The use of actuarial risk-need assessment tools is increasing in many Western penal, social welfare, health, and human service contexts. Although considerable variation exists in how these tools are applied in criminal justice settings, many international jurisdictions are beginning to use risk instruments to structure, inform, and/or determine an increasing range of corrections-management practices including arrest, diversion, bail, pre-sentence reports, sentencing, prison classification, and parole decisions. The most common use of these tools is to predict recidivism in cases involving violent, youth, or sexual offenders (Cunningham and Reidy, 2002; Dutton and Kropp, 2000; Grann and Langstrom, 2007; Harris and Rice, 2007; Lodewihka, Doreleijers, and De Ruiter, 2008; Looman and Abracen, 2010; Lucken and Bales, 2007; Manchak, Skeem, Douglas, and Siranosian, 2009; Maurutto and Hannah-Moffat, 2007; Monahan, 2006; Murrie and Balusek, 2008; Rettenberger, Matthes, Boer, and Eher, 2010; Roehl and Guertin, 2000; Schwalbe, 2007; Scott, 2009; Vess, 2009; Viljoen et al., 2008). In addition to these applications, some jurisdictions have also incorporated risk tools into new sentencing guidelines for non-violent offenders (Kleiman, Ostrom, and Cheesman, 2007; Ostrom, Kleiman, Cheesman, Hansen, and Kauder, 2002). This new application reflects a broader movement toward evidence-based penalty.

Internationally, the current penal context is characterized by rapidly growing prison populations, declining resources, and rising demands for public accountability and security. Within this context, actuarial tools are being promoted as offering a rational, objective, and empirically sound method of predicting recidivism. The emphasis on actuarial risk prediction tends to accompany demands for evidence-based “smart” sentencing (Etienne, 2009), alternatively known as “crime prevention jurisprudence” (Andrews and Dowden, 2007), which relies on the use of social scientific evidence to determine a sentence that is more likely to improve public safety” (Chanenson, 2009; Heilbrun, 2009; MacKenzie, 2001; Marcus, 2009a; Marcus, 2009b; Warren 2007; Wolfe, 2008). Particularly relevant to sentencing is evidence about
recidivism and correctional interventions that reduce criminal propensities. The evidence and techniques informing this sentencing framework are being imported from a particular type of correctional research and practice. Proponents of evidence-based sentencing and corrections argue that research can identify ‘what works’ to reduce the risk of recidivism, and that lower risk offenders can be efficiently, economically, and effectively managed in community settings (Andrews and Bonta, 2003; Andrews et al., 1990; Clements, 1996; Cullen and Gilbert, 2001; Etienne, 2009; Mcguire, 1995; 2002). They also maintain that penal resources should be redirected into correctional programs “proven” to work and that all interventions should be accredited and evaluated to ensure that they facilitate the reduction of recidivism. These perspectives are highly influential in correctional policy sectors.

Although the law, policy and principles of sentencing vary considerably by jurisdiction and country, the introduction of risk into sentencing is part of an established effort to curtail judicial discretion and regulate sentencing. The indeterminate sentencing policies popular throughout the 1970s gave judges broad discretionary power and parole boards considerable flexibility in determining release. Concerns about the global effects of these practices produced significant changes to sentencing practices, which were introduced by the late 1970s: (1) “determinate sentencing;” (2) the abolition of parole in many jurisdictions; (3) mandatory sentencing laws; (4) “truth” in sentencing; and (5) sentencing guidelines. The general disillusionment with rehabilitation and the popularised view that “nothing works” that characterized this era contributed to additional shifts in sentencing theory and policy. By late 1999, several jurisdictions had adopted sentencing guidelines that sought to: (1) reduce judicial disparity; (2) promote consistent sentencing; (3) prioritize and allocate correctional resources; (4) adjust punishments for certain categories of offenders; (5) reduce prison overcrowding; and (6) encourage the use of non-incarceration sanctions. The most recent sentencing trend in Canada and most US states follows the recommendations of risk assessment advocates (Bonta, 2008b; Marcus, 2009a; Wolfe, 2008) to increase and institutionalize the use of formal actuarial risk assessment instruments by purchasing those already on the market, adopting those in the public domain, or developing jurisdictional specific tools (Etienne, 2009).

The logic built into risk tools is different from the subjective professional or clinical knowledge that previously informed the decision making of judges, correctional authorities,
police, and parole boards (Ericson and Haggerty, 1997; Feely and Simon, 1992; 1994; Gray, Lanig and Noakes, 2002; Hannah-Moffat, 2004 O’Malley, 2004; Simon, 1993). Actuarial risk uses aggregate statistics to categorize defendants’ risk levels and then matches interventions to those levels. The use of actuarial technologies promises improved safety/security, a more efficient allocation of resources, and superior decision-making practices (Andrews and Bonta, 2003). Consequently, both the public and practitioners tend to “embrace risk” (Baker and Simon, 2002) as a technology for improving decision making and limiting discretionary powers (Hannah-Moffat et al., 2009; Ostrom et al., 2009). Risk assessment tools are appealing because of their perceived ability to engender reliable, valid, and objective determinations of future risks and to enhance managerial accountability; however, there are numerous conceptual and methodological concerns associated with their use in corrections, including concerns about racial and gender discrimination and the individualization of social problems (Hannah-Moffat, 2009).

The use of risk tools in sentencing is especially problematic because when used in courts they may offend moral and legal norms as well as country-specific constitutional values (cf. Etienne 2009; CHRC, 2003; Hannah-Moffat and Shaw, 2001). Sentencing reports and statements about an offender’s risk level remain with the offender for the entirety of his or her sentence, affecting a range of correctional decisions from levels of surveillance and intervention to eventual parole release. Because the tools classify and promote interventions based on categories of offender risk (i.e. low, medium, high), risk technologies tend to de-individualize punishments and can shift and reorient sentencing practices in unanticipated ways. Moreover, legal and correctional professionals who use risk information in decision making are unlikely to have considered the documented limitations about the science of risk, and frequently have only a limited understanding of the actuarial technologies they are using². The trend toward using risk instruments in all sectors of the criminal justice system, therefore, merits further theoretical deliberation and empirical study.

Most research on actuarial risk is highly compartmentalized; limited meaningful dialogue takes place between risk theorists, legal scholars, sentencing and government researchers, policy makers, and practitioners. This means that concerns about legal, ethical, and policy implications arising from the use of risk-based instruments do not receive the benefit of cross-sectoral dialogue. Consequently, my primary objectives in this paper are to highlight a number of these
concerns in order to: (1) stimulate discussion about the potential impact of incorporating actuarial risk logic into sentencing processes; and (2) identify questions requiring further empirical examination.

I used multiple sources of data to inform the analysis I present in this paper. These included: the international research on risk; criminal justice policy; a selection of Canadian legal cases; a content analysis of common risk tools, their user manuals, and interview guides; and interviews with 127 practitioners (crown attorneys; defense lawyers; policy makers; probation officers; prison staff; and risk instrument developers, trainers, and researchers). My focus in this paper is on how sentencing decisions may be affected by actuarial risk instruments that are predictors of general recidivism. While some risk tools can be administered only by psychiatrists or psychologists, general risk tools can be used by any professional (i.e., probation officer, social worker, police, and program personnel) who has completed required training. A considerable number of examples used in this paper reference the Level of Service Inventory–Revised (LSI–R). The LSI–R is one of the most common tools used in Canada and the United States (Maurutto and Hannah-Moffat, 2006; Peters and Warren, 2006) and is perhaps one of the most researched instruments for predicting general recidivism.

In this paper, I first briefly outline the emergence and entrenchment of probabilistic reasoning in criminal justice decision making and the more recent extension and application of actuarial risk logic to sentencing. Then, acknowledging that risk has managerial and organizational benefits, I argue that risk can be used to generate a culture of penality that relies less heavily on incarceration and recognizes its criminogenic effects. Prison minimization is a laudable goal, but using actuarial risk to achieve it is not without consequence. Before risk is adopted as an organizing principle for sentencing, I argue that further research about the following is required: (1) the methodological structure and varied logics of risk; (2) the effect of actuarial risk models on individuals and groups of criminal defendants; (3) the legal relevance and epistemological basis of risk; and (4) the organizational and policy impact of risk-need technologies.
Risk is a flexible construct that can support seemingly incompatible principles (Ewald, 1991; O’Malley, 2004; Hannah-Moffat and O’Malley, 2007). It is easily aligned with outwardly incongruous managerial and jurisprudential concerns about public safety, rehabilitation, justice, and efficiencies (Hannah-Moffat, 2004). Thus risk has considerable institutional appeal; in fact, it has become a focal point in the criminal justice system. Risk instruments are used for a range of different purposes. How risk is used in a given jurisdiction is determined by law, policy and sentencing purposes. For example, in Canada sentencing laws for youth, Aboriginals and dangerous offenders are different as is the assessment of ‘risk’ (Maurutto and Hannah-Moffat 2006; 2007). Accordingly, the use of ‘risk’ produces an assortment of problems. The nature and severity of these complications will vary by, and within, the jurisdiction (or sentencing regime).

The assessment of risk and risk tools has significantly changed over the past 40 years. Although current practices of risk assessment are influenced by local needs and law, it is possible to identify four generations of risk assessment tools within this time frame. The first generation of risk tools was based on clinical prediction—the unstructured clinical judgment of skilled practitioners. This method was discredited as subjective and unempirical, and critiqued for its poor predictive accuracy. Research supporting the use of actuarial instruments continues to consistently report that actuarial risk scoring provides more accurate assessments of risk than clinical judgments based on professional training and experience (Andrews and Bonta, 2006).

In the 1970s, new “evidence-based” risk technologies that relied on statistical prediction emerged. These tools assign a quantitative risk score to an offender by assessing individual factors (e.g., history of substance abuse, age at first offence) that have been statistically linked to the risk of recidivism in correctional populations. These tools were developed using research on large population samples. These second-generation risk assessments, which are still used in many jurisdictions, use static historic factors, such as an offender’s age, gender, and number and type of convictions in order to make predictions about the offender’s likelihood of recidivism. The presence of a risk factor equates with a score of one; its absence with a score of zero. Factors are tallied and the higher the score, the higher the probability of reoffending. Examples of these tools currently being used are the Salient Factor Score (in the United States), Static 99 (United
States, Canada) the Statistical Inventory on Recidivism (SIR) (in Canada), and the Risk of Reconviction (in the United Kingdom). 6

While these tools are typically evaluated as “better predictors of recidivism than clinical judgments” (Ægisdóttier, White, Spengler et al., 2006; Andrews, Bonta, and Wormith, 2004; Grove, Zald, Lebow, Snitz, and Nelson, 2000), criminal justice and treatment professionals critiqued them because of their rigidity and over-reliance on inert offence-based risk criteria. They were thought to produce a “fixed” prediction of risk based on accumulated historic and immutable factors. Static risk logic implies that an offender’s risk level cannot be reduced because the variables used to predict it do not change. This conceptualization of risk and the offending subject limited the scope of correctional management and provided little guidance about interventions to correctional professionals (Hannah-Moffat, 2004). Nonetheless, many static (second generation) risk instruments are being developed and used to inform decision making.

Don Andrews (1989:5-6), a pioneer of third-generation risk assessment, claimed that improving the predictive accuracy of risk assessments required that the tools move beyond their reliance on static risk factors to include offender characteristics and circumstances that can be changed during incarceration. This proposition, which is rooted in the presumption that some correctional treatment actually works, stimulated a third generation of risk research. The logic of ‘what works’ rehabilitative rehabilitation is central to the design of third generation risk tools that focus on areas of an individual’s life that can be altered through effectively delivered correctional treatment. Although some static factors, such as criminal history, remain important features of third-generation tools, additional factors in an offender’s life that can change over time have been added to the tools. For example, the new tools attempt to capture and categorize data on present employment (one can lose a job or find a job), criminal friends (one can make new friends and lose old friends), and family relationships (which can change from supportive to unsupportive or vice versa) (Bonta and Andrews, 2007:4). The changeable factors highlighted in third generation tools are known as “dynamic risk factors” or “criminogenic need”. Some tools, such as the Youth Level of Service Inventory, include “protective factors,” which are positive influences that can improve the lives of individuals or enhance public safety. These factors can decrease the likelihood of recidivism and counterbalance risk factors. Some examples of
protective factors include stable employment or housing, access to social services, positive self-esteem, and positive attitudes, values, or beliefs. Additional factors for youth are parental supervision and strong parenting skills, social support, and positive role models and peer groups.

Third-generation risk instruments are commonly referred to as “risk-need” instruments: However, since offenders’ needs are predetermined as factors shown to be statistically co-related with recidivism in aggregate offender populations, the needs assessment in these tools is not “individualized” or self-reported. These tools are designed to align risk prediction with the management of offender needs that are empirically designated as treatable; by default, this logic categorizes some needs as illegitimate targets for intervention. In other words, a distinction is made between criminogenic and non-criminogenic needs. While non-criminogenic needs (i.e., poor health) may be important, they have not been demonstrated in research to be related to recidivism and are consequently considered low priority for intervention except for “humane” reasons (see Hannah-Moffat 2004). The LSI–R, SARA (Spousal Abuse Risk Assessment), SAVRY (Structured Assessment for Violence Risk among Youth), HCR-20, PCL–R (Psychopathy Checklist revised), and SONAR (Sex Offender Need Assessment Rating) are examples of these tools. These tools, which incorporate changeable risk and need factors, have been gaining popularity since their introduction in the mid-1980s.

Before discussing the newest fourth generation risk-need technologies, it will be useful to briefly examine the logic of dynamic risk (criminogenic need), which remains a central feature of fourth-generation risk tools. The concept of dynamic risk is derived from the Risk-Needs Responsivity Model (RNR) (Andrews and Bonta, 2006). This model is considered central to the delivery of “effective” correctional treatment programs (Andrews, Bonta, and Hoge, 1990:19), whereby “effective treatment” is understood to produce a measureable reduction in the propensity for recidivism. Within the RNR model, the risk principle is an endorsement of the premise that criminal behavior is predictable and that treatment services (often cognitive behavioral interventions that claim to “teach” and not “treat,” as previous rehabilitative connotations suggest) need to be matched to an offender’s level of risk; “high risk” offenders are targeted for the greatest number of interventions. The needs principle targets for treatment an offender’s dynamic risk factors, or criminogenic needs. Correctional researchers established a set
of criminogenic needs, some of which I identified earlier, by identifying variables that prior research had empirically correlated with recidivism and that are amenable to intervention.

Proponents of this model argue that “evidence of dynamic validity, that is, changes in risk scores signal changes in the likelihood of committing a new offence, is immensely important for correctional programs and the staff charged with managing offender risk. The third-generation risk-need instruments offer a way of monitoring the effectiveness, or ineffectiveness, of programs and supervision strategies” (Bonta and Andrews, 2007). Knowledge of dynamic needs allows correctional officials to target their interventions and prioritize scarce correctional resources.

The targeting of interventions is critically linked to the *responsivity principle*, which refers to the matching of styles and modes of intervention to the learning styles and abilities of offenders (Andrews et al., 1990:20). It requires attentiveness to: (a) how diverse populations respond to various treatment options; and (b) “specific” responsivity factors (i.e., self-esteem, motivation, personality traits, life circumstances, and therapeutic relationships) that may facilitate or impede an individual’s response to intervention (Ogloff and Davis, 2004:233). The concept of responsivity is also critical to how non-white and female offenders are managed. As I discuss in detail below, risk instruments rarely distinguish between racialized or ethno-cultural groups of men and women—gender, culture, ethnicity, and race are only considered within the context of responsivity (i.e., how an offender from a particular ethno-racial group and/or gender may respond to an intervention based on population-level data) (Hannah-Moffat, 2009).

The responsivity principle features more prominently in the newest convention in risk assessment and classification, which uses strategies and tools that “systematically bring together information about an offender’s history and needs to develop a treatment plan and assign levels of supervision” (Bonta, 2002:1). These *fourth-generation tools* (i.e., LSI–CMI) adopt the same basic approach as third-generation tools, but refine assessments of risk and need so that they align more directly with case management priorities (Andrews, Bonta, and Wormith, 2004; Bonta and Wormith, 2008; Maurutto and Hannah-Moffat, 2006). The fourth-generation risk assessments are yet not as widely used (Andrews, Bonta, and Wormith, 2006), but a considerable amount of risk-need research is now devoted to determining, measuring, and categorizing responsivity factors in keeping with their centrality to this newest risk logic.
With the evolution of RN-based risk instruments, some authors claim that there is “little justification for the continued use of professional judgment to make decisions related to risk” (Bonta, 2008a:1) and that “any correctional agency that has the goal to reduce recidivism should use, at a minimum, third generation risk-needs assessment instruments” (Bonta and Wormith, 2008). These authors also claim that risk-need tools can help corrections staff efficiently and effectively allocate resources for case management. This logic has been accepted at a policy level in a number of jurisdictions, resulting in the integration of the LSI–R (or variations) into the preparation of pre-sentence reports and in sentencing guidelines that explicitly stress risk-need assessment as an evidence-based technology.

These types of actuarial instruments enable new ways of understanding risk and “knowing the offender.” In particular, third- and fourth-generation risk instruments reassert the premise that an offender’s risk of recidivism can be changed if knowledge of their needs is integrated into assessment technologies and then “correctly” targeted in correctional interventions. Importantly, the RNR model emerged when few supported the continuance of rehabilitation. The RNR model has informed the development and refinement of a series of administrative processes and risk tools currently used to target treatment, direct case management plans, accredit treatment programs, and make decisions about institutional resources. The RNR has played a pivotal organizational role in offender management by advancing a necessarily narrow and targeted view of rehabilitation.

At the same time, there are significant problems with RNR derived risk logic. Both third- and fourth-generation risk tools are fundamentally different from second-generation static risk assessments because they purport to guide the “treatment or rehabilitation” of the offender to prevent reoffending, rather than simply predict recidivism. These tools embed the assessment of risk and need in a utilitarian theory of punishment. The application of sanctions under this risk model should support the reduction of recidivism. For example, probation conditions can be aligned with the results of risk assessment and enable the probation office to stream an offender into a program that targets his/her areas of risk-need.

Using this risk–need logic, sanctions reinforce the principle of effective correctional interventions by using the results of the risk-need assessment to target treatment regimes that have empirically demonstrated reduction in criminogenic need, and thus probabilities of
recidivism. This approach to risk **differs importantly** from the correctional use of static risk (second generation tools) for preventive or selective incapacitation or to deter recidivism through the administration of harsh penalties, which is another possible outcome of risk-based sentencing. Nonetheless, defendants who score high on third generation (rehabilitative oriented) risk-need scales will still continue to endure incapacitation, especially if they are designated as non-responsive to treatment and unwilling to participate in regimes “empirically proven” to reduce their probable risk. Offenders are required to participate in programs designed to address their risk-need area(s). Their success in these programs, which hinges on their ability to “gain insight” into their criminogenic factors and to demonstrate a capacity for change and risk management, plays a significant role in their correctional placement and access to parole (Hannah-Moffat and Yule, forthcoming; Hood and Shute, 2000a; 2000b; Padfield, 2007; Padfield and Liebling, 2000; Padfield, Liebling, and Arnold, 2003; Shute, 2007).

*The Limitations of Risk in Sentencing*

In this section, I argue, the fact that risk instruments do not use general population data suggests that it is difficult, if not impossible, to determine the extent to which other underlying contributors such as the social and economic disadvantage of certain groups, economic status, or some other factor is erroneously influencing the risk prediction. Actuarial risk (both rehabilitative of incapacitation oriented) de-individualizes the assessment of risk by categorizing offenders on the basis of unalterable group characteristics. This means that decisions about community or custodial punishments, the conditions of probation, and levels of supervision are determined based not on what offenders did, but rather on how closely they “are” approximates subgroups of an offender population. Categorizing individuals as risky in comparison with an aggregate group contradicts the jurisprudential value of individualism (Simon, 1988:776).

*Seduced by Science: A Misguided “Trust in Numbers”*

Risk scores impart a moral certainty and legitimacy into the classifications they produce, “allowing people to accept them as normative obligations and therefore scripts for action” (Ericson and Haggerty, 1997:7). Yet, most scholars agree that our present risk knowledge does
not allow us to provide an absolute statement about an offender’s likelihood of recidivism or the timing of potential recidivism. Nor can risk scores tell us with certainty how an offender will recidivate, whether violently, sexually, or simply as a violation of a condition. The possibility of making a prediction error (false positive or false negative) using a risk tool is probable, but not easily determined—a problem that is attracting considerable empirical testing (see Netter 2007). Several scholars have additionally questioned the ethics of punishing someone for potential future behavior (Harcourt, 2007; Hudson, 2003; Maurutto and Hannah-Moffat, 2006 Monahan, 2006; Netter, 2007; Simon, 1993). The danger is that we shift to what Silver and Miller (2002) and Reichman (1986) label “statistical justice,” wherein dispositions are determined on the basis of how closely an offender matches an actuarial profile, with less significance being given to other relevant legal criteria.

The extent to which various tools can reliably predict recidivism within specific offending categories is a key research focus (Kleiman et al., 2007; Schwalbe, 2007; Silver and Chow-Martin, 2002), with the majority of studies focusing on the predictive validity of tools assessing violent recidivism (Gendreau, Goggin, and Smith, 2002). Different risk assessment instruments have been developed to address specific offending categories, but not all instruments appear equally capable of discriminating between high- and low-risk offenders. A risk tools’ empirical foundations and psychometric properties determine the type and quality of predictions possible with that tool.

The use of aggregate statistics to apply individualized punishments has been critiqued on theoretical, methodological, and ethical grounds. The fact that actuarial risk assessments are typically created from the case files of a subpopulation of incarcerated offenders raises concerns about the ability of any instrument to make an unbiased prediction of risk. Prison populations are not random (Netter, 2007); they are the products of past sentencing policies and patterns and they disproportionately represent Blacks, Aboriginals, and other socially disadvantaged groups (Blumstein, 1982; Bushway and Smith, 2004). Some researchers claim that the prison population does not accurately reflect who is at risk of reoffending and question the predictive ability of risk instruments that are based on these populations. These issues are important to the of ethics decision making because the base rate estimates for recidivism may actually be lower in the general offender population than what is predicted on risk assessment instruments. This may
result in the possibility that a more severe penalty is administered on the basis of a risk assessment tool that inflates the actual risk posed by certain groups of offenders. Logically the converse may also occur. Advocates of empirically based sentencing have positioned recidivism as a central organizing principle of sentencing,¹⁰ and risk-need assessments as necessary sentencing aids. These tools are variously described as using hard data, social science and empirical evidence, and actuarially sound data to make predictions relevant to sentencing. However, there appears to be considerable misuse (and little understanding) of the “social science” and “hard data,” which can lead practitioners using the tools to reproduce some of the same difficulties the tools are purported to remedy.

Risk scholars have yet to examine how tools designed to predict recidivism function with the general population. One of the intrinsic difficulties in these tools is that recidivism itself is a notoriously slippery concept that is difficult to operationalize and reliably measure (Maltz, 1984). Within the pool of risk tools, recidivism is variably defined as rearrest, reconviction, or reincarceration, and does not always refer to the original offence(s). The general definitions of recidivism used in many actuarial risk instruments do not differentiate between types of recidivism. For example, tools such as the LSI–R offer a generic prediction of recidivism that cannot reliably indicate if the defendant will breach a probation condition or commit a serious or violent offence.

A noteworthy slippage: predict to cause.

The issue of risk and prediction is further complicated by practitioners’ unfamiliarity with probability statistics and the general tendency to conflate correlations with causality. My interviews with criminal justice practitioners demonstrated that few understand and appropriately interpret probability scores. Despite receiving training on these tools and their interpretation, practitioners tended to struggle with the meaning of the risk score and the importance of the items contained in the assessment tools (Hannah-Moffat and Maurutto, 2004; O’Malley 1998). Rather than understanding that an individual who obtains a high risk score shares characteristics of an aggregate group of high-risk offenders, the individual is likely to become known as a high-risk offender. Instead of being understood as correlations, risk scores are misconstrued in court submissions, pre-sentence reports, and the range of institutional file narratives that ascribe the
characteristics of a risk category to the individual. Significantly, these reports follow the offender through the system and can stick with them for the entirety of their institutional careers, and risk–based characterizations can have significant managerial effects (e.g., classification levels, institutional placement, treatment access, parole release, number and type of conditions, etc.).

In practical terms, correlation becomes causation and potential risk is translated into an administrative certainty. When used at the pre-sentence stage, the courts may assume that a “high-risk” offender poses a greater danger to society and sentence accordingly. Risk scores, however, merely identify who is more likely to reoffend and, in the case of third-generation tools, identify treatment targets for correctional programming that may reduce the likelihood of recidivism. This slippage positions “some” areas of an individual’s life as legitimate targets for corrective evidence-based interventions. Risk scores and accompanying quantifications of “need” are no longer probability scores, but are instead a prescription for correctional treatment. The “abstract” risk score is converted into a correctional artefact that enables the efficient management of correctional populations and resources and defines a person’s criminal justice experiences.

Risk scores have a range of purposes: targeting and streaming various types of offenders into different kinds of programs; reducing penal costs; and regulating entry into alternative measures, diversion, community sanctions, and custody. Risk schemes can be used by criminal justice officials to systematically target groups of high to low offenders. For example, the Virginia Criminal Sentencing Commission developed a method for diverting 25% of non-violent, prison-bound offenders into alternative sanction programs using risk assessment to identify the lowest-risk offender to help inform possible diversion options for non-violent offenders (Kleiman, Ostrom, and Cheesman, 2007).

Concerns about prediction have not impeded the expansion, development, and use of risk instruments. Actuarial instruments continue to appeal because of their purported ability to classify offenders based on a set of statistically relevant factors, without a reliance on clinical discretion and because they offer the prospect of “colonizing the future.” (Giddens 1991) Proponents of these tools deflect some critiques by stressing that the practice of identifying an individual’s level of risk based on their membership in a classification group results in superior predictive validity of the tools over clinical judgement alone (Kim et al., 2008). However, they
have not conclusively demonstrated that these “classification groups” are free from biases or other non-normalised contributors to recidivism (i.e., arrest bias, practitioners’ interpretations, gender, race or other social factors). On the level of correctional practice, James Bonta, a staunch advocate for risk assessment, cogently observes that, “It is one thing for scientists to demonstrate that a risk instrument or a treatment program can work but it is a very different matter to make it work in correctional agencies with a diverse work force in terms of education, values and experience, conflicting criminal justice policies, and management practices that are not conducive to selecting and training of staff in effective assessment techniques” (Bonta, 2008:np). Further sentences predicted on risk will not guarantee the availability of correctional programs. Regrettably, there is considerable evidence of governments’ not providing resources for correctional programming (Quigley, 2008) and, on a very basic level; it is faulty to assume that a cohesive and directive evidence-based rehabilitative framework on how to “treat” offenders, guide practitioners, and organize correctional regimes exists.

**Gender, Race, and “Evidence-Based Risk”**

Since risk instruments cannot predict who will recidivate, what should be addressed and minimized are the factors that contribute to the causes of crime. This is certainly easier said than done, as evidenced by the considerable academic debate on what causes crime (and who is detected, prosecuted etc.). Nevertheless, social context—gender, race, and economic and socio-structural factors—are generally agreed to play a role. Risk assessment does not and, by virtue of its internal logic and structure, cannot account for these variables. This fact has contributed to the extensive debate about whether conventional risk assessment instruments can be usefully, or ethically, applied.

I do not think that risk instruments such as the LSI–R are universally applicable and suggest that even if their “validity” were demonstrated, these instruments may not be the most appropriate method for identifying and responding to the therapeutic needs of female offenders (cf. Vanhooris et al 2010; Hannah-Moffat 2009). Several scholars (Morash 2009; Belknap and Holsinger 2006; Blanchette and Brown 2006; Bloom, Owen, & Covington, 2003; Hannah-Moffat and Shaw 2001; Morash, Bynum, & Koons, 1998) fault risk instruments for over-classifying women, ignoring the risk factors and needs most relevant to women offenders, and
for having little regard for the validity of specific risk instruments for women\textsuperscript{11} (Van Voorhis et al. 2010). The predictive reliability of existing instruments for women and racialized populations is unclear because the criminogenic factors included in generic risk tools are derived from statistical analyses of aggregate male correctional population data and are based on male-derived theories of crime; a gender/race problem is therefore built into the tools (Van Voorhis & Presser, 2001). Blanchette and Brown’s (2006:140) comprehensive analysis of “some of the most commonly used mathematical risk assessment tools clearly demonstrated that the science of recidivism prediction of girls and women lags far behind that for males.” Feminist researchers have repeatedly critiqued mainstream criminology for treating females as “afterthoughts,” and for the uncritical use of male norms in the management and treatment of incarcerated women; yet, these critiques have been ignored in the development and application of risk tools. An analogous race problem stems from a lack of attention to the racialized nature of offending and imprisoned populations and to the specific needs of non-white offenders. The science of risk assessment as it pertains to racialized populations is underdeveloped and, consequently, using conventional risk tools to predict recidivism with these populations is of questionable legal and ethical value.

Characterizing men and women as having essentially the same universal risks and needs reproduces the male normative criteria that feminist (and critical race) researchers have theoretically and empirically contested.\textsuperscript{12} The ‘sameness’ argument demonstrates little regard for, or understanding of, the intellectual breadth and depth of feminist and critical race theory, the sophisticated feminist and critical race critiques of methodology, or the ample empirical and theoretical academic literature documenting how women’s crime is qualitatively and quantitatively different from men’s, even if they commit the same type of crime. Research has demonstrated differences in the motivational factors that lead to women’s use of violence, involvement in drug and property crimes, patterns of substance abuse, and how factors such as drug use are connected in gender-specific ways to initial and continued prostitution and other crimes (for example, see Blanchette and Brown, 2006; Bloom and Covington, 2003; Daly, 1992, 1994; Hannah-Moffat and Shaw, 2001; Clarke, Monahan, and Silver, 2003; Heimer and Kruttschnitt 2006; Moretti, Odgers, and Jackson, 2004). International data also clearly shows that women commit few violent crimes, are infrequently repeat offenders, and that when they do
reoffend, their crimes tend not to escalate in severity (Kong and AuCoin, 2008). Empirical analyses of risk tools, including the LSI–R, show that the criteria for establishing levels of risk routinely pay little attention to gender, racial, or ethnic differences, or to the differing social, economic, and political contexts in which these tools are deployed (Maurutto and Hannah-Moffat, 2006). For example, a recent study by Reisig, Holffreter, and Morash (2006), reported that

the LSI–R was found to correctly classify female offender only when: (1) they did not follow gendered pathways into offending; (2) they offended in contexts similar to that of male offending; (3) they occupied a relatively advantaged social position in society. … The LSI–R misclassified a significant portion of female offenders who were socially and economically marginalized and who offended within typical gendered contexts, particularly female offenders who were drug-connected or who could be classified as harmed and harming women. (p. 384)

Concerns about gender and risk assessment are also relevant to tools such as the PCL–R, which has been relatively well researched on samples of women (Brown and Blanchette, 2006). In this case, research suggests that some gender differences in the risk factors included in the PCL–R and their salience are evident (Blanchette and Brown, 2006:73; also see Warren et al., 2003). Researchers continue to test whether risk assessment tools developed for male offenders can be used accurately for female offenders (Harer and Langan, 2001). Conventional risk assessment scholars are beginning to go beyond concerns with predictive reliability and validity to closely examine the construct validity of non-gender-specific needs for female populations (Brennan and Austin, 1997; Brennan, 1998; Blanchette and Brown, 2006; Van Voorhis, Salisbury, Bauman, and Wright, 2007; Van Voorhis et al. 2010; VanVoorhis, 2005). At present, the debate on the suitability of tools such as the LSI-R for women is significantly polarized;13 this warrants caution about their use at sentencing.

The uncritical adoption of risk instruments is also problematic from the standpoint of race and social inequality. Risk research has not adequately vetted tools for race. Research that has examined the suitability of available risk instruments for racialized offenders reports that racial minorities have different needs (Gavazzi, Yarcheck, and Lim, 2005; Petersilia and Turner, 1987;
Mitchell, 2005) and that race and ethnicity can influence practitioners’ attribution of risk factors to racialized offenders (Hudson and Bramhall, 2005). These studies raise concerns about the extent to which “risk assessment and risk prediction is a transcultural, transracial” phenomenon (Severson and Duclos, 2003). Race differences in exposure to risk and vulnerability are not presently accounted for in empirical risk studies. The logic of risk assessments is premised on a set of assumptions that underestimates and devalues social and racial inequality, while simultaneously decontextualizing the criminal act, the offender, and his or her history, and needs to be carefully considered before these tools become further embedded into sentencing practices (Hannah-Moffat and Maurutto, 2010; Harcourt, 2007; Hudson and Bramhall, 2005; Pridemore, 2004; Severson and Duclos, 2003; Durrance and Williams, 2003; Bhui, 1999).

Racialized social economic structures and context are relevant to the production and composition of the offender population. Proponents of risk instruments often overlook the fact that both static and criminogenic risk variables cannot be easily abstracted from the socio-political, economic, and cultural specificity of individuals. For example, if one examines the general risk-need factors and compares these factors with the lived reality of Blacks and Hispanics in the United States or Aboriginal people in Canada (Maurutto, unpublished), it is clear that these marginalized groups will unavoidably score higher on risk instruments because of their elevated exposure to risk, racial discrimination, and social inequality—and not necessarily because of their criminal propensities or the crimes perpetrated. Marginalized individuals’ lives tend to be mired by a range of criminogenic and other needs, and consequently risk scores reflect systemic factors. High risk scores are associated with custodial sentences and/or a greater number of conditions attached to their disposition, making them more vulnerable to breach, increased surveillance, and further criminalization.

Race and gender are complex social constructs that cannot simply be reduced to binary variables and tested for significance (predictive validity and reliability) in risk instruments (Hannah-Moffat and O’Malley, 2007). An emphasis on narrowly “validating” risk tools for use on diverse populations that were developed using theories and research mainly about Caucasian men’s crime may produce disparity in sentencing. When the inner logic of risk is exposed, important theoretical and methodological concerns about the relationships between race, gender, and social inequality and specific risk factors become evident.
The science supporting risk tools is contested and insufficiently advanced to claim that these tools do not replicate or produce forms of systemic discrimination, or worse, as Harcourt (2007:3) theorizes “a ratchet effect,” wherein “the profiled populations become an even larger portion of the carceral population” with highly determined consequences on their employment, education, family and social outcomes. Individuals who are racialized, live in poverty, are unemployed, and/or struggle with mental illness are potentially disadvantaged by risk-need criteria. Although generic risk - need factors (i.e., substance abuse, marital family difficulties, and employment) are presented in existing research as “relevant” or “predictive” for women and diverse offenders (see Vose, Cullen, and Smith, 2008), these factors are experienced differently and have different effects. As critical race theorists aptly illustrate, it is impossible to treat individuals fairly if they are treated as abstractions, unshaped by the particular contexts of social life. The nuances of gender, race and social inequality are of paramount importance to the production of a just and equitable sentence. The failure to acknowledge and meaningfully account for differences (beyond risk) among individuals is a significant jurisprudential concern.

Discrimination and the structure of “knowledge”

In a 2003 report, the Canadian Human Rights Commission\(^\text{16}\) (CHRC) criticized the use of third-generation risk-need instruments and their content, arguing that some common risk factors are systemically linked to prohibited grounds of discrimination. This is concerning, as an incorrect assessment may produce significant burdens, barriers, or missed opportunities for offenders.\(^\text{17}\) Ultimately, the CHRC (2003) argued that risk-need assessment technology was important for classifying offenders for treatment, but that the instruments should be revised to ensure their suitability for their intended subjects. The CHRC stated that “assessment and testing processes must be responsive to the populations to which they are applied and properly crafted to meet the purpose they are intended to achieve. Where assessment tools do not meet these requirements, they are blunt instruments that tend to lead to unjustifiable differential treatment. In the absence of adequate testing and modification, these instruments should not be used [emphasis added] on women or Aboriginal offenders” (CHRC, 2003:np).

“Empirical knowledge” is rarely positioned in a broader discussion of the range of relevant approaches to the reduction of offending available to the court. Instead, RNR and related
risk-need approaches can be misrepresented as superior to, or more scientific than, other approaches. Feminist/race critiques undermine the “empirical” bases of many conventional risk instruments and raise doubts about their use with certain groups. Alternate gender or race-based holistic approaches are equally legitimate and can produce alternative understandings of and managerial approaches to risk. For example, reforms to Canadian sentencing law in 1996, and the Supreme Court of Canada decision *R. v. Gladue* (1999), opened the door for a new normative set of practices intended to reconstitute the racial legal subject (Rudin and Roach, 2002; Daubney, 2002; Stenning and Roberts, 2001; Vancise and Healy, 2002, Maurutto and Hannah-Moffat, 2010 -LSA). The court had to address systemic racial discrimination and redress the disproportionate incarceration of Aboriginal minorities. The *R. v. Gladue* decision requires that all criminal courts consider the broader collective social history of Aboriginal defendants; this requires a different way of understanding the risk-need of Aboriginal offenders. Hannah-Moffat and Maurutto’s (2010) article on race, risk, and pre-sentence reports argues that although risk-based pre-sentence reports recognize issues related to race, their structure and emphasis on actuarially based risk assessments frames race and risk differently from specialized Aboriginal reports (e.g., Gladue reports). Gladue reports use of more holistic approaches and cultural impact factors to analyze and contextualize risk-need. The authors maintain that the conceptualization and relevance of race is limited by actuarial risk logic. This is a significant consideration for jurisdictions (i.e., United States, Canada, Australia) with a disproportionate racial representation in prison populations which adopt evidence-based sentencing practices that do not reflect the range of evidence on race (and gender), which could compromise the “objectivity” risk-need practices.

The use of recidivism risk assessment instruments can only be as helpful as they are accurate. More empirical research is needed about: the composition of tools; the effects of various risk variables on specific populations; the use and impact of risk technologies on the management of offenders; and ancillary system effects. There is a limited assessment of the current “scientific” evidence about the validity and reliability of various tools entering into legal discourse. Nor is there a healthy debate about the types of risk instruments that should be integrated into sentencing practices. Even research about risk instruments with a high degree of “capital,” such as the LSI series of tools, has reported competing findings. Yet, few scholars
have examined the impact of risk information on sentence outcomes or applied knowledge about
differential impacts of criminal histories or race and sentencing to the study of the composition
and results of risk instruments. It is critically important to do so, because if the characteristics
associated with an assessment of “risk” are not normalized for bias (i.e., in initial arrests, crime
for which no arrest was made, racial targeting, race and gender difference in crimes committed
etc.) then the prediction (risk score) is inherently flawed.

Neutrality and Transparency

The empirical evidence of “what works” that informs risk-need is presented as morally
neutral from a realistic perspective. Conversely, critical scholars and other stakeholders are
questioning the assumptions and techniques of regulation that underpin risk frameworks and
their impact on due process, justice, and governance (O’Malley, 2004; Kemshall, 1998; Hudson,
2003; Ericson and Doyle, 2003).

Risk instruments’ score and structure tend to “black-box” and obscure the subjective and
arbitrary nature of questions and judgements associated with risk-factor scoring. For example,
risk tools’ operational manuals outline normative questions designed to determine if particular
risk-need areas are criminogenic. The nature and type of questions vary across tools, but the
questions are designed to gather information about relevant relationships\(^{18}\) and circumstances, so
that practitioners may assess risk of recidivism and treatment needs. For example, regardless of
the offence, the LSI–R interview guide for the risk factor “employment” asks a series of
questions about employment and school history, including whether a defendant “get’s along
with” teachers, employers, and colleagues. The assessment questions also address sources of
income, fiscal concerns, banking and credit histories (i.e., if personal cheques bounced or were
returned for non-sufficient funds), if the defendant has a budget and follows it, whether the
defendant and significant others fight about money and/or children, whether there was an
‘unwanted pregnancy’ or infidelity, if divorce has been contemplated, and a series of additional
questions about personal and emotional issues, leisure habits, and substance use (Andrews and
Bonta, 1995). Answers to these questions form the basis of the risk-need factor score, but the
only aspect of the risk-need score visible and known to the court is the score and the
recommendation derived from the score. Neutrality is assumed in the questions, the way the
questions are asked, and in the way that the responses are interpreted and recorded. However, as my research showed (Hannah-Moffat et al., 2009), practitioners administering risk instruments such as LSI–R “adjust the assessment of criteria in order to control the final score, rather than relying on formal overrides, which are an option that allow for final risk scores to be adjusted for individual cases” (405). Risk manuals acknowledge that adjustments (overrides) may be required, but state that these should not exceed 10% of cases. When and if a practitioner elects to use an override, they must document the reasons for not conforming to the outcome of the risk score. This level of accountability can deter the use of overrides and favours “criteria tinkering.” It is also clear from the nature of these questions that assessing risk through these actuarial tools is not a neutral statistical exercise. Assessments of risk are produced using an investigative and interpretive process that is largely concealed within the process. Practitioners still rely on their discretion to selectively use responses to interpret, target, and isolate facts about past experiences, and to make claims about the probability of reoffending.

In risk assessments, the simplification and organization of facts can lead to prisoner categorization according to crude typologies; small details tend to form the foundations of the construction of stereotyped identities and agency. Through the use of risk assessment instruments, the potential complexities and subtleties of meaning are compacted. Actuarial methods can accentuate prejudices and biases that are built into law and criminal law enforcement (Harcourt, 2007) because the opaque structure of assessment instruments makes broader discriminatory practices less visible and contestable.

Although the information acquired from risk tools is often useful to correctional personnel, its use in court opens a legal and evidentiary “can of worms” and raises a host of important sentencing considerations. For example, in the United States, the federal sentencing guidelines and case law explicitly direct that race, sex, religion, national origin, socioeconomic status, and a disadvantaged upbringing “are not relevant” in the determination of a sentence [sic] in addition education, vocational skills, employment record, family ties, age, mental and emotional condition, and substance abuse are not ordinarily relevant in the determination of a sentence (Monahan, 2006:397).
Monahan (2006:397) notes that with the single exception of criminal history, “virtually all of the variables that potentially could be used as scientifically valid risk factors of violence under a forward looking consequentialist ‘crime control’ theory of punishment are explicitly excluded from consideration in federal sentencing guidelines.” Arguably, even criminal history is gendered and racialized; it is affected by the offender’s age, as well as policing and charging practices (Bushway and Morrison Piehl, 2007). Implicitly, risk-need instruments make certain variables relevant to sentencing because they are the foundation of the score.

Pre-sentence reports (PSRs) frame legal subjectivities for the court; they inform the court about the offender’s risk and treatment potential, and contribute to the type, length, and conditions attached to the final sentences imposed. Some of the data from risk assessments is of questionable legal relevance to sentencing and are obtained via “collateral contacts,” without the benefit of legal protections, cross-examination, or authentication. The use of such collateral contacts (who are not subject to cross-examination), subjective interpretations, and extralegal criteria by probation officers to compile the risk score and to formulate recommendations about sentencing can undermine broader principles of sentencing jurisprudence (Cole and Angus, 2003:17). This problem is particularly acute when practitioners’ recommendations do not disclose the use of a risk instrument. For example, in Ontario, Canada, probation officers are mandated by policy to use the LSI–R to complete a PSR, but they are also told to not disclose the score or the fact that they used the tool in their court reports, an omission intended to prevent probation officers from being called to the stand for cross-examination on the risk instruments. Yet, research demonstrates that the assessor’s choice of informants and their interpretation of the authenticity of the informants’ claims determines what information “counts” and becomes part of the “official” record (Ballucci, 2008; Maurutto and Hannah-Moffat, 2006). At minimum, courts should require the disclosure of the use of risk instruments in the reports and formulation of sentence recommendations, particularly because research reports high rates of concordance between the recommendations of pre-sentence reports and actual sentences (Cole and Angus, 2003; Hannah-Moffat and Maurutto, 2010).

Legal Standards and Quality of Evidence

Wellford (2007) reminds us that
“[o]ur understanding of sentencing has been limited by the narrow view we have taken of what and who influences sentencing [...] United States v. Booker [U.S. 220 (2005)] and Blakely v. Washington [124 S. Ct. 2531 (2004)] draw attention to the fact that whatever is used in sentencing should be subject to review and debate [emphasis added]. Although the justices think of this as a legal standard to be met, we as researchers should recognize that our goal is the same—understanding all of the factors and actors that actually determine criminal sentences.” (p. 400)

Courts are swayed by empirical evidence and expertise. Litigation in some jurisdictions notes that aggravating factors that expose a defendant to higher penalties must “be proven by a higher standard of evidence, beyond a reasonable doubt before a jury rather than a preponderance of evidence before a judge” (Hunt, 2007:438). In a similar vein, Quigley, a Canadian lawyer (2008:531) argues that, “[w]hile the accumulation of evidence is common to all of these disciplines, law assesses evidence in a context bounded by values, rights (particularly constitutional rights), rules, and principles that are broader, and, therefore, because of rules concerning the inadmissibility of certain types of evidence, more constraining than those governing scientific inquiry.” This rigor should apply to risk-need assessment because the score can have significant consequences for the offender in terms of the opportunities provided, levels of supervision, and the conditions imposed. Although contested in our adversarial culture, expert knowledge is established and attributed authority and legal relevance. What is remarkable is that few court actors are aware of, or concerned about, how risk scores are produced and interpreted, or the contested science of risk prediction, especially as it applied to non-specific offending categories and “general” recidivism.23

The courts in various jurisdictions generally accept these technologies and are reluctant to debate their validity. Canadian court cases that raise concerns about the LSI–R have typically sidestepped the debate on jurisdictional grounds and have deferred to the administrative authorities of probation departments to independently determine the suitability and type of risk instrument to use during assessments (Hannah-Moffat and Maurutto, 2010; 2007; Cole and Angus, 2003; Cole, 2008). Although the American court context is considerably more complex, similar tendencies exist. For example, in June 2010 the Indiana Supreme Court ruling in Malenchik v. Indiana24 on the probation department’s use of the LSI–R held that “legitimate
offender assessment instruments do not replace but may inform a trial court’s sentencing determinations and that, because the trial court’s consideration of the defendant’s assessment model scores was only supplemental to other sentencing evidence that independently supported the sentence imposed. “This judgement, like many others, sidesteps broader questions of the impact and relevance of ‘general’ risk-need determinations at sentencing.

Scholars, practitioners, and law and policymakers are more regularly claiming (without careful examination) that accurate prediction of risk can be achieved and is useful to a sentencing decision. This infers that the use of risk assessment technologies contributes to the production of a “just sentence.” The concept of justice is complex. Scholars on both sides of the risk and evidence-based sentencing debates mobilize justice-based concerns to validate their positions. More legal and empirical analysis is required to tease out how risk may enhance correctional practices in some cases, but compromise jurisprudential ideals of justice in others. One salient debate to consider is the extent to which differences and social context are important to sentencing.

Few courts have entertained an analysis of these instruments or require expert testimony on the psychometric properties of tools and their supporting research, even in cases where local tools, which are not universally accepted or empirically vetted, are used. Too little empirical and socio-legal research examines how factors such as race, gender, educational attainment, and mental health impact sentencing practices, and how forms of discrimination can be reproduced and possibly magnified through the uncontested use of actuarial risk instruments as exceptions (Harcourt, 2007). Greater transparency and thought should be devoted to how practitioners communicate risk information to legal decision makers. Thus far it appears that the courts and others are accepting a relatively weak standard of scientific evidence when using risk-need instruments.

A Piece of a Puzzle: Weighing Risk Information

Research has yet to investigate how judges and practitioners use and weigh risk information in decision making. Interestingly, research reports that judges and legal practitioners are generally supportive of risk instruments, but they know little about risk technologies (Bonta, Bourgon, Jesseman, and Yessine, 2005). Research reports that judges often will use the clinical
opinions that fit their perceptions of risk (Scott, 2008), and that they use the information that is available to them to predict risk and assign punishments (Bushway and Smith, 2007), but little is actually known about what factors judges consider in their own professional assessments of risk (Vigorita, 2003:361). Judicial decisions, including risk decisions, are likely based on a small number of factors, either to simplify the decision-making process or because certain factors are the attributes associated with the typical case (Vigorita, 2003:364). If individual factors could be identified, it would be essential to determine how they are actually used in courts.

I have challenged the assumptions that lawyers, judges, and other relevant professionals are familiar with the limits and strengths of actuarial risk instruments and that they can correctly interpret and apply the findings to make recommendations for offenders and sentence them. Part of the problem is that actuarial risk assessment instruments are created from subjective processes that rely on a particular body of largely uncontested social scientific evidence. I am not implying that there are no debates in the literature on risk and “effective correctional intervention,” but rather that thus far, research on evidence-based sentencing has not closely studied the epistemological architecture of risk-need evidence. It is clear that risk-need models conflict with other equally important ethical and legal commitments to race- or gender-responsive correctional treatment (Hannah-Moffat, 2009; Hannah-Moffat and Maurutto 2010).

In totality, risk instruments demonstrate an increasingly refined capacity to sort and classify criminalized populations. Even in light of the recognition that actuarial risk assessment instruments may be more effective than clinical assessment, there remain some unanswered questions in the literature. There also remains a gap with regard to how much influence these instruments actually have in determining an offender’s principle sentence. Since risk practice should result in a redistribution of penal populations, a full appreciation of its impact would require an integrated study of sanctions at multiple jurisdictional levels. In the United States, this would require an examination of the use of risk-based sentencing practices in federal and state sentencing courts (Wolfe, 2008). It would be useful to examine if the flow and outcome of cases in various types of sanctions (probation, imprisonment, and diversion) change with the introduction of evidence-based sanctioning. However, caution should be exercised in any attempt to study the outcome of programs into which offenders are pre-selected through their
risk-need score, because there is an obvious selection bias created by risk-need-based selection into programs or sanctions.

Unresolved Debates: Risk and Competing Theories of Punishment

The actuarial risk instruments are highly structured, empirical assessments that target potential rather than actual crime and offenders. Accordingly, actuarial sentencing stresses recidivism risk. Scholars have identified a number of important conceptual difficulties associated with a reliance on actuarial risk at sentencing. For example, some argue that the incorporation of risk technologies into sentencing prioritizes forward-looking theories of punishment and masks a range of concerns about disparity, discrimination, and “just punishments.” Critical scholars have characterized actuarial technologies as a negative development, which can result in racial targeting, deindividualization, social exclusion, a prioritization of recidivism and future conduct, and the devaluing of social context (Simon, 1988; Feeley and Simon, 1992; 1994; Hudson, 2003; Hannah-Moffat, 2004; Maurutto and Hannah-Moffat, 2003; 2006; 2008; Monahan, 2006; Harcourt, 2007; Netter, 2007)

O’Malley (2004; 2008), however, argues that the focus on risk’s negative side fails to adequately account for how risk embodies a heterogeneous array of practices with diverse effects and implications. Following François Ewald (1991), he argues that risk is an abstract technology, which is always shaped and given effect by specific social and political rationalities and environments (O’Malley, 2008:453). Extending this point, I believe that law and criminal justice contexts play a pivotal role in shaping understandings of risk and how it should or should not be incorporated into practices such sentencing. A quick survey of the use of risk instruments in Canada, the United Kingdom, Australia, and the United States reveals a general lack of consensus on the suitability, use, and actual role played by actuarial instruments in sentencing. Some jurisdictions provided judges with risk instruments and explicitly required the use of risk score to determine appropriate sanctions. In other jurisdictions, the use of risk instruments is most pronounced in sexual or violent-offender cases where time and severity of offence both allow for, and require, a detailed consideration of the offender’s future conduct. Risk instruments are commonly used by probation officers to write pre-sentence reports and make sentence recommendations, as well as to facilitate the post-sentence management of offenders. Parole and
correctional officials have a long-standing interest in the use of actuarial tools for classification and decision making.

Actuarial risk is a seemingly progressive penal practice that operates in accordance with the institutional conditions in which it is realized. Despite their conceptual and methodological deficiencies, it is unlikely that the emphasis on risk will dissipate. Consequently, the appeals to actuarial sentencing can simultaneously support two distinct utilitarian goals of sentencing: incapacitation and reformist intervention. The later reformist logic, although not without technical, ethical, and conceptual limits, can shift sentencing practice.

I would like to consider how the appeal of risk thinking can enable a return to a more therapeutic approach to sentencing and possibly reduced reliance on incapacitation, which is in danger of spiralling out of control. This is not, however, to suggest that the framing of “treatment” in a risk model is unproblematic (see Hannah-Moffat, 2004; 2007; Maurutto and Hannah-Moffat, 2006) or that the jurisprudential concerns about how actuarial risk undermines just desserts and, more precisely, the principle of proportionality is insignificant. To the contrary, there are a number of persuasive and astute arguments for limiting the role of actuarial risk logics in sentencing decisions due to its conflict with existing principles and theories of “just” punishments (Simon, 1988; 2005; Hudson, 2003; Maurutto and Hannah-Moffat, 2006; 2007; Monahan, 2006; Harcourt, 2007). This is particularly the case when a jurisdiction explicitly prioritizes proportionality in sentencing (i.e., Canadian Youth Criminal Justice Act; Hannah-Moffat and Maurutto, 2004; 2008).

Putting aside the importance of proportionate sentencing for a moment, it is worth considering how certain types of risk assessment logic might operate to change the type of sanction an offender receives and the consequences of these shifts. For example, if the risk-need principle underpinning third-generation risk assessments could be judiciously implemented, then a portion of offenders could theoretically benefit from a “less severe sanction.” Accordingly, a risk score could help determine who can be diverted out of the system or sent to a specialized court or treatment program. Recall that the risk principle states that treatment services need to be matched to an offender’s level of risk; offenders who present a high risk are targeted for the greatest number of therapeutic interventions, and those who are low risk are targeted for the least. Some evidence provided by proponents of the RNR suggests that too much intervention
with low-risk offenders can escalate risk (Andrews, 1991:11). Thus the low-risk offenders could be diverted out of the system in the interest of reducing recidivism, and offenders with concrete interventions and supervision plans are more likely to be characterised by the court as “manageable risks.” This possibility is supported by recent evidence suggesting that judges are more inclined to impose a community sanction when they receive information on how to manage an offender’s risk (Dolores and Redding, forthcoming; cited in Maurutto and Hannah-Moffat, 2010). Further, Dumanis (2009:25) argues that “as a culture, we have taken the rehabilitation out of prison and focused on punishment, creating an untenably large prison population. Now, we are forced to rethink our way of imposing justice. Evidence-based sentencing merges punishment with rehabilitation. Imposing a sentence with appropriate conditions based on the defendant’s individual risk to reoffend and need for treatment or programming.” The use of risk-need can reinvigorate rehabilitation; however, it should not be done in a vacuum. There is ample empirical evidence of the unintended consequences of well-meaning “rehabilitation.” Risk-need informed probation conditions and treatment requirements evoke the causal slippage discussed earlier and make the marginalized offenders more susceptible to increased numbers of conditions and thus surveillance and possible breaches.

To use an imperfect risk instrument to “ratchet up” sentences of high-risk offenders (i.e., sex offenders) or to impose indefinite sentences is considered by some to be a “misuse of evidence-based sentencing” (Etienne, 2009:59; also see Harcourt, 2007). But, methodological capacities aside, the use of risk instruments to seemingly customize sentences through the provision of targeted interventions and clear strategies of risk management is persuasive. Some evidence suggests (Dolores and Redding, 2009) that judges were more likely to release to the community when the risk assessment included information on risk management than when it only provided a prediction of risk level. Risk-need assessment is being popularized as a reasonable way of restricting custodial populations, reinvigorating rehabilitation, and enhancing public safety through “anticipated” reductions of “recidivism.”

However, is the fact that the reduction of recidivism is anticipated but not certain, important in this context? Answering this question will require more empirical evidence to determine if these intentions are realizable and with what effect. Given the limitations of existing risk instruments, it is not clear whether or not the administration of risk will reproduce existing
racial inequalities, unfairly target specific groups, and result in increasingly layered and severe penalties.

Risk can offer a politically defensible way out of mass incarceration, but part of the empirical dilemma with risk is that it is fluid and interpretive. Risk levels and thresholds often vary from institution to institution and sometimes indiscriminately vary over time even when the same risk instrument is deployed (i.e., Virginia, LSI local population norms). Wormith (1997) argues that “the number of risk levels in any scale or instrument is decided arbitrarily by the developer or the agency using it.” As resources contract or expand, the tools can be adjusted accordingly. Correctional management scholars (White, 2004; Kreamer, 2004; Dal Pra, 2004) indicate that the development and adoption of particular risk tools (actuarial risk versus risk-need) are decisions that should be made in concert with an evaluation of correctional mission statements and resources. Such claims confirm that risk technologies can be and are used instrumentally by correctional organizations. Clearly, corrections organizations select the risk technologies that “fit” their agency’s vision and mission (Maurutto and Hannah-Moffat, 2006; 2003). Averting disaster has become a political imperative for governments anxious to safeguard their reputation against adverse fallout that comes if harm happens (Zedner, 2008). It is possible that the supporters of risk approaches do not share the goal of reducing crime or of matching offenders with the programs that “work,” but rather seek only to defensibly manage populations.26

Actuarial risk has both supporters and detractors who argue “against prediction” (Harcourt, 2007), cautioning law and policy makers about some of the assumptions and ambiguities of actuarial technologies. To what end could risk assessments be jurisprudentially relevant to sentencing for judges concerned with crime prevention, recidivism and effective interventions? In this context, a probabilistic statement of risk and systematic weighing of risk factors may help a judge craft a sentence and apply meaningful conditions. New risk configurations inherent in tools like the LSI are important not only for theorizing the concept of risk, but also for understanding new and evolving penal strategies (Maurutto and Hannah-Moffat, 2006).

Pitfalls and Potentials: Questions Left Unanswered
Courts embracing risk are joining other institutions that believe “the world can be made more secure by ever more perfect knowledge of risk. This leads them to search incessantly for whatever rules, formats, and technologies will allow them to feel that they are closer to perfection” (cf. Ericson and Haggerty, 1998 on police). But imperfect rules, technologies, and predictions will remain. The fact that risk appears to be an efficient, empirical alternative ought not to insulate its normative dimensions. Questioning the technical efficiency of a risk approach is important, but it invites improvements or technical refinements and does not tackle the prior question of whether or not it is justifiable (Zedner 2008:360).

Actuarial risk instruments systematically organize a diverse range of information about an offender to guide practitioners through a logical and simple process to itemize and score that information. Theoretically, risk templates ensure that nothing is overlooked or missed when reviewing an offender or case history. Risk instruments are structured to produce a managerial form of defensible, consistent decision making. They ensure that to a certain degree all practitioners use the same variables to make decisions. Interviews with practitioners revealed that risk instruments are embraced because the tools standardize decision-making criteria, enhance the defensibility of decisions, and ensure that all the players in the system are working with the “same information,” making case files easier to transfer (Hannah-Moffat, Maurutto, Turnbull, 2009). The use of risk instruments can have considerable cache with “elected” judges and prosecutors who must defend their decisions to an electorate concerned with security. Interviews with practitioners in discretionary decision-making contexts consistently showed that they believed actuarial risk scores can neutralize politics. Institutionalization of risk insulates practitioners who follow policy guidelines, scapegoats those who do not, and creates new forms of organizational accountability. Risk instruments, regardless of their flaws, foster greater professional confidence in the system because they appear objective, rational, and empirical. But are they?

The uncritical acceptance of science and related risk technologies can jeopardize due process, produce disparities and discrimination, undercut proportionality, escalate the severity of sentences, and punish individuals for crimes that they have not committed. However, risk (and some evidence-based practices) may facilitate a reduction in penal populations, and over time lead to the application of different and perhaps more constructive interventions. Acknowledging
this possibility, I remain concerned about how introductions of risk into sentencing will shape punishment and impact already disadvantaged and unmotivated or treatment-resistant defendants and further how the use of actuarial methods can accentuate the prejudices and biases that are built into law, punishment, and criminal law enforcement (cf. Harcourt, 2007).

This discussion has prompted a number of questions for which there are presently insufficient answers:

1. To what extent should risk predictions inform sentencing practices?
2. How should risk predictions be balanced with wider sentencing goals?
3. What is the impact of using risk prediction in conjunction with other sentencing guidelines (i.e., mandatory minimums) and priorities (i.e., due process, just desserts, etc.)?
4. Are the risk instruments used in given sentencing jurisdiction responsive to the populations to which they are applied?
5. Are the instruments used properly crafted to achieve their intended purposes? This question raises concerns about using risk instruments in sentencing without first considering their compatibility with broader sentencing jurisprudence and guidelines.
6. Are racialized populations disproportionately represented in risk categories?
7. Does risk-need assessment increase sentence uniformity?
8. How does risk assessment data compare with general population data?
9. What are the implications of not using the risk assessment device when tools exist? If a tool is used, then how does one reconcile conflicting understandings about how it should be used to predict and/or sentence? For example, some of the research supporting risk tools argues that the combination of actuarial scores with clinical judgments inevitably produces lower accuracy than actuarial scores alone, and that “unaided” clinical judgment is less accurate than actuarially aided judgments. This raised the interesting paradox of whether judges are, or even should be, concerned with the accurate prediction of recidivism and the legal relevance and weight of that information.
10. Given the technical and conceptual limits of risk, how and when can these tools be meaningfully used?
The structural integration of risk into sentencing guidelines can restrict judicial discretion to varying degrees. However, it also simultaneously shifts discretion to the organizations and individuals producing risk templates and providing courts with risk scores. These organizations are less occupied with the normative concerns of law and justice. Although law is occasionally labelled a science, it is also seen as an “imperfect art” (Quigley, 2008:531). If sentencing is to assume an “evidence-based” focus, then it would be prudent for legislatures and legal scholars to integrate the range of evidence on topics such as risk and correctional intervention. I share the scepticism of socio-legal scholars (Slobogin, 2005; Hudson, 2003; Harcourt, 2005; Simon, 2005; Monahan, 2006) who in varying ways advocate for courts and legal scholars to pay greater attention to how, for what reasons, and based on what “evidence” we can justify the deprivation of liberty based on risk scores. Arguably, we should pause to reflect on the complexities of risk-need assessments and concordant calls for and against evidence-based risk jurisprudence.

References


Cases Cited:

---

1 I would like to thank Linn Clark, Tony Doob, Amy Klassen, Paula Maurutto and Shawn Bushway and reviewers and commentators for their contributions to earlier drafts.

2 Interview data.

3 There are a plethora of risk instruments available. For example, tools like the Level of Service Inventory – Revised (LSI–R), Static 993, and more recent domestic violence risk assessment guides (Ontario Domestic Assault Risk Assessment (ODARA) and Domestic Violence Risk Appraisal Guide (DVRAG))3, which are designed for and administered and interpreted by non-psychological/psychiatric professionals. There are a multitude of offender and offence (female, youth, violent) “risk instruments” that exist or are being developed. Most risk instruments share a common structure and focus on comparable “risk” factors because they draw on the similar empirical literatures. What varies among these tools is the degree of emphasis placed on a particular set of factors; cut-off scores for the categorization of risk, or nee; and the targeted offender population (youth, domestic violence, violent or sexual offenders). Consequently, the risk-assessment industry is influential and extensive, and has produced a variety of assessment tools. In the context of sentencing, risk instruments are being built into sentencing guidelines (i.e., Virginia), and used in the preparation of pre-sentence reports.

4 One of the most common risk instruments for the prediction of general criminality is the Level of Service Inventory – revised (LSI–R). LSI–R is currently being utilized within Canada and the United States to guide sentencing decisions, placement in correctional programs, institutional assignments, and release from institutional custody. This instrument’s widespread use and modification makes it a useful exemplar for this discussion. The LSI was originally developed in Ontario, Canada, in the late 1970s and quickly developed international notoriety. Currently, the tool is used in jurisdictions throughout Canada, the United States, the United Kingdom, and Australia, among others. The tool, originally written in English, is available in Spanish, Croatian, and French (French European and French Canadian), and it is in the process of being translated into Dutch and Icelandic. Multi Health Systems—the company that markets the LSI–R—indicates that more than 600 agencies in the United States currently use this risk-need tool (Lowenkamp et al., 2004) It is one of the most extensively researched offender classification instruments; consequently, it is less vulnerable to some of the methodological critiques of less rigorously tested local risk instruments. Nonetheless the LSI–R, similar to all risk tools, has limits, and its current...
use in sentencing generates a complicated debate, as well as a need for further empirical research.

5 For a discussion of the criminogenic effects of imprisonment see Craig Haney (2006).
6 Andrews and Bonta (2006) provide a detailed description of these tools and this development.
7 Andrews and Bonta report that the research on providing any treatment to offenders as a function of risk shows that “providing intensive services to low risk offenders may actually increase criminal behaviour and also that these services can lead to a significant decrease in recidivism when delivered to higher risk offenders. For example, Bonta, Wallace-Capretta and Rooney (2000) in an evaluation of a Canadian program found that low risk offenders who received minimal levels of treatment had a recidivism rate of 15% and low risk offenders who received intensive levels of services had more than double the recidivism rate (32%). In the same study, the high risk offenders who did not receive any intensive treatment services had a recidivism rate of 51% but the high risk offenders who did receive intensive services had almost half the recidivism rate (32%). The risk principle calls for intensive treatment services to be reserved for the higher risk offender” and that “Treatment interventions that do not adhere to any of the three principles (that is, they target the non-criminogenic needs of low risk offenders using non-cognitive-behavioural techniques) are actually criminogenic! This situation is particularly exacerbated when the treatment is given in residential/custodial settings (we presume because the offender cannot escape from the well-intentioned but poorly designed treatment)” (Andrews and Bonta 2007:15-16).

8 See Ward and Maruna (2007) for a fuller critique of RNR. The critical literatures on punishment include a lively debate about the “what works” analyses of the logic of RNR and cognitive behavioralism (Ward, 2003; Ward and Stewart, 2003; Ward and Brown, 2004).


11 For a more comprehensive discussion of this debate and present developments see Van Voorhis et al. 2010.

12 For an elaboration of this argument see Hannah-Moffat (2009). Such analyses highlight that men and women have different needs and represent different kinds of risk, and note the importance of considering “gender-specific needs” and the specific needs of minority women, (for example, see Belknap and Holsinger, 2006Blanchette and Brown, 2006; Brennan, 1998; Hollin and Palmer, 20062007; Holsinger and Holsinger, 2005Holtfreter and Morash, 2003; Nesbitt and Argento 1984; Reisig et al., 2006Van Voorhis and Presser, 2001Wright, Salisbury, and Van Voorhis, 2007Some studies also demonstrate that tools such as the LSI–R function differently for men and women. For example, a growing international scholarly body of research has been questioning the limits of risk assessment and the potential for bias and discrimination. A number of researchers have questioned the convergent,
concurrent, and discriminate validity and reliability of established risk instruments when used on women and racially diverse and economically disadvantaged offenders.

13 As Van Voorhis et al. 2010: 262) document a number of studies have found dynamic risk assessments, such as LSI-R “valid for women (see Andrews, Dowden, & Rettinger, 2001; Blachette & Brown, 2006; Coulson, Ilacqua, Nutbrown, Giulekas, & Cudjoe, 1996; Dowden & Andrews, 1999; Holsinger, Lowenkamp, & Latessa, 2003; Simourd & Andrews, 1994; Smith, Cullen, & Latessa, 2009). Others studies have produced conflicting results (see Blachette, 2005; Law, Sullivan, & Goggin, in press; Olson, Alderden, & Lurigio, 2003; Reisig, Holtfreter, & Morash, 2006; Salisbury, Van Voorhis, & Spiropoulis, 2009).”

14 That is, education/employment; family/marital relations; leisure/recreational involvement; criminal acquaintances; attitudes toward crime; and substance abuse which are six of the eight risk/need factors in the LSI–OR)

15 A long tradition of sociological research that demonstrates social stratification and poverty are related to both criminality and criminalization.


22 Although evidence on concordance between PSR and sentencing is contested, highly nuanced, and difficult to empirically unravel (Haines and Morgan, 2007; Tata et al., 2008), existing research demonstrates that the PSR plays a central interpretive role in sentencing and among criminal justice professionals (see also Tata, 2010; Wandall, 2010).

23 In cases where individuals are legally designated “dangerous offenders” there is considerable more debate among experts on how opinions about “danger and risk” are formulated. However they are also considerable debates in this instance about the suitability of risk instruments. For additional discussion see Monahan 2010.


25 Some authors contend that this conflict is not as problematic as it seems, see for example Marcus (2009).

Some scholars maintain that the transition to risk-based penalty has led to “deskilling,” “scientification,” and the “erosion of professional discretion” (Robinson, 2003:33; see also Baker, 2005; Fitzgibbon, 2007, 2008; Schneider et al., 1996), or even to the elimination of discretion among criminal justice practitioners (Hannah-Moffat and Maurutto, forthcoming).

See Mental Health Centre Penetanguishene for discussion of research on the use of Violence Risk Appraisal Guide (VRAG) and Sexual Offender Risk Appraisal Guide (SORAG).