

# RÉCIDIVE : TROIS ANS APRÈS LA CONFÉRENCE, POURQUOI IL N'Y A TOUJOURS PAS DE CONSENSUS

par **Sacha Raoult**

Maître de conférences HDR à Aix-Marseille Université (AMU, LDPSC, EA 4690)

Des résistances s'élèvent toujours contre la « modernisation » des méthodes d'évaluation du risque de récidive en France. Ces résistances ne sont pas typiquement françaises, elles reflètent également une tension interne au débat scientifique anglophone. Derrières les aspects techniques, c'est une question fondamentale qui pose problème à de nombreux chercheurs.

Cet article fait un point sur la littérature scientifique récente relative à l'évaluation moderne du risque de récidive<sup>1</sup>. D'abord, il identifie et mesure l'absence de consensus dans la littérature anglophone. Puis, il passe en revue deux catégories de problèmes sur lesquels les chercheurs défavorables aux méthodes modernes insistent ces deux dernières années : l'incertitude sur la précision des outils et les conséquences douteuses de leur mise en œuvre. En conclusion, il émet l'hypothèse que le désaccord fondamental sur cette question résulte du fait que deux communautés de chercheurs opèrent avec une vision différente du rôle que doit assumer la recherche par rapport à l'administration pénitentiaire et au système de justice pénale en général.

(1) On désigne les méthodes *evidence-based* par l'expression « méthodes modernes » dans le reste de l'article, que celles-ci soient purement actuarielles et ou sous la forme plus souple du jugement professionnel structuré.

(2) Sur les résistances à la modernisation de l'évaluation du risque de récidive en France V. la conférence de *disensus sur la récidive : exercices critiques sur une production de vérité* organisée par P. Poncela et B. Harcourt et disponible en audio sur Carceral Notebooks, Vol. 8, 2012 <http://www.thecarceral.org/journal-vol8.html> ; V. égal. É. Dubourg et V. Gautron, *La rationalisation des méthodes d'évaluation des risques de récidive*, Champ pénal/Penal field [En ligne], Vol. XI, 2014.

(3) La plupart des arguments que l'on trouve débattus aujourd'hui en France le sont aussi à l'étranger. Une exception notable est l'argument de Faugeron et Le Boulaire selon lequel une augmentation du taux de récidive traduit un succès de l'institution et non un échec. Pour une version actuelle de cet argument P. Combessie, À quels espaces assigner les condamnés ? Quels espaces observer ?, *Espaces et sociétés* 162.3 (2015) : 167-174.

(4) Y. Min., S. CP Wong et J. Coid, The efficacy of violence prediction : a meta-analytic comparison of nine risk assessment tools, *Psychological bulletin*, 136, n° 5 (2010) : p. 754.

(5) Il s'agit du projet de recherche *Social Position / Knowledge Production* conduit à University of Chicago depuis juill. 2014.

(6) La banque de données fut construite avec l'assistance de B. Peppers (*University of Chicago*). Elle fut faite à partir d'un ensemble de requêtes *Google Scholar* du 9 juill. au 11 août 2014. Pendant la construction de cette banque de données, un point avait été fait relativement à trois remarques préliminaires qui me semblaient expliquer une partie des problèmes dans cette littérature, S. Raoult, Évaluation du risque de récidive : l'expert, le politique et la production du chiffre, *RSC*, 2014. 655.

(7) Les 54 universitaires restant n'avaient pas d'opinion claire.

(8) S. Raoult, Des méthodes et des hommes. La production sociale du savoir sur l'efficacité de la peine de mort, *Déviance et Société*, 2015/1, vol. 39, p. 99-122.

(9) Dans le test statistique de Student, *p*. donne la probabilité que le résultat obtenu soit le fait d'un hasard d'échantillonnage.

## ■ Identification et mesure de l'absence de consensus dans la littérature anglophone

On peut avoir l'impression que l'absence de consensus sur la récidive en France est une spécialité locale<sup>2</sup>. Elle correspond en réalité aux tensions qu'on retrouve dans la littérature anglophone, où se dessinent les mêmes types d'arguments et de controverses<sup>3</sup>.

L'absence de consensus international en matière d'évaluation du risque de récidive, de violence ou de délinquance sexuelle est un élément est bien connu des psychiatres. Ainsi, par exemple, dans une revue de l'*American Psychiatry Association*, une équipe a récemment présenté la question de l'évaluation du risque de violence comme « l'un des problèmes les plus complexes et controversés des sciences du comportement »<sup>4</sup>. Dans le cadre d'un projet plus large<sup>5</sup> nous avons donc examiné l'ensemble des publications en langue anglaise sur les méthodes actuarielles et les jugements professionnels structurés de janvier 2012 à juillet 2014 (140 publications)<sup>6</sup> afin de mesurer et de comprendre l'étendue de la controverse. Sur les 174 universitaires de notre banque de données, 59 avaient une opinion favorable aux méthodes modernes, 28 une opinion sceptique et 33 une opinion défavorable<sup>7</sup>. Si l'on compare cette répartition avec celle des opinions chez les non-universitaires, on est frappé par la différence. La question des méthodes modernes est bien moins polarisante chez ce public, en large adhésion avec les modèles *evidence-based* (V. tableau 1). En miroir, plus on monte dans le prestige scientifique, plus la question pose problème : 33 % des universitaires en poste dans le top 50 du *Times of Higher Education* (Oxford, Yale, Columbia etc.) ont une opinion défavorable aux outils les plus récents, 17 % une opinion sceptique. Ce lien entre la reconnaissance académique et une opinion scientifique rappelle d'autres domaines (comme les recherches sur la peine de mort, où les universités les mieux classées hébergent les auteurs les plus critiques<sup>8</sup>).

**Tableau 1. Opinions des chercheurs sur les méthodes modernes d'évaluation du risque de récidive**

| Catégorie d'auteur                | favorable | sceptique | défavorable |
|-----------------------------------|-----------|-----------|-------------|
| non universitaire (82)            | 65 %      | 20 %      | 15 %        |
| tout universitaire (120)          | 49 %      | 26 %      | 25 %        |
| universitaire dans le top 50 (24) | 50 %      | 17 %      | 33 %        |

Source des données : Banque de données *Social Position / Knowledge Production*

La différence entre tout universitaire et non universitaire est très significative ( $p < 0.01$ ), la différence entre universitaire dans le top 50 et les autres n'est pas significative du fait de la petitesse de l'échantillon ( $p = 0.17$ )<sup>9</sup>.

Une des raisons pour lesquelles les spécialistes de la justice pénale qui ont une opinion défavorable sur ces questions tendent à être légèrement moins présents que ceux qui ont une opinion favorable est qu'ils ne multiplient pas les études empiriques, car ils se concentrent sur des problèmes fondamentaux. Les problèmes de source ou des conséquences possiblement ambiguës de « la justice actuarielle » sur la criminalité sont bien connus<sup>10</sup>. Au-delà deux problématiques ont monopolisé les recherches critiques ces dernières années : la précision des outils et leur mise en œuvre.

## ■ Les problèmes liés à la précision des outils

Les méthodes modernes se sont imposées car elles ont été pré-sentées dans des études empiriques comme plus précises que les méthodes cliniques<sup>11</sup>. Des travaux récents revisitent en partie cette histoire. Certains auteurs suggèrent que la précision des outils a en fait été exagérée, et d'autres qu'elle n'a en réalité pas été étudiée selon le bon critère.

### L'exagération de la précision

Les méthodes modernes présentent une précision reconnue comme généralement élevée, même si la marge d'erreur est parfois importante<sup>12</sup>. Néanmoins, une littérature récente a découvert un certain nombre de biais dans la recherche empirique, qui amènent à penser que cet acquis a pu être exagéré.

D'abord, une étude publiée dans *PlosOne* montre que les travaux conduits par les auteurs ou promoteurs d'un outil tendent à les trouver, en moyenne, deux fois plus utiles que ceux effectués par des chercheurs indépendants<sup>13</sup>. Cette découverte est gênante, dans la mesure où elle signifie que la plupart des méta-analyses produites sur le sujet ont tendance à fabriquer des « moyennes » tirées vers le haut.

À ce biais d'auteur se superpose un biais de qualité. On avait remarqué il y a quelques années, au sujet des modèles fondés sur les principes du risque, des besoins et de la réceptivité (modèles RBR), que les travaux de moindre qualité tendaient à donner des résultats plus favorables aux modèles que les travaux de meilleure qualité<sup>14</sup>. Cette remarque se vérifie dans notre banque de données. En assignant un score à chaque publication en fonction de ses conclusions (-1 pour défavorable aux outils, 0 pour sceptique, +1 pour favorable aux outils) on se rend compte que les publications de la littérature grise, c'est-à-dire les travaux publiés hors des revues à comité de lecture ou dans des revues inconnues, sont en moyenne bien plus favorables aux outils qu'ils étudient que la littérature publiée dans des revues à comité de lecture. Cette différence est indiquée au tableau 2, et s'accroît dans les revues les plus prestigieuses. Plusieurs chercheurs remarquent d'ailleurs que « la recherche primaire étant d'une qualité variable, il nous faut aujourd'hui plus de recherche indépendante de haute qualité »<sup>15</sup>.

**Tableau 2. Résultats de la littérature scientifique selon le prestige du support de publication**

| Type de support  | Nombre de publications | Score moyen |
|--|------------------------|-------------|
| Littérature grise<br>(aucun H index du support)            | 37                     | 0,40        |
| Littérature blanche<br>(H index du support > 0)            | 103                    | 0,22        |
| Revues les plus prestigieuses<br>(H index du support > 40) | 14                     | 0,08        |

Source des données : Banque de données Social Position / Knowledge Production

Chaque publication défavorable est comptée (-1), sceptique (0) et favorable (1). Le *H index* du support est l'indicateur de prestige des revues selon le classement *Google Scholar Metrics 2014*. Malgré la petitesse de l'échantillon, la différence entre les revues les plus prestigieuses et la littérature grise est plus significative ( $p=0.06$ ) qu'entre la littérature blanche et la littérature grise ( $p=0.18$ ).

### Le mauvais critère de précision : à l'assaut du ROC

Le deuxième problème est le choix du critère de précision. Les outils d'évaluation se jugent par deux critères, les faux positifs (nombre de « dangereux » non récidivistes) et les faux négatifs (nombre de « non-dangereux » récidivistes). Or, les méthodes cliniques sont généralement plus efficaces en termes de minimisation de faux négatifs, les méthodes actuarielles de faux positifs<sup>16</sup>. De plus, quelle que soit la méthode dont on dispose, on peut toujours imaginer une méthode plus « précise » en termes de faux positifs mais moins précise en termes de faux négatifs<sup>17</sup>. Pour considérer une méthode comme plus efficace « tout court », il fallait donc trouver un équivalent entre faux positif et faux négatif, entre un récidiviste identifié à tort comme non dangereux et un non-récidiviste identifié à tort comme dangereux<sup>18</sup>. Cet équivalent a été trouvé en important une méthode initialement utilisée pour sélectionner les meilleurs modèles de radar, la méthode de l'aire sous la courbe des caractéristiques opératoires du récepteur (courbe de ROC). Cet indice de ROC (aussi appelé AUC pour *area under the curve*) est celui qui permet principalement de classer les méthodes actuelles par niveau de précision et donc de juger si elles sont suffisamment précises. Selon les contextes, on juge généralement qu'une méthode est acceptable si l'indice de ROC est supérieur à 0,7 ou à 0,75. Le ROC fut utilisé sans trop susciter de questionnements dans les années 1990 et 2000, et il a permis aux méthodes modernes, qui avaient de très bons indices de ROC, de s'imposer ; mais cette approche commence aujourd'hui à être remise en cause.

Une première équipe britannique a suggéré d'évaluer les méthodes modernes non pas par l'indice de ROC mais par des diagrammes de fréquence naturelle<sup>19</sup>. Ce diagramme indique la fréquence de récurrence ob-

(10) V. par ex. B. Harcourt, *Against Prediction*, University of Chicago Press, 2007 ainsi que le débat qu'il a suscité lors de sa parution dans *Champ pénal*.

(11) Litwack, R. Thomas, Actuarial versus clinical assessments of dangerousness, *Psychology, Public Policy, and Law* 7.2 (2001) : 409.

(12) Hart, D. Stephen, C. Michie, and D. J. Cooke, Precision of actuarial risk assessment instruments, *The British Journal of Psychiatry* 190.49 (2007) : s60-s65.

(13) J. Singh, M. Grann, S. Fazel, Authorship Bias in Violence Risk Assessment ? A Systematic Review and Meta-Analysis. *PLoS ONE* 8(9).

(14) R. K. Hanson, G. Bourgon, L. Helmus, S. Hodgson, A Meta-Analysis Of The Effectiveness Of Treatment For Sexual Offenders : Risk, Need, And Responsivity, *Criminal Justice And Behavior*, 2009, n° 36, p. 865 – 891.

(15) R. Tully, C. Chou, K. Browne, A systematic review on the effectiveness of sex offender risk assessment tools in predicting sexual recidivism of adult male sex offenders. *Clinical Psychology Review* 33(2), 287-316

(16) Litwack, art. préc.

(17) M. Hamilton, *Adventures in Risk : Predicting Violent and Sexual Recidivism in Sentencing Law*, Arizona State Law Journal, sous presse.

(18) Pour des développements sur ce point v. Raoult, art. préc. (2014).

servée pour chaque catégorie de score de risque (bas, moyen, élevé, très élevé). Or, le taux d'erreurs qu'ils observent pour les sujets à risque « très élevé » est étonnant : dans l'outil qu'ils examinent (la Risk Matrix 2000, qui était validée en termes d'indice de ROC) seuls 5 sujets jugés très dangereux sur 70 ont récidivé. Dans la catégorie « risque élevé », il n'y en avait que 9 sur 256<sup>20</sup>. Ce taux d'erreurs est masqué par la méthode usuelle « pour qui la différence de taux récidive entre deux individus à bas risque vaut autant [de points] que la différence entre un individu à bas risque et un individu à haut risque »<sup>21</sup>. Pour ces auteurs, l'indice de ROC n'est rien d'autre qu'un « rituel statistique qui a pour fonction de faire paraître des résultats hautement informatifs » alors qu'ils ne le sont pas<sup>22</sup>. Une autre équipe, de l'université d'Oxford, a ajouté à l'indice de ROC six autres indices statistiques dans une méta-analyse sur 13 pays<sup>23</sup>. Cette technique leur a permis de remarquer que la valeur prédictive négative (proportion de non dangereux non récidivistes) était beaucoup plus élevée que la valeur prédictive positive (proportion de dangereux récidiviste). Cette remarque les amène à penser que si les outils actuels sont « modérément intéressants » pour informer les décisions de traitements et de gestion, leur intérêt dans le domaine pénal ou sécuritaire est réduit<sup>24</sup>. Partant du constat soulevé par ce type de travaux<sup>25</sup>, plusieurs chercheurs se demandent si, au final « ces méthodes apportent une information véritablement

utile ? »<sup>26</sup> « Les données essentielles qu'il faudrait pour prendre ce type de décisions ne sont la plupart du temps pas rapportées au sujet des outils utilisés »<sup>27</sup>, concluent deux autres auteurs.

## ■ Les problèmes dans la mise en œuvre des outils

Les méthodes modernes ont eu pour objectif de confisquer à l'agent une part de son jugement discrétionnaire pour permettre un traitement non seulement plus efficace (c'est ce qui vient d'être évoqué) mais aussi plus égalitaire des sujets<sup>28</sup>. Néanmoins, des travaux récents insistent d'une part sur les disparités d'usage dans la pratique de ces outils, qui réintroduisent une subjectivité dans des items pourtant apparemment objectifs) mais aussi sur les discriminations structurelles liées à la réplification, dans les prises de décisions pénales, des données issues de statistiques administratives.

### Les disparités d'usage dans la pratique

Dès les années 2000, deux travaux non publiés avaient suggéré qu'en pratique, il y avait dans l'usage des outils modernes des disparités importantes<sup>29</sup>. La recherche sur ce point ne s'est cependant intensifiée que très récemment<sup>30</sup>. Par exemple, une étude de 2012 suggère que pour des outils pourtant simplistes comme Static-99, les scores assignés par des pairs d'évaluateurs varient dans 45 à 60 % des cas<sup>31</sup>, et ce alors même que les évaluateurs avaient parfois connaissance du score donné par leur pair. Cette différence est d'autant plus étrange que les questions posées par Static-99 sont objectives ou objectivables : nombre de condamnations antérieures, sexe de la victime, âge, etc. Cela amène les auteurs de cette étude à suggérer que la variation entre deux scorages doit être encore plus importante quand les évaluateurs n'ont pas connaissance d'un score déjà donné par un pair et que l'outil en question emploie des critères plus subjectifs (par exemple des questions de personnalité)<sup>32</sup>. Évidemment, cette variation qui commence à peine à être mesurée n'a jamais été comparée à la variation des jugements cliniques. Il faut aussi réfléchir à ce qu'elle implique en termes d'évaluation des outils : une étude qui valide un outil au New Jersey valide à la fois les items de l'outil sur la population du New Jersey et une part du jugement subjectif des agents de probation du New Jersey sur la période où l'étude a été effectuée.

### Les discriminations structurelles liées à la réplification des statistiques administratives dans les prises de décision pénales

La dernière catégorie de problèmes soulevée par les travaux récents est l'aspect préjudiciable des méthodes modernes d'évaluation du risque de récidive en termes de discrimination. Alors que ces méthodes avaient été mises en place pour assurer un traitement égal aux détenus, certains suggèrent que dans la pratique, le recours plus appuyé aux outils statistiques aux États-Unis a eu tendance à augmenter l'écart Blancs/Noirs en défaveur de ces derniers, en prison comme en hôpital psychiatrique : le risque devenant au final « l'approximation de la race »<sup>33</sup>. L'explication proposée résiderait dans les sources employées pour construire les statistiques informant les outils modernes<sup>34</sup>. Les taux différentiels d'arrestation et de retour en prison étant fort différents d'un groupe social à l'autre (problème que nous connaissons également en France<sup>35</sup>), les outils qui prétendent prédire la récidive à partir de taux de retour en prison différentiels reproduisent ce différentiel social d'arrestation une seconde fois. Il est certain que ce différentiel initial n'est pas identique pour chaque type d'infraction. Un des grands débats des années 1980 a conduit à suggérer que le différentiel d'arrestation et de retour

(19) D. Cooke & C. Michie, The generalizability of the Risk Matrix 2000 : On model shrinkage and the misinterpretation of the area under the curve. *Journal of Threat Assessment and Management* 1(1), 42-55.

(20) *Id.*, p. 48.

(21) *Id.* p. 49.

(22) *Ibid.*

(23) S. Fazel, J. Singh, H. Doll, M. Grann, Use of risk assessment instruments to predict violence and antisocial behaviour in 73 samples involving 24,827 people : systematic review and meta-analysis. *BMJ* 2012. 345.

(24) *Id.* p. 4.

(25) Et d'autres travaux du même type comme une thèse de doctorat sur les mineurs délinquants sexuels, aucun des sujets jugé à risque n'a récidivé, A. Houle, An Examination of the Predictive Validity of Court Assigned Risk Levels on Recidivism Rates among Juvenile Sex Offenders, th. Auburn University, ou encore P. Falzer, Valuing Structured Professional Judgment : Predictive Validity, Decision-making, and the Clinical-Actuarial Conflict. *Behavioral Sciences & the Law* 31, 40-54.

(26) M. Large, C. Ryan, S. Callaghan, M. Paton, S. Singh, Can violence risk assessment really assist in clinical decision-making ? *Aust N Z Journal of Psychiatry* 48(3), 286-288.

(27) D. Neller & I. Frederick, Classification Accuracy of Actuarial Risk Assessment Instruments. *Behavioral Sciences & the Law* 31, 141-153. Pour des remarques similaires G. Szmukler, Risk assessment for suicide and violence is of extremely limited value in general psychiatric practice, *Australian and New Zealand Journal of Psychiatry* 46(2), 173-174.

(28) En ce sens, V. par ex. Kaas, Rob, et al. *Modern actuarial risk theory : using R*. Vol. 128. Springer Science & Business Media, 2008.

(29) Ces travaux non publiés des années 2001 et 2003 sont évoqué par M. Boccacini, D. Murrie, C. Mercado, S. Quesada, S. Hawes, A. Rice, E. Jeglic, Implications of Static-99 Field Reliability Findings for Score Use and Reporting. *Criminal Justice and Behavior* 39(1), p. 42.

(30) J. Lanterman, D. Boyle, L. Ragusa-Salerno, Sex Offender Risk Assessment, Sources of Variation, and the Implications of Misuse. *Criminal Justice and Behavior* 41(7), 822-843.

(31) Boccacini, et al, art. préc.

(32) Art. préc., p. 55.

(33) Harcourt, Risk as a Proxy for Race, *John M. Olin Law and Economics Working Paper* n° 535, p. 1-14.

(34) K. Hannah-Moffat, Actuarial Sentencing : An « Unsettled » Proposition. *Justice Quarterly* 30(2), 270-296.

(35) L'étude la plus récente en ce sens porte sur le contrôle d'identité à Paris, N. Jounin et autres, « Le faciès du contrôle », *Déviante et Société* n° 39(1), 2015, p. 3-29.

en prison d'un groupe social à l'autre à infraction égale était beaucoup moins important pour la délinquance sexuelle ou sérieuse que pour la petite délinquance<sup>36</sup>, mais certains chercheurs pensent que les délinquances sexuelles subissent également des effets de tri important, qui résultent du fait que chaque type de délinquance n'étant pas pris au sérieux ou rapporté avec la même intensité, ni traité sur les mêmes circuits pénaux, cela déforme les données administratives en faisant passer pour plus dangereux ceux qui ont une délinquance traitée plus intensément sur le « circuit prison »<sup>37</sup>. Il faut également noter que les tables de prédiction de la grande délinquance s'intéressent la plupart du temps au passé petit délinquant du sujet comme prédicteur de risque, et font donc rentrer ces critères qui réintroduisent des informations criminologiques socialement stratifiées, toutes choses égales par ailleurs, dans le pronostic de dangerosité<sup>38</sup>.

## ■ En conclusion

À partir de l'examen de ces 140 publications, il nous semble que l'absence de consensus se caractérise de la façon suivante.

Il y a au final peu de désaccords dans la littérature anglophone sur la liste des points qui posent problèmes à proprement parler. Les auteurs les plus favorables aux outils modernes reconnaissent aisément l'existence et l'importance de ces problèmes, au moins dans le contexte des revues et ouvrages scientifiques<sup>39</sup>. Mais là où le bât blesse c'est sur le poids qu'il faut donner à ces problèmes.

Les chercheurs favorables aux outils sont dans une posture où ils répondent à une demande de l'administration pénitentiaire. L'administration se retrouve quant à elle avec le devoir d'effectuer une évaluation du risque, du fait cette fois d'une demande politique<sup>40</sup>. Les chercheurs sont alors satisfaits d'offrir des outils à même de classer les détenus avec une précision (en termes d'indice de ROC) apparemment plus grande que les méthodes cliniques. Ils ont conscience des problèmes soulevés : ils répondent à la disparité dans les utilisations par une exigence d'intensification de la formation et aux risques de discrimination par la nécessité d'une veille sur l'utilisation des outils<sup>41</sup>. Je résumerais la pensée de cette recherche apologétique de la façon suivante : mieux vaut ça que rien.

À l'inverse, la littérature critique des méthodes modernes ne se fait que rarement l'avocat des méthodes cliniques ou de toute autre méthode alternative. La façon dont j'analyse cette position est : mieux vaut rien que ça. Le point de départ de cette position est la faible valeur prédictive de l'outil (quand, par exemple dans la publication vue plus haut, seul 5 détenus jugés très dangereux sur 70 récidivent) au vu des risques de mésusage et des risques de surévaluation de leur efficacité. La littérature critique admet la possibilité qu'il vaille mieux abandonner entièrement la tâche d'évaluation du risque de récidive et basculer les priorités du système pénal soit vers une logique éducative, soit vers une logique rétributive, alors que la recherche apologétique n'admet pas cette possibilité. Cette approche est à replacer dans le contexte plus général d'une critique de la place de la personnalité et des questions virtuelles dans le traitement pénal au détriment de la gravité des actes jugés et des questions factuelles<sup>42</sup>.

Ce qui fait qu'il ne peut pas y avoir de consensus sur la récidive, c'est qu'à mon sens le débat n'est pas placé au bon niveau. La question qui doit être débattue est plus profonde que « comment mieux évaluer ? », car on ne peut pas se mettre d'accord sur ce que signifie ce « mieux ». Si l'on veut que l'entièreté de la recherche participe au débat, le débat doit porter sur le fonctionnement et les dysfonctionnements de la justice pénale en général, contexte à l'intérieur duquel l'évaluation du risque de récidive (et la question « récidive » tout court) n'est qu'un aspect. Et il n'est pas certain que la réforme prioritaire du système pénal porte sur ce dernier.

(36) A. Abbot revient sur ce débat ancien dans *Chaos of disciplines*. University of Chicago Press, 2001, p. 60 à 89.

(37) Larcombe, W (2012). Sex Offender Risk Assessment : The Need to Place Recidivism Research in the Context of Attrition in the Criminal Justice System. *Violence Against Women* 18(4), 482-501, en particulier les p. 491 à 494 sur les différentiels de traitement policiers et judiciaires en matière de viol qui se répercutent sur les taux de « récidive ».

(38) Hannah-Moffat, K. art. précit.

(39) Un bon exemple de cet accord est la conclusion de l'ouvrage de R. Berk *Criminal Justice Forecasts in Risk : A Machine Learning Approach* (Springer, 2012).

(40) L'avènement de cette demande est très bien analysé dans le texte classique de Simon et Feeley, « La nouvelle pénologie », expliqué en français par Slingeneyer, Thibaut. « La nouvelle pénologie, une grille d'analyse des transformations des discours, des techniques et des objectifs dans la pénalité. » *Champ pénal/Penal field* 4 (2007).

(41) R. Berk, *loc. cit.*

(42) La National Academy of Science, organe le plus scientifique le plus prestigieux outre-atlantique, vient d'ailleurs de rendre un rapport en ce sens v. Travis, Western et Reburn (dir.) *Growth of Incarceration in the United States : Exploring Causes and Consequences*, National Academies Press, 2014.