficiary's appearance on the comptrollers' priority list.

The steps of the process have now been identified: from theoretical models developed nationwide to beneficiary-targeting modes on the ground, and from the generalized attribution of a score to the local selection of beneficiaries to be controlled. In spite of some difficulties in appropriating data mining, and some local resistance to the generalization of this statistical tool, the process lead to increased targeting of the most vulnerable populations.

5.4.2 The Most Vulnerable Are Overexposed to Monitoring

From the identification of internal errors and fraudulent activity associated with certain benefits based on field officers' experience, we have arrived at a generalized method of risk prediction associated to the situational characteristics of some welfare beneficiaries. We are thus led to postulate a possible shift from a risk management policy to a form of at-risk population management. This is not a deliberate focus on the most vulnerable populations, who are also the most dependent on social benefits (Dubois 2010), but it would seem that the shift in the control-triggering logic is leading, *de facto*, to the most vulnerable beneficiaries being overexposed to monitoring – a trend that is increasing gradually. This over-verification can be defined as a greater probability, compared to the sum total of beneficiaries, of being targeted for a check. This gap grows in proportion to the intensity of economic and social hardships, particularly the frequent changes of situation that are characteristic of vulnerable populations (*INSEE* 2017).

In keeping with the diachronic perspective of our study, it is possible to show that this trend of targeting the most disadvantaged populations has increased during the course of the period under study, especially as regards welfare beneficiary home checks. The data collected enable us to compare the characteristics of the study population for the years 2006, 2010, 2013 and 2014. Although the trend toward increased targeting of the most vulnerable welfare beneficiaries can be seen in the requests for data sharing or document-based controls, it is most clearly shown in home checks. This appears to be in line with the goals of national monitoring plans, inasmuch as the productivity gains related to the generalized use of data mining were most expected with this type of check (see above). Staff involved in monitoring and anti-fraud were already focusing their efforts on the most vulnerable beneficiaries in the early 2000s (Dubois 2003). Nevertheless, the data collected enables us to show how this trend has grown and toward which beneficiaries specifically. These are seen to be targeted according to household composition, their age and the age of their child/children, and their income.

Regarding the welfare beneficiary household composition, it is clearly shown that single-parent families (essentially single mothers) tend to be increasingly over-monitored: there is a 29.2 gap in percentage between the proportion of single-parent families monitored at home over the year 2004 and the population of welfare beneficiaries as a whole, compared to only 16.4 in 2006. The age of beneficiaries and that of their child/children also appears as a determining factor, but the increase in spot checks is most striking in households with children aged 18–25 – in other words old enough to work but who have been flagged as liable to do not declare their income (only 2.2 percentage points with other beneficiaries compared to 24.4 in 2013 and 23.7 in 2014). Significant differences in the probability of home checks can also be observed according to professional status. Unemployed beneficiaries, or beneficiaries not in active employment, tend to be controlled more often, while the trend is reversed for people in employment. The same evolution can be noted for income-related checks. A major divide exists between beneficiaries with a monthly income under €500 per household consumption unit – which places them as earning below half of the poverty threshold at 60%[4] – and the others. The poorest were already over-monitored in the mid-2000s, but this trend has seen a sharp increase, notably in the case of home checks; the difference with the sum total of beneficiaries going from 25.1 percentage points in 2006 to 33.6 in 2014. (** Shift in 2015 over-monitoring of household between 500 and 1000).

[4] In 2015, the poverty threshold was 60% of the median standard of living, i.e. €987 per month. A growing number of households, however, are dropping away from this threshold, their income per consumer unit falling below 30% of the median standard of living (Fontaine and Sicsic 2015).

Distribution of home checks in 2006, 2010, 2013 and 2014 (in percentage points)

	2006	2010	2013	2014
Family situation				
Single and childless	9	5.4	−9	−11.6
Childless couple	1.2	0.4	−0.6	−0.9
Couple with children	−28.9	−27.1	−17	−17.9
Single-parent family	16.4	18.1	25.2	29.2
Age of head of household				
Under 30	5.6	3.9	−6.6	−3.7
30–39	−4.2	−2.5	−3	−2.6
40–59	0.6	0	8	7
60 and over	−4.9	−4.8	−5.1	−6.6
Age of child/children				
0–13	−13.2	−10.6	0.7	3.7
14–17	−5.1	−3.7	7.3	7.4
18–25	2.2	1.6	24.4	23.7
Professional status of head of household				
Employed	−21.9	−19.4	−21	−22.7
Self-employed	−0.6	0.2	1.7	1
Extended sick leave	−0.9	−0.1	1.2	1.2
Pensioner	−4.7	−4.8	−5	−6.5
Student	−3.3	−3.8	−5.5	−5.8
Unemployed	11.1	7.3	9	8.2
No status	18.6	18.7	20.7	26.1
Monthly income per consumer unit (in euros)				
0–500	25.1	21.6	31.1	33.6
500 and over	−29.3	−22.6	−30.8	−32.2

Source: CNAF – ALLNAT monthly files
N.B. *The proportion of single and childless beneficiaries within the audited population is 9.9 percentage points higher than that found in the total population of beneficiaries*
Field: *the entire population of beneficiaries in 2006, 2010, 2013, 2014*

Certainly, there are still checks following suspicious behaviour during a CAF centre visit, the doubts of an advisory officer, flags by partner administrations or private reports of benefit fraud. Arguably, this checking is more frequent in the most disadvantaged populations as people living in more unstable conditions are more prone to mistakes and lateness when providing information. However, the generalization of data mining is the element that most systematically tends to orient a monitoring that is targeted at beneficiaries with the most acute social and economic hardships. The application of statistical instruments according to socio-demographic characteristics logically leads to a situation in which the beneficiaries most often monitored are those who receive resource-based welfare benefits and among these, the solidarity benefit recipients. This is not an *a priori* decision, as in the case of the localized targeting of RSA recipients – with the advent of data mining, these operations become all the more outdated and give a negative image – but because statistical models that are outwardly neutral and field-tested identify more risks factors in the populations concerned.

5.5 Conclusion

The use of data mining, and the risk management policy of which it is a key element, fits into a rationalization process that associates instrumental and value-rational action, as defined by Weber. The objective is twofold: to combat social fraud, the immorality of which is regularly denounced publically as the antithesis of a work ethic, and to develop new ways of limiting social spending. This rationalization process is profoundly political because of the highly political nature in which it was triggered, or at least accelerated. Indeed, there was the calling into question of the traditional forms of the welfare state criticized for being lax, and the corresponding accusations of welfare profiting. Though intrinsically political, this process appears to be technical. It combines (i) legal reasoning implying strict application of the rules to ensure the proper and rightful payment of benefits (Buchet 2013), (ii) management reasoning implying the correct allocation of resources, and (iii) statistical reasoning serving the first two and integrating the probabilistic approach into welfare institutions (Desrosières 1993).

This rationalization, of which data mining is both the product and vehicle, is conducted in the name of accuracy and efficiency and is not socially neutral. It leads to a situation that was created unintentionally, but not without reason, in which the underprivileged have become particular targets for monitoring that should not be confused with the fight against fraud and penalties, but that are a big part of it nonetheless. The current rationalization therefore goes hand in hand with coercive measures toward the most vulnerable. The increase in data mining is accompanied by an increase in the judicialization of fraud, a development that is part of a larger movement to judicialize social policies. The social impacts of this judicialization of social welfare benefits fraud must by analysed, all the more so since its target population increasingly seems to be predominantly composed of vulnerable welfare beneficiaries, and since it is extremely difficult for the underprivileged to access justice (Galanter 1974). As in tax matters, there are more stringent penalties and more resources put towards the surveillance and detection of fraudulent behaviours in low-income households than in high-income ones (Spire 2012).

As we have seen, a statistical instrument such as data mining does not determine by itself its uses, which are defined in its social and political implementation process. The conditions in which data mining was transferred into the French social welfare system have turned it into a technology geared towards risk management, one that is primarily defined by the imperative to control and to limit social spending. Nonetheless, there is nothing preventing it from being used to improve access to social rights, and to combat the non-claiming of social benefits by targeting potential beneficiaries through non-claiming predictive factors (Warin 2016). This is done in Great Britain and in the Netherlands (Zuurmond 2008) and it has its advocates in France,[5] but it would need to overcome the highly unfavourable social, political and financial conditions.

[5] See for example this recent parliamentary report: Gisèle Biémouret, Jean-Louis Costes, Rapport d'information sur l'évaluation des politiques publiques en faveur de l'accès aux droits sociaux, n°4158, Assemblée Nationale, Paris, 2016.

References

Baudot, P.-Y. (2011). L'incertitude des instruments. L'informatique administrative et le changement dans l'action publique (1966–1975). *Revue française de science politique, 61*(1), 79–103.

Buchet, D. (2013). Le paiement à "bon droit" des prestations sociales des CAF. *Informations sociales, 178*, 97–103.

Castel, R. (1981). *La Gestion des risques. De l'anti-psychiatrie à l'après-psychanalyse*. Paris: Minuit.

Collinet, P. (2013). Focus – Le data mining dans les CAF: une réalité, des perspectives. *Informations sociales, 178*, 129–132.

Desrosières, A. (1993). *La politique des grands nombres*. Paris: La Découverte.

Desrosières, A. (2014). *Prouver et gouverner. Une analyse politique des statistiques publiques*. Paris: La Découverte.

Dubois V. (2003), *Les conditions socio-politiques de la rigueur juridique. Politique de contrôle et lutte contre la fraude aux prestations sociales*, CNAF, Dossier d'études, n° 48.

Dubois, V. (2010). *The bureaucrat and the poor: Encounters in French welfare offices*. London: Routledge.

Fontaine, M., & Sicsic, M. (2015). *« Des indicateurs précoces de pauvreté et d'inégalités – Résultats expérimentaux pour 2014*. Insee Analyses, 23.

Galanter, M. (1974). Why the Haves come out Ahead? *Law and Society Review, 9*(1), 95–160.

Harcourt B. (2005). *Against prediction: Sentencing, policing, and punishing in an actuarial age*. University of Chicago Public Law & Legal Theory Working Paper, 94.

Harcourt B. (2014). *Digital security in the expository society: Spectacle, surveillance, and exhibition in the neoliberal age of big data*. Columbia Public Law Research Paper, 14–404.

INSEE. (2017). Les allocataires contrôlés par les caisses d'Allocations familiales en 2015, *L'essentiel*, *169*.

Lazarus, J. (2012). Prévoir la défaillance de crédit: l'ambition du scoring. *Raisons politiques, 48*, 103–118.

Lipsky, M. (1980). *Street-level bureaucracy: Dilemmas of the individual in public services*. New York: Russell Sage Foundation.

Ollion, É. (2015). L'abondance et ses revers. Big data, open data et recherches sur les questions sociales. *Informations sociales, 191*, 70–79.

Parton, N. (2008). Changes in the form of knowledge in social work: From the social to the informational? *British Journal of Social Work, 38*(2), 253–269.

Peretti-Watel, P. (2005). La culture du risque, ses marqueurs sociaux et ses paradoxes. Une exploration empirique. *Revue économique, 56*(2), 371–392.

Spire, A. (2012). *Faibles et puissants face à l'impôt*. Paris: Raisons d'agir.

Wacquant, L. (2010). Crafting the neoliberal state: Workfare, prisonfare, and social insecurity1. *Sociological Forum, 25*(2), 197–220.

Warin, P. (2016). *Le non-recours aux politiques sociales*. Grenoble: Presses universitaires de Grenoble.

Zuurmond, A. (2008). Bureaucratic bias and access to public services: The fight against non-take-up. In J. De Jong & G. Rizvi (Eds.), *The state of access. Success and failure of democracies to create equal opportunities* (pp. 148–166). Washington, DC: Brookings Institution Press.