SJC-13197

IN THE COMMONWEALTH OF MASSACHUSETTS
SUPREME JUDICIAL COURT

JOSE RODRIGUEZ,
Plaintiff- Appellant,

v.

MASSACHUSETTS PAROLE BOARD,
Defendant-Appellee.

ON APPEAL FROM AN ORDER OF THE MIDDLESEX SUPERIOR COURT

BRIEF OF *AMICUS CURIAE* ELECTRONIC PRIVACY INFORMATION CENTER (EPIC) IN SUPPORT OF APPELLANT

Benjamin Winters (*pro hac vice pending*)
Counsel of Record
Caitriona Fitzgerald (BBO #673324)
ELECTRONIC PRIVACY INFORMATION CENTER
1519 New Hampshire Ave. NW
Washington, DC 20036
(202) 483-1140
winters@epic.org

Counsel for Amicus Curiae *EPIC*

TABLE OF CONTENTS

TABLE OF AUTHORITIES

STATUTES

OTHER AUTHORITIES

CORPORATE DISCLOSURE STATEMENT

Pursuant to Supreme Judicial Court Rule 1:21, *amicus curiae* Electronic Privacy Information Center ("EPIC") states that it is a District of Columbia corporation with no parent corporation or publicly held company with a 10 percent or greater ownership interest. EPIC is a non-profit, non-partisan corporation, organized under section 501(c)(3) of the Internal Revenue Code.

PREPARATION OF *AMICUS* BRIEF DECLARATION

Pursuant to Appellate Rule 17(c)(5), *amicus* declares that:

(a) No party or party's counsel authored this brief in whole or in part;

(b) No party or party's counsel contributed money to fund preparing or submitting the brief;

(c) No person or entity other than the *amicus curiae* contributed money that was intended to fund preparing or submitting a brief; and

(d) Counsel has not represented any party in this case or in proceedings involving similar issues, or any party in a case or legal transaction at issue in the present appeal.

INTEREST OF THE *AMICUS CURIAE*

The Electronic Privacy Information Center ("EPIC") is a public interest research center in Washington, D.C., established in 1994 to focus public attention on emerging civil rights and liberties issues.[1] EPIC regularly participates as *amicus curiae* in federal and state courts in cases concerning privacy and civil liberties. EPIC advocates for transparency, oversight, and regulation of predictive analytical tools and AI systems.

EPIC has long urged jurisdictions to ensure that the use of algorithms is transparent and fair. In 2017, EPIC filed a Freedom of Information Act lawsuit against the Department of Justice to compel production of records related to predictive policing and risk assessment tools. *EPIC v. DOJ* (D.C. Cir.) (18-5307).[2] EPIC also successfully compelled production of documents related to tools used by Customs and Border Protection and the Department of Homeland Security through FOIA lawsuits. In 2020, EPIC published *Liberty at Risk: Pre-trial Risk Assessments in the United States*, which analyzed productions from several public records requests around the country and made policy recommendations.

In 2018, EPIC, as part of a coalition called The Public Voice, helped establish and develop the Universal Guidelines for Artificial Intelligence,[3] which have been

---

[1] EPIC Law Fellow Thomas McBrien contributed to this brief.
[2] https://epic.org/documents/epic-v-doj-criminal-justice-algorithms/
[3] https://thepublicvoice.org/ai-universal-guidelines/

endorsed by over 70 organizations and 300 individuals.[4] EPIC has urged government agencies using AI tools to adopt these guidelines, which include mandated transparency, accuracy and reliability requirements, testing requirements, and an obligation for all systems to be fair."[5]

---

[4] https://thepublicvoice.org/AI-universal-guidelines/endorsement/

[5] EPIC Comments on Regulations Under the California Privacy Rights Act of 2020 https://epic.org/documents/comments-of-epic-and-three-organizations-on-regulations-under-the-california-privacy-rights-act-of-2020/; EPIC Comments on Implementation Plan for a National Artificial Intelligence Research Resource, https://archive.epic.org/apa/comments/EPIC-Comment-NAIRR-Oct2021.pdf; EPIC Comments on Artificial Intelligence Risk Management Framework, https://archive.epic.org/apa/comments/EPIC-NIST-AIRiskManagementFramework-Aug2021-Comments.pdf; EPIC Comments on Financial Institutions' Use of Artificial Intelligence, Including Machine Learning, https://archive.epic.org/apa/comments/EPIC-Financial-Agencies-AI-July2021.pdf; EPIC Comments on the Equal Opportunity Act and Regulation B, https://archive.epic.org/apa/comments/EPIC-CFPB-Oct2020-AI-ML.pdf; EPIC Comments on the National Security Commission on Artificial Intelligence Reporting, https://archive.epic.org/apa/comments/EPIC-comments-to-NSCAI-093020.pdf; EPIC Comments on a Draft Memorandum to the Heads of Executive Departments and Agencies, https://archive.epic.org/apa/comments/EPIC-OMB-AI-MAR2020.pdf; EPIC Comments on Housing and Urban Development's Implementation of the Fair Housing Act's Disparate Impact Standard, https://archive.epic.org/apa/comments/EPIC-HUD-Oct2019.pdf

SUMMARY OF ARGUMENT

The Massachusetts Parole Board's use of LS/CMI on juvenile lifers is dangerous because predictive analytical tools provide inaccurate and biased results when used on groups for which they were not designed, especially when details about these tools are kept secret. LS/CMI is one version of a predictive analytical tool that claims to predict recidivism in a general population of parole applicants, which Massachusetts has been licensing from MultiHealth Systems, Inc. since 2013.[6] In LS/CMI, the assessor can give the parole applicant "points" for each of 54 factors that the developers believe are tied to recidivism. A higher point total results in a higher risk score. Available information shows that the tool is likely to yield inaccurate results for parole applicants generally and juvenile lifers in particular. But Massachusetts' lack of transparency around its use of LS/CMI prevents the public from fully understanding the tool's accuracy and potential for bias. The lack of transparency also prevents juvenile lifers like Mr. Rodriguez from understanding how the tool decided their recidivism risk and whether those decisions were accurate.

Like every predictive analytical tool, LS/CMI is built by humans who make potentially faulty assumptions and decisions while designing the tool. These design

---

[6] Berkman Klein Center, *Risk Assessment Tool Database* (2022), https://criminaljustice.tooltrack.org/tool/16747.

decisions shape the scope of the tool's usefulness. An understanding of those design decisions is necessary to assess the context in which the tool may provide accurate predictions. Outside of that context, there is no reliable indicator of the tool's accuracy without additional testing.

LS/CMI's developers set out to develop a broad-based tool that could be used to determine the parole risk of individuals within the general population of parole applicants, but available data shows that the tool falls far short of its goal. The developers studied a general sample of 160,000 Canadian and American incarcerated people and decided which factors to rely on when calculating recidivism risk based on theories of criminality and interviews with probation and parole officers. Across this broadly defined group of parole applicants, LS/CMI has not scored highly on measures of accuracy. For discrete subgroups that have different characteristics that impact recidivism risk, accuracy is likely even lower, particularly for juvenile lifers.

Applying LS/CMI to juvenile lifers such as Mr. Rodriguez is misguided because juvenile lifers do not have the same characteristics as the general group of parole applicants, but instead have unique experiences and circumstances that the LS family of tools do not consider. Juvenile lifers are usually significantly older than the average incarcerated person when eligible for parole, and advanced age has been shown to reduce recidivism risk. Also, because of their age and life circumstances, many of the factors that predict recidivism in younger prisoners are irrelevant for juvenile lifers. For example, a related system called LSI-R considers marital status to be a strong factor predicting recidivism, but studies show this not to be the case for juvenile offenders. LS/CMI does not take these important differences into account when generating risk scores for juvenile lifers and, as a result, likely assigns juvenile lifers artificially inflated risk scores. A system with such questionable accuracy should not be used to inform decisions about juvenile lifers' liberty.

Compounding matters, parole applicants in Massachusetts, such as Mr. Rodriguez, are only given a highly redacted version of their LS/CMI scoresheet and score. Both parole applicants like Mr. Rodriguez and the public are unable to access even a full blank Massachusetts LS/CMI scoresheet or critical information that explains how the scoresheet is used, such as a scoring guide. The Parole Board should not be able to outsource parole decisions to a third-party contractor and use the outsourcing as an excuse not to reveal information about how the tool works. Juvenile parole applicants have a due process right to a meaningful opportunity for parole, and the inability to confront or understand their scores fundamentally undermines this right.

ARGUMENT

I. LS/CMI's predictions are unlikely to be accurate for juvenile lifers.

Predictive analytical tools like LS/CMI attempt to explain the present and predict the future by studying patterns in past data. The keyword here is "attempt." These tools are not crystal balls. Behind every tool is a set of decisions and assumptions made by fallible humans. When predictive analytical tools attempt to model complex subjects where there is little scientific consensus, such as human behavior, they are especially prone to error. And when a tool's predictions are based on data that is skewed by a history of discrimination, the tool may simply encode and perpetuate discrimination.[7]

When a predictive analytical tool is used to decide people's liberty, it is crucial to understand how the tool works and how it may err. But the Massachusetts Parole Board ("MPB") and MultiHealth Systems, Inc. ("MHS"), the company that developed the tool, have continuously blocked hopeful parolees and the public from accessing information about the MPB's use of LS/CMI. Information from other states, however, indicates that LS/CMI does not accurately predict the risk of granting parole to juvenile lifers.

---

[7] Vincent M. Southerland, *The Intersection of Race and Algorithmic Tools in the Criminal Legal System*, 80 Maryland L. Rev. 1, 8 (forthcoming 2022), *available at* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3797102.

a. A tool's predictive validity depends on the decisions and assumptions of its human developers.

One of the most important aspects of a predictive analytical tool is its ability to predict outcomes accurately, also known as the tool's predictive validity. Any tool's usefulness depends on the context in which it is used: Even a precisely crafted chef's knife will struggle with a bowl of soup. A tool's human developers determine its useful context through a set of design decisions. Specifically, developers decide what population to make predictions about, what data to base the tool on, and which factors are important in explaining past trends and predicting new ones. Taken together, these decisions establish the context in which the tool's predictive validity will be highest (whether or not it is objectively "high" in that context is another question). These decisions also represent points during which developers can accidentally introduce bias. When an agency like the MPB adopts the LS/CMI, it is adopting the decisions the contractor made when they designed the tool, including the factors they think affect risk and the weight that each factor should receive.

First, to understand predictive tools and predictive validity, a few concepts require explanation: "population," "sample," and "factor." Predictive analytical tools attempt to make predictions about individuals within a specific group of people. The group is called the "population,"[8] for example, "all Americans over 35" or "all drivers." A "sample" is a representative subset of individuals within the population whose data will be measured and used to make the tool.[9] The sample is smaller than the population—often a lot smaller. For example, LS/CMI's developers used a sample of about 160,000 North American youth and adult offenders — about 100,000 from Canada and 60,000 from the United States — to represent the approximate 2.1 million offenders on which the tool might be used in a given year.[10] A developer must also decide which characteristics, or "factors," to measure. These are generally the factors that the developers think impact the outcome they are studying. Examples of factors in the LS/CMI context include relationship status and education level.[11]

After choosing a population, sample, and factors, the developer then applies statistical analysis to the sample data in an attempt to spot useful patterns, e.g., the impact a parole applicant's level of education has on their likelihood to recidivate. These statistical techniques can be complicated or simple. For example, in more

---

[8] William J. Lammers & Pietro Badia, *Fundamentals of Behavioral Research* 7–2 (2004), *available at* https://uca.edu/psychology/files/2013/08/Ch7-Sampling-Techniques.pdf.
[9] *Id.*
[10] Berkman Klein Center, *supra* note 6; World Prison Brief, *Canada*, https://www.prisonstudies.org/country/canada; World Prison Brief, *United States of America*, https://www.prisonstudies.org/country/united-states-america.
[11] Idaho LSR-R Scoring Guide, *available at* https://archive.epic.org/EPIC-19-11-21-ID-FOIA-20191206-ID-lsi-scoring-guide.pdf.

simple “checklist”-type tools, researchers can decide what weight to assign different factors based on the researcher’s theories of cause and effect.[12] Alternatively, developers can create tools with an advanced statistical technique called “machine learning” in which the tool itself “learns” over time which factors should have the greatest weight and automatically assigns that weight when making its calculations.[13] In the context of risk assessment, a developer could assign a certain weight to the education category based on previous studies of education and recidivism, or a machine-learning tool could “notice” a connection between education and recidivism and apply its own weight to the factor—though, either way, a developer decided to measure education level. To the extent we have clarity, the LS/CMI tool seems to follow the checklist model, with assessors assigning scores between 0 and 3 on different factors that are subsequently summed as part of a risk score calculation.[14] Every time a developer selects a population to study, a factor to measure, and a subgroup to sample, they make explicit and implicit decisions that impact a tool’s accuracy, usefulness, and level of bias.

---

[12] Megan Stevenson, *Assessing Risk Assessment in Action*, 103 Minnesota L. Rev. 303, 315, 329 (2018).
[13] *Id.*
[14] Idaho LSR-R Scoring Guide, *available at* https://archive.epic.org/EPIC-19-11-21-ID-FOIA-20191206-ID-lsi-scoring-guide.pdf.

How a developer defines a tool's population constrains the context in which the tool is appropriate.[15] But if the population is defined too broadly, the tool may not make accurate predictions about certain subgroups of the population because the factors that affect outcomes for the subgroups are different than for the overall population.

Consider a hypothetical tool developed to predict the likelihood of success for a surgery. The population is defined as all patients no matter their medical histories. Factors such as age, weight, and blood pressure may be important for predicting the outcome for most patients in the broadly defined population, but these variables will not be predictive for patients with hemophilia, a condition that inhibits blood clotting and makes almost any surgery dangerous without special precautions.[16] A tool that averaged the two subgroups could severely underrate the risk for hemophiliacs, and it could overrate the risk for non-hemophiliacs if the portion of hemophiliacs in the sample is significant enough. To maximize accuracy, people with hemophilia should be excluded from the population, and the tool should not be used to predict their outcomes. This hypothetical shows why, generally, a tool created for a broadly defined population should be tested to ensure

---

[15] Lammers & Badia, *supra* note 8, at 7-2; Kathleen E Grady & Barbara Strudler Wallston, Research in Health Care Settings 24, 66–67 (1988).
[16] Elias Bastounis et al., *General Surgery in Haemophiliac Patients*, 76 Postgraduate Med. J. 494 (2000).

that it makes accurate predictions for distinct subgroups within that population.[17] The tool should also be tested over time to ensure changing real-world circumstances do not render its predictive power obsolete.[18]

Selecting a sample that accurately represents the population is crucial to ensure predictive validity and to avoid bias. For the purposes of this brief, a tool is "biased" if it systematically assigns higher risk scores to defendants from a particular racial or ethnic group than their true risk warrants. There are two important ways in which bias can infect a tool: through biased sample data and through inadequate sampling.

A predictive analytical tool can only learn from the data it is fed. If that data reflects biased decision-making in the criminal legal system, then the tool will learn to reproduce that data—but now, with a veneer of empiricism and objectivity.[19] As an example, in 2016, the Human Rights Data Analysis Group ("HRDAG") reproduced a predictive policing tool that—similar to LS/CMI—purported to predict crime.[20] They fed the tool with crime data from Oakland, California and asked it to forecast potential drug crime. The algorithm advised the

---

[17] Melissa Hamilton, *The Biased Algorithm: Evidence of Disparate Impact on Hispanics*, 56 Am. Crim. L. Rev. 1553, 1560 (2019).
[18] D.G. Mayer & D.G. Butler, *Statistical Validation*, 68 Ecological Modelling 21, 21 (1993).
[19] Southerland, *supra* note 7, at 8.
[20] *Id.* at 17-18.

police to target low-income, minority neighborhoods despite the reality that drug use had been demonstrated to occur more evenly throughout the city. HRDAG explained that this disparity was because officer bias in enforcement plagued the records used to feed the tool.[21] When informed by a discriminatory sample, predictive analytical tools will provide discriminatory results.

Bias and poor predictive validity can also enter through the sampling process when a tool's developer fails to consider racial and ethnic groups specific circumstances such as "behavioral practices and expectations, health beliefs, social/environmental experiences, phenomenology, illness narratives, deviant conduct, and worldview."[22] In these circumstances, developers failed to focus on factors that are meaningful for the racial or ethnic group. This type of failure has caused serious harm when the tool is used on individuals from groups that are underrepresented in the sample. For example, experts have shown that a majority of facial recognition systems—which are trained using similar statistical techniques as predictive analytical tools—perform worse on Black Americans and women because the systems' creators used samples with a disproportionate number of white male faces and an insufficient number of Black faces.[23] As a result of this

---

[21] *Id.* at 18.
[22] Hamilton, supra note 17, at 1560-61.
[23] Larry Hardesty, *Study Finds Gender and Skin-Type Bias in Commercial Artificial-Intelligence Systems*, MIT News (Feb. 11, 2018),

problem in one facial recognition program, Robert Williams, a Black man in Detroit, was wrongfully identified as a thief and subsequently arrested in front of his family.[24] Every time a tool is used, it is crucial to establish that the sample on which it was trained is representative of the subject of the decision. This is especially true for LS/CMI, as researchers have shown that LS family of tools performs worse on racial and ethnic minorities.[25]

Finally, choosing the correct factors to measure is important because a tool can only "learn" to recognize patterns in the factors that its developers measure.[26] Developers often choose factors based on preexisting assumptions about the domain in which they are trying to make predictions. For example, LS/CMI's

---

https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212; N. Hanacek, *NIST Study Evaluates Effects of Race, Age, Sex on Face Recognition Software*, NIST (Dec. 19, 2019), https://www.nist.gov/news-events/news/2019/12/nist-study-evaluates-effects-race-age-sex-face-recognition-software.

[24] Drew Harwell, *Wrongfully Arrested Man Sues Detroit Police Over False Facial Recognition Match*, Wash. Post (Apr. 13, 2021), https://www.washingtonpost.com/technology/2021/04/13/facial-recognition-false-arrest-lawsuit/.

[25] Joselyne L. Chenane et al., *Racial and Ethnic Differences in the Predictive Validity of the Level of Service Inventory–Revised Among Prison Inmates*, 42 Crim. Justice & Behavior 1 (2014), https://journals.sagepub.com/doi/abs/10.1177/0093854814548195.

[26] Georg Heinze, Christine Wallisch & Daniela Dunkler, *Variable Selection – A Review and Recommendations for the Practicing Statistician*, 60 Biometrical J. 431 (2018).

developers identified 43 factors across eight categories[27] to measure based on interviews with parole officers and "a general . . . theory of criminal behavior."[28] These are intensely political decisions given that the question of what causes crime is long-standing and hotly debated.[29] As just one example, LS/CMI's developers assume that measuring an applicant's *attitudes and behavior* alone is the proper paradigm to predict risk of recidivism instead of the *policing practices and history of discrimination* in their jurisdiction. These decisions will impact the tool's predictions greatly. For example, in the surgery prediction hypothetical, whether someone had hemophilia is a crucial factor to which the tool was blind because it was not measured by the tool's developers.

---

[27] Electronic Privacy Information Center, *Documents Obtained by EPIC Show Idaho's Use of Subjective Categories in Calculating Risk* (Dec. 11, 2019), https://epic.org/documents-obtained-by-epic-show-idahos-use-of-subjective-categories-in-calculating-risk/; MHS, Inc., LS/CMI QuikScore Form, *available at* https://faculty.uml.edu/jbyrne/44.203/documents/LSCMIblankpaperversion.pdf.

[28] Mark E. Olver, Keira C. Stockdale & J. Stephen Wormith, *Thirty Years of Research on the Level of Service Scales: A Meta-Analytic Examination of Predictive Accuracy and Sources of Variability*, 26 Psych. Assessment 156, 157 (2013), https://www.researchgate.net/publication/258920739_Thirty_Years_of_Research_on_the_Level_of_Service_Scales_A_Meta-Analytic_Examination_of_Predictive_Accuracy_and_Sources_of_Variability; https://www.ojp.gov/pdffiles1/Digitization/89859NCJRS.pdf.

[29] See, e.g., Per-Olof H. Wilkström & Kyle Treiber, *Social Disadvantage and Crime: A Criminological Puzzle*, 60 Am. Behavioral Sci. 1232, 1232 (2016); Lindsay M. Monte, & Dan A. Lewis, *Desperate or Deviant? Causes of Criminal Behavior Among TANF Recipients*, 3 Poverty & Pub. Pol'y 1 (2012).

b. There is no indication that LS/CMI's developers considered juvenile lifers' unique context when developing the tool.

Because there are no public studies testing LS/CMI's predictive validity for any Massachusetts parolees, much less Massachusetts juvenile lifers, EPIC's assessment of LS/CMI's predictive validity is based on statistical principles and analogies to data from LS tools in other states. Based on this information, LS/CMI's predictions are likely to be inaccurate for juvenile lifers. Like any predictive analytical tool, LS/CMI's predictions for juvenile lifers will only be accurate if (1) it is appropriate to include juvenile lifers in the general population of prisoners when predicting recidivism, (2) the sample on which LS/CMI was built appropriately represented juvenile lifers, and (3) all of the variables measured are appropriate for estimating juvenile lifers' risk of recidivism, and no crucial variables were excluded. To the extent that we have transparency into LS/CMI's construction and use, the tool likely fails on all three fronts. Given juvenile lifers' general lack of recidivism and the different factors that lead them to recidivate, they should be considered a different population. Any tool that was not trained specifically for juvenile lifers is unlikely to return accurate predictions about them.

Juvenile lifers should not be included in a general population of potential parolees because they experience imprisonment and release differently.[30] The main

[30] This will be true of most people with long sentences handed down while they are relatively young, not strictly juvenile lifers.

reason is age: The average incarcerated person serves a prison term of approximately two years and is released at 35 years old,[31] whereas the average juvenile lifer is incarcerated for decades and released at a more advanced age.[32] Juvenile lifers are thus less likely to recidivate simply because older people are less likely to commit crime.[33] A recent study of Philadelphia juvenile lifers showed that fewer than 1% recidivated[34] compared to 53.4% of the general population of Pennsylvanians.[35] A U.S. Sentencing Commission Report explains that "[o]lder offenders were substantially less likely than younger offenders to recidivate following release" and, if they did recidivate, they usually did so for much less serious crimes than younger parolees who recidivated.[36] But LS/CMI does not take age into account. Thus, the tool ignores a factor strongly indicating low recidivism risk—in fact, an extremely strong one that would justify putting juvenile lifers and

---

[31] Marieke Liem & Jennifer Garcin, *Post-Release Success Among Paroled Lifers*, 3 Laws 798, 799 (2014).
[32] *Id.*
[33] Dana Goldstein, *Too Old to Commit Crime?*, The Marshall Project (Mar. 30, 2015), https://www.themarshallproject.org/2015/03/20/too-old-to-commit-crime.
[34] Tarika Daftary-Kapur & Tina M. Zottoli, Resentencing of Juvenile Lifers: The Philadelphia Experience 2 (2018), *available at* https://www.msudecisionmakinglab.com/philadelphia-juvenile-lifers.
[35] World Population Review, Recidivism Rates by State 2021 (2021), https://worldpopulationreview.com/state-rankings/recidivism-rates-by-state.
[36] U.S. Sentencing Commission, The Effect of Aging on Recidivism Among Federal Offenders 3 (2017), *available at* https://www.ussc.gov/sites/default/files/pdf/research-and-publications/research-publications/2017/20171207_Recidivism-Age.pdf.

other older parole applicants in a distinct subgroup. As in the surgery prediction hypothetical, LS/CMI may be right to ignore age in most circumstances because most parole applicants are younger. But it is wrong to ignore this factor when predicting juvenile lifers' risk of recidivating.

Factors other than age also call for treating juvenile lifers as a distinct subgroup for which LS/CMI should not be used. Juvenile lifers' circumstances such as marital status and education are different than the general population of parole applicants because juvenile lifers are locked up indefinitely during important formative years, but LS/CMI is blind to this reality.[37] For example, the Idaho LSI-R tool places significant weight on marital status.[38] It explains that "[s]trong marital ties have been identified as one of the strongest protective factors," and it instructs assessors to rate parole applicants as higher risk if they are single and "lonely or frustrated by single status."[39] This is potentially dire news for juvenile lifers, many of whom find it difficult to form intimate partner relationships while locked up indefinitely during their 20s and 30s.[40] But despite lacking "one of

[37] *See generally* Liem & Garcin, *supra* note 31.
[38] LSI-R and LS/CMI are two evolutions of the same system. While they are not exactly the same, they are substantially similar. *See* Pamela M. Casey et al., Offender Risk & Needs Assessment Instruments: A Primer for Courts at A-32–33 (2014), https://www.ncsc.org/__data/assets/pdf_file/0018/26226/bja-rna-final-report_combined-files-8-22-14.pdf.
[39] Idaho LSR-R Scoring Guide 12, *available at* https://archive.epic.org/EPIC-19-11-21-ID-FOIA-20191206-ID-lsi-scoring-guide.pdf.
[40] Liem & Garcin, *supra* note 31, at 807.

the strongest protective factors," juvenile lifers rarely recidivate,[41] implying that this factor is largely irrelevant when predicting juvenile lifer recidivism. In a study of Massachusetts juvenile lifers and recidivism, many of the juvenile lifers who entered stable relationships during or shortly after their imprisonment recidivated, while many who remained single avoided recidivating.[42] In fact, for juvenile lifers, entering a relationship could prove counterproductive: The strong desire to "make up for lost time" meant that "none of the successes, but rather some of the failures, were attributed to the role intimate partners played in [juvenile lifers'] lives post-release."[43]

Similarly, LS/CMI inflates juvenile lifers' risk scores by ignoring that certain educational programs are systematically refused to juvenile lifers. LS tools instruct users to inflate applicants' risk scores if they have failed to graduate from 10th grade and 12th grade, but it considers prison GED or HSE programs as analogous to high school education.[44] In Massachusetts, however, there is a waitlist for these educational programs, and "enrollment is based on proximity to

---

[41] *See, e.g.*, Daftary-Kapur & Zottoli, *supra* note 34; Susan Samples, *Crime By 'Juvenile Lifers' After Prison 'Very Rare,' State Says*, Target 8 (Aug. 8, 2021), https://www.woodtv.com/news/target-8/crime-by-juvenile-lifers-after-prison-very-rare-state-says/.
[42] Liem & Garcin, *supra* note 31, at 816.
[43] *Id.* at 807.
[44] Idaho LSR-R Scoring Guide 6, *available at* https://archive.epic.org/EPIC-19-11-21-ID-FOIA-20191206-ID-lsi-scoring-guide.pdf.

release."[45] This creates a catch-22 for juvenile lifers: their risk scores will be higher until they can complete educational programs, but it is difficult to enroll in the programs until they are rated sufficiently low risk to warrant a release date. As funding for these education programs continues to plummet, many juvenile lifers may be stuck with higher risk scores through no fault of their own.[46] Not only is this unfair, but it improperly lumps together juvenile lifers who want to (but cannot) enroll in these programs with short-term prisoners who have no desire to do so. These are two different subgroups who may have very different chances of recidivism.

These crucial, nuanced distinctions between juvenile lifers and the general population of parolees are exactly what can make a predictive analytical tool useful for one population and not for the other. But—from what we can tell—the MPB does not recognize these distinctions when using LS/CMI on juvenile lifers. Instead, LS/CMI assessors seem to ratchet up a juvenile lifer's parole risk score with no contextual inquiry. Because the LS/CMI tool is not sensitive to juvenile

[45] Liem & Garcin, *supra* note 31, at 811–12; Email from Jennifer Gaffney, Deputy Commissioner of the Massachusetts Department of Correction, to Lisa Polk on Oct. 20, 2021 (on file with author).

[46] Benjamin Fordman & Michael Widmer, *Getting Tough on Spending: An Examination of Correctional Expenditure in Massachusetts* 11 (2017), *available at* https://massinc.org/wp-content/uploads/2017/05/Getting-Tough-on-Spending-1.pdf.

lifer status, juvenile lifers' risk scores are consistently, artificially, and unfairly inflated.

Even if LS/CMI had been designed to recognize juvenile lifers' unique experiences and circumstances, it is unlikely that juvenile lifers were adequately represented in the tool's sample. LS/CMI's developers sampled 160,000 people in the general population of imprisoned Canadian and American offenders.[47] While some juvenile offenders were included in the sample, we do not know what portion of the sample they represented or how many were juvenile lifers.[48] But given the proportion of juvenile lifers to non-juvenile-lifers in the United States, it is safe to presume that juvenile lifers were a vanishingly small subset of the sample. In 2016, only 0.002% of United States prisoners were juvenile lifers.[49] Assuming the proportion was the same for the LS/CMI sample, that represents about 190 juvenile lifers in the sample of 160,000 people. Any tool would struggle to recognize how a factor impacts 190 people in a sample set that large. LS/CMI has been shown to have performance issues with much larger subgroups. For example, LS/CMI has been shown to perform worse on U.S. residents compared to the Canadians it was

---

[47] Berkman Klein Center, *supra* note 6.
[48] *Id.*
[49] The Sentencing Project, *Youth Sentenced to Life Imprisonment* (2019), *available at* https://www.sentencingproject.org/publications/youth-sentenced-life-imprisonment/ (juvenile lifers); Statista, *Adult Correctional Population in the U.S. 2005–2019* (2021), *available at* https://www.statista.com/statistics/253024/adult-correctional-population-in-the-united-states/ (general prisoners).

predominantly trained on despite the fact that many Americans were included.[50] And researchers have shown that LS tools perform worse on racial and ethnic minorities.[51] Given these concerning struggles with groups that were better represented in the sample data, it is likely that LS/CMI is even less accurate when applied to juvenile lifers.

Mr. Rodriguez's recidivism factors line up with juvenile lifers,' even if he also committed crimes while a young adult. What distinguishes juvenile lifers is what also distinguishes Mr. Rodriguez: an indefinite prison sentence that lasted through his 20s and 30s, and release at an advanced age.

c. LS/CMI predicts recidivism poorly even for the typical parole applicant.

LS/CMI has low predictive validity even for the general population for which it was designed. Studies confirm that the LS system has weak predictive validity—not rare for a system that attempts to make predictions in the messy realm of human behavior but unacceptable for a tool that is used to determine a person's liberty. Academics, journalists, and activists have shown many instances

[50] Mark E. Olver, Keira C. Stockdale & J. Stephen Wormith, *Thirty Years of Research on the Level of Service Scales: A Meta-Analytic Examination of Predictive Accuracy and Sources of Variability*, 26 Psych. Assessment 156, 170 (2013), https://www.researchgate.net/publication/258920739_Thirty_Years_of_Research_on_the_Level_of_Service_Scales_A_Meta-Analytic_Examination_of_Predictive_Accuracy_and_Sources_of_Variability.
[51] *See* Chenane et al., *supra* note 25, at 1.

in which risk assessment tools have had low predictive validity. With respect to one risk-assessment algorithm used in Florida, "only 20 percent of the people predicted to commit violent crimes actually went on to do so."[52] LS/CMI's developers have stressed that the tool must be carefully tested "for *all populations* it is designed to serve,"[53] but there is no evidence that MPB has tested the tool for its general pool of parole applicants or for juvenile lifers.

Predictive validity can be measured with statistical tools such as the Area Under the Receiver Operating Characteristic Curve ("AUC"). In the context of risk assessment, AUC is the likelihood the tool correctly gave higher risk scores to people who ended up recidivating.[54] An AUC of 50% would mean the tool was only as good as a coin flip—half the recidivists incorrectly received lower scores than non-recidivists. An AUC of 100% would mean the system assigned higher scores to every single recidivist in each random pairing. Between the two extremes of a coin flip and a perfect predictor, the definition of a "good" AUC is contingent on circumstances and is even contested among experts within the same industry.

---

[52] Julia Angwin et al., *Machine Bias*, ProPublica (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[53] James Lant, *Risk Assessment FAQ*, LinkedIn Pulse (Dec. 31, 2019), https://www.linkedin.com/pulse/risk-assessment-faq-james-lant/.

[54] Christopher T. Lowenkamp & Kristin Bechtel, *The Predictive Validity of the LSI-R on a Sample of Offenders Drawn from the Records of the Iowa Department of Corrections Data Management System*, 71 Fed. Probation 1, 5 (2007), *available at* https://www.uscourts.gov/sites/default/files/71_3_4_0.pdf.

Often, different industries will have different acceptable AUCs: a marketing tool's AUC of 60% might be acceptable whereas a medical tool's AUC can never dip below 95% because medical decisions involve life-or-death choices that marketing decisions do not.[55] While some risk assessment scholars refer to AUCs of 56%, 64%, and 71% as the thresholds for low, medium, and high predictive validity, these benchmarks are far too low for decisions impacting liberty.[56] Some propose not using any such rules of thumb to define a "good" AUC score, but instead giving the standard definition each time and leaving it to the policymaker to decide whether that is an acceptable outcome. For example, jurisdictions must decide whether it's acceptable to use a tool when the probability that a randomly selected recidivist had a higher risk score than a randomly selected non-recidivist is only 64%.

Although LS tools inform decisions that affect a person's liberty, they have low AUC scores. Studies show LS tools to have AUCs of 64 to 69% for general parolee or probationer populations in different jurisdictions, which is concerningly close to a coin flip given the stakes.[57] In other words, many people who did not

---

[55] Statology.org, *What Is Considered a Good AUC Score?* (Sept. 9, 2021), https://www.statology.org/what-is-a-good-auc-score/.
[56] Hamilton, *supra* note 17, at 1566-67.
[57] Pamela M. Casey et al., *Offender Risk & Needs Assessment Instruments: A Primer for Courts* (2014), *available at* https://www.ncsc.org/__data/assets/pdf_file/0018/26226/bja-rna-final-report_combined-files-8-22-14.pdf.

recidivate still received higher risk scores than people who did, and vice versa. If LS/CMI is not accurate for the population it was designed for, it is particularly unlikely to be accurate for juvenile lifers, a population it was not designed for.

Predictive tools must be validated for each population they will be used on. That analysis should be done by an independent third party, and the results should be made publicly available. No studies have examined LS/CMI's predictive validity for parole applicants in Massachusetts, particularly juvenile lifers. LS/CMI should not be used to determine whether a person remains incarcerated at least until such a study has been conducted and made available to the public.

II. Parole applicants and the general public need transparent accounting of how the Commonwealth uses risk assessment tools.

The opacity surrounding the MPB's use of LS/CMI is unacceptable, particularly given the important interests at stake and the likelihood that LS/CMI is wrong about both juvenile lifers and general parole applicants. The parole applicant is only provided a redacted version of their LS/CMI report and isn't given information about the sources of data that went into their assessment, the logic of the tool, or what role the report played in their parole decision. When EPIC requested a blank scoresheet, scoring guides, training manuals, and validation studies from the Parole Board and the company, we were flatly denied.[58]

[58] Massachusetts Parole Board and MHS, Inc., Denial of EPIC FOIA Request, https://epic.org/wp-content/uploads/2022/02/EPIC-Denial-LSCMI-Parole.pdf.

Lack of transparency around this tool disadvantages the public, but particularly parole applicants like Mr. Rodriguez, whose liberty is being determined by a tool with known accuracy and bias problems.

a. A parole applicant must be given unredacted LS/CMI scoresheets along with any necessary information to fulfill the parole applicant's due process right to a meaningful opportunity for parole.

The LS/CMI tool directly affects the liberty of an individual in that it contributes to a decision about whether parole is granted. Mr. Rodriguez has a due process right to a meaningful opportunity for parole.

Highly redacted scoresheets and blank comment sections do not give parole applicants in Massachusetts a sufficient understanding of the assessment LS/CMI performed on them and fails to fulfill the requirement of a meaningful opportunity for parole. The MPB claims that the "bar chart" and partially redacted "strengths and comments/notes" given to Mr. Rodriguez, which do not include the full LS/CMI scoresheet along with essential information necessary to understand it,[59] provides him with enough information to understand his score. But Parole applicants like Mr. Rodriguez cannot understand their score from the documents provided. Questions about which subjective factors may have been erroneously computed, or which factors were overridden, or which factors should not be used

[59] *See, e.g.* Idaho LSI-R Training Manual, MultiHealth Systems, Inc. (2017) https://archive.epic.org/EPIC-19-11-21-ID-FOIA20191206-LSI-R-Training-Manual.pdf.

to determine applicants' suitability for parole cannot be answered based on these records. The paucity of information provided also puts an unreasonable burden on the parole applicant to piece together vital information about what was fed into the LS/CMI calculation, with no assurance of accuracy or completeness. The state and the Parole Board both acknowledge the obvious shortcomings of the LS/CMI for individuals like Mr. Rodriguez and concede that when overrides are used in the LS/CMI, the predictive validity of the model is further reduced.[60] Still, the Board maintains that parole applicants like Rodriguez are not entitled to all information required to fully interpret each factor behind their denial of parole.[61] The inherent fallibility of risk assessment tools demands that parole applicants be able to meaningfully interrogate them.

Finally, it is not only important for a parole applicant to know what factors were used to assess their recidivism risk, but also how those factors were weighed, how each factor was determined in their case, and if any overrides affected their final risk score. Secrecy around the use or function of automated tools is, as scholar Natalie Ram explains, "in tension with, if not in violation of, parole

[60] Def.-Appellee's Resp. Br. 47.
[61] Def.-Appellee's Resp. Br. 48.

applicants' ability to vindicate their due process interests throughout the criminal justice process, as well as their confrontation rights at trial." [62]

b. The Commonwealth should make LS/CMI scoresheets, training guides, and other developmental documents available to the public.

Parole applicants and the public are entitled to all information necessary to understand and interrogate the Massachusetts Parole Board's use of LS/CMI. But both the agency and MHS, Inc. have refused to release information about the tool, claiming that the information is proprietary. The Commonwealth should not be able to shield important information about a tool that impacts individuals' liberty simply because it was developed by a private company. If the Commonwealth wishes to use a tool to determine a person's liberty, it must be transparent about its use.

As explained in section I, the LS/CMI has limited capability to accurately assess the risk and needs of people eligible for parole. Risk assessment tools are based on a set of decisions about what makes someone more deserving of parole. Adopting a tool means adopting the decisions a contractor made when they designed the tool, including the factors they think affect risk and the weight that

---

[62] Natalie Ram, *Innovating Criminal Justice*, 112 Northwestern U. L. Rev. 659, 692 (2018), https://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?article=1322&context=nulr.

each factor should receive. The public and parole applicants deserve to know more about these value-laden, and potentially error-prone, decisions.

The status quo in Massachusetts, however, blocks that transparency. The Parole Board and MHS Inc., Inc assert that the details of LS/CMI are proprietary and privileged, preventing parole applicants and the public from understanding which factors are included in the LS/CMI risk score and how each factor impacts their score. Melissa Hamilton explains succinctly that

> Too often, claims by tool developers or criminal justice agencies that tools are proprietary have meant that little information on scientific validity is publicly available. Even when some data are released, a critical eye can quickly uncover reasons to be skeptical. Validation studies may offer inflated estimates when performed by interested parties with allegiance bias or when based on research not conducted in real-world settings. Typically, these studies are not placed in peer-reviewed journals, which means that they are not vetted by independent referees.[63]

Both parole applicants and the general public would benefit from this transparency. Transparency tells parole applicants how they were scored, allows third-party testing to detect biases, promotes changes to improve the accuracy of the system, and helps root out conflicts of interest.

---

[63] Melissa Hamilton, *Judicial Gatekeeping on Scientific Validity with Risk Assessment Tools*, 38 Behavioral Sciences & The Law 226 (2020), https://onlinelibrary.wiley.com/doi/10.1002/bsl.2456.

The Commonwealth is particularly secretive; most other jurisdictions are more transparent about their use of risk assessments. In Idaho, for example, the public and parole applicants have transparency rights, and the state has a law banning "builder[s] or user[s] of a pretrial risk assessment tool [from] assert[ing] trade secret or other protections . . . in a criminal or civil case."[64] West Virginia also commissioned a study testing LS/CMI's predictive validity for the state's population.[65]

The public should have access to blank LS/CMI scoresheets, scoring guides, training manuals, data sharing and retention policies, and localized validation

---

[64] Idaho H.B. No. 118, available at https://legislature.idaho.gov/wp-content/uploads/sessioninfo/2019/legislation/H0118A2.pdf. ("PRETRIAL RISK ASSESSMENT TOOLS. (1) All pretrial risk assessment tools shall be transparent, and:
(a) All documents, data, records, and information used to build or validate the risk assessment and ongoing documents, data, records, information, and policies surrounding the usage of the risk assessment shall be open to public inspection, auditing, and testing; (b) A party to a criminal case wherein a court has considered, or an ex-pert witness has relied upon, a pretrial risk assessment tool shall be entitled to review all calculations and data used to calculate his own risk score; and (c) No builder or user of a pretrial risk assessment tool may assert trade secret or other protections in order to quash discovery in a criminal or civil case.")

[65] Maria M. Orsini, Stephen M. Haas & Douglas H. Spence, *Predicting Recidivism of Offenders Released from the West Virginia Division of Corrections: Validation of the Level of Service/Case Management Inventory*, West Virginia Office of Research & Strategic Planning in the Criminal Justice Statistical Analysis Center (2015), available at https://djcs.wv.gov/ORSP/SAC/Documents/JCEBP%20LSCMI%20Validation%20DOC%202015.pdf.

studies. Withholding all documents related to the Parole Board's use of LS/CMI is antithetical to the public's right of access to information and prioritizes corporate profit over fair and consistent administration of the law. Just as Mr. Rodriguez and similarly situated parole applicants can't access an unredacted version of the risk assessment scoresheets that impacted their parole status, the public is barred from getting blank versions through Massachusetts Public Records law.[66] The criminal justice system can affect anyone at any time, and the public has a right to know how that system is administered.

At minimum, the Commonwealth should ensure the following is available or proactively published: who developed the tool, the stated purpose of the tool, the input data, what factors are used, how those factors are interpreted, how much weight each factor holds, the decision-making matrix, and any data sharing and retention policies.[67]

CONCLUSION

The developers of predictive analytical tools promise more efficient and less biased decision-making, but that is not necessarily what they deliver. Poor tool design, careless implementation, and secrecy can lead to the opposite of what these

---

[66] Massachusetts Parole Board and MHS, Inc., Denial of EPIC FOIA Request, https://epic.org/wp-content/uploads/2021/12/EPIC-FOIA-MassDOC-Rejection-06102021.pdf.

[67] EPIC, *Liberty at Risk* at 15 (2020), *available at* https://archive.epic.org/LibertyAtRiskReport.pdf.

tools promise, entrenching past patterns of bias and producing inaccurate predictions. The LS/CMI should not be used by the Massachusetts Parole Board to estimate the risk and needs of juvenile lifers like Mr. Rodriguez.

Dated: February 14, 2022

*/s/ Benjamin Winters*__________
Benjamin Winters (*pro hac vice pending*)
Counsel of Record
Caitriona Fitzgerald (BBO #673324)
Electronic Privacy Information Center
1519 New Hampshire Ave. NW
Washington, DC 20036
(202) 483-1140

CERTIFICATE OF COMPLIANCE

I hereby certify that the above brief complies with the rules of court that pertain to the filing of brief, including, but not limited to: Rule 16(a)(13); Rule 16(e); Rule 18; Rule 20; and Rule 21. This brief complies with the type-volume limitation of Rule 20(2)(C) because it contains 6,857 words, excluding the parts of the brief limited by the rule. It complies with the type style requirements of Rule 20 because it has been prepared in proportionally spaced typeface using Microsoft Office Word in 14-point Times New Roman style.

Dated: February 14, 2022

*/s/ Benjamin Winters*__________
Benjamin Winters (*pro hac vice pending*)
Counsel of Record
Caitriona Fitzgerald (BBO #673324)
Electronic Privacy Information Center
1519 New Hampshire Ave. NW
Washington, DC 20036
(202) 483-1140

CERTIFICATE OF SERVICE

I hereby certify that on February 14, 2022, this brief was electronically filed with the Clerk of the Court for the Supreme Judicial Court, was served by emailing PDF copies of each document to Shara Benedetti at shara.benedetti@state.ma.us, and served upon the following counsels of record through the Electronic Filing System:

Counsel for the Plaintiff-Appellant:
Melissa Allen Celli (Melissa@MelissaCelliLaw.com)
Law Office of Melissa Allen Celli
31 Trumbull Road – Suite B
Northampton, MA 01060

Counsel for the Defendant-Appellee:
Todd M. Blume (todd.blume@mass.gov)
Assistant Attorney General
Office of the Attorney General
Criminal Bureau
One Ashburton Place
Boston, MA 02108

Dated: February 14, 2022

*<u>/s/ Benjamin Winters</u>*___________
Benjamin Winters (*pro hac vice pending*)
Counsel of Record
Caitriona Fitzgerald (BBO #673324)
Electronic Privacy Information Center
1519 New Hampshire Ave. NW
Washington, DC 20036
(202) 483-1140