FIRST, DO NO HARM

Ethical Guidelines for Applying Predictive Tools Within Human Services

SUPPORTED BY
THE ANNIE E. CASEY FOUNDATION
Introduction (Why is this an important issue?)

Predictive analytical tools are already being put to work within human service agencies to help make vital decisions about when and how to intervene in the lives of families and communities. The sector may not be entirely comfortable with this trend, but it should not be surprised. Predictive models are in wide use within the justice and education sectors and, more to the point, they work: risk assessment is fundamental to what social services do, and these tools can help agencies respond more quickly to prevent harm, to create more personalized interventions, and allocate scarce public resources to where they can do the most good.

“There is a misapprehension that sometimes evolves, that somehow involving predictive analytics in this process can eliminate structural bias. To the contrary, it may just make those problems less conspicuous by intermediating them with a computer.” —Logan Koepke

Governments, in particular those with constrained resources, are looking for better tools to be able to identify where services are going to be needed and when.”

—Andrew Nicklin

There is also a strong case that predictive risk models (PRM) can reduce bias in decision-making. Designing a predictive model forces more explicit conversations about how agencies think about different risk factors and how they propose to guard against disadvantaging certain demographic or socioeconomic groups. And the standard that agencies are trying to improve upon is not perfect equity—it is the status quo, which is neither transparent nor uniformly fair. Risk scores do not eliminate the possibility of personal or institutional prejudice but they can make it more apparent by providing a common reference point.

That the use of predictive analytics in social services can reduce bias is not to say that it will. Careless or unskilled development of these predictive tools could worsen disparities among clients receiving social services. Child and civil rights advocates rightly worry about the potential for “net widening”—drawing more people in for unnecessary scrutiny by the government. They worry that rather than improving services for vulnerable clients, these models will replicate the biases in existing public data sources and expose them to greater trauma. Bad models scale just as quickly as good ones, and even the best of them can be misused.

The stakes here are real: for children and families that interact with these social systems and for the reputation of the agencies that turn to these tools. What, then, should a public leader know about risk modeling, and what lessons does it offer about how to think about data science, data stewardship, and the public interest?

“If used incorrectly, these tools can let people off the hook—to not have to attend to the assumptions they bring to the table about families that come from a certain socioeconomic background or are of a particular race and ethnicity.” —Tashira Halyard
Audience, Purpose, and General Outline

This report is intended to provide brief, practical guidance to human service agency leaders on how they can mitigate the risks that come with using predictive analytics. It is structured around four principles—engagement, model validation, review and transparency—each with specific cautions and recommendations, and includes discussion of several foundational strategies related to data governance, crisis communications, and partnership that are important to any new application of predictive risk modeling.

In order to focus on high-level principles that we recommend should guide every human service agency’s thinking about ethical dimension of predictive analytics, this report touches only lightly on topics that others have explored in great detail. We will take the opportunity throughout the text to refer you to their work.

Finally, because this is a fast-evolving field, we recommend readers to the latest guidance available from the individuals and organizations who contributed to this report.

Thanks to this group of data scientists, policy advocates and public leaders for being so generous with their time and expertise.

Robert Brauneis, George Washington University
Alexandra Chouldechova, Carnegie Mellon University
Brian Clapier, City of New York, Administration for Children’s Services
Erin Dalton, Allegheny County, Department of Human Services
Ellen P. Goodman, Rutgers University
Tashira Halyard, Alliance for Racial Equity in Child Welfare at the Center for Study of Social Policy
Carter Hewgley, District of Columbia, Department of Human Services
Bill Howe, University of Washington
Eric Miley, Annie E. Casey Foundation
Andrew Nicklin, Center for Government Excellence
Martin O’Malley, 61st Governor of Maryland & 47th Mayor of Baltimore
Kathy Park, National Council on Crime & Delinquency
David Robinson and Logan Koepke, Upturn
Jennifer Thornton and Amber Ivey, Pew Charitable Trusts
Jeff Watson, Hoonuit

A digital version of this report with resource web links can be viewed and downloaded as a PDF here: http://bit.ly/2hdJH4U
Four Principles for Ethically Applying Predictive Tools Within Human Services

Engage p.6

Internally:
- Enlist leaders throughout the system
- Connect your “quants” and your domain experts
- Train your front-line staff
- Do not delegate this training to technologists, consultants or vendors

With the Community:
- Begin early — ideally, before now
- Include your most skeptical partners
- Ensure that the public understands what you are trying to achieve with PRM
- Be explicit about how data is being used and protected
- Encourage feedback on the process

(Pre) Validate the Model p.8
- Use your agency’s own historical data
- Interrogate that data for bias
- Ask for multiple measures of validity and fairness
- Treat variables like race with great care—but do not eliminate them from the dataset
- Pilot the model and expect the unexpected

(Re)view p.11
- Build evaluation into the project plan and budget
- Clearly assign responsibility for using, monitoring and refreshing the PRM
- Create a system of checks and balances
- Check that software is augmenting—not replacing—social workers’ professional judgement
- Make changes

Open Up p.12
- Include a plain-language explanation of the agency’s intent
- Publish a description of design choices
- Use your leverage during procurement
- Get the log file
- Insist on free access to the model
- Be proactive in sharing all of this information
- Prefer models that generate explainable results
This report takes it as well-established that predictive analytics can be extremely useful to public leaders and that they will become increasingly integrated into human services agencies in the coming decade. But predictive modeling is only one of many ways that human service agencies leverage information to improve their result—and it is one of the more narrow and complex.

Some of these strategies fall approximately into the category of “data science” or what the Administration for Children’s Services in New York City calls “advanced analytics”—statistical modeling, randomized trials, comparison testing of different program designs and so forth. However, many fundamentally important uses of evidence to improve agency decision-making do not require this kind of sophistication. Human services agencies wanting to build a culture that values data can start simple: connect information across departments, get feedback to front-line staff, use business intelligence tools to analyze trends, and evaluate existing programs. One of the most celebrated examples of a local government using data to drive better results, Baltimore’s CitiStat, was initially built from nothing fancier than the Microsoft Office suite, a podium, and a projector.

The bottom line is that every use of predictive analytics comes with its own significant costs, ethical questions, and technical pre-requisites. It is worth an agency spending time at the outset to check that it has chosen the most appropriate tool for the job it has in mind. For a “back of the envelope” judgement of whether a predictive tool is a good fit, consider whether it meets the tests on page 6.

**WHAT IS PREDICTIVE RISK MODELING?**
PRM uses statistical methods to identify patterns across large, diverse datasets and then uses those patterns to estimate the likelihood of a more positive or negative outcome for a client, or group of clients with similar characteristics. These models can be more or less sophisticated, and are might be labeled “early warning systems”, “risk assessments”, “rapid safety feedback tools”, or something else, depending on the specific application, field and vendor. Their increasing relevance within the public sector is a consequence of several trends, including greater digitization and sharing of government records, advances in statistical techniques such as machine learning, and increased pressure on agencies to use data to achieve better and more cost-effective results.1

The need arose from engagement with staff, not from a sales pitch. Adapting a new predictive model for agency use is a major undertaking and is much more likely to be successful if the tool solves a clear problem for program managers and front-line workers. Consider these staff your customers and be skeptical of solicitations from outside of the agency.

The agency’s objective for PRM is reasonable and modest. Predictive analytics can improve the odds of a good decision; it cannot replace the professional judgment of agency staff or mitigate the government’s responsibility for human error or negligence. Agencies trying to implement PRM to shift or eliminate legal liability are likely to be disappointed.

The problem PRM is being considered for as a solution is well-described, with a clear decision point, evidence-based intervention in mind, and an articulated measure of success. For example, a child protective services agency might apply predictive techniques to try to more accurately screen incoming calls reporting possible abuse and neglect (as within Allegheny County) or to predict the likelihood of injury to children already in the system (as with Florida’s use of the Eckerd Rapid Safety Feedback® tool). The feasibility, efficacy, and possible ethical consequences of the underlying model are extremely dependent on the details of the use case and cannot be evaluated until an agency is ready to be specific.

The agency has—or can partner to get—the support it needs. Because human service agencies fundamentally own responsibility for the consequences of the models they put into service, it is crucial for agency leaders to confirm that they have the resources—expertise, commitment from leadership, access to quality data, etc.—not only to do this, but to do this right. The remainder of this document is a guide to what that entails.

“The real game changer, in some cases, is academic partnership. We realized here in DC that there was an alignment between what a university wanted to accomplish with its students and research agenda, and what the city needed in terms of its expertise—there was something we could give them to chew on that served both our purposes. I’m seeing more and more examples of this kind of success.” —Carter Hewgley
Engage

Agencies cannot design and implement a predictive risk model without actively soliciting the contributions of both agency and community stakeholders. Agency leaders and their data team do not, by themselves, have the expertise and context to ask all the right questions to root out the potential for bias and possible misuse of these tools. Their success depends on staff buy-in, which can only be achieved through an iterative process of design, testing, and consultation. Additionally, community and family advocates, if not authentically engaged as partners in the development and oversight of new predictive tools, may well use their influence to have these tools thrown out.

It is up to public leaders to clearly explain what predictive risk modeling will do, what it will not do, how the data will be protected, and to directly acknowledge some of the concerns associated with data bias.

---

“I think you have to open the model up to criticism, and you need to do that face-to-face. Because the most-informed person about the practice that you’re trying to predict is probably not a computer scientist. It’s somebody who has been doing this work for 20 years, is in the field, knows these clients’ names, and knows inherently where the biggest risks are. They’ve just never realized they’re doing that math formula in their head.” – Carter Hewgley

---

This starts at home, within your agency:

- **Enlist leaders throughout the system** to ensure that they are aware of the project as it develops and that they are briefed on the ethical dimensions of using predictive tools. While senior leadership may be necessary to get a new initiative off the ground, “everything lives or dies in middle management” and civil rights organizations will want assurances from the C-suite to be institutionalized through departmental policies.

- **Connect your “quants” and your domain experts.** Without a regular line of communication between your data scientists and the managers and front-line staff who generate the data they rely on, the resulting predictive tool is unlikely to be useful or used. Create opportunities for your model designers to test and refine their assumptions, and to ferret out decision points where a naïve use of the tool might introduce or replicate existing bias.
Train to reduce the deskilling of your front-line staff. PRM is intended to augment, not replace, for the professional judgement of caseworkers; staff who have contextual knowledge, relationships and ethical responsibility that cannot be delegated to an algorithm. To exercise that judgement, they need more than a risk score—they need to know how the model works and what its limits are. They need to know how much more likely a client with a risk score of ‘7’ is to become homeless than a client scored a ‘5’, and to have a sense of the distribution of these risk scores across all of their clients. Finally, they need policy guidance about when they should override the recommendation of a PRM, and how exercising that discretion will be factored in to discussions about their job performance.

Do not delegate this training to technologists, consultants or vendors. Be sure you have someone with deep knowledge of the relevant human system prepared to interpret these new tools for staff. PRM is not something that should just be treated as a data project.

TO ENGAGE COMMUNITY PARTNERS:

Begin early—ideally, before now. This conversation about predictive risk modeling should not be the first time an agency is engaging with the community. If it is, strongly consider a “pause” to develop those relationships before testing them with such a complex and sensitive topic.

Include your most sceptical partners, such as from civil rights organizations and privacy advocates. Validate their concerns and demonstrate your seriousness in addressing them as part of your model’s design. If you need to look beyond your local partners, reach out to national advocates with credibility and experience in this space.

Ensure that the public understands what you are trying to achieve with PRM. (See “Open Up”, below.) Work with community and family advocates to craft straightforward explanations of the problem these tools are helping to solve, stripped of technical jargon and “bureaucracy speak.”

Be explicit about how data is being used and protected, and how you are negotiating the question of consent for any data sharing that is taking place. This will help reduce the likelihood of reputational damage to the agency. (See the “Crisis Communications” section.)

Encourage feedback on the process, keep a phone line open, and communicate twice as often as you think you need to with important stakeholders. Projects like this need a public champion, though this sometimes runs contrary to the instincts of public agencies to “go dark” while in the process of developing new systems and practices.

For example, the Alliance for Racial Equity in Child Welfare Services at the Center for Study of Social Policy, National Council on Crime and Delinquency, and the Leadership Conference of Civil and Human Rights at bigdata.fairness.io.
(Pre) Validate the Model

The first test for any predictive model bound for use in human services is whether it can provide a more accurate and fair assessment of risk than the status quo. This is an excellent time for public managers to brush up on Mark Twain’s three kinds of lies (lies, damn lies, and statistics) and consider several fundamental design choices that will affect whether the model they are building is trustworthy.

- **Use your agency’s own historical data to train and test the model.** There is no such thing as a “validated tool”—only validated applications of a tool. Differences in an agency’s client characteristics, departmental policies, and record-keeping will all affect the reliability and fairness of a model’s risk core. The more complex the model, the truer this is. Beware anyone who tries to sell you a tool without careful consideration of how to tailor it to your local community.

- **Interrogate that data for sources of bias.** “BIBO (Bias In, Bias Out)” is the predictive equivalent to “GIGO (Garbage In, Garbage Out).” If a risk model is trained against data that is not representative of the phenomena it is meant to evaluate—if the data is systematically inaccurate or oversamples certain groups—the model is likely to replicate those biases in its own risk predictions. Measures of statistical parity can help catch these problems, but the surest way to uncover them is through conversation with line staff and client advocates. Be suspicious of any conversation about PRM in human services that does not demonstrate a recognition of the historical biases present in these administrative data systems.

- **Ask for multiple measures of validity and fairness.** Data scientists habitually use a single measurement—the area under the ROC curve (see next page)—as the most important indicator of a model’s validity. Public leaders should understand the limits of this “AUC” score and be sure they understand more deeply how risk is distributed and classified at the point where they plan to intervene. How effective is the model at distinguishing among the 25% of the population most at risk, for example, and what are the specific consequences of false-positive to a client and to the agency?

“To do this work, you have to acknowledge the biased nature of administrative data. Administrative data reflect both structural and institutional bias and are also influenced by individual cognitive bias, which drives certain families into the system more than others.” —Kathy Park
A ROC curve is a measurement of how likely a model is to correctly identify risk. Data scientists—particularly the machine learning community—frequently use an estimate of the area under this ROC curve (or “AUC”) as a way of comparing different models to one another. You can think of the AUC as the probability that a model will correctly order the risk of two cases, assigning a higher score to the riskier of the two. It is a measure of a model’s ability to discriminate. It does not, however, tell us anything about the level or distribution of risk in a population.

For example, a model applied to the three risk distributions illustrated here might generate the same AUC score—their risk order is identical. But the level and dispersion of that risk is so different that, very likely, each of these would demand different government responses. This calibration of risk is just as important as a model’s discrimination.

Remember, predictive validity is assessed by two measures: discrimination and calibration. An AUC score only speaks to how well a model discriminates between levels of risk, not how well it is calibrated, and public leaders cannot rely on this single measure to evaluate whether a model is a good fit for the purpose they have in mind.
Treat variables like race with great care—but do not eliminate them from the dataset. Protected attributes include both personal characteristics that have been vectors of historical discrimination—like race, gender, poverty, or disability—and behaviors the government wants to be careful not to stigmatize, such as seeking mental health services. Data scientists cannot simply avoid biasing their model by excluding these underlying data, as risk models are extremely likely to “rediscover” protected variables through their close correlation with other factors. (For example, in many communities, a client’s race will be highly correlated with her zip code of residence.)

Where possible predictive models should always capture these variables and report on their relative treatment by the risk model and subsequent service decisions made by staff, in order to strengthen the government’s ability to detect and correct for latent institutional biases in the data. Data scientists should exercise extreme caution, however, in using any of these variables as a risk factor, and do so only after discussion with the affected communities. Including protected variables as predictive factors can sometimes increase a model’s accuracy and occasionally improve its fairness, but this should be carefully tested and monitored.1

“At the end of the day, if the algorithm doesn’t treat key protected classes equally—however measured—it almost doesn’t matter how the bias crept in.” —Bill Howe

Pilot the model and expect the unexpected. How a model is used by front-line staff is the final test of its validity, before any agency-wide rollout. Usually, it will be important that any feedback the provided by the model can fit naturally into case workers’ established routines and not require extensive retraining. Agencies should test how staff assess risk for the same clients both with and without the new risk model, using this “burn in” period to look for patterns that indicate changes that may need to be made in how feedback from the model is introduced into agency decision-making. A model’s fairness cannot be evaluated separate from its use.

PRINCIPLES

1 Protected attributes may, in some cases, capture real cultural differences, and these cultural differences may actually be important in understanding both the risk and the needs of the population. Models tailored for majority populations may need to be retested for validity against minority groups.
(Re)view

There is no such thing as a “set it and forget it” approach to predictive analysis in human services. At the most basic level, models must routinely incorporate new information collected by agency programs. More pointedly, the very act of using a predictive model to intervene with clients and change their outcomes is likely to alter the underlying distribution of risk and, over time, change or invalidate the original model. Only if they actively monitor and “tune” these tools can agencies to ensure they remain accurate and fair.

- **Build evaluation into the project plan and budget.** Public leaders interested in PRM must not only build and deploy them, but also measure and refine these models over time. Structuring the project to include assessment as a significant activity is important, and might require as much of 50% of the total project’s resources. Consider calling a halt to development half-way through the “resource burn” and ceasing the development of new tools and analyses to focus on testing and evaluation. And that evaluation should begin immediately.

- **Clearly assign responsibility for using, monitoring, and refreshing the PRM.** Both data scientists and family advocates worry about the potential for predictive analytics to inadvertently “sanitize or legitimize” bias by making accountability for these important decisions ambiguous. If there appears to be a systemic problem with the results of the model, is the fault with the underlying data? The software? Or the agency at fault for not catching the error? Who, ultimately, can the public rely on for redress? Predictive analytics should be implemented as part of a larger governance strategy that treats data as a strategic asset and assigns responsibility for overseeing predictive tools to a group that includes senior public leaders and public advocates with access to technical expertise. (See “Data as a Strategic Asset” Sidebar)

- **Create a system of checks and balances.** At a minimum, ensure one or more people with the necessary expertise are in a position to evaluate gains in efficiency and decision-making, scan for signs of disproportionate impact on certain communities, and make this information available under controlled circumstances for scrutiny from advocates and stakeholders. Take precautions to ensure the independence of these periodic reviews by contracting with university staff either outside of the agency or with the private firm responsible for the model’s development. Negotiate in advance to allow agency-designated researchers the access to the model and outcomes they need to conduct this kind of review.

“It’s one thing to hope and assume these projects will go as planned. It’s another thing to systematically measures and cross-check our assumptions about what the impact of these tools is going to be.”  – David Robinson

Private auditors may increasingly be an option for public agencies concerned about the impact of these predictive models. E.g., see [O’Neil Risk Consulting & Algorithmic Auditing](https://oneilrisk.com), recently launched by the author of *Weapons of Math Destruction*.
Check that software is augmenting—not replacing—social workers’ professional judgement. Monitoring is important both because the characteristics of the population being evaluated for risk will change over time and also because these predictive tools themselves change how agencies make decisions. Staff may over-credit the reliability of machine-generated risk scores or, in extreme cases, begin to make decisions about care “by rote”. Processes for regularly training and soliciting feedback from staff can guard against an otherwise well-designed model failing due to poor implementation.

Make changes. Work with your implementation partners to periodically use this feedback to revisit the model to adjust how it weighs risk, presents feedback to staff, and benchmarks its accuracy and fairness. This regular review not only verifies that your model is delivering fair results right now, but creates the strongest possible case for its continuation into future administrations.

“Opacity in algorithmic processes, when they have real welfare effects, is a problem in all contexts. But in the case of public governance, it poses particular dangers to democratic accountability, to the efficacy and fairness of governmental processes, and to the competence and agency of government employees tasked with doing the public’s work.”

– Algorithmic Transparency for the Smart City by Ellen P. Goodman and Robert Brauneis in the forthcoming Yale Journal of Law and Technology

Open Up – Model Transparency

One of the core concerns about the expansion of predictive analytics within the public sector is the fear that citizens will lose effective oversight of some of the most important decisions government makes about their welfare. Predictive models are (fairly) criticized for often being “black boxes” whose inner workings are either protected by proprietary code or, increasingly, created through machine learning techniques which cannot easily be described by even their creators. Everybody with a stake in the debate—policy makers, government officials, advocates, data scientists and technology companies—agrees and acknowledges that greater transparency about the use of predictive risk modeling within human services is important. But disagreements remain about how that transparency should be defined and observed.
Focusing on algorithmic transparency is insufficient. The math that is implemented by software code is just one element of a predictive risk model, and not the most important. Agency leaders should ask for it—and some insist on it—but access to source code alone does not expose the more likely sources of bias and error in predictive analytics to scrutiny.

“It’s really about transparency of the whole system that you’re building, not just the software. It’s the data that goes in, it’s how you’re processing that data, it’s how you’re using that predicting to make decisions and it’s the impact you’re having on clients.” – David Robinson

Instead, governments should embrace an approach to transparency in predictive risk modeling that accounts for the broader development and use of these tools. This involves negotiating with software developers to ensure that agencies retain maximum ownership of these models and their derivatives, explaining trade-offs made in each model’s design, and reporting results to make a public case for how these tools are improving outcomes for all clients. Local governments have a number of practical steps they can take to be open about both of the critical elements of this work: the predictive model in use, and the process that created and governs it.

- **Include a plain-language explanation of the agency’s intent** (for example, the specific outcome the agency is trying to achieve through its use of predictive analytics) in all solicitations, contracts, operating rules, authorization legislation, and so forth. Take every opportunity to provide context and framing for this work.

- **Publish a description of design choices**, including tradeoffs made to balance model accuracy against ethical concerns. Document the data used to design the model, its basic operation, and oversight. Include a discussion of major policy decisions; for example, about whether or not to include protected attributes or to use machine learning.

- **Use your leverage during the procurement process** to push back against claims of trade secrecy and to limit the scope of non-disclosure agreements. Require contractors to explain in writing their key model design choices, model validation, and protections against unintended or biased outcomes. Agencies have more power to insist on openness and transparency than they may think – but they must exert it.

For a particularly comprehensive example from child welfare services, see the report created by Allegheny County’s Department of Health and Human Services, *Developing Predictive Risk Models to Support Child Maltreatment Hotline Screening Decisions*. A paper forthcoming in the *Yale Journal of Law and Technology* by Ellen P. Goodman and Robert Brauneis, provides an excellent list of elements for public leaders to ask be disclosed. The authors have also used open records acts requests to states and counties to demonstrate the strikingly different levels of transparency government agencies have been able to negotiate from vendors—and even from the same vendor.
“Without free access to the model, how can you understand how its predictions would have differed from the decisions that your agency actually made? How do you know that what you’re building or buying is any better than what you have?”
- Alexandra Chouldechova

- Get the log file. Every model should create records documenting its use, including information on how agency users interact with the decision support tool. Define the elements of this log file in advance with your data team and vendor, bearing in mind what aggregate performance measures your governance body may need to periodically evaluate the performance of the model. Require that these records belong to the government and can be shared with important stakeholders without any encumbrance from trade secrecy claims.

- Insist on free access to the model, if it is being operated by a commercial vendor. Ensure your agency can run its historical data against the model to test how its predictions would have differed from the decisions that were actually made by agency staff.

- Prefer models that generate explainable results. Though data scientists are working to create ways for people to better-interpret the output of machine learning algorithms, the bottom line is that it remains “stupid hard” to explain the decision-making criteria of many models designed this way. This matters, particularly when the government is accountable for defending or providing redress for a decision that substantially affects a person’s well-being, such as whether they are eligible for parole. In addition, many human services interventions are non-binary: it may be necessary not only to identify clients at risk of homelessness, for example, but to understand why a client is vulnerable in order to successfully intervene. Despite these drawbacks, it may still be preferable to use machine learning techniques where they are much more accurate. But there are good reasons to prefer explainable models, in close cases, or to take additional steps to reverse engineer a sense of the specific factors influencing a model’s results.

- Be proactive in sharing all of this information with close partners and community advocates. Fulfilling open records or Freedom of Information Act requests to answer these questions is a generally a sign that government has failed its duty to be transparent. The scrutiny these policies receive from many eyes is a second “failsafe” against the chance these tools will do unintended harm to vulnerable populations. Government should welcome this scrutiny.

For an in-depth exploration of these issues, we recommend the work of a growing community of researchers and practitioners organized under the banner of “Fairness, Accountability, and Transparency in Machine Learning (FAT/ML)” at www.fatml.org.

2 See The State of Explainable AI by Jillian Schwiep
Smart engagement with the public about government’s use of predictive techniques is not only about building a better and fairer risk model—it’s about heading off the possibility of reputational damage to your agency. The embrace of “big data” in related sectors has been mixed at best, spawning unsubtle headlines in the press (Courts are using AI to sentence criminals. That must stop now.) and, within the education sector, several firings, failed philanthropic initiatives, and a flurry of state legislation.

Validate the concern. Position your agency as an active advocate for the oversight and auditing of advanced analytical tools like this. The most important message to share is that “we’ve heard you.” Reacting defensively to accusations of negligence or bias is the most natural government response—and also the most destructive to public trust.

Create a plain language description of the initiative and ensure all public-facing staff can cite it. This short document should include what the agency is hoping to achieve, why it chose this model, a simple description of how the model works and of the safeguards in place to protect client privacy and create fairer results. Words matter—use language that is relatable to your field and that is not likely to trigger unmerited concern: for example, “early warning” in education. Emphasize the ultimate autonomy of public employees; these tools augment the discretion of social workers, they do not replace it.

Be an enthusiastic conduit for the public to get as much additional information as they need (but make it easy on yourself). Expect open records requests and have a process for easily meeting them; expect ethical/legal challenges and be prepared to outline how contracting process was conducted, explain what assurances were provided, provide aggregate information on the actual use of the model, and so forth. Having a credible data governance group helps to reduce and prioritize complaints, and is an important oversight measure in its own right.

Don’t oversimplify or overpromise. Be honest about the limits and risks associated with a particular implementation of PRM, and be clear that this implementation is part of a process that will need to adapt and improve. Consider what assurance the agency can offer that the benefits to vulnerable children and families are worth this risk.

Finally, have a rapid response plan, and be sure your senior leaders and public information officers know it. Even a very good risk model will produce false negatives, just as the previous system did, and a report to child protective services that was screened out by the new system may nevertheless precede injury to a child. Plan how to respond to that, and to address concerns about the accountability of the agency to negative outcomes as well as positive.

These backlashes were not primarily a consequence of bad government behavior. They reflected a failure of public leaders to effectively communicate that they are behaving responsibly, and to gracefully negotiate the legitimate concerns and fears that the public and policymakers may have about the data collected by government is used. Human services agencies may have the chance to learn from others’ mistakes here. Anticipate these critical questions and:
Predictive analytics are just one example of a broader set of innovations within government that rely on smarter use of data. In most cases, these data are already collected by different agencies as a part of funding, managing, and accounting for public programs. But, as the Pew Charitable Trusts points out, “collecting data is not the same as harnessing it” and public leaders have only recently begun to use this information as an important resource for forward-looking decision-making rather than an accidental byproduct of compliance activities.

Risk modeling, outcomes-oriented contracting, social impact bonds, and using behavioral insights to increase program uptake—all of these innovations are built on a foundation of policies that answer critical questions about the ownership, management, sharing, integration and protection of information across the government enterprise. The data governance structure established by an agency clearly identifies who is accountable and who is responsible for uses of public data, and for ensuring the development of new tools like predictive risk models are guided by a set of shared principles (like the four suggested in this report).

“Treat your data like an enterprise asset, in some ways as important as your people. Key decisions about how to protect it—retain appropriate access, oversight and control over it—follow from this underlying approach.” – Carter Hewgley

For recommended data governance practices in local government and the human services sector, see:

**Focus Areas for a Data Governance Board**, from What Works Cities.

**IDS Governance: Setting Up for Ethical and Effective Use of Data**, from Actionable Intelligence for Social Policy at the University of Pennsylvania.

About MetroLab

MetroLab Network introduces a new model for bringing data, analytics, and innovation to local government: a network of institutionalized, cross-disciplinary partnerships between cities/counties and their universities. Its membership includes more than 35 such partnerships in the United States, ranging from mid-size cities to global metropolises. These city-university partnerships focus on research, development, and deployment of projects that offer technologically- and analytically-based solutions to challenges facing urban areas including: inequality in income, health, mobility, security and opportunity; aging infrastructure; and environmental sustainability and resiliency. MetroLab was launched as part of the White House’s 2015 Smart Cities Initiative.

In 2017, MetroLab launched its Data Science and Human Services Lab, as an effort to bring together academics, city and county practitioners, and non-profit leaders to consider the issues at the intersection of technology, analytics solutions, and human services deployment. MetroLab’s activity in this domain, including the development of this report, is supported by the generosity of the Annie E. Casey Foundation. This report addresses the use of predictive risk modeling in human services, an emergent practice that aims leverage technology and analytics to supplement human service providers in their work. The report is meant to serve as a guide for practitioners, a resource for local government leaders, and a conversation-starter for advocates and stakeholders.

This report was written by Christopher Kingsley, with support from Stefania Di Mauro-Nava.