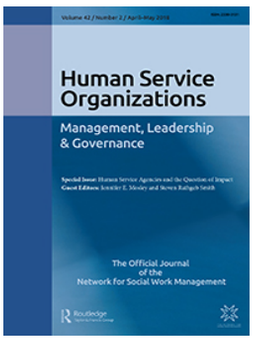

Routledge
Taylor & Francis Group

Human Service Organizations: Management, Leadership & Governance

ISSN: 2330-3131 (Print) 2330-314X (Online) Journal homepage: https://www.tandfonline.com/loi/wasw21

What counts? quantification, worker judgment, and divergence in child welfare decision making

Emily Adlin Bosk

To cite this article: Emily Adlin Bosk (2018) What counts? quantification, worker judgment, and divergence in child welfare decision making, Human Service Organizations: Management, Leadership & Governance, 42:2, 205-224, DOI: 10.1080/23303131.2017.1422068

To link to this article: https://doi.org/10.1080/23303131.2017.1422068

Accepted author version posted online: 08 Jan 2018.
Published online: 24 Jan 2018.

HUMAN SERVICE ORGANIZATIONS: MANAGEMENT, LEADERSHIP & GOVERNANCE
2018, VOL. 42, NO. 2, 205–224
https://doi.org/10.1080/23303131.2017.1422068

Routledge
Taylor & Francis Group

What counts? quantification, worker judgment, and divergence in child welfare decision making

Emily Adlin Bosk

School of Social Work, Rutgers University, New Brunswick, NJ, USA

ABSTRACT

In an effort to manage risk under chronic resource constraints, information uncertainty, and accountability pressures, U.S. child welfare organizations have embraced the structured decision-making (SDM) model, which combines actuarial-based risk assessment with clinical decision making. Although 33 states have adopted the SDM to impose greater rationality and precision to child welfare decision making, little is yet known about how actuarial-based risk assessment interacts with child welfare workers' own judgment in implementation, and to what effect. Drawing on original data from a case study of four child welfare agencies in one state, this article examines the nature of this interplay, and its implications for the quality of worker decision-making, child welfare, and worker job satisfaction.

KEYWORDS

Decision-making; child welfare; quantification; standardization; risk assessment; worker judgment

Introduction

Errors in judgment in child welfare can have lasting consequences for the children the system intends to protect and the families it is charged to serve. At their most severe, flawed assessments allow children to remain in unsafe situations that can result in severe injury or death (false negatives [FN]) or lead to unwarranted separations between children and their caregivers with all the attendant trauma that accompanies them (false positives [FPs]). Over the past 20 years, child welfare agencies have embraced quantified decision-making strategies under intense pressure to enhance performance, reduce error, and avoid the accompanying tragic outcomes that have often been associated with the low reliability of worker clinical judgment (Camasso & Jagannathan, 2013; Gainsborough, 2010).

To date, more than 33 states and multiple countries have now implemented the structured decision-making model (SDM), a series of standardized tools to improve child welfare decision making. The SDM addresses the low reliability of clinical decision making through the integration of standardized and actuarial-based risk assessments with clinical judgment (Children's Research Center (CRC), 2012). By systematizing the process, the SDM creates a template from which it can be expected that any worker assessing a specific case will arrive at a similar conclusion.

Actuarial-based risk assessments (RA) are considered the gold standard for conducting child welfare assessments because they use a validated algorithm to determine the likelihood that a child who has been found to be maltreated will reexperience abuse or neglect in the future (Baird & Wagner, 2000; Baird, Wagner, Healy, & Johnson, 1999; Schwalbe, 2008). This probabilistic determination of risk is based on analysis of case characteristics that are highly associated with the reoccurrence of maltreatment. Because one of the most fundamental goals of risk assessment in child protective services is to minimize the potential for decision-making errors that lead to fatality or severe harm to a child, developers of the SDM's RAs have been transparent that the balance of the tool is tilted toward FPs (Baird & Wagner, 2000). That is, when it errs, the RA may be more likely to

CONTACT Emily Adlin Bosk eab276@ssw.rutgers.edu 390 George Street, Room 714, School of Social Work, Rutgers University, New Brunswick, NJ, USA

find that a child is at a high risk of future maltreatment, when she is not (FP), than to err by finding a FN.[1] According to its developers, however, the RA's bias toward FPs should be "moot" because it is designed to classify cases according to levels of risk, and is not meant to be used as a sole means for predicting outcomes (Baird & Wagner, 2000, p. 850). That is, integrating "clinical"—or workers'—judgment into the use of the RA mitigates the chances of FPs and thus optimizes decision making as a whole (Baird & Wagner, 2000).

Prominent child welfare scholars also describe the need for comprehensive systems of decision making that balance a rational, systematic approach with the application of worker judgment to reduce errors (Pecora, Chahine, & Graham, 2013; Shlonsky & Gambrill, 2001). Despite this widespread recognition, however, there are no formal guidelines for agencies or their workers on how specifically to go about integrating the RA tool with workers' own sense of a given case as derived from their combination of training and experience (Schwalbe, 2008). At the same time, social and economic demands for increased productivity and accountability, the desire to create transparent bases for decision making, and the fiscal constraints that necessitate more streamlined operating procedures in child welfare agencies, all exert organizational pressure for increased dependence on quantified forms of decision making that are efficient and clearly defensible (Espeland & Vannebo, 2007; Porter, 1995; Power, 1997).

In this article, I examine three key questions. First, what are the sources and nature of conflicts arising between risk that is automatically scored by these newer, highly rationalized decision-making tools (like RA) and risk as identified by workers' own judgment? Second, how do workers handle those conflicts, by potentially adjusting (downgrade or upgrade) risk assessments and in making final decisions about the trajectories cases will take? And finally, how do workers feel about their jobs in the face of navigating such conflicts? In other words, how do workers integrate their judgment into the use of standardized risk assessment tools in practice, and to what effects?

In line with new research on the decision-making ecology in child welfare (Fluke, Yuan, Hedderson, and Curtis, 2014), I draw on data from a case study of child welfare decision making involving 35 workers affiliated with four urban and suburban, midwestern child welfare agencies. Using this case as a foundation, I examine two particular scenarios of decision making that arise when RA scores and worker judgments conflict: first, when RAs indicate high risk but workers assess low risk (worker-perceived "FPs"), and, second, when RAs indicate low risk but workers assess high (er) risk (worker-perceived "FNs"). The midwestern state in which the case study was completed offers a particularly helpful context for this investigation. As will be elaborated on below, the SDM approach and RA tool are legally mandated, thus enforcing the integration of a highly rationalized, quantified risk assessment strategy with workers' instinct and experience, with limited accompanying guidance on that integration.

Emergence of standardized decision making in child welfare

Research on child welfare decision making prior to the introduction of actuarial-based assessments highlights the critical and consistent deficits of clinical decision-making, demonstrating that case outcomes were more dependent on which worker (and agency) a family was assigned to than other relevant case factors. Lindsey's (1992) research on the reliability of clinical decision making found that, when given an identical set of facts, child welfare workers agree on a case trajectory only 25% of the time. In separate studies, Rossi, Schuerman, and Budde (1996) and Davidson-Arad and Benbenishty (2010) also found large levels of disagreement among child welfare workers evaluating the same case. More broadly, ample research in psychology details the kinds of errors to which clinical decision making is prone, such as the use of heuristic biases and other cognitive shortcuts that widely lead to incorrect assessment (Gilovich, Griffin, & Kahneman, 2002; Kahneman & Frederick, 2002; Tversky & Kahneman, 1974). These heuristics include confirmation bias where

1. Conversely, a false negative (FN) finds a child is at a low risk of future maltreatment when she is not.

ambiguous evidence is interpreted as supporting their initial belief, the representativeness heuristic where risk is under- or overestimated based on a misunderstanding of how common an actual occurrence is, and the conjunction fallacy where events that are not necessarily related are assumed to be linked (Tversky & Kahneman, 1974). Incomplete or missing information, as is common in child welfare cases, can lead a worker to draw a faulty conclusion (Gambrill & Shlonsky, 2001; Van de Luitgaarden, 2009). Agency characteristics such as the task environment and workers' personal orientation to casework also influence decision making, leading researchers to conclude that it is really a subjective rather than objective process (Chu & Tsui, 2008; Loughlin, 2008; Regehr, Bogo, Shlonsky, & LeBlanc, 2010). In the absence of standardized criteria, Stein and Rzepnicki (1983) observed, "it appears that vague laws and knowledge deficits create a void for decision-makers, which they tend to fill by interjecting personal values and biases" (as cited in Wilson and Morton, 1997 p. 3). Rossi et al. (1996) were early in those calling for new modes of decision making in child welfare, stating, "It seems reasonable to assert what happens in cases should not be a crap shoot".

Taken together, these studies offer a decisive picture of the weaknesses of clinical decision making and how, without additional aid, clinical decision making is likely to result in error. In response to these issues, actuarial-based assessments have come to prominence (Gambrill & Shlonsky, 2000, 2001; Schwalbe, 2004). Quantifying child welfare decision making through actuarial-based RAs offers the potential to enhance the consistency and evidence base of child welfare decisions. More specifically, actuarial-based RA explicitly seek to improve low reliability and reduce subjective or biased judgments that contribute to inconsistent case outcomes and clinical mistakes by providing schemas for decision making grounded in an analysis of case characteristics that are highly associated with the risk of reabuse. Easy to review and track, actuarial-based RAs also provide a clear system for holding individual workers and organizations accountable for the decisions they make. The broad consensus is that actuarial-based decision-making outperforms clinical judgment across fields (Baird & Wagner, 2000; Baird et al., 1999; CRC, 2008; Dawes, 1988; Dawes, Faust, & Meehl, 1989). A meta-analysis of more than 100 studies found that "mechanical prediction techniques were about 10% more accurate than clinical predictions" (Grove et. al 2009, p. 19). As a result, actuarial-based risk assessments are considered to be the "gold standard" for case decisions and are now used in the majority of child welfare cases (Schwalbe, 2008, p. 205).

Child welfare decision making under uncertainty

Child welfare decision making comprises multiple smaller decisions, each made in an overall context of uncertainty (Gambrill & Shlonsky, 2001). Figure 1 describes the distinct decision-making points in each case and their possible consequences.

The focus of this case is decision making related to the integration of worker judgment and the actuarial-based RA for cases that have been substantiated (the allegation of child maltreatment is found to be true). The purpose of the RA is to accurately classify cases, specifically those that are at a high risk for future maltreatment, so appropriate intervention can take place with families and fatal or severe injury to children avoided. Some of the mechanisms utilized to protect children in cases deemed high risk are removal from the home, court ordered in-home supervision where any noncompliance by the caregiver is grounds for subsequent removal, intensive treatment of the caregiver, placement of the caregiver on a Central Registry for Abuse and Neglect (explained in more detail below) and frequent visits from Child Protective Service (CPS).

The RA comprises a list of items related to neglect (in one column) and abuse (in a separate column) that are that are empirically associated with the risk of reabuse. Points are assigned to each item a worker determines to be accurate and then summed. The highest number of points in either the neglect or abuse column is then used to determine the final risk score. The neglect and abuse inventories include a mix of objective and subjective indicators. Objective factors include demographic items about the caregiver such as number of children in the home and information about the caregiver's previous involvement in the child welfare system. These items are understood by workers

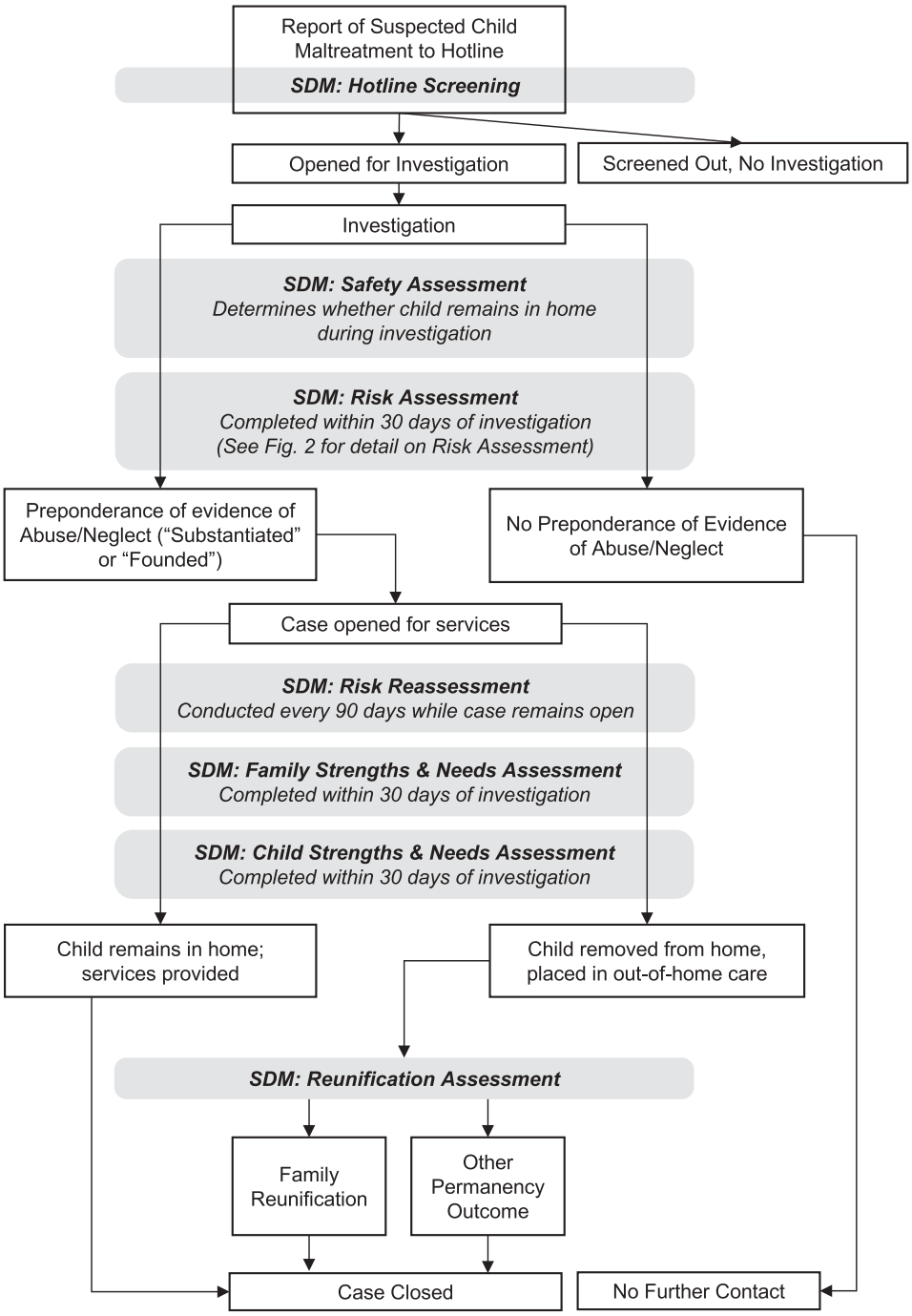

Figure 1. Child welfare decision-making points in relationship to the SDM .

to be fixed, requiring little of their own judgment in their assessment. More subjective factors included in the RA are related to current case information such as whether the caregiver is cooperating with CPS. Workers' understand these items to require some judgment. Although the RA is not used to determine whether a case is substantiated, it is intended to be filled out during the

investigative process. When a case is denied (the allegation found to be untrue), the RA does not inform its trajectory because the case is automatically closed. Functionally, this means that a case could score as High or Intensive risk without these categorizations being pursued.

Designing risk assessments involves a balance between considerations of sensitivity and specificity which, in turn, necessitates embedding tradeoffs in decision-making rules. In the context of the RA, "sensitivity refers to the degree to which true cases of maltreatment are correctly identified and specificity refers to the degree to which non-abusing families are correctly identified" (Shlonsky & Wagner, 2005, p. 416). Choices about whether to favor sensitivity or specificity necessarily reflect choices about what types of error (FPs or FNs) to avoid, and which types of error should be afforded a greater degree of tolerance. The fact that these decisions are grounded in empirical data about risk factors for child maltreatment does not remove the values embedded in choices about cutoff points.

Validation studies of the RA demonstrate that its design privileges avoiding FNs (Baird & Wagner, 2000). Using a hypothetical example, Baird and Wagner (2000) note that a high rate of FPs on the RA, even of "56%," would not be problematic as long as the tool improves the identification of cases that need more intensive services (p. 850). Baird and Wagner state:

> Despite the high proportion of FPs, cases that were rated high risk experienced maltreatment at a 44% rate, while only 5.3% of those rated at lower risk levels had subsequent maltreatment reported. The ratio of "failures" in the high risk group to "failures" in the low risk group is more than 8: 1. Such results help agencies identify which families are more likely to abuse or neglect their children. In addition, 11 of the 15 cases (73.3%) where subsequent maltreatment occurred were correctly identified (a relatively high rate of specificity). (p. 850)

The benefits of weighting RAs to prevent fatalities or serious harm are clear. However, an unintended consequence of this tradeoff is that when FP errors occur, the costs to families are largely unexamined and unacknowledged. Scheff (1963) highlights that these unintended consequences are an unavoidable result of basing decision-making rules on averting one kind of suffering. When doing so, other kinds of suffering will always be obscured. Camssasso and Jagannathan (2013) assert that risk assessments in child welfare policy in general, and in child welfare decision making in particular, are formulated around the principle of avoiding FPs. They note that lawsuits filed against child welfare agencies and a "zero tolerance" policy for errors where children are severely harmed have caused more cases to be screened into the system, and higher rates of FPs at every decision point in a case.

As noted above, worker judgment is a mechanism intentionally built into the SDM to avoid error in general and FPs in particular. When the RA score is integrated with worker judgment, four separate scenarios are possible as indicated in Table 1.

An RA score and a worker's sense of risks in a case can diverge in two key ways: where the RA identifies high risk, but a worker interprets low risk (a worker-perceived potential FP); and conversely, where the RA scores as low risk, but a worker believes there to be high (or higher)

Table 1. Potential Convergence and Divergence of Risk Assessment (RA) Score and Worker Judgment.

		Worker Judgment	
		Low	High
RA score	Low	Convergence: no need to negotiate between RA score and worker judgment	Divergence: need to negotiate between RA score and worker judgment (worker-perceived potential false negative)
	High or Intensive	Divergence: need to negotiate between RA score and worker judgment (worker-perceived potential false positive)	Convergence: no need to negotiate between RA score and worker judgment

risk (a worker-perceived potential FN). Resolving any worker-perceived potential FP (or FN) is dependent on how this divergence is negotiated.

Although there is broad agreement that actuarial-based RA improves decision making, few studies have examined how RAs are used in practice in a U.S. context (Schwalbe, 2008). As a result, we have little knowledge about what workers and agencies do—by harnessing their clinical expertise—to make sense of divergence in RA scores in light of their potential for high rates of FPs. Prominent child welfare scholars all make clear that the success of risk assessment instruments is predicated on the integration of professional judgment alongside them (Gambrill & Shlonsky, 2000; Munro, 2004; Pecora et al., 2013; Shlonsky & Wagner, 2005). However, work from Australia suggests that the SDM is being used in ways that blunt critical thinking in case assessments or that worker judgment is not used at all (Gillingham, 2006; Gillingham & Humphreys, 2010). In an experimental study, Regeher and colleagues (2010) found that use of an actuarial-based RA did not standardize worker assessments of the same case, and that decisions were still variable. Additionally, Munro (2004) documented that workers are not adequately trained in probability theory, leading them to overestimate the accuracy of actuarial-based RAs and an insufficient understanding of their FP rates. This study offers unique insight into the conditions under which divergence between the RA and worker judgment happens, identifying recurring patterns of chronic stress between RA and worker assessments.

Method

This article uses the case study method to examine child welfare decision making utilizing the SDM in one state. Case studies are useful for sorting out the complex relationship between a policy's theoretical grounding, intentions, and implementation (Flyvbjerg, 2006; Greenwood & Lowenthal, 2005). Employing a critical case approach allows for an examination of the layered manner in which divergence between worker judgment and the RA are managed in the specific policy environment in which both operate.

Case selection: The SDM in midwest state

I chose to investigate Midwest State (a pseudonym) because of the state policies that connect child welfare decision-making to quantified processes (Flyvbjerg, 2006: 230). Midwest State's legislature mandates the use of the SDM for all child welfare casework and relies on the RA score to determine a case's trajectory once it has been substantiated. Figure 2 below details the relationships between risk score, case category, and the use of overrides:

In practice, this Midwestern State policy ties the risk score to a number of case decisions such as whether a parent or caregiver is placed on the Central Registry for Abuse and Neglect (CRAN, a centralized list of offenders of child maltreatment searchable by public and private employers and others), the intensity of services a family will receive, whether those services will be voluntary or mandated, and the frequency of contacts with the CPS. For some caregivers, RAs that are rated high do not result in the removal of their children but do result in their placement on the CRAN. Placement on the CRAN is automatic for all cases that score as intensive or high and can result in termination from employment as well as prohibit participating in activities with their children (such as attending school functions), so there are serious consequences that can last long after a maltreatment case has been concluded. Tying the risk score to case trajectory runs contrary to the intentions of SDM's developers who stress that the SDM is not designed to replace clinical judgment but rather "help to structure decisions by bringing objective information to bear on these critical questions" (CRC, 2008, p.16).

In Midwest State, there is no policy mechanism to lower risk scores and the subsequent trajectory that the score determines. The inability to lower a risk score functionally means that there is no mechanism to remove a caregiver from the CRAN. Over-rides to raise the risk score can be made, in contrast, and are either discretionary or mandatory. Discretionary over-rides permit a worker to

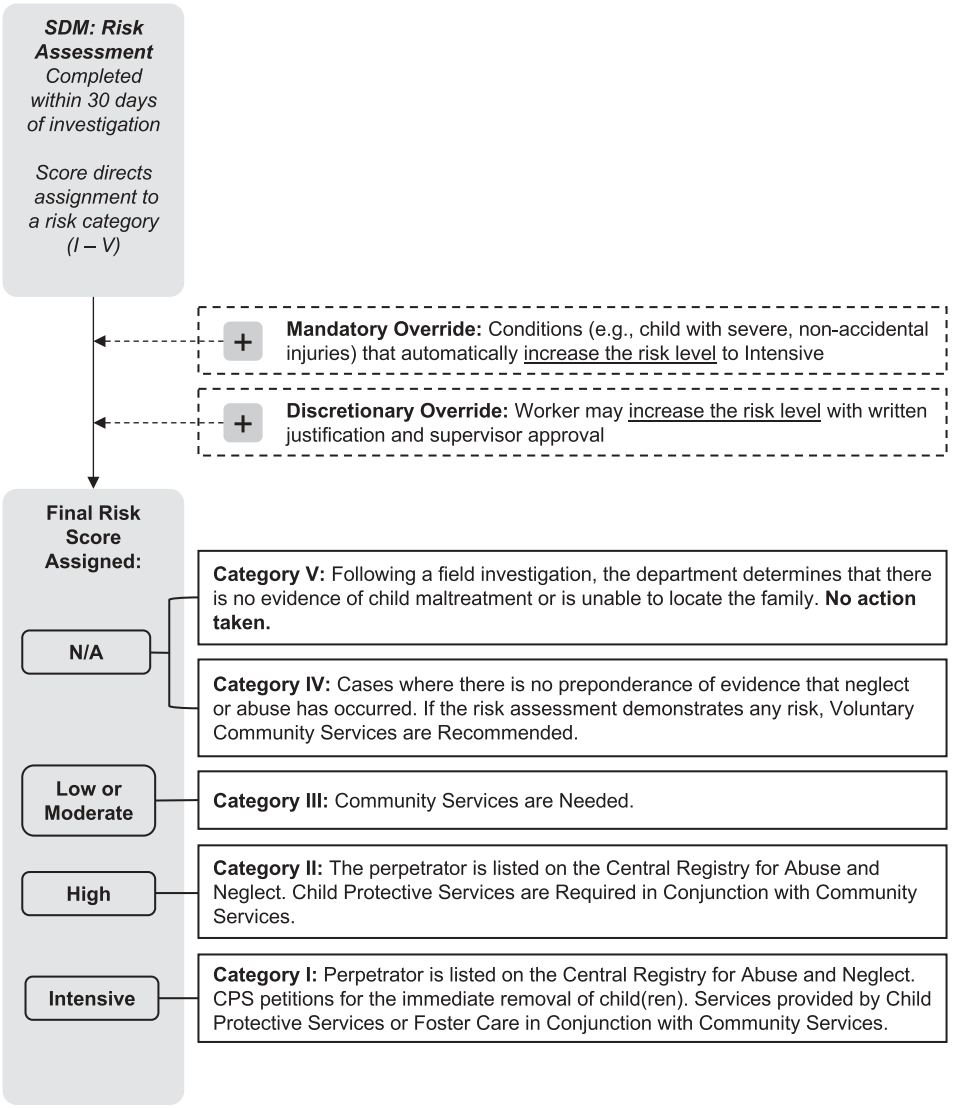

Figure 2. The relationship between risk score, case categorization, and case trajectories in Midwest State.

increase the risk level with written justification and supervisor approval. Mandatory over-rides are a series of policy conditions that require the risk level to be raised to intensive (which would automatically mean the child is removed from the home during the investigation period), such as a child with severe nonaccidental injuries. When it comes to cases that require a child to be quickly removed from their caregiver due to safety concerns, failures of the RA are assumed to occasionally occur.

In sum, Midwest State stands out from the other states for its explicit connection of the RA score to how a case is managed and is therefore helpful for understanding the process and the unintended consequences of child welfare decision making that combines highly structured risk assessment tools with workers' judgment—and to what effects on the process for determining case actions.

Data and data analysis

Data come from interviews, documents, and brief observation related to the decision-making procedures. Interviews were conducted with 35 CPS workers working at four agencies across the state who were responsible for the initial case investigation and following families until a case was closed. The agencies were selected to represent a mix of urban and suburban counties. Workers were recruited via e-mail, a presentation at an all agency staff meeting, fliers, and at (CEU, Continuing Education Unit) trainings geared toward child welfare workers, and interviews were conducted in a confidential location of the worker's choosing. On average, interviews were 90 minutes in length and ranged from 45 minutes to 3 hours, all were single sessions. Interviews were recorded and professionally transcribed. Review of documents included texts related to the case site such as the CPS manual for the state, the legal statute for use of the SDM in Midwest State, and all publicly available materials about the SDM released by its developers. Brief observations took place when workers allowed me to sit with them at their desk and witness their administrative work.

The semistructured interview protocol for this study was designed to ground data about participants' beliefs through their behaviors (Charmaz, 2006). Following an initial discussion, I asked participants "to walk me through" their most recent case, a challenging case, and a case where they felt good about their decision. By collecting data on decision making across a continuum of affective states and challenges, I was able to get a rich sense of how a worker made case decisions (Weiss, 1995). After collecting these accounts, I turned the focus of the interview to the SDM tools. Once again, I collected information on attitudes through questions about behaviors. I asked workers how they used these tools in their everyday practice and asked each worker to "walk me through" the RA for any or all of the cases they described. These questions allowed me to collect data on the weight these assessments were given in the final disposition of the case and how they influenced decision making. Throughout the interview, I followed up every question with probes specific to the conversation. All names of offices, agencies, and people are pseudonyms.

I began analyzing data during the observation and interview phase of my fieldwork (Charmaz, 2006). Using NVivo 10, I coded interviews and documents based on the emergence of patterned regularities within them as well as my knowledge of the literature (Emerson, Fretz, & Shaw, 1995). That is, my reading of the data was informed by current debates about the strengths and weaknesses of actuarial-based and clinical decision making, the goals of the SDM as stated by its developers, and sociological work on quantification. Therefore, I was sensitized to the tensions inherent in integrating actuarial and clinical judgment and alert to these themes during coding but also allowed for the emergence of other, unrelated themes. Once I identified core themes, I then coded interviews at the aggregate level.

Throughout, memoing was used as a strategy for recording observations and insights gained through all types of data collection. To ensure confirmability and independence of the findings, two research assistants analyzed the data deductively using a codebook based on the themes and categories I initially identified. Any discrepancies between coders were noted and resolved in a consensus coding process. The data presented below represents larger themes found in the data, except where noted. Quotes from the data are edited for clarity (e.g., repeated "Ums" removed), but no words have been changed or reordered. Institutional Review Board approval was granted for this study.

Findings

Results examine patterns of chronic divergence and points of stress in worker and RA understanding of cases. Two key scenarios of conflicted decision making that workers faced in the agencies studied are considered: divergence in risk assessment in which the RA scored high but a worker rated the risk lower and divergence in which the RA scored low but a worker observed greater risk. Findings consider how these divergences occur. Broadly, workers report that divergence originates from two

sources: (1) items on the RA that they believe are not sufficiently nuanced in relationship to presenting case information and (2) state policy that prohibits workers from downgrading a risk score. The final section describes the impact on morale when case workers believe decisions are being made in error with limited options for correcting them.

When risk diverges: How it happens

Demographic factors

The RA in Midwest State contains a series of questions on a set of demographic factors, such as number of children in the home and children's age, used to assess risk for child maltreatment based on established research. Focus on demographic factors means that it is possible for an RA to score high through the presence of these indicators rather than the behaviors of the caregiver in question. Angela describes how this can occur:

A: Well sometimes with the outcome of the RA I don't agree with; like 'cause you know when it comes to high and it's open, it's gonna be a Category 2. And sometimes I feel like it shouldn't be 'cause sometimes it's just for having four or more kids in the home…
E: So like that might be something that throws it over the edge?
A: That throws it over the edge and it shouldn't; like or either there'll be like there were four or more kids in the home and they had… and the children had mental health issues. Then that throws it over the edge to make them be a Category 2 and put them on the unfit list; even when it's like a first-time offense.

As it is currently constructed, the Midwest State RA weighs having more than 2 children, children under the age of 7 and a previous child maltreatment complaint (not necessarily a substantiation) as factors that will raise the risk level for a family. This risk factor is based on clear evidence demonstrating that more children and younger children in the household are at higher risk for maltreatment (MacKenzie, Kotch, & Lee, 2011). However, many workers reported feeling that families who, in their estimation, were low risk for child maltreatment have often been penalized on the basis of demographic rather than relational or behavioral factors related to the complaint.

Alternately, demographic items that seem without any controversy as to their relationship to child maltreatment, such as N2. Number of prior assigned neglect complaints and/or findings and A2. Number of prior assigned abuse complaints and/or findings, can be more problematic in practice. The more past complaints (whether denied or substantiated) a caregiver has, the riskier the current situation is likely to be and the greater the probability of future maltreatment (Hussey, Chang, & Kotch, 2006). However, workers object to measuring the number of complaints rather than only the number of "substantiated" complaints. Investigators report that reasons for multiple complaints vary and are not necessarily indicative of the parent or caregiver's past actions. Frequently, workers described how parents engaging in custody battles can use the CPS system as a way to bolster their case for receiving custody. Filing an anonymous complaint with CPS is also a way of settling scores between neighbors, and of exacting revenge when relationships between parents change (breakups, divorce) and when new people (girlfriends, boyfriends,) enter the picture. Because unsubstantiated complaints are treated the same as substantiated ones, they can and do inflate the risk score. As Ashley describes:

> You'll have some families that the neighbors will continually call on them or there'll be like one RP [Reporting Person] that continually calls on them and makes up a new allegation every time and we have to go out and investigate and deny the case repeatedly. But let's say they have four kids and one's young. So then they got a bunch of previous CPS history, the young kid and the four kids in the home, they're already at a very high-risk level. So even if it's a small issue because of all those risk points, there's a very good chance they'll be put on Central Registry and Cat 2.

In general, workers believed that the RA treats the suspicion of maltreatment by others and the presence of confirmed incidents of maltreatment as the same event or at least as events that predict the same amount of risk, when they should be kept separate in practice.

The presence of the past: How caregivers' personal history impacts current assessments

Playing out within the RA form are larger discussions about the role that a parent's experience of maltreatment has to the current complaint. Question A5 asks workers to record whether either caretaker was abused or neglected as a child. If the question is endorsed, 2 points are added to the RA. Workers assert that, in combination with other demographic questions, adding past history into the mix can be enough to move caregivers into a high-risk category and onto the CRAN without regard to other factors that should influence the case such as the severity of the substantiated maltreatment and how those under investigation understand and are addressing the problem. Jennifer asserts:

> I mean there's cases all the time that score out as high, you have parents participating throughout the whole thing. … You know one of the biggest questions we have to ask in policy is; does the family … have you ever been abused or neglected as a child? They self report that. Well again, if you clicked yes, parent self reports they were sexually abused as a minor, you know again that just raised their risk level, but that doesn't mean that they're at [risk] … it could've been over something of a dirty house you know.

Although a parent's history of child maltreatment is a risk factor, Jennifer notes that it might not affect the current complaint in a meaningful way. Workers note that as the item is written there is no way to gain a more sophisticated understanding of how a parental history of maltreatment as a child affects the current situation. The inventory thus interprets a history of maltreatment uniformly to be negative and static, collapsing events that workers' understand to be dynamic and multilayered into a singular meaning. Lauren recounts how an RA scored as high even when her investigation revealed no risk to the children:

> There was a family who had previous CPS history a lot of it, which bumps it up on the risk scale. You know here was previous domestic violence, previous substance abuse, and the father I think was abused or neglected as a child. So the risk level kind of scored through the roof, but the allegations we got were that the children they didn't have food or clothing. And I went to the house and everything was perfect. All the kids, they didn't disclose anything concerning and there was plenty of food; they had tons of clothes. So it was kind of bogus allegations I think, so obviously I wasn't going to substantiate the family, but the risk level did score through the roof because of everything that had happened in the past … (short laugh) … so and you can't bump it down a risk level, you can only bump it up so in those cases you just got to leave it as a high risk and deny it.

Although a negative outcome for the family is prevented because the case was denied, the family would have had a very different outcome if Lauren had found maltreatment that needed to be addressed but was not severe. This leaves little room for workers to intervene in a maltreatment complaint that might need to be substantiated, but where the severity of maltreatment does not, in their opinion, match the imposed consequences for being high risk (according to state policy).

When analyzed collectively, one of the main critiques that CPS workers in Midwest State have about the RA is the way that it can penalize caregivers for being "a victim of circumstance." Workers view demographic items that document facts or events differently than those that encode agency or behaviors, interpreting them to be more "unfair" because they may not be related to the maltreatment allegation. As a result, CPS workers assert that the resulting case actions can unfold in ways they feel are unjustified. Ironically, the demographic factors on the risk assessment are considered by researchers to be the fairest way to assess risk because they are objective indicators that do not require subjective assessment that opens the door to all the issues inherent in clinical decision making that the RA is designed to avoid.

Subjectivity within mechanical objectivity

In addition to the demographic questions, the RA also includes questions that necessitate some degree of interpretation given their inherent subjectivity. N11. Primary caretaker able to put child needs ahead of own is one such question, answering which requires a different level of skill than a demographic question such as A2. Number of prior assigned abuse complaints and/or findings. Questions that are open to interpretation have the potential to lead to exactly what the RA is designed to avoid: disparate assessments of the same case by different workers. Rob believes that individual workers understand the same evidence differently, producing wide variation in case categorization:

> You know I tell people this all the time. If you had one CPS case and it has a borderline complaint, you know could be, couldn't be. And you gave it to 50 different workers, all throughout the state with different supervisors, you'll probably get 20 Cat 4s, 20 Cat 3s, a couple Cat 2s and then maybe one or two guys will give it a Cat 5.

Justine details how different scoring of the same question can occur:

> I mean its how you interpret that question you know…. You know like "primary caretaker puts child's needs ahead of her own." That night she didn't but say two days later she's saying all the right things and she's showing, she's willing to put her kids above others … a worker might say "okay, she's willing to do what she did but that one instance she didn't so"; it's kind of like depending on how the worker interprets it all.

As Justine's observations highlight, all RA questions are cross-sectional data points, recording information about a caregiver at one moment in time. Workers who attempt to carefully complete the RA can struggle with how to endorse items where the caregiver has behaved differently throughout the investigation. Lila describes the dilemma:

L: I put no [to whether the caregiver has a substance abuse problem].
E: Okay, you seem embarrassed?
L: Right, well I don't know because I'm like thinking to my supervisor should it be like, you know she tested positive? Like I know, but I tested her again and she tested negative so I don't think she has a problem because if she did she would've tested [positive].

Lila wants to be as accurate as possible in her assessment, but there is no way to record the varied nature of the caregiver's engagement with substances during the course of the investigation. Whether the caregiver has a substance abuse problem is open to some amount of interpretation, with some workers using the evidence of a single positive drug screen to confirm and others using the evidence of a single negative drug screen to deny a finding of substance abuse. Divergence occurs between workers and the RA and across workers depending on how they manage issues of interpretation in their assessment.

To manage issues of subjectivity as they relate to scoring the risk assessment, Midwest State has created a policy manual that outlines exactly how questions should be interpreted and what criteria should be used in answering each question. Policy manuals are accessible by clicking on a question when a worker is filling out the risk assessment and through the smart phones that workers carry with them during the workday. Even with these measures and a significant amount of training on how to complete the RA, some workers report that they do not ever consult the policy manual, leaving a wide range of responses to the RA open and multiple possibilities for divergence. Workers who do consult the policy manual also recognize that there are individual differences in how workers respond to questions that affect whether the risk assessment is truly standardized. In practice, the presence of a policy manual that answers these questions is not functionally enough to standardize worker interpretation of questions (Bosk, 2015).

Keeping it simple is not less complicated

Kelly finds that the lack of context that Casey refers to on the RA makes it difficult to complete it accurately:

> It's either—it is this or it isn't this. There's no grey area you know, so it's like, and that's what makes it hard going through this. You know it's like; am I gonna want to put another thing on the RA or is it not to that point where I really need to put it on the RA? Its just not so straightforward, you know.

Like all standardized assessments, the RA in Midwest State is designed to mediate the gray areas of cases by creating a more objective reading of the facts of the case. William describes how flattening the complexity of cases can actually invite more complication and produce divergence between worker and RA assessments:

> We have kids that beat up their parents. I know we're here to protect the kids, but we, when I in foster care we had kids that were [in] 7 and 8 old in residential placements. They come back to the community they're out of control. Their parents are sitting on them holding their hands down because they've busted everything up and they're hurting everyone in the house, and they don't know what else to do, and the child may have a bruise on his wrist from being held down. That's a bruise, that's substantiation. We need the discretion to say; how did that bruise get there? Was mom protecting the child? Was mom protecting another child from a child? Those things go on.

William is asserting that cases presenting as straightforward (e.g., corporal punishment by a parent that results in a mark, which is not only grounds for a mandatory substantiation but also against the law) are often anything but clear in practice. An unintended consequence of abstracting concepts into numerical forms can be an outcome that violates the principle of the rule, prompting divergence between worker and RA assessment.

Lack of a transactional perspective on risk

Although research clearly identifies the individual risk factors associated with a higher likelihood of future maltreatment, what is less clear is how these individual factors come together to affect risk. The risk assessment captures the cumulative burden of risk (e.g., more risk factors, equal a higher likelihood for future maltreatment) but is less able to account for the transactional and interactive effects of the items on the risk assessment. Pecora et al. (2013) note that: "safety and risk assessment tools generally contain discrete factors, yet, it is the interactions of these factors … that are likely to figure in the causal processes leading to the lethal assaults of young children" (p. 145).

Workers believe that collecting information only on the presence or absence of risk factors without a more dynamic understanding of their relationship to the presenting complaint can overestimate risk. In its current form, the RA itself does not distinguish between risk factors and protective factors (e.g., a parent that is addressing substance abuse issues vs. one that is not; or a caregiver that was abused as a child and has received treatment for this experience vs. one who has not). Because clinical judgment is meant to account for the ways that the RA by nature is probabilistic not deterministic, these issues are not taken into account in its design. In an implementation of the SDM that uses the RA to drive case decision making, the absence of more finely grained assessments of risk have profound implications and make divergence between worker and RA assessment more likely.

Policy limitations as a driver of divergence

As Shlonsky and Wagner (2005) observe, child welfare decision making is highly dependent on context. The construction of the RA and the items described above have far more limited impact on decision making when the RA risk score is not determinative for decisions about removal, in-home supervision, and placement on the CRAN, and when workers are permitted to lower those scores based on their clinical judgment. Midwest State's decisions to tie the risk score to case trajectories and to not allow worker over-rides to lower the risk score establish a particularly strict context for SDM's implementation (and one that that runs counter to the intent of SDM's designers). That is, in other states where workers can infuse their clinical judgment into case decisions by either adjusting

automatically generated RA scores or simply taking the RA score into account: the two sources of information about risk—RA and worker expertise—are able to be far more integrated in the process of tool implementation. This is not so in Midwest State child welfare decision making.

Casey explains her perception of how punitive consequences can escalate a case when context is excluded from the assessment in Midwest State:

> It was a neglect case and that's one of the ones I ended up having to do in-home ward with the courts. Mom is, I believe, between 24 and 26 and she has mental health issues of her own, and she comes from a very difficult relationship with her parents; a difficult relationship with her family. I think her mother is also undiagnosed bipolar so I can only imagine how bad they, you know don't get along. And she has 3 children, an 8-year-old, a 5-year-old and a 1-year-old, I think she's like 13 months. And she was living with the father of the two youngest and for some reason she decided to let the oldest boy live with her parents and so she was only living with the two youngest and the father.
>
> And it was a domestic violence relationship and the house was in disarray partially because the landlord hadn't cared for, like the water main broke. So between that and a little bit of them knocking around and throwing things and her being, her mental health issues, … when the police came out for a domestic violence incident they filed because the house was dirty.
>
> I came out and she ended up getting evicted from the house within a couple of weeks because they weren't paying rent, I guess because the landlord wasn't fixing the issues. And when, you know, I looked at the case because of mom's mental health issues, because she has three children and one of, her youngest baby was special needs. She was a domestic violence victim, she had CPS history as a victim; all of these things made her have a high risk and made us have to go to the point where we had to file in-home (supervision).
>
> But she was completely willing to work with services. She was so happy to have someone to help her get out of the domestic violence situation. To help her get her own housing because she's always relied on her mom or a man and this was her first opportunity to support herself and we were helping her and she went to counseling and you know, she went with Families First and this was all before we filed. And just because her RA was so high; we had to file.

In this case, Casey believes that while multiple risks exist, the outcome for the case is unnecessarily punitive. The divergence occurs not between Casey and the RA's assessment of risk but in her disagreement with Midwest State's policy of how that risk should be managed. Casey observes that the mother in this case is working hard to attend to the issues present in the referral. In her estimation, these efforts are enough to mitigate the higher risk assessed by the RA. Casey sums up:

> And I totally agreed with the allegations and that somebody needed to you know address her mental health, address why she is doing this, and give her parenting classes to address the fact that she can't parent her kid and she just wants to medicate them to deal with them.
>
> But I didn't think that it was to the level that we had to take her job away and make it so she couldn't care for her four children, because then if she can't, you know that's just going to make the whole situation worse. And, I mean, I don't know I just felt that the way that, if she is a risk to children I feel like she should be a risk to children because of the actions and decisions she made, not because of the things that happened to her.

An unintended consequence of linking the risk score to the CRAN is that placement on the CRAN might actually make it more likely that child maltreatment will occur in the future, by leading the mother to lose her job and in turn creating the conditions for physical neglect. Gambrill and Shlonsky (2000) clearly discuss the importance of incorporating individualized risk assessment into case judgments to ensure that these kinds of nuances are taken into account in case decision making.

Although the RA is designed to protect families from the harmful consequences of subjective assessments that are incorrect, rooted in bias, or based on erroneous beliefs, the flip side of an implementation of the SDM that limits discretion from casework as described here is that when, in the worker's opinion, the assigned case trajectory diverges from the case dynamics, caseworkers have almost no recourse to redress the situation (with the exception of denying a case).

A risk score that is deemed inappropriately high by a caseworker would not be of much consequence if there were processes in place to manage the case in another way (such as not placing a parent on the CRAN) or to confirm that the caseworker was not erroneously interpreting the score as being too high.

Two sides to every story: Contested approaches to practice and restrictive policy

Another consequence of prohibiting workers to lower the risk score is that this policy limits responses to divergence between the RA and worker judgment when there are debates about how certain types of case should be handled. Cases that involve a child's exposure to domestic violence (DV) exemplify this issue. Ample empirical evidence supports the fact that the presence of DV presents a risk of future child maltreatment (Fantuzzo & Lindquist, 1989; Levendosky, Huth-Bocks, Shapiro, & Semel, 2003). At the same time, there is an ongoing debate about whether substantiating women for child maltreatment due to the presence of DV is itself a process of blaming the victim (Rivett & Kelly, 2006; Rogerson, 2012). This issue has not reached an empirical or policy consensus, making it far from settled. Kate believes that the RA treats this complex and unresolved concern in the field as overly simple leading to unnecessarily high risk scores and their attendant consequences for domestic violence victims—typically mothers:

> Like there's a lot of times like I'm surprised like when it comes out as high. I think that it's not so accurate 'cause there's two different questions regarding domestic violence; one if there's a history; one if the person's in a harmful relationship. So if there's any domestic violence they're scored negatively on two areas, whereas everything else is just one question. So I feel like it's a little skewed when it comes to that.

The way intimate partner violence is captured on the risk assessment may actually serve to create confusion about how to evaluate the role of intimate partner violence in a case. As Kate explains, official policy for the substantiation of cases is supposed to take a more nuanced view of DV cases that incorporates current debate about the issue:

> I mean there's a couple pages of very specific things and how you can't substantiate a parent. Like if there's domestic violence, a lot of people don't realize like if Mom's allowing the dad to beat her instead of the children, that's a protecting measure and you can't substantiate Mom for that. She's doing what she can within her means to protect the kids. . .. A lot of people interpret that differently because obviously you're still exposing your child to domestic violence, but that's an effort that she's making to protect them.

In practice, workers report that substantiated failure to protect cases may result in punitive consequences for these caregivers because of the weight that DV is given on the RA.

Neglecting emotional neglect and abuse

Although the case examples above highlight patterns of divergence that workers perceive to result in a potential FP not all divergence occurs in this direction. I find that policy requiring independent validation of mental injury (as represented on the RA) from external professionals works to lower risk scores and therefore is one area where divergence occurs that might lead to worker-perceived potential FNs for emotional abuse or neglect.

Although some items on the RA are heavily weighted, such as DV and past history of maltreatment, other potential sources of risk are functionally left out of the assessment process despite their official presence on the RA tool. One such item is emotional abuse and neglect. Workers are explicitly directed to assess on risk for this form of maltreatment in the RA through the following question: A1: Current complaint and/or finding includes mental injury. Midwest State's policy regarding mental injury is that only a mental health professional can identify it. Because of the inconvenience this requirement imposes, workers report that they rarely, if ever, endorse a finding of

mental injury, answering the question negatively as a matter of routine. Casey and Monica detail this process;

M: Mental injury; No. I don't think I've ever marked that one.

C: And so for the other one mental injury. Mental injury, we don't assess that it's kind of going be assessed by psychiatrist, yeah and so anytime we have an allegation of mental injury we have to send them for an assessment and we wait for their [evaluation]… so I almost always hit zero on this one.

The majority of workers report endorsement of mental injury by a professional is important because cases involving emotional maltreatment are difficult to identify and outside their expertise. Casey articulates, "I don't think that we should be able to assess that … I'm not qualified to make that assessment…. I definitely think we should refer it out to other people." However, workers also report that they almost never actually seek out a psychiatrist to do this assessment. As a result, assessments of mental injury are omitted from regular consideration during the investigative process even though their inclusion on the RA asks workers to attend to its presence or absence. This means that emotional neglect and abuse are rarely substantiated or addressed by CPS workers in Midwest State.

William is one of the few workers who objects to having a mental health professional assess the presence of mental injury because he believes that the lack of ability to do so leads to exclusion of the presence of emotional maltreatment in risk assessments. William describes a case where he disagrees with his supervisor about whether mental injury is occurring:

> A case where the mother, I think as part of discipline to get her kids in line, "I'm sending you back to foster care and I'm never taking care of you again, and you're never gonna see me again." That's mental injury. I also think that that's a poor use of discipline to try to get her kids in line. Right?
>
> [The supervisor says:] "No, that's not inappropriate discipline and I can't call it mental Injury because it's gotta be substantiated by a therapist, who says this particular thing has really made this child emotionally unstable. All these kids aren't even in therapy for the most part. Why do they have mental injury? Call it something else."
>
> I think it's mental injury, [DHHS administrators are] saying "no, you can't use it." And I'm like, when a kid hears I'm going to foster care, and has experience in foster care that maybe wasn't so good, what else could it be? Do I really need a therapist to confirm that?

Ultimately unable to convince his supervisor of the presence of mental injury, William is forced to omit that factor in assessing risk in this case.

When risk diverges: How workers feel

Lacking official recourse to address cases that are scored as higher risk than workers believe is warranted, or facing a mandatory assessment of high risk for cases that sets of a service trajectory that feels unnecessarily punitive, workers' morale suffers. Approximately two-thirds of the participants described the negative impact that relying on the RA risk assessment to the exclusion of their own clinical judgment to determine case trajectories had on their feelings about their jobs. Angela summarizes a general sentiment, "Sometimes you just feel bad. That's all … you just feel bad and you try to look … like you try to go over the risk assessment again and see maybe if you did something wrong if you clicked on the wrong button." Casey articulates how the ability to lower risk scores would resolve many of her issues with the RA:

> But I feel like our overrides should be able to able to go both ways. I feel like, cause I know people, you know in my own personal life that meet a lot of these things, and maybe they'll be a high, intensive risk, but you know, just because you're a victim of domestic violence or because your children have issues or because you have, you know four children or mental injury, you know things like that; you don't always have to fall into the statistic of what's likely to happen. You know you can overcome those odds and be a good person and be a good parent. So I just wish that we had the ability to override both ways, that's all.

Policy that bases consequences for substantiated cases solely on risk score left multiple workers questioning the utility of the RA in their work. Without discretion related to case context, the majority of workers reported feeling their decision making had devolved to a computational process that they carried out but could not influence, with the exception of denying a case (Bosk, 2015). It is important to note that one third of workers felt extremely positively about the RA and the SDM, seeing it as a critical intervention against subjective decision making, which also removed the responsibility (and terror) of making a mistake (Bosk, 2015). These workers reported few, if any, instances where their assessment diverged from the RA.

Discussion

The cases described here identify patterns of divergence between the RA and worker judgment. The majority of divergence occurs under two conditions: (1) when workers perceive the RA is overestimating risk or (2) when workers perceive the consequences of the risk score to be more more punitive than the situation warrants. In the first scenario, divergence primarily happens when workers believe that the RA score is an FP. Many workers perceive these FPs to occur due to the static nature of questions and their inability to take more nuanced contextual information into account. The phrasing of questions on the RA could be altered to address workers concerns and be more dynamic. For example, questions on the presence of substance abuse could be reframed from a binary to to be more reflective of a continuum. Possibilities for revisions could include (1) primary caretaker has had a positive substance abuse screening during the course of the investigation (with points given to reflect the number of positive screenings), (2) primary caretaker is currently in treatment to address substance abuse treatment, (2b) if primary caretaker is not currently in treatment to address substance abuse, primary caretaker is willing to participate in services. Demographic questions could also be reformulated to contextualize the presence or absence of documented risk factors within the current case. For example, questions about the number of children in the house could be revised as follows: There are more than three children in the house Y/N. If Y, Primary caretaker is overwhelmed by meeting the needs of all the children. Phrasing questions in this way has the potential to respond to workers' concerns by making the presence of risk factors directly related to the current complaint and the assessment process. When it comes to emotional maltreatment, the question and the policy could be adjusted. Instead of having a professional verify the presence of mental injury, the RA could instead assess for the presence of emotionally abusive language.

Although the SDM developers highlight the need for further discussion when workers' assessments diverge from the RA score, Midwest State's policy does not allow for these conversations to take place. Prohibitions to lower risk scores functionally mean that the RA supercedes worker judgment when they perceive a FP. Agency policy works to account for these situations through mandatory over-rides, yet no such accounting is made for scenarios where the RA understands the situation to be more severe than a caseworker believes. In this way, state policy views the RA to be infallible when it comes to overestimating risk and fallible when it comes to underestimating it. State policy that prohibits workers from lowering a risk score, whereas building in procedures to raise it reflects the organizational and moral costs of a FP. At a moral level, a child suffering a fatality or serious harm is always to be avoided. At an organizational level, high-profile fatalities often lead to very public excoriations of child welfare agencies. It makes sense, then, that a larger organizational focus is on catching FNs. An unintended consequence of this construction is that there are no structural mechanisms in place to account for the sensitivity of the RA and no organizational policy to guide workers when they perceive a FP has taken place.

Although FN cases receive a high level of scrutiny, the opposite is true for the impact of FP cases. Yet the impact of FPs are also consequential for those who experience them. FPs become real positives in any future case. Removals when they are unwarranted can be devastating to children and their caregivers. For young children, disruptions in care may profoundly alter relationships with life-long consequences for their emotional, social, and cognitive development

(Bowlby, 1973, 1979, 1980). For caregivers, extensive involvement with the child welfare system comes with costs related to time, money, energy, self-esteem, and well-being. Competent legal representation may be necessary to facilitate a reunification, and access to these services may not be available to all who require it. The child welfare system also loses when families are incorrectly sorted into risk categories. Intensive services for high-risk families are costly and time consuming. Caseworker time may be inappropriately spent attending to needs that do not require such a high level of services.

In Midwest State, when workers diverge from the RA's assessment because of the punitive consequences that accompany a high risk score, they often discuss the impact of the CRAN. Midwest State's decision to utilize the RA score to determine placement on the CRAN is likely a conflation of concepts of risk. The RA is designed to consider the likelihood of a reoccurrence of maltreatment and not issues related to whether the allegation should be substantiated, or, when it is, the severity of the offense. In practice, the risk score is being used as a proxy for maltreatment severity, and as a result two separate concepts are merged into a single policy. Much of workers' discomfort with the RA could be resolved by using another mechanism to establish placement of a caregiver on the CRAN.

Conclusion

In this article, I identify six distinct areas of the RA that likely contribute to divergence in worker and RA assessment. These areas of divergence primarily cause concern among workers that a FP will occur. To strengthen the integration of actuarial-based decision making and clinical judgment, it may be useful to develop structured questions related to these areas, which are informed by clinical evidence and validated through further empirical testing. Specifically, supervision of caseworkers could include a more nuanced (1) assessment of the meaning of demographic factors in relation to case evidence, (2) discussion of the impact of the caregivers' history of previous maltreatment on their caregiving, (3) assessment of any disagreement a worker has with the risk score based on contextual case factors, (4) consideration of how domestic violence is being accounted for and understood by the caregiver and worker, (5) discussion of how the caseworker interpreted questions on the RA and what evidence the caseworkers used to arrive at their endorsement of specific items, and (6) probing for the presence of abusive language in parent-child interactions.

Structured conversations between supervisors and caseworkers that explicitly address areas of concern from caseworkers have the potential to be an important site for making sense of the risk scores in the context of case information. Further, standardizing how workers are asked to make meaning of the RA in relationship to their assessment would create a clear process for grounding how divergences are handled in research about risk factors and ecological evaluation of cases. Procedures to strengthen the integration of clinical and actuarial-based judgment will also necessitate the creation of policies for reviewing divergence when it arises. As discussed above, over-ride review panels composed of child welfare experts could be one way to ensure that CPS workers are not erring in their objection to the RA's assessment.

The results of this study should be interpreted with care as these findings are based on a small sample of workers in a specific policy setting. Many of the case examples presented here represent instances where caseworkers believe the risk score was high in error. However, there is no way to know whether their assessments are correct. It may be that some of these examples demonstrate the reason that structured risk assessment is so important: because workers may downplay specific risk factors to which they should pay close attention. Additionally, the issue of FP and FN are not restricted solely to divergence between worker and RA assessments. There may be instances in which there is convergence but still a FP or FN error.

Debates about the the utility of mechanical versus clinical approaches have been at the center of social work for the last 20 years. Recognition that mechanical and clinical decision making are

mutual rather than exclusive has, perhaps unsurprisingly, not yet fully translated into their reconciliation on the ground. Hirschman, Berrey, and Rose-Greenland, (2016) note that "numbers, even bad numbers, tend to drive out no numbers. Once in place, quantification does not seem to yield easily except to perhaps a 'better' quantification no matter how bad the critics allege the system to be" (p. 8). Other case studies of decision making across fields demonstrate that numbers become "self-vindicating" (Porter, 1995, p. 45). When a measure is established as important, it not only evaluates but also shapes behavior related to the original indicator (Espeland & Sauder, 2007; Porter, 1995).

Scholarship on quantification quite clearly predicts how difficult it will be for clinical judgment to be integrated alongside actuarial-based assessments like in Midwest State. In the context of a policy culture that incentivizes mechanical forms of decision making and accountability, more research is needed on how actuarial-based tools are being combined with worker judgment in practice across settings and fields. Further investigation is also needed about what specific practices would support integration of worker judgment with actuarial-based assessments and how frontline workers and child welfare agencies can be guided in analyzing cases when the RA and worker are not in agreement.

Acknowledgements

The author gratefully acknowledges Anna Haley, Karin Martin, Renee Anspach, Mary Ruffolo, Michael MacKenzie, Jerry Floersch, Donna Van Alst, Sheridan Miyamoto, Holly Thurston, Ethan Schoolman and the two anonymous reviewers, for their keen comments. Additionally, the author thanks Christopher Luoma and especially Morgan Purrier, for research assistance.

Funding

Funding for this work was provided by the National Science Foundation: Law and Social Sciences Division (SES – 1323916), The Doris Duke Fellowship for the Promotion of Child Wellbeing, and the Fahs-Beck Fund for Research and Experimentation.

References

Baird, C., & Wagner, D. (2000). The relative validity of actuarial and consensus based risk assessment systems. *Children and Youth Services Review*, *22*, 839–871.

Baird, C., Wagner, D., Healy, T., & Johnson, K. (1999). Risk assessment in child protective services: Consensus and actuarial model reliability. *Child Welfare*, *78*(6), 723–748.

Bosk, E. A. (2015). *All unhappy families* (doctoral dissertation). University of Michigan, Ann Arbor.

Bowlby, J. (1973). *Attachment and loss: Separation: Anxiety and anger* (Vol. 2). New York,NY: Basic Books.

Bowlby, J. (1979). *The making and breaking of affectional bonds*. New York, NY: Basic Books.

Bowlby, J. (1980). *Attachment and loss: Loss sadness and depression* (Vol. 3). New York, NY: Basic Books.

Camasso, M. J., & Jagannathan, R. (2013). Decision making in child protective services: A risky business? *Risk Analysis*, *33*(9), 1636–1649.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Thousand Oaks, CA: Sage.

Chu, W. C., & Tsui, M.-S. (2008). The nature of practice wisdom in social work revisited. *International Social Work*, *51* (1), 47–54.

CRC. (2008). *The structured decision-making model: An evidenced-based approach to human services*. Madison, WI: Author.

CRC. (2012). *Structured decision-making (SDM). The SDM system in child protection*. Madison, WI: Author.

Davidson-Arad, B., & Benbenishty, R. (2010). Contribution of child protection workers' attitudes to their risk assessments and intervention recommendations: A study in Israel. *Health & Social Care in the Community*, *18*, 1–9.

Dawes, R. M. (1988). *Rational choice in an uncertain world*. Orlando, FL: Harcourt, Brace, Jovanovich.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical vs. actuarial judgment. *Science*, *243*, 1668–1674.

Emerson, R., Fretz, R., & Shaw, L. (1995). *Writing ethnographic fieldnotes*. Chicago: University of Chicago Press.

Espeland, W., & Sauder, M. (2007). Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology*, *113*(1), 1–40.
Espeland, W., & Vannebo, B. (2007). Accountability, quantification, and law. *Annual Review of Law and Social Science*, *3*, 21–43.
Fantuzzo, J. W., & Lindquist, C. U. (1989). The effects of observing conjugal violence on children: A review and analysis of research methodology. *Journal of Family Violence*, *4*(1), 77–94.
Fluke, J., Yuan, Y., Hedderson, J., & Curtis, P. (2003). Disproportionate representation of race and ethnicity in child maltreatment: Investigation and victimization. *Children and Youth Services Review*, 25.
Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry*, *12*(2), 219–245.
Gainsborough, J. (2010). *Scandalous politics: Child welfare policy in the States*. Washington, DC: Georgetown University Press.
Gambrill, E., & Shlonsky, A. (2000). Risk assessment in context. *Children and Youth Services Review*, *22*(11/12), 813–837.
Gambrill, E., & Shlonsky, A. (2001). The need for comprehensive risk management systems in child welfare. *Children and Youth Services Review*, *23*(1), 79–107.
Gillingham, P. (2006). Risk assessment in child protection: Problem rather than solution? *Australian Social Work*, *59* (1), 86–98.
Gillingham, P., & Humphreys, C. (2010). Child protection practictioners and decision-making tools: Observations and reflections from the front line. *British Journal of Social Work*, *40*(8), 2598–2616.
Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Huerstics and biases: The psychology of intuitive judgment*. New York, NY: Cambridge University Press.
Greenwood, D., & Lowenthal, D. (2005). Case study as a means of researching social work and improving practitioner education. *Journal of Social Work Practice*, *19*(2), 181–193.
Hirschman, D., Berrey, E., & Rose-Greenland, F. (2016). Dequantifying diversity: Affirmative action and admissions at the University of Michigan. *Theory and Society*, *45*(3), 265–301.
Hussey, J. M., Chang, J. J., & Kotch, J. B. (2006). Child maltreatment in the United States: Prevalence, risk factors, and adolescent health consequences. *Pediatrics*, *118*(3), 933–942.
Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and Biases: The Psychology of Intuitive Judgment*, *49*, 49–81.
Levendosky, A. A., Huth-Bocks, A. C., Shapiro, D. L., & Semel, M. A. (2003). The impact of domestic violence on the maternal–child relationship and preschool-age children's functioning. *Journal of Family Psychology*, *17*(3), 275–287.
Lindsey, D. (1992). Reliability of the foster care placement decision: A review. *Research on Social Work Practice*, *2*(1), 65–80.
Loughlin, M. (2008). Reason, reality and objectivity–Shared dogmas and distortions in the way both 'scientistic'and 'postmodern'commentators frame the EBM debate. *Journal of Evaluation in Clinical Practice*, *14*(5), 665–671.
MacKenzie, M. J., Kotch, J. B., & Lee, L.-C. (2011). Toward a cumulative ecological risk model for the etiology of child maltreatment. *Children and Youth Services Review*, *33*(9), 1638–1647.
Munro, E. (2004). A simpler way to understand the results of risk assessments. *Children and Youth Services Review*, *26*, 873–883.
Pecora, P. J., Chahine, Z., & Graham, J. C. (2013). Safety and risk assessment frameworks: Overview and implications for child maltreatment fatalities. *Child Welfare*, *92*(2), 143–160.
Porter, T. (1995). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press.
Power, M. (1997). *The audit society: Rituals of verification*. Oxford, England: Oxford University Press.
Regehr, C., Bogo, M., Shlonsky, A., & LeBlanc, V. (2010). Confidence and professional judgment in assessing children's risk of abuse. *Research on Social Work Practice*, *20*(6), 621–628.
Rivett, M., & Kelly, S. (2006). 'From awareness to practice': Children, domestic violence and child welfare. *Child Abuse Review*, *15*(4), 224–242.
Rogerson, S. (2012). Unintended and unavoidable: The failure to protect rule and its consequences for undocumented parents and their children. *Family Court Review*, *50*(4), 580–593.
Rossi, P. H., Schuerman, J. R., & Budde, S. (1999). Understanding decisions about child maltreatment. *Evaluation Review*, *23*(6), 579–598.
Scheff, T. J. (1963). Decision rules, types of error, and their consequences in medical diagnosis. *Systems Research and Behavioral Science*, *8*(2), 97–107.
Schwalbe, C. (2004). Re-visioning risk assessment for human service decision making. *Children and Youth Services Review*, *26*, 561–576.
Schwalbe, C. (2008). Strengthening the integration of actuarial risk assessment with clinical judgment in an evidence based practice framework. *Children and Youth Services Review*, *30*, 1458–1464.
Shlonsky, A., & Gambrill, E. (2001). The assessment and management of risk in child welfare services. *Children and Youth Services Review*, *23*(1), 1–2.

Shlonsky, A., & Wagner, D. (2005). The next step: Integrating actuarial assessment and clinical judgment into an evidence-based practice framework in CPS case management. *Children and Youth Services Review*, *27*, 409–427.
Stein, T., & Rzepnicki, T. (1983). Decision-making in child welfare: Current issues and future directions. In: B. McGowan & W. Meezan (Eds.), *Child welfare: Current dilemmas and future directions*. Itasca, IL: Child Welfare League of America.
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainity: Hueristics and biases. *Science*, *185*, 1124–1131.
Van de Luitgaarden, G. M. (2009). Evidence-based practice in social work: Lessons from judgment and decision-making theory. *British Journal of Social Work*, *39*(2), 243–260.
Weiss, R. (1995). *Learning from strangers: The art and method of qualitative interview studies*. New York, NY: Free Press.
Wilson, D., & Morton, T. D. (1997). Issues in CPS decision making. Decision making in children's protective services: Advancing the state of the art, 1–16. Atlanta Georgia: Child Welfare Institute.