

SYLLABUS

This syllabus is not part of the Court's opinion. It has been prepared by the Office of the Clerk for the convenience of the reader. It has been neither reviewed nor approved by the Court and may not summarize all portions of the opinion.

State v. Michael Olenowski (A-56-18) (082253)

Re-Argued June 1, 2023 -- Decided November 15, 2023

SABATINO, P.J.A.D. (temporarily assigned), writing for the Court.

In State v. Olenowski (Olenowski I), 253 N.J. 133 (2023), the Court adopted for criminal cases a non-exclusive, multi-factor test for the reliability of expert testimony patterned after the standard established in Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993). The Court now considers whether Drug Recognition Expert (DRE) testimony is reliable and admissible under that standard. The Court also considers the appropriate standard of review for Daubert-based expert reliability determinations in criminal appeals.

N.J.S.A. 39:4-50 prohibits impaired driving, whether the impairment is caused by alcohol or one or more drugs. A driver whose blood alcohol concentration (BAC) level exceeds the 0.08% limit prescribed by that statute is guilty -- per se -- of driving while intoxicated. But there is no equivalent per se violation in this state for persons who drive with impairment-causing drugs in their system.

Detecting and proving that a driver ingested and was under the influence of drugs while behind the wheel can be challenging. To enable such detection, law enforcement officials and researchers developed a twelve-step protocol:

- (1) a breath alcohol test;
- (2) an interview of the arresting officer;
- (3) a preliminary examination and first pulse check;
- (4) a series of eye examinations;
- (5) four divided attention tests;
- (6) a second examination and vital signs check;
- (7) a dark room examination of pupil size and ingestion sites;
- (8) an assessment of muscle tone;
- (9) a check for injection sites;
- (10) an interrogation of the driver by the DRE;

- (11) a final opinion, based on the totality of the examination, about whether the driver is under the influence of a drug or drugs; and
- (12) a toxicological analysis.

Defendant Michael Olenowski was convicted of drug-impaired driving based in part on DRE evidence. His convictions were upheld on appeal, and the Court granted certification to determine whether DRE testimony is admissible under the “general acceptance” admissibility standard established in Frye v. United States, 293 F. 1013 (D.C. Cir. 1923). 236 N.J. 622 (2019).

Finding that the record was not sufficient to make that determination, the Court asked a Special Master to conduct a hearing. 247 N.J. 242, 244 (2019). The Special Master concluded that DRE evidence should be admissible under Frye.

In subsequent briefing to the Court, several counsel focused upon error rates associated with DRE evidence. Because error rates are expressly considered under Daubert, but not Frye, the Court asked for supplemental briefing on “whether this Court should depart from Frye and adopt the principles of Daubert in criminal cases.” Both parties and nearly all of the amici advocated that the Court adopt the Daubert standard, similar to its previous adoption of Daubert-based principles for civil cases in In re Accutane Litigation, 234 N.J. 340 (2018).

In Olenowski I, the Court adopted a “Daubert-type standard” for determining the reliability of expert evidence in criminal and quasi-criminal cases and remanded this matter to the Special Master to apply that standard. 253 N.J. at 153, 155. The Special Master concluded that the twelve-step DRE protocol satisfies the reliability standard of N.J.R.E. 702 when analyzed under the methodology-based Daubert-Accutane standard. The Court now considers that conclusion.

HELD: Daubert-based expert reliability determinations in criminal appeals will be reviewed de novo, while other expert admissibility issues are reviewed under an abuse of discretion standard. Here, the extensive record substantiates that DRE testimony sufficiently satisfies the Daubert criteria to be admissible, with the following four limitations and safeguards:

- * The DRE may opine only that the evaluation is “consistent with” the driver’s ingestion or usage of drugs, not that it was actually caused by drugs.
- * If the State fails to make a reasonable attempt to obtain a toxicology report without a persuasive justification, the DRE testimony must be excluded.
- * The defense must be afforded a fair opportunity to impeach the DRE.
- * Model instructions to guide juries about DRE evidence should be considered.

1. Most evidentiary rulings are reviewed for an abuse of discretion. In Accutane, the Court held that trial courts' expert reliability determinations should be reviewed under that standard in civil matters. See 234 N.J. at 392. In criminal law, however, a trial court's reliability determination under Frye -- i.e., its determination of whether the relevant scientific community generally accepts a scientific theory, test, or technique -- was accorded less deferential review than other evidentiary decisions. Going forward, in New Jersey criminal and quasi-criminal cases in which the trial court has admitted or excluded an expert witness based upon Daubert reliability factors, appellate courts shall review that reliability determination de novo. However, other case-specific determinations about the expert evidence -- such as whether the witness has sufficient expertise, whether the evidence can assist the trier of fact, and whether the relevant theory or technique can properly be applied to the facts -- should be reviewed for an abuse of discretion. (pp. 44-60)

2. The United States Supreme Court identified in Daubert a list of four factors for assessing reliability of an expert's methodology under Fed. R. Evid. 702:

- (1) whether the scientific theory or technique can be, or has been, tested;
- (2) whether it has been subjected to peer review and publication;
- (3) the known or potential rate of error as well as the existence of standards governing the operation of the particular scientific technique; and
- (4) general acceptance in the relevant scientific community.

Daubert made clear that the factors are non-exclusive and that the reliability inquiry is "flexible," signaling that other considerations may also be pertinent. See 509 U.S. at 594. For ease of discussion in this particular case, the Court reorganizes the Supreme Court's listing of Daubert factors in a few ways and applies them in this sequence: (A) adequacy of standards; (B) publication and peer review; (C) testability and error rate; and (D) general acceptance. (pp. 60-63)

3. Adequacy of Standards. The twelve-step DRE process is elaborate and standardized. It is grounded in a program that has been used across the nation and abroad for decades and is periodically modified. The Court reviews counter-arguments, including the concern that DREs are neither physicians nor medical professionals, and explains why they do not alter its conclusion. (pp. 63-68)

4. Peer Review and Publication. The Special Master appropriately considered not only the existence of roughly two dozen studies but also their substantive content and conclusions. He determined that they "support the State's position that the DRE protocol has consistently been found to be a reliable method for detecting impairment by drugs." Although the studies have certain limitations, the Court holds that they meet the Daubert factor of publication and peer review. (pp. 69-77)

5. Testability and Error Rate. “Ordinarily, a key question to be answered in determining whether a theory or technique is scientific knowledge that will assist the trier of fact will be whether it can be (and has been) tested.” Daubert, 509 U.S. at 593. The term “ordinarily” conveys that a judge’s findings of testability and reasonably low error rates from test results are expected -- but not always required -- elements of a proponent’s reliability showing. As the Special Master recognized, there are inherent practical limitations within the DRE program that complicate efforts to test the program results empirically and to obtain meaningful error rates. Constitutional, ethical, and practical constraints make the DRE program less “testable” and the error rate less “knowable” than the ideal. After reviewing the New Jersey data in the record, the Court concludes that the testability and false-positive error rate aspects of the Daubert analysis are largely inconclusive but finds that the inconclusiveness should not categorically bar admission of this useful evidentiary source. The Court rejects the assertion that testability and error rates are categorically the most important Daubert factors. (pp. 78-90)

6. General Acceptance. For many years, the DRE protocol has been widely and regularly used across this country and abroad. No state has discontinued it, and no state’s highest court has nullified it. The protocol has been studied multiple times and periodically revised and enhanced. Although it has imperfections, the protocol has stood the test of time in its widespread acceptance. (pp. 90-95)

7. Many facets of the DRE protocol weigh in favor of its reliability, but the protocol has several weaknesses as well. It does not establish that a driver is actually impaired, or that the drug categories identified by the DRE are definitively the cause of any such impairment. And there are palpable risks of confirmation bias when a DRE officer administers the protocol, particularly in the more subjective aspects of the examination. Thus, although Court finds DRE testimony sufficiently reliable to be admitted in our courts, it adopts several limitations on the admissibility and probative use of a DRE’s opinion in criminal and quasi-criminal cases:

First, a DRE is only allowed to opine in court that the protocol has presented indicia that are “consistent with” the driver’s usage of certain categories of drugs. The DRE’s expert opinion testimony must not go further than that. Proof of consistency can be pertinent as one component within the totality of the evidence to support an inference that drugs caused a driver’s impairment.

Second, a toxicology report corroborating a DRE’s opinion is important evidence. DRE officers must make a reasonable attempt to obtain a toxicology report when it is feasible to do so -- and preferably to obtain a blood sample rather than a urine sample -- when their protocol indicates at Step 11 an opinion of consistency with drug use. If the court finds no reasonable attempt was made, despite its feasibility, the DRE evidence shall

be excluded. However, if the State establishes a reasonable justification for the lack of a toxicology report, then the DRE evidence is admissible, subject to defense impeachment and counterproofs.

Third, if the trial court admits DRE evidence for the State, the defense shall have a fair opportunity to impeach or rebut it through cross-examination of the DRE and with counterproofs.

Fourth, it may be beneficial for the court to provide jurors with an explanatory instruction about the DRE evidence, such as the consistency limitation. The Court refers this subject to the Model Criminal Jury Charges Committee for its consideration.

A positive DRE opinion at Step 11 is not dispositive of a driver's guilt of driving under the influence of drugs. Unlike a BAC reading of .08% or more in a drunk driving case, the DRE's opinion is not used as a per se test of guilt. Instead, the DRE testimony is just one part of the evidence as a whole, and it can be amplified or rebutted. The State would have a much steeper burden to prove a driver's guilt when it lacks corroborating proof from a toxicology report. (pp. 96-107)

The reports and findings of the Special Master are ADOPTED AS MODIFIED. Olenowski's convictions are VACATED.

JUSTICE PIERRE-LOUIS, dissenting, explains that the Court adopted the Daubert standard for criminal cases in Olenowski I as a means to ensure reliability through concentration on “the soundness of the methodology used to validate a scientific theory or technique, the strength of the reasoning underlying it, and the accuracy of the theory or technique in practice.” 253 N.J. at 150 (emphasis added). Under Daubert, Justice Pierre-Louis notes, the Court's charge is not to create safeguards to try to preserve the use of techniques that cannot withstand rigorous scrutiny, but rather to ensure that if evidence is given the weight of an expert's endorsement, that evidence has “a sufficient scientific basis to produce uniform and reasonably reliable results.” Ibid. By altering the Daubert factors here, Justice Pierre-Louis writes, the majority not only reaches a determination of reliability that is not supported by the test, it also upends the clear guidance in Olenowski I regarding placing the focus of these expert reliability determinations on testing, peer review, and error rates. Justice Pierre-Louis would hold that DRE evidence is not admissible under N.J.R.E. 702.

JUSTICES PATTERSON, SOLOMON, WAINER APTER, and FASCIALE join in JUDGE SABATINO's opinion. JUSTICE PIERRE-LOUIS filed a dissent, in which CHIEF JUSTICE RABNER joins.

SUPREME COURT OF NEW JERSEY

A-56 September Term 2018

082253

State of New Jersey,

Plaintiff-Respondent,

v.

Michael Olenowski,

Defendant-Appellant.

On certification to the Superior Court,
Appellate Division.

Supplemental
Special Master Report
April 13, 2023

Re-Argued
June 1, 2023

Decided
November 15, 2023

Margaret McLane, Assistant Deputy Public Defender, argued the cause for appellant (Joseph E. Krakora, Public Defender, attorney; Margaret McLane, of counsel and on the supplemental brief).

Sarah C. Hunt, Deputy Attorney General, argued the cause for respondent (Matthew J. Platkin, Attorney General, attorney; Sarah C. Hunt, of counsel and on the supplemental brief, and Adam D. Klein, Deputy Attorney General, on the supplemental brief).

Alexander Shalom argued the cause for amici curiae American Civil Liberties Union of New Jersey and statistics experts Alicia Carriquiry, Kori Khan, and Susan VanderPlas (American Civil Liberties Union of New

Jersey Foundation, attorneys; Alexander Shalom and Jeanne LoCicero, on the supplemental brief).

John Menzel argued the cause for amicus curiae New Jersey State Bar Association (New Jersey State Bar Association, attorneys; Jeralyn L. Lawrence, President, of counsel, and John Menzel, on the supplemental brief).

Steven W. Hernandez submitted a supplemental brief on behalf of amicus curiae National College for DUI Defense (The Hernandez Law Firm, attorneys; Steven W. Hernandez, of counsel and on the supplemental brief).

Jeffrey H. Sutherland, Cape May County Prosecutor, submitted a supplemental brief on behalf of amicus curiae County Prosecutors Association of New Jersey (Jeffrey H. Sutherland, President, attorney; Jeffrey H. Sutherland, Joseph Paravecchia, First Assistant Hunterdon County Prosecutor, Laura Sunyak, Assistant Mercer County Prosecutor, Gretchen Pickering, Assistant Cape May County Prosecutor, and David M. Liston, Assistant Middlesex County Prosecutor, of counsel, and Monica do Outeiro, Assistant Monmouth County Prosecutor, of counsel and on the supplemental brief).

Aidan P. O'Connor submitted a supplemental brief on behalf of amicus curiae Association of Criminal Defense Lawyers of New Jersey (Pashman Stein Walder Hayden, attorneys; Aidan P. O'Connor and Marc M. Yenicag, of counsel and on the supplemental brief).

Vito A. Gagliardi, Jr., submitted a supplemental brief on behalf of amicus curiae New Jersey State Association of Chiefs of Police (Porzio, Bromberg & Newman, attorneys; Vito A. Gagliardi, Jr., of counsel, and David L. Disler and Thomas J. Reilly, on the supplemental brief).

Evan M. Levow submitted a supplemental brief on behalf of amicus curiae DUI Defense Lawyers Association, Inc.

(Levow DWI Law, attorneys; Evan M. Levow, on the supplemental brief).

JUDGE SABATINO (temporarily assigned)
delivered the opinion of the Court.

Table of Contents

Introduction 4

I. New Jersey’s Driving Under the Influence Statutory Scheme 7

II. Factual Background and Procedural History 15

 A. Drug Recognition Experts (DREs) and the Drug Evaluation and Classification Program15

 1. The 12-Step DRE Protocol15

 2. Development of the DRE Protocol27

 B. Defendant’s DRE-Based Convictions and Ensuing Appeals28

 C. This Court’s Review and the Special Master Proceedings30

 1. Special Master Hearing and SM Report I: Admissibility Under the Frye Standard31

 2. Olenowski I: Adoption of the Daubert Admissibility Standard36

 3. SM Report II: Applying the Daubert Standard.....37

III. Arguments of the Parties and Amici 42

IV. Standard of Review of Expert Reliability Rulings 44

V. De Novo Application of the Daubert Factors 60

 A. Adequacy of Standards63

 B. Peer Review and Publication69

 C. Testability and Error Rate78

 D. General Acceptance90

VI. Analysis 96

 A. The “Consistency Only” Limitation98

B. The Absence of a Toxicology Report (“Step 12”).....	102
C. Fair Opportunity for Defense Impeachment and Counterproofs	104
D. Jury Instructions.....	105
VII. Impact on the State’s Burden of Proof	106
VIII. Conclusion	107

Introduction

Driving by drug-impaired persons is a critical and growing public safety problem in New Jersey and across the nation. Such drivers commonly have reduced perception and slowed reaction times. They are prone to cause accidents -- sometimes fatal ones.

Our state laws make it illegal to operate a motor vehicle “while under the influence of intoxicating liquor, narcotic, hallucinogenic, or habit-producing drug.” N.J.S.A. 39:4-50. Repeat offenders can be sentenced to jail. However, unlike driver impairment caused by alcohol, which can be proven by a prohibited blood alcohol concentration (BAC) level established by statute, New Jersey statutes do not have comparable “per se” standards for driving under the influence of drugs (DUID), or drugged driving. Many other states similarly lack numerical standards for DUID.

Detecting and proving that a driver ingested and was under the influence of drugs while behind the wheel can be challenging. To enable such detection, law enforcement officials and researchers began in the 1970s to develop a

protocol now known as the Drug Evaluation and Classification Program. The protocol is used today in all fifty states, the Canadian provinces, and other countries.

The protocol consists of twelve steps administered by specially trained officers known as Drug Recognition Experts (DREs). The twelve steps entail interviewing and observing the driver, checking vital signs, administering standardized field sobriety tests, and other information-gathering measures. At the end of the protocol, the DRE, guided by a standardized matrix, reaches an opinion about whether the driver is under the influence of drugs from one or more of seven categories and is thereby unable to operate a motor vehicle safely.

The DRE protocol, which the Public Defender and other amici have challenged as unreliable and inadmissible under N.J.R.E. 702, is the subject of this appeal. This Court referred the dispute to a Special Master who conducted forty-two days of extensive hearings with sixteen witnesses.

In his initial 332-page report, the Special Master concluded that the DRE protocol is generally accepted as reliable and thus admissible under the then-applicable admissibility test of Frye v. United States, 293 F. 1013 (D.C. Cir. 1923). Thereafter in State v. Olenowski (Olenowski I), 253 N.J. 133 (2023), this Court prospectively adopted a non-exclusive, multi-factor test of expert

reliability patterned after the standard in Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993). We remanded the matter to the Special Master to reconsider his findings under the Daubert factors. On remand, the Special Master issued a 57-page report determining that DRE testimony is reliable and admissible under a Daubert-type standard.

With the benefit of thoughtful advocacy from the parties and amici on both sides, we adopt the Special Master's conclusions with significant modifications and limitations.

First, we unanimously hold that Daubert-based expert reliability determinations in our criminal appeals should be reviewed de novo, while other expert admissibility issues are to be reviewed under an abuse of discretion standard.

Second, we conclude the extensive record substantiates that DRE testimony sufficiently satisfies the Daubert criteria to be admissible with important limitations. We reach that conclusion based on the totality of factors, particularly standardization, publication and peer review, and general acceptance. Although the factors of testability and false positive error rate are largely inconclusive due to the skewed composition of the sample of stopped drivers, the record as a whole justifies the admission of DRE testimony, with the following four limitations and safeguards:

- The DRE testimony must be confined to an opinion that the evaluation is “consistent with” the driver’s ingestion or usage of one or more of the identified drug categories. The DRE may not present opinions as to whether the driver’s observed impairment was actually caused by such drugs and, if so, to what extent.
- If feasible, the State must make a reasonable attempt to obtain a toxicology report based on a blood or urine sample from the driver. If the State fails to make such a reasonable attempt without a persuasive justification, the DRE opinion testimony must be excluded.
- The defense must be afforded a fair opportunity to impeach the DRE and present competing proofs.
- Model instructions to guide juries about DRE evidence should be considered.

With those limitations and safeguards, we adopt the Special Master’s findings, as modified.

I. New Jersey’s Driving Under the Influence Statutory Scheme

For context, we begin with a brief review of the motor vehicle statute that prohibits impaired driving in this state, N.J.S.A. 39:4-50. The statute declares that “[a] person who operates a motor vehicle while under the influence of intoxicating liquor, narcotic, hallucinogenic or habit-producing drug, or operates a motor vehicle with a blood alcohol concentration of 0.08% or more” is subject to certain penalties after each violation, including fines, detainment, and imprisonment. N.J.S.A. 39:4-50(a) (emphasis added); see also State v. Chun, 194 N.J. 54, 71 (2008) (discussing drunk driving and the DWI

penalty scheme); State v. Bealor, 187 N.J. 574, 576, 588 (2006) (discussing DUID).

The statute has a vital purpose. It “seeks to prevent the operation of motor vehicles by those whose faculties are so impaired as to present a danger to the safety of others as well as themselves.” State v. DiCarlo, 67 N.J. 321, 325 (1975).¹ In enacting N.J.S.A. 39:4-50, “[t]he obvious intention of the Legislature was to prescribe a general condition, short of intoxication, as a result of which every motor vehicle operator has to be said to be so affected in judgment or control as to make it improper . . . to drive on the highways.” State v. Johnson, 42 N.J. 146, 164-65 (1964) (noting that even “the smallest amount of alcohol has some slight effect or influence on an individual” and that being “absolutely ‘drunk’” is not a statutory requirement).

¹ In 2016, the Governors Highway Safety Association, based on data from the Fatality Analysis Reporting System of the National Highway Traffic Safety Administration (NHTSA), reported that 43.6% of drivers with known drug test results who were fatally injured in car accidents nationwide were drug-positive and 37.9% of those drivers were alcohol-positive. Jim Hedlund, Governors Highway Safety Ass’n, Drug-Impaired Driving: Marijuana and Opioids Raise Critical Issues for States 7 (May 2018). That report, which was admitted into evidence during the Special Master’s hearings, highlights that, for those drivers killed in car accidents from 2006 to 2016, drug-positive rates have substantially increased while alcohol-positive rates have decreased. Ibid. An expert witness for the State, Thomas E. Page, also testified before the Special Master about the prevalence of drug-positive drivers in car accidents in which drivers were fatally injured.

The present appeal concerns a driver's impairment caused by the ingestion of drugs, not alcohol. Yet the legal consequences are the same. "The driving while intoxicated statute expresses the Legislature's desire to prohibit driving while intoxicated; whether the cause of intoxication is alcohol or narcotics, hallucinogens or habit-forming drugs is largely irrelevant." Bealor, 187 N.J. at 588.

The statute does express one critical difference: per se liability. Specifically, a driver whose BAC level exceeds the 0.08% limit prescribed by N.J.S.A. 39:4-50 is guilty -- per se -- of driving while intoxicated. N.J.S.A. 39:4-50(a). There is no equivalent per se violation in this state for persons who drive with impairment-causing drugs in their system. Compare Bealor, 187 N.J. at 583, 588-92 (noting in DUID cases a lack of evidence that specific levels of marijuana consumption have a uniform effect on a driver), with Chun, 194 N.J. at 64 ("[D]rivers whose breathalyzer test results demonstrate the requisite statutorily-imposed BAC are guilty per se of [DWI].").²

² By contrast, at least fifteen other states have such laws. John Lacey et al., U.S. Dept. of Transp., Drug Per Se Laws: A Review of Their Use in States 1 (2010). Three of those states have set cutoff blood concentration levels for certain prohibited drugs. See id. at 1; see also Nev. Rev. Stat. § 484C.110; Va. Code Ann. § 18.2-266; Ohio Rev. Code Ann. § 4511.19. The twelve remaining states prohibit driving with any amount of any prohibited drug in the body. See Drug Per Se Laws, at 1. One of those states, Pennsylvania, has set minimum cutoff levels for blood test results to be admissible in a DWI

Consequently, in prosecutions under N.J.S.A. 39:4-50 involving drugs (or alcohol below .08% BAC), the State must present evidence on a case-by-case basis that the driver actually was under the influence of such drugs. See Bealor, 187 N.J. at 587-91 (discussing the proof required to convict for DUID).

The critical phrase “under the influence” within N.J.S.A. 39:4-50 is “not self-defining and [has] required judicial ascertainment of the legislative intent.” Johnson, 42 N.J. at 164. Case law has provided guidance concerning the phrase’s intended meaning.

In State v. Tamburro, this Court explained that, “[g]enerally speaking,” “[t]he language ‘under the influence’ used in [N.J.S.A. 39:4-50] . . . means a substantial deterioration or diminution of the mental faculties or physical capabilities of a person whether it be due to intoxicating liquor, narcotic, hallucinogenic or habit-producing drugs.” 68 N.J. 414, 420-21 (1975); see also Bealor, 187 N.J. at 589-90. Tamburro cited an earlier opinion of this Court, which held that a driver “was under the influence of a narcotic drug within the meaning of N.J.S.A. 39:4-50(a) if the drug produced a narcotic

prosecution. See 75 Pa. Cons. Stat. § 1547(c)(4); Pa. Dept. of Health, Minimum Levels of Controlled Substances and/or Their Metabolites in Blood, <https://www.health.pa.gov/topics/Labs/Pages/Minimum-Levels.aspx> (last visited Nov. 8, 2023).

effect ‘so altering his or her normal physical coordination and mental faculties as to render such person a danger to himself as well as to other persons on the highway.’” 68 N.J. at 421 (quoting DiCarlo, 67 N.J. at 328); see also Bealor, 187 N.J. at 589-90. Bealor also adopted a definition of “under the influence [that] means ‘a condition which so affects the judgment or control of a motor vehicle operator as to make it improper for him to drive on the highway.’” 187 N.J. at 589 (quoting Tamburro, 68 N.J. at 421).

To establish guilt under N.J.S.A. 39:4-50, the State must prove beyond a reasonable doubt that the defendant was driving while “under the influence” of at least one of the specified types of substances. In particular, the State must prove both the “facts of intoxication” and the “cause of intoxication.” Id. at 588, 591. “[A] conviction for driving while under the influence of alcohol will be sustained on proofs of the fact of intoxication -- [as shown by] a defendant’s demeanor and physical appearance -- coupled with proofs as to the cause of intoxication -- i.e., the smell of alcohol, an admission of the consumption of alcohol, or a lay opinion of alcohol intoxication.” Id. at 588.

Likewise in DUID cases, facts of intoxication must be linked to proofs of the cause of intoxication. For instance, proofs of “erratic and dangerous driving,” “slurred and slowed speech,” “‘bloodshot and glassy’ eyes,” “droopy eyelids,” a “‘pale and flushed’ face,” “fumbl[ing],” “sagging knees,” and an

“emotionless stare” may be linked with physical evidence of an intoxicating drug in the car or in the driver’s control, and “the presence of [an intoxicating drug] in [the] blood stream at the time of the arrest and its likely source.” Id. at 589-90 (first alteration in original).³ Hence, the State must prove in DUID cases that (1) the defendant was intoxicated and (2) the cause of the intoxication was either narcotics, hallucinogens, or habit-producing drugs.

With regard to proving the cause of intoxication in DUID cases, N.J.S.A. 39:4-50 does not define, nor has this Court imposed, a minimum quantum of narcotics, hallucinogens or habit-producing drugs to establish guilt under the statute. Id. at 589. In fact, the State is not required to identify “the particular narcotic[, hallucinogen or habit-producing drug].” Ibid. (alteration in original) (quoting Tamburro, 68 N.J. at 421).

For many years, this Court has permitted lay persons to testify that alcohol was the cause of a driver’s intoxication because “[t]he symptoms of that condition have become such common knowledge that the testimony is admissible.” State v. Smith, 58 N.J. 202, 213 (1971); accord Bealor, 187 N.J. at 587. However, the Court treated marijuana differently, observing in Smith that “[n]o such general awareness exists as yet with regard to the signs and

³ Apart from these examples of evidence of impairment, other evidence of drug use may be present, such as the smell of burnt marijuana or injection track marks on a driver.

symptoms of the condition described as being ‘high’ on marihuana.” 58 N.J. at 213.

Thirty-five years after Smith, this Court clarified in Bealor that lay testimony about the fact of a driver’s intoxication is always admissible, whereas lay testimony ascribing the cause of intoxication is admissible only when the alleged cause is alcohol. 187 N.J. at 577. In Bealor, the defendant was prosecuted for driving under the influence of marijuana. Id. at 581-82. Although we acknowledged that “much ha[d] changed in the intervening years since” Smith, we concluded the State failed to meet its burden of showing that the signs and symptoms of marijuana intoxication had become common knowledge. Id. at 587.

Although Bealor deemed lay testimony about the cause of non-alcohol intoxication inadmissible, this Court’s opinion did not go so far as to require the State to present expert testimony on the subject. Id. at 591. Instead, Bealor held that expert testimony confirming the presence of marijuana in defendant’s blood stream (such as toxicology evidence), in addition to other evidence in the case, would suffice to prove that fact. Id. at 590. To hold “that the nexus between the facts of intoxication and the cause of intoxication can only be proved by expert opinion [] impermissibly impinges on the traditional role of the factfinder and is explicitly disavowed.” Id. at 591.

The circumstances in Bealor illustrated that point. In that case, the arresting officer offered fact testimony as to the defendant's impaired conduct and noted that he smelled burnt marijuana and found a smoking pipe with marijuana residue in it at the time of the driver's arrest. Id. at 578, 590. Forensic scientists testified that the pipe did indeed contain marijuana and that the defendant's blood at the time of arrest contained marijuana. Id. at 581. We held that the aggregate of such proofs would be "more than sufficient" to establish that the driver was under the influence of marijuana through the reasonable inference of the factfinder "to connect the objective facts of intoxication with the proven presence of a cause of intoxication." Id. at 590-91.

Even so, we observed in Bealor that "expert testimony remains the preferred method of proof of marijuana intoxication." Id. at 592 (emphasis added). We noted that, "[i]n view of their training, police officers in this State are eligible to qualify as experts on marijuana intoxication under N.J.R.E. 702." Ibid. More broadly, we acknowledged the training of police officers in "detecting drug-induced intoxication" was "a required course of study" for trainees. Id. at 592-93 (emphasis added).

In this appeal, the State has asked this Court to consider whether expert testimony by a DRE is admissible to prove the cause of intoxication in DUID cases. Bealor did not address that critical issue now before us.

II. Factual Background and Procedural History

With that backdrop, we turn to the facts⁴ and procedural history of this matter. For brevity, we generally incorporate by reference the details meticulously set forth in the Special Master’s 332-page initial report and his 57-page supplemental report. We highlight certain details here.

A. Drug Recognition Experts (DREs) and the Drug Evaluation and Classification Program

1. The 12-Step DRE Protocol

The Drug Evaluation and Classification Program (DECP)⁵ has developed a twelve-step protocol to assess whether a person suspected of drugged driving

⁴ Many facts in the massive record, such as those concerning the history of the DRE program and its twelve component steps, are uncontroverted. We also note the Office of the Public Defender, which served as lead defense counsel for most of the hearings, has expressly advised the Court that it does not dispute the Special Master’s credibility findings about the sixteen witnesses who testified.

⁵ The original name for the program, the “Drug Recognition Expert program,” was eventually changed to the DECP. Int’l Ass’n of Chiefs of Police, The International Drug Evaluation & Classification Program: About the DECP, <https://www.theiacp.org/projects/the-international-drug-evaluation-classification-program> (last visited Nov. 8, 2023).

is impaired by a certain category or multiple categories of drugs. The protocol is administered by trained police officers known as DREs. The protocol is widely used in all fifty states, the District of Columbia, and the Canadian provinces.⁶

To reach an opinion concerning a stopped motorist's condition, DREs combine their general observations, ordinary police work and investigative tactics, standardized field tests, and observations of medically-related manifestations of drug ingestion.

The twelve steps in the DRE protocol⁷ consist of (1) a breath alcohol test; (2) an interview of the arresting officer; (3) a preliminary examination and first pulse check; (4) a series of eye examinations; (5) four divided attention tests; (6) a second examination and vital signs check; (7) a dark room examination of pupil size and ingestion sites; (8) an assessment of muscle tone; (9) a check for injection sites and a third pulse reading; (10) an interrogation of the driver and documentation of statements made by the driver

⁶ Int'l Ass'n of Chiefs of Police, States and Countries with DREs: DEC Program States, <https://www.theiacp.org/states-and-countries-with-dres> (last visited Nov. 8, 2023).

⁷ Although not all twelve steps were performed in Olenowski's cases, we stress that the best practice is to reasonably attempt all of the first eleven steps, and, unless it is demonstrably infeasible, to complete the toxicology analysis of step twelve. See our discussion below at Part VI (B).

as well as any other observations; (11) a final opinion based on the totality of the examination; and (12) a toxicological analysis. See Special Master's Report of Findings and Conclusions of Law 4 (Aug. 22, 2022) (SM Report I).

We describe each of those steps in more detail.

Step 1: The breath alcohol test determines the driver's blood alcohol concentration. The test is used to determine whether alcohol may be the sole or contributing cause of any observed signs of the driver's impairment. The DRE training materials emphasize that many drivers who are under the influence of drugs also have alcohol in their systems.

Step 2: The specially trained DRE's interview with the arresting officer occurs because the specially trained DRE examining the driver usually will not be the same officer who stopped or arrested that person. During this step, the DRE can obtain information from the arresting officer that might be indicative of the drug or drugs the driver has ingested. The information from the arresting officer may include any observations of the driver's behavior and of the scene of the arrest, any statements offered by the driver during any questioning, and any relevant physical evidence, such as drugs or drug paraphernalia seized at the scene.

Step 3: The preliminary examination is the first opportunity for the DRE to observe the driver closely. DRE training materials emphasize that a primary

purpose of this preliminary examination is to determine whether the driver has an injury or other medical condition that may be related to drug use or observed impairment. During the examination, the DRE takes the first of three pulse measurements.

As noted in the training materials, the DREs at this step can also perform some preliminary eye-related assessments and initial estimations preceding the eye tests to be performed fully later in Steps 4 and 7. This step is intended to help the DRE decide whether to continue with the evaluation, to proceed instead with a drunk driving charge, or to refer the driver for medical treatment. Drivers have the right to refuse to proceed with the DRE process.

Step 4: The eye examinations in this step are conducted because some drugs produce observable effects on the eyes. The examinations conducted are designed to assess equal tracking by the eyes and equal pupil size in both eyes. Three tests are involved.

First, there is the horizontal gaze nystagmus (HGN) exam, which checks for the lack of the smooth eye pursuit, sustained eye jerking at maximum deviation (where the eye is turned as far to the side as possible), and the angle of onset at which the eyes first begin to jerk, all while tracking the eyes in a horizontal path following a stimulus. Second is the vertical gaze nystagmus (VGN) exam, which tests the same factors as above, but by tracking the eyes

in a vertical path. Third, there is the lack of convergence (LOC) exam, which checks how the driver's eyes move together by tracking their coordinated convergence when the DRE moves a finger or penlight towards the driver's nose (a point of convergence) until one or both eyes drift outward toward the side instead.

Several of the State's testifying experts asserted that the eye examinations can reveal whether the driver has ingested drugs that may detrimentally affect such things as the driver's visual acuity, contrast sensitivity, glare sensitivity, ability to track objects, coordination, and other driving-related skills.

Step 5: The divided attention tests in Step 5 are comprised of four psychophysical tests, including two of the three standard field sobriety tests (SFSTs)⁸ developed to detect drunk driving.

The tests in this step consist of (1) the modified Romberg balance test, which tests balance by requiring the driver to stand feet together, head tilted back, with eyes closed and asked to estimate when thirty seconds has passed; (2) the walk-and-turn test, which tests a driver's ability to balance while

⁸ The three SFSTs were deemed the most accurate tests for determining alcohol-caused impairment. The two psychophysical tests in this step that are part of the three-test SFST battery are the walk-and-turn and one-leg-stand tests. The third SFST is the HGN test, which is an eye exam covered previously in Step 4.

standing heel-to-toe with the driver's arms at the side and while the driver takes heel-to-toe steps along a straight line pivoting to turn back to the starting point; (3) the one-leg-stand test, which tests the driver's balance on one leg with the other leg raised and arms at the side, all while the driver counts out loud while looking at the raised foot (done twice, once standing on each foot); and (4) the finger-to-nose test, which involves the driver putting a fingertip to the tip of the nose, while the driver's eyes are closed and head is tilted back.

DREs are familiar with the components of Step 5 from prerequisite training.⁹ The rationale for performing the tests in drugged driving cases is that any drug that impairs driving ability will also impair the driver's ability to perform divided attention tests, which help evaluate a driver's "psychomotor" skills. Among other things, divided attention deficits may impair a driver's ability to maintain lane position on a roadway while monitoring the surrounding environment.

Step 6: In this step, a second examination and vital signs check adds another data point to the DRE's evaluation. Although repetitive of Step 3, Step 6 is undertaken because the effects of drugs on the body may arise at any

⁹ These psychophysical field sobriety tests were all developed to detect drunk driving. The most robust training that DREs receive in administering them is during their DWI Detection and SFST and Advanced Roadside Impaired Driving Enforcement training sessions, which are prerequisites to being admitted to the DRE training program.

time during the overall DRE protocol. The training materials emphasize that blood pressure, pulse, and internal body temperature are “reliable indicators of drug influence” that DREs should measure.

Step 7: This step entails a dark room examination. While in the dark room, the DRE will measure the dilation of the driver’s pupil in response to a light stimulus in three different lighting conditions. The examination at this step involves changing the amount of light entering the driver’s eyes using a penlight so that the DRE can observe the pupil’s appearance and reaction to the light.

The training materials note that some drugs (such as marijuana) may cause the pupils to widen, or dilate, and others (such as opioids) may cause the pupils to narrow or constrict. The DRE will also examine areas of the body where drugs are commonly ingested (nasal area and oral cavity) for signs of drug use using a penlight.

Step 8: In this step, muscle tone examinations measure whether the muscles in the driver’s arms are tense or, conversely, flaccid. The training materials highlight that certain drugs tend to cause rigidity in the muscle, while others are known to cause a “rubbery-like” flaccidity in the muscle. Although this examination is sequenced at Step 8, the training materials also inform DREs that muscle tone can be observed at many points of the

examination, including when taking the driver's vital signs in Steps 3, 6, or 9, when checking for injection sites in Step 9, or during the Step 5 divided attention tests.

Step 9: The examination for injection sites in Step 9 enables the DREs to observe any indications that the driver has injected drugs through hypodermic needles, most commonly associated with heroin use. Injection of certain drugs can cause lengthy scars, called "tracks," and produce observable sores or bruising. The injection sites typically examined appear on the driver's neck, forearms, wrists, and backs of the hands.

Step 10: At this step, the DRE interviews the driver as a means to confirm or dispel any reasonable suspicions or opinions the DRE may have about the driver's impairment being caused by drug use, and any category of drug that the indicators point towards. The training materials reflect that the interview "can proceed only in conformance with formal admonition and strict observance of the driver's Constitutional rights." The materials also note that the interview procedures vary with the DRE's suspicions and opinions about the potential drug categories involved.

Throughout the examination, including this step, the DRE has been recording the information gathered on a drug influence evaluation form, known as a facesheet. DREs are also trained to record all spontaneous

statements and any responses by the driver, and to ask follow-up questions, as appropriate, at any step.

Step 11: After conducting the first ten steps of the protocol, the DRE reaches an opinion about whether the driver is under the influence of a drug or drugs, and if so, the probable category or categories of drugs that are causing the impairment. The opinion is “[b]ased on all of the evidence and observations gleaned from the preceding steps.” The DRE consults with a printed and standardized matrix that provides a reference tool for matching the symptoms and indicators observed to seven categories of drugs: CNS depressants, CNS stimulants, hallucinogens, narcotic analgesics, dissociative anesthetics, inhalants, and cannabis. For each category, there is a grid with a list of indicators and measurements that can be observed.

The DRE matrix disclaims that indicators listed on the matrix are only those that “are the most consistent with the category” but “that there may be variations due to individual reaction, dose taken and drug interactions.” The matrix also notes that normal measurements and observations refer to the population averages. The training materials additionally provide that the DRE must record a narrative summary of the facts that form the basis for the opinion.

Relatedly, DREs are trained to identify polydrug use, or the ingestion of two or more drugs or a combination of drugs and alcohol. The training materials instruct that “[i]t is actually more common for a [DRE] to encounter polydrug users than single drug users.” For example, the materials identify marijuana and alcohol as the most common polydrug mix and note that alcohol was often found in combination with one or more drugs.¹⁰

The training materials also discuss four effects of drug combinations on clinical indicators of drug use that might impact the DRE’s use of the matrix: the null effect, the overlapping effect, the additive effect, and the antagonistic effect.

The null effect denotes a situation in which neither drug in the combination affects a particular indicator of drug use, such that the combination will also not affect that indicator.

The overlapping effect refers to a situation in which one drug does not affect an indicator of drug use, but another drug does, and so the effect of the latter drug appears.

¹⁰ Other common combinations listed in the materials include cocaine and cannabis, cocaine and heroin, PCP and cannabis, PCP and heroin, crack and PCP, and Xanax and methamphetamine. Many of the combinations have common street names that the DREs are trained to recognize -- which may be of particular significance in understanding any potential admissions made by drivers.

The additive effect describes the result of two drugs that independently affect an indicator in the same way, such that the effect may be reinforced and appear more obviously or to a higher degree.

Lastly, the antagonistic effect describes the result of two drugs that oppositely affect an indicator. In those cases, the DRE training materials instruct that the two drugs “tend to try to override or compete” and that the result is “unpredictable” -- typically, the drug which is more psychoactive at the time determines which effect a DRE will observe, and usually that means the drug with the longest duration of effect on the indicator is observable. The DRE materials also instruct that the indicators of HGN, VGN, LOC, and reaction of the eyes to the light will not show any antagonistic effect.

The DREs are provided with some examples of drug combinations and their effect on certain indicators. They are also instructed that they will receive additional examples through the model DRE evaluations that they study in their training. For instance, the DRE materials provide the example of the polydrug combination of CNS stimulants and CNS depressants, which has an antagonistic effect on pulse rate.

Step 12: At this stage, after the DRE reaches an opinion, a toxicological sample is requested from the driver, if the driver did not already provide a specimen prior to the DRE’s arrival or during the examination (e.g., if a

bathroom break was necessary). Drivers have the right to refuse to provide a specimen, and the training materials instruct DREs to not allow that refusal to affect the evaluation or opinion reached. The materials also note that DREs should follow the departmental policies on sample collection. Often the arresting officer is the person who collects and retains the sample to be sent for toxicological testing.

DREs are instructed that “[t]he toxicological examination is a chemical test or tests designed to obtain scientific, admissible evidence to support the DRE’s opinion.” As several of the State’s witnesses attested, the toxicology exam does not, by itself, establish that the driver has been impaired by drug usage. Instead, the toxicology exam evidences only the presence of drugs in the driver’s body and might thereby “corroborate” the DRE’s clinical findings. By the same token, a negative toxicology report does not necessarily mean the driver was not impaired while behind the wheel. As we discuss further in this opinion, drugs can dissipate within the body before a sample is taken.

Toxicology labs cannot test for all drugs.¹¹ Some people may be impaired by the presence of drugs below a toxicology lab’s numerical cutoff levels.

¹¹ Among other things, the record indicates that (1) there are no tests available for certain designer drugs and synthetic cannabinoids because those drugs are continually being developed; (2) the State laboratory began testing for fentanyl only in 2019, and the test is often not informative because fentanyl can be

2. Development of the DRE Protocol

The DECP and DRE protocol was created in the 1970s by the Los Angeles Police Department (LAPD) with the assistance of the NHTSA. The protocols were designed to combat a growing problem of “drugged driving” which more easily evaded law enforcement detection than drunk driving. SM Report I 76-79, 85. As drunk driving was also the subject of growing public concern, the NHTSA made efforts to research reliable methods of testing sobriety, including the three-test battery that comprises the SFSTs in the DRE protocol. Id. at 79-86.

After two key studies (the 1985 “Bigelow” study and the 1986 “Compton” study)¹² evaluating the reliability of the entire DRE protocol were conducted in the 1980s, the NHTSA and LAPD, in consultation with doctors, toxicologists, and emergency nurses, among other professionals, developed a curriculum to train DREs. Id. at 86-90. Those efforts culminated in the first symptomatology matrix to serve as a reference tool for DREs in rendering their opinions. Id. at 91.

toxic at low concentrations; and (3) the State laboratory also does not test for MDMA (ecstasy) or LSD because it is ingested in such small quantities.

¹² See G.E. Bigelow et al., Identifying Types of Drug Intoxication: A Laboratory Evaluation of Subject-Examination Procedures (1985); Richard Compton, Field Evaluation of the Los Angeles Police Department Drug Detection Procedure (1986).

In 1987, the International Association of Chiefs of Police (IACP) began to participate in the development and national expansion of the DECP and, at NHTSA's request, to oversee the credentialing of DREs. Id. at 91-92.

In 1988, again at the NHTSA's request, the IACP established the Technical Advisory Panel (TAP), which develops criteria for training and certifying DREs, and continually improves the DRE protocol. Ibid. The TAP typically consists of a physician, a behavioral optometrist, and a toxicologist, as well as DREs and educational institutions. Ibid.

The DECP expanded outside of Los Angeles in 1987 and began in New Jersey in 1991. Id. at 93. As of December 2022, the IACP has certified over 400 DREs in New Jersey, the second most of any state in the nation.¹³ Presently, there are over sixty certified DRE instructors in this State who train officers in the protocol.

B. Defendant's DRE-Based Convictions and Ensuing Appeals

The Special Master's appointment was prompted by the prosecutions of defendant Michael Olenowski, who was twice subjected to the DRE protocol, for suspected drug-impaired driving.

¹³ DEC Program States, <https://www.theiacp.org/states-and-countries-with-dres>.

In February 2015, Olenowski was pulled over by a municipal police officer for not wearing a seatbelt. The officer conducting the stop detected the odor of alcohol, triggering field sobriety tests and a full DRE examination at police headquarters by a certified DRE. The DRE concluded that Olenowski “[wa]s under the influence of a CNS Depressant, CNS Stimulant and Alcohol.”

A separate incident involving Olenowski occurred in August 2015. Police officers were dispatched to the scene of an accident in which a car, driven by Olenowski, had run off the road, striking a telephone pole and sustaining “significant” damage. The responding officers detected signs of possible impairment, including slurred speech, balance trouble, and a lack of responsiveness. After speaking with the officers about the circumstances of the accident, Olenowski agreed to perform field sobriety tests. Eventually, he was arrested and transported back to headquarters where a different certified DRE conducted the protocol. The DRE concluded that Olenowski “[wa]s under the influence of a CNS Depressant, as well as a CNS Stimulant.”

At each of the municipal court trials for the two DUID charges, the prosecutor introduced DRE testimony to prove that Olenowski had been driving while under the influence of a central nervous system depressant and stimulant. After a Frye hearing during the first trial, the municipal court admitted the DRE testimony as reliable but acknowledged that Olenowski had

presented “impressive” evidence that it should not be accepted. Olenowski had called his own expert in rebuttal to criticize the DRE protocol, and he relied on that testimony in both trials. In both cases, the municipal court convicted Olenowski.

The Law Division upheld the admissibility of DRE evidence under Frye and affirmed each of the convictions after a de novo trial. The Appellate Division affirmed the convictions, finding they were supported by sufficient evidence even without the DRE testimony, but also holding it appropriate for the municipal court and Law Division to rely on the DRE evidence and agreeing with the Law Division analysis that DRE evidence was generally accepted under Frye.

C. This Court’s Review and the Special Master Proceedings

This Court first granted certification in this case to determine whether DRE testimony is admissible under the Frye “general acceptance” admissibility standard. 236 N.J. 622 (2019). We heard oral argument in October 2019, after which we concluded in an order that “the existing factual record [wa]s inadequate to test the validity of DRE evidence.” 247 N.J. 242, 244 (2019). We then designated the Honorable Joseph F. Lisa, a retired Presiding Judge of the Appellate Division on recall, as Special Master. Ibid. We asked him through our order of appointment to conduct “a plenary hearing

to consider and decide whether DRE evidence has achieved general acceptance within the relevant scientific community and therefore satisfies the reliability standard of N.J.R.E. 702.” Ibid.

1. Special Master Hearing and SM Report I: Admissibility Under the Frye Standard

The Special Master conducted an extensive Frye hearing over forty-two days, including hearing testimony from sixteen witnesses from both parties and amici.¹⁴ Hundreds of exhibits were admitted during the hearing, creating the fulsome record now before us in this appeal.

Part of the discovery process leading up to that hearing included the Public Defender’s request for the State to produce statewide records of all DRE cases for statistical review. That New Jersey “retrospective data” spanning from 2017 to 2018 plays a significant role in the resolution of this appeal. The collected data encompassed 5,855 DRE reports. Of that total, only 2,551 were non-training cases that included a toxicology report for corroboration of the DRE conclusion. That particular segment of the data was

¹⁴ We commend the Special Master, counsel, and the court staff for their extraordinary cooperative and dedicated efforts in taking part in these proceedings, at times remotely or with social distancing measures, for two years from January 2020 to January 2022 through the peak of the COVID-19 pandemic.

at the forefront of the case. In about 27% of the 5,855 cases, there was no toxicology report for various reasons.

During the early stages of the proceedings before the Special Master, Olenowski passed away. At that point, all parties agreed that the proceedings to develop a record for the Court concerning DRE admissibility should nonetheless proceed, given the public importance of the issue. This Court instructed the Special Master to continue with the hearing despite the apparent mootness of Olenowski's appeal. See Malanga v. Township of West Orange, 253 N.J. 291, 307 (2023) (quoting Redd v. Bowman, 223 N.J. 87, 104 (2015)); State v. Cassidy, 235 N.J. 482, 491 (2018). Later, Olenowski's personal attorney discontinued his participation in the hearings. The Public Defender maintained the lead role as advocate for the defense throughout the hearing, as a prerequisite to hiring experts in the case.

In August 2022, the Special Master released a 332-page report concluding that DRE evidence should be admissible under the Frye standard of reliability of expert testimony. SM Report I 331.

After his review of the extensive record, the Special Master concluded that the DRE protocol and its various components were generally accepted in the relevant scientific communities (specifically, medicine and toxicology) because (1) "the DRE protocol replicates generally accepted medical practices

for identifying the presence of impairing drugs and their likely identity”; (2) “the DRE matrix comports with matrices . . . generally accepted and used in the medical field”; and (3) “the training DREs receive is comparable to that received by medical technicians.” Ibid.

In the course of his decision, the Special Master expressly commented on the credibility of each of the sixteen witnesses who testified. In general, he regarded the State’s experts as more credible than the defense witnesses, although he recognized that many of the witnesses on both sides had biases in favor of or opposed to the DRE program. The Special Master explained in detail why he considered the testimony of each witness especially credible or less credible. See id. at 20-76.¹⁵

The Special Master summarized his initial findings about the reliability of individual components of the protocol as follows. First, he reiterated that DRE testimony implicates two aspects of expertise: (1) “specialized knowledge that DREs acquire that enable[s] them to reliably administer the tests and make the observations and gather the information required by the DRE protocol and . . . to determine whether the driver is impaired by drugs

¹⁵ Significant portions of the expert witnesses’ testimony will be explored later in Part V. The Special Master also supported his credibility findings with more detailed information about the witnesses and the substance of their testimony. See SM Report I at 20-76.

and if so by which category or categories in the DRE matrix”; and (2) “scientific expertise” underlying “[t]he validity of the DRE matrix and the procedures and methods for applying it.” Id. at 307-08.

Second, the Special Master determined that “the appropriate scientific communities are medicine and toxicology,” not the traffic-safety research community. Id. at 309-10. He concluded that those communities -- although unfamiliar with the DRE protocol -- have “impliedly generally accepted” the protocol because “it is in all material respects the same as [their protocol used in emergency medicine for toxidrome recognition], including the level of training required.”¹⁶ Id. at 310.

The Special Master further noted that Steps 1 and 12 of the protocol are “clearly scientific in their entirety” and that some steps are “scientifically based” (e.g., checking pulse and vital signs, eye examinations), while other steps are “clearly not scientific” (e.g., checking for injection marks and drug paraphernalia, interrogating the driver and others). Id. at 310-15. He concluded that all the steps were reliable. Ibid.

¹⁶ A “toxidrome” is another term for a “toxic syndrome,” which refers to a combination of findings and symptoms that is “suggestive of a diagnosis” of a condition caused by a toxin, like a drug. SM Report I 128-30. Toxidrome recognition is “‘widely used in most medical specialties,’ including emergency medicine in particular.” Id. at 129.

Third, the Special Master found that a DRE “would never form an opinion that would be accepted as reliable based upon any one or even a few isolated factors.” Id. at 315. Rather, “[a]ll of the observations must be taken into consideration and assessed together.” Ibid. In so finding, he also concluded that toxicological testing -- Step 12 of the protocol -- should not be a prerequisite to the admission of a DRE opinion because “toxicology is not considered by the DREs and plays no role in forming their opinions.” Id. at 318. Instead, he wrote that toxicology is merely “another piece of evidence for the factfinder that corroborates, or fails to corroborate, the DRE opinion -- potentially affecting the weight accorded [to] the opinion but not affecting its admissibility.” Ibid.

In addition, the Special Master identified several limitations of the DRE protocol, including (1) the various scientific limitations associated with toxicological testing and (2) situations in which the driver refuses to provide a urine sample. Id. at 315, 317. He found that those limitations were just other factors for the factfinder to consider in a particular case, rather than an indictment of the entire protocol’s reliability. He noted the DRE evidence would be subject to credibility assessments and weight allocations as the factfinder deems appropriate. Id. at 319, 331.

2. Olenowski I: Adoption of the Daubert Admissibility Standard

Following the first report, we accepted additional briefing from the parties and amici. In those briefs, as in submissions to the Special Master before the release of his first report, several counsel focused upon error rates associated with DRE evidence. Because error rates are expressly considered under Daubert, but not Frye, we asked the parties and amici to submit supplemental briefing on “whether this Court should depart from Frye and adopt the principles of Daubert in criminal cases.” Both parties and nearly all of the amici advocated that we adopt the Daubert standard, similar to our previous adoption of Daubert-based principles for civil cases in In re Accutane Litigation, 234 N.J. 340 (2018).

In our first opinion dated February 17, 2023, we prospectively adopted a “Daubert-type standard” for determining the reliability of expert evidence in criminal and quasi-criminal cases. Olenowski I, 253 N.J. at 153. Identifying a number of difficulties with maintaining the Frye approach, we concluded that “Daubert offers a superior approach to evaluate the reliability of expert testimony.” Id. at 139, 150-52. The Daubert analysis involves multiple factors, which we discuss in depth below.

We consequently remanded this matter to the Special Master to apply the Daubert-type standard, directing that, in his discretion, he could “rule on the

basis of the existing record, or ask for and accept additional evidence, briefing, and argument from the parties and amici.” Id. at 155.

After a case management conference at which all counsel agreed the record was complete, the Special Master ordered that the record would not be reopened. He further ordered the parties, and permitted amici, to “file supplemental briefs regarding the application of Daubert principles to the evidence presented in the Special Master proceeding.”

3. SM Report II: Applying the Daubert Standard

After briefing, the Special Master released on April 13, 2023 a 57-page Supplemental Report of Findings of Fact and Conclusions of Law (SM Report II). In that second report, the Special Master concluded that the State “clearly established that the DECP and the twelve-step DRE protocol satisfy the reliability standard of N.J.R.E. 702 when analyzed under the methodology-based Daubert-Accutane standard.” Id. at 56. He further concluded that “DREs can be and are adequately trained to reliably perform the steps in the protocol,” thereby satisfying N.J.R.E. 702’s requirement that the witness have “sufficient expertise” to offer the testimony. Id. at 6, 57. The Special Master’s Daubert analysis largely cross-referenced the analysis in his earlier report, and he directed at the outset of the second report that it “must be read in conjunction” with the first. Id. at 1, 55.

The Special Master analyzed each of the Supreme Court’s Daubert factors as follows. First, he found that the DRE protocol had been “studied extensively over a number of decades.” Id. at 22. He concluded that “Daubert factor one provides substantial support for [his] finding of reliability of the DRE protocol.” Id. at 26. The Special Master rejected the Public Defender’s argument that the DRE protocol had never been tested for its ability to detect drug-impaired drivers because it had only been tested as to whether it could identify drug presence in drivers (i.e., tested against toxicology results), but not drug-induced impairment. Notwithstanding that distinction, the Special Master was satisfied that sufficient relevant testing had supported the reliability of DRE opinion testimony.

As to the second Daubert factor, the Special Master concluded that many of the key studies on the reliability of the DRE protocol -- which he found to be supportive of its reliability -- had been published in peer reviewed journals. Id. at 26-28. He also found that earlier studies -- although published by NHTSA rather than a peer reviewed journal -- had been “reviewed by other scientists as part of the internal agency review process before publication.” Id. at 27. Thus, he concluded that the second Daubert factor “provides substantial support” for his reliability finding. Id. at 29.

Acknowledging that the third Daubert factor contains two components -- (1) the known or potential rate of error, and (2) the existence of standards governing the operation of the particular scientific technique -- the Special Master concluded that both components weighed in favor of reliability. Id. at 29, 38-39.

As to the error rate component, the Special Master made several observations. He noted that the alleged false positives -- which comprise instances in which a DRE opined impairment, but the toxicology result was negative -- were not necessarily “errors,” due to the limitations of toxicological testing. Id. at 30. He found that a negative toxicology result does not prove that the DRE opinion was “wrong.” Ibid. He concluded that “in this context, error rates . . . can at best be described as a conservative metric. . . . The error rate might actually be lower.” Id. at 32.

Further, the Special Master highlighted the apparent low error rates and high accuracy rates reflected in three key peer reviewed studies, known as the Beirness/Canada, Vaillancourt, and Hardin studies.¹⁷ Id. at 34. Regarding the

¹⁷ See Douglas Beirness et al., The Accuracy of Evaluations By Drug Recognition Experts in Canada, 42 Can. Soc. Forensic Sci. J. 75 (2009); Lucie Vaillancourt et al., Drugs and Driving Prior to Cannabis Legalization: A Five-Year Review from DECP (DRE) Cases in the Province of Quebec, Canada, 149 Accident Analysis & Prevention (2021); Glenn G. Hardin et al., Minn. Dep’t of Pub. Safety, Minnesota Corroboration Study: DRE Opinions and Toxicology Evaluations (Apr. 1993).

New Jersey retrospective DRE data, the Special Master acknowledged that a false positive rate could not be reliably calculated due to the small number of cases that produced negative toxicology results. Id. at 35. However, he concluded that the “low number of false positives is a favorable factor in supporting the reliability of the process.” Ibid. He considered that the low number of false positives might result from the high-prevalence population of drivers evaluated, namely drivers who both “(1) displayed affirmative signs of impairment sufficient to provide probable cause to arrest for driving under the influence, and (2) have a BAC level . . . that shows either no or limited alcohol consumption.” SM Report I at 273; see also SM Report II at 35-36. However, he deemed that concern overblown. SM Report II at 35-36. The Special Master also recognized that although testing the general driving population might allow for the calculation of a reliable false positive rate, subjecting all drivers to such testing -- including those who could not be arrested for drunk driving for lack of probable cause -- would trigger constitutional and practical constraints. SM Report I at 272-73; SM Report II at 27, 36.

As for the latter component of the third Daubert factor -- standardization -- the Special Master found the history of the DECP and DRE protocol and the rigorous training requirements for DRE certification significant. SM Report II

at 38. He also observed that the Public Defender's brief had not focused its arguments upon this particular component of the third factor. Ibid.

Overall, the Special Master determined that both components of the third Daubert factor supported and corroborated his finding of reliability, subject to the limitations of the toxicology in the protocol. Id. at 38-39.

Finally, the Special Master concluded that the fourth Daubert factor -- general acceptance -- provides "substantial support" for his reliability determination. Id. at 54. He reiterated his finding from the first report that the DRE protocol had been generally accepted by the medical and toxicological communities by implication. See id. at 39, 41-42, 53. He specifically found that the "methodology" relied on by DREs is generally accepted within those professional communities. Id. at 53. As he stated, "it is the methodology that is dispositive, rather than knowledge of the overall DRE protocol." Id. at 53-54.

The Special Master rejected the Public Defender's argument that there could not be implied general acceptance because the DRE protocol does not adhere to the differential diagnosis process, which the Public Defender characterized as "the only reliable method used in the medical and toxicological fields to determine drug impairment as the cause of observed signs and symptoms." Id. at 40-41, 49-50. He explained that the context in

which a differential diagnosis typically occurs (i.e., an individual in significant distress seeks immediate medical treatment from an emergency physician) differs from the context in which a DRE evaluation occurs (i.e., an individual displays affirmative signs of impaired driving; is stopped, observed, questioned, and searched by the police; is discovered to have a “zero BAC or very low BAC that is inconsistent with the driver’s behavior”; and is then arrested and taken to the stationhouse for a DRE evaluation). Id. at 42-43. The primary difference is that the driver who is subject to the DRE protocol is not ordinarily seeking medical attention. Id. at 44-45. Hence, the driver’s condition “is typically sufficiently differentiated to point the DRE in the direction of probable drug use as the cause of impairment.” Ibid.

The Special Master further noted that DREs, although they are not physicians or medical professionals, are trained to be familiar with potential non-drug causes of impairment (e.g., bipolar disorder, diabetes, head trauma, seizures) and required to ask questions about the driver’s medical conditions and prescription medications. Id. at 43-44.

III. Arguments of the Parties and Amici

Following the release of SM Report II, the Court ordered the parties and amici to submit briefs addressing the report and specifically requested that the briefs also “address the appropriate standard of review for appeals from a

determination under the Daubert standard.” Those briefs, and the second oral argument that followed, focused on both the standard of review and the merits of the Special Master’s analysis of the Daubert factors.

Briefly summarized, the Public Defender and the associated amici¹⁸ essentially contend that a de novo standard of appellate review should apply to a Daubert-based determination of reliability in New Jersey criminal cases. In opposition, the State and the prosecution-aligned amici¹⁹ contend that the standard of review should be whether the reliability determination below was an abuse of discretion.

As for the merits, the Public Defender and the defense-aligned amici maintain that the Special Master erred in finding the DRE testimony is sufficiently reliable under the Daubert factors and urge that it instead must be excluded in all DUID prosecutions. As a fallback position, the Public Defender contends that, if we are unpersuaded by their arguments for complete

¹⁸ The defense-aligned amici curiae are the New Jersey State Bar Association; the National College for DUI Defense; the Association of Criminal Defense Lawyers of New Jersey; the DUI Defense Lawyers Association; and the American Civil Liberties Union of New Jersey, joined by three statistics experts.

¹⁹ The prosecution-aligned amici curiae are the County Prosecutors Association of New Jersey and the New Jersey State Association of Chiefs of Police. The Attorney General had previously been amicus curiae in the appeal but replaced the Morris County Prosecutors Office in representing the State.

exclusion, we should allow DRE testimony to be admitted only with various strict limitations, several of which we will discuss in Part VI. In response, the State and the prosecution-aligned amici contend that the Special Master correctly applied the Daubert factors and that we should therefore affirm his findings of reliability and admissibility without imposing any constraints.

IV. Standard of Review of Expert Reliability Rulings

In our earlier decision in this case adopting Daubert factors to evaluate the admissibility of expert testimony in New Jersey criminal cases, we did not address the novel question of what standards should guide the appellate review of such Daubert-based rulings, reserving that for a later day. That day has now arrived. We have the benefit of additional briefing of the parties, as well as the specific context of this appeal as an illustrative opportunity to apply those review standards.

We preface our resolution of this issue with a short discussion of our past traditions. Most evidentiary rulings by New Jersey trial judges have been reviewed on appeal by considering whether the judges abused their discretion in admitting or excluding proofs. Such rulings are generally upheld “unless the evidentiary ruling is ‘so wide of the mark’ that it constitutes ‘a clear error in judgment.’” State v. Allen, 254 N.J. 530, 543 (2023) (quoting State v. Garcia, 245 N.J. 412, 430 (2021)). That highly deferential general standard is

subject to the appellate court’s obligation to provide relief in situations in which the trial court’s decision was “clearly capable of producing an unjust result.” R. 2:10-2.

As noted, before our February decision in this case, the Frye standard of general acceptance guided the admissibility of scientific or other expert opinion testimony in New Jersey criminal cases. 253 N.J. at 143. But even on appeal of such Frye-based rulings in criminal cases, our courts traditionally did not apply an unqualified abuse of discretion standard of review. Instead, our case law recognized that a trial court’s reliability determination under Frye -- i.e., its determination of whether the relevant scientific community generally accepts a scientific theory, test, or technique -- ought to receive less deferential review than other evidentiary decisions. See State v. Harvey, 151 N.J. 117, 167-68 (1997).

As we stated in Harvey, “[t]o the extent that [reliability] focuses on issues other than a witness’s credibility or qualifications, deference to the trial court is less appropriate.” 151 N.J. at 167. In Harvey, we acknowledged that appellate courts generally review a trial court’s admissibility determinations for an abuse of discretion. Id. at 166. However, we concluded that this usual standard of review may not be appropriate for Frye reliability determinations for two reasons.

First, “[u]nlike many other evidentiary issues, whether the scientific community generally accepts a methodology or test can transcend a particular dispute.” Id. at 167. “In determining the general acceptance of novel scientific evidence in one case, the court generally will establish the acceptance of that evidence in other cases.” Ibid.

Second, “[l]ike trial courts, appellate courts can digest expert testimony as well as review scientific literature, judicial decisions, and other authorities.” Ibid. Appellate courts routinely “scrutinize the record” in appeals and can “independently review the relevant authorities.” Ibid.

We reiterated the propriety of such a less deferential approach in a later criminal appeal, State v. Torres, observing that “the appellate court need not be as deferential to the trial court’s ruling on the admissibility of expert scientific evidence as it should be with the admissibility of other forms of evidence.” 183 N.J. 554, 567 (2005).

More recently, we went further in State v. J.L.G. and instructed that Frye reliability determinations in criminal cases should be reviewed de novo. 234 N.J. 265, 301 (2018) (“Whether expert testimony is sufficiently reliable to be admissible under N.J.R.E. 702 is a legal question we review de novo.”).

Our tradition in civil cases has been different. In Accutane, we held that trial courts’ expert reliability determinations in civil matters are to be reviewed

under the abuse of discretion standard. 234 N.J. at 392. Our opinion in Accutane pointed out that none of the key civil cases²⁰ applying a methodology-based approach such as the Daubert reliability factors “[spoke] to any such less-deferential standard.” Id. at 391.

The Court’s adoption in Accutane of an abuse of discretion review standard for Daubert rulings in civil appeals coincides with the standard used in federal appeals prescribed by the United States Supreme Court in General Electric Co. v. Joiner. See 522 U.S. 136, 139 (1997). The Supreme Court provided limited reasoning for choosing this standard of review. See id. at 141-42. Noting that the change from Frye to Daubert did not fundamentally alter the gatekeeping role of the trial court in admitting expert testimony, the Court announced “that abuse of discretion is the proper standard of review of a district court’s evidentiary rulings.” Id. at 142.

Joiner did not explain why Daubert reliability determinations should be reviewed under the same standard of review as other types of evidentiary decisions. In fact, the question was not squarely litigated before the Court, as the parties in Joiner agreed that abuse of discretion was the appropriate

²⁰ See Rubanick v. Witco Chem. Corp., 125 N.J. 421 (1991); Landrigan v. Celotex Corp., 127 N.J. 404 (1992); Kemp ex rel. Wright v. State, 174 N.J. 412 (2002).

standard of review. Id. at 141. They differed only about whether a more robust form of abuse of discretion review was appropriate in the case. Ibid.

Other Jurisdictions

Some federal circuit courts have placed a gloss on Joiner's abuse of discretion approach.²¹ For instance, the Seventh Circuit uses a “two-step standard of review.” C.W. ex rel. Wood v. Textron, Inc., 807 F.3d 827, 835 (7th Cir. 2015). First, it “review[s] de novo a district court’s application of the Daubert framework.” Ibid. Second, if it determines that “the district court properly adhered to the Daubert framework,” it “review[s] [the district court’s] decision to exclude (or not to exclude) expert testimony for abuse of discretion.” Ibid.

In Textron, the Seventh Circuit’s de novo review of the district court’s application of Daubert delved into the details and design choices of the studies underlying the expert opinion. See id. at 836. Then in the second step, the circuit court found that it was not an abuse of discretion to exclude an expert when there was “too great an analytical gap between the data and opinion

²¹ Legal scholarship has spotlighted how the federal circuits have differed in the level of scrutiny applied to trial court Daubert determinations. See, e.g., Sean Ryan, Backfire: Abandoning the Abuse of Discretion Standard of Review for Daubert Rulings Shoots Trial Courts in the Foot, 47 U. Tol. L. Rev. 349, 365-67 (2016); Victor E. Schwartz & Cary Silverman, The Draining of Daubert and the Recidivism of Junk Science in Federal and State Courts, 35 Hofstra L. Rev. 217, 262-66 (2006).

proffered’ such that the opinion amounts to nothing more than the ipse dixit of the expert.” Id. at 837 (quoting Joiner, 522 U.S. at 146).

The Tenth Circuit, by comparison, engages in a more deferential form of a two-step standard of review. As the court explained in Norris v. Baxter Healthcare Corp., it first “review[s] de novo whether the district court applied the proper standard in determining whether to admit or exclude expert testimony -- that is, whether the district court properly performed its role as ‘gatekeeper’ pursuant to Federal Rule of Evidence 702 and Daubert.” 397 F.3d 878, 883 (10th Cir. 2005) (citation omitted). Second, the Tenth Circuit “then review[s] the manner in which the district court ‘exercises its Daubert “gatekeeping” role in making decisions whether to admit or exclude testimony’ for an abuse of discretion.” Ibid. (quoting Bitler v. A.O. Smith Corp., 391 F.3d 1114, 1119 (10th Cir. 2004)).²²

Conversely, several states have rejected an abuse of discretion review standard for Daubert-based reliability determinations. Most notably, in State v. Sharpe, the Alaska Supreme Court adopted a hybrid standard of review in

²² More specifically, at the first step, the Tenth Circuit considers whether the Daubert test was indeed applied and is “not necessarily concerned with . . . ‘exact conclusions reached to exclude or admit expert testimony.’” Ibid. (quoting Bitler, 391 F.3d at 1119). The second step analyzes the reliability determination, recognizing the trial court’s “wide discretion both in deciding how to assess an expert’s reliability and in making a determination of that reliability.” Ibid. (quoting Bitler, 391 F.3d at 1120).

the context of a criminal appeal. 435 P.3d 887, 889 (Alaska 2019). It held that the trial court’s preliminary factual determinations should be reviewed in a deferential manner for clear error. Ibid. However, the trial court’s eventual decision on whether the scientific theory or technique is sufficiently reliable under Daubert “is a question of law to which [appellate courts should] apply [their] independent judgment.” Id. at 889-90, 900. Rulings on other “case-specific” requirements for the admissibility of expert evidence should be reviewed for an abuse of discretion. Id. at 900 n.89. Such rulings include whether the evidence is helpful to the trier of fact, and whether the relevant theory or technique can properly be applied to the facts in issue. Ibid.

In adopting that hybrid standard of review, the Sharpe Court overturned an earlier decision, State v. Coon, 974 P.2d 386 (Alaska 1999), which had adopted an abuse of discretion review standard for Daubert reliability determinations. 435 P.3d at 896, 899. The Sharpe Court relied heavily on the dissenting opinion in Coon, which similarly advocated for a hybrid standard of review. See Coon, 974 P.2d at 403 (Fabe, J., concurring in part and dissenting in part). That dissent drew a qualitative distinction between evidentiary rulings that are case-specific (such as relevance) and should be reviewed on appeal for abuse of discretion, and rulings that are not (such as the validity of scientific theories or techniques), which should be reviewed de novo. Ibid.

The dissent in Coon cautioned that an abuse of discretion review on Daubert reliability questions would lead to inconsistent rulings and thereby undermine predictability in the law and public confidence in the justice system. Id. at 404-05.

The Sharpe Court echoed the Coon dissent's concerns in emphasizing that abuse of discretion review "raises at least the appearance of arbitrariness, i.e., the appearance that the outcome of a Daubert determination . . . depends more on which judge was assigned to the case than on the objective application of law to the evidence presented." 435 P.3d at 898. Such an appearance, the Court highlighted, "has the potential to raise serious questions in the eyes of the public about the integrity of [the state's] judicial system." Ibid. Further, the Court found arbitrariness to be especially problematic "in the context of serious criminal proceedings." Ibid. (emphasis added).

Sharpe predicted that a less deferential standard of review "would allow trial courts and parties to avoid repeatedly relitigating the validity of scientific evidence, saving the court and parties the time, effort, and cost of a Daubert hearing -- at least absent new or previously overlooked research and evidence." Id. at 899. Additionally, Sharpe reasoned that appellate courts would often have more time than trial courts to "careful[ly] study . . .

secondary sources such as scientific treatises and surveys of academic literature in the relevant field.” Ibid.

Apart from Alaska, several other states apply a stricter standard of review than abuse of discretion to Daubert-based reliability determinations. See, e.g., Lee v. Martinez, 96 P.3d 291, 296 (N.M. 2004) (determining that both the special master’s legal conclusions and findings of fact were subject to de novo review); Taylor v. State, 889 P.2d 319, 331-32 (Okla. Crim. App. 1995) (holding that “independent” review “not limited by deference to the trial judge’s discretion” was appropriate for Daubert determinations); State v. Dahood, 814 A.2d 159, 161-62 (N.H. 2002) (noting that review of evidentiary determinations is generally deferential but that, “[w]hen the reliability or general acceptance of novel scientific evidence is not likely to vary according to the circumstances of a particular case, . . . we review that evidence independently”); State v. Beard, 461 S.E.2d 486, 492 n.5 (W. Va. 1995) (noting that the proper standard of review for a determination about the admissibility of scientific evidence is de novo); State v. Lyons, 924 P.2d 802, 805 (Or. 1996) (same); see also People v. Doolin, 198 P.3d 11, 53 (Cal. 2009) (noting that appellate courts defer to a trial judge’s factual and credibility findings, but decide admissibility as a “matter of law” based on those findings).

Other states, as we did in J.L.G., 234 N.J. at 301, review reliability determinations under the Frye standard de novo. See State v. Cauthron, 846 P.2d 502, 505 (Wash. 1993) (en banc) (determining that “the proper standard of review of the trial court’s decision [of admissibility under Frye] is de novo”), abrogated in other part by State v. Buckner, 941 P.2d 667 (Wash. 1997) (en banc); In re Commitment of Simons, 821 N.E.2d 1184, 1189 (Ill. 2004) (overturning a previous decision that had adopted abuse of discretion review); Goeb v. Tharaldson, 615 N.W.2d 800, 814 (Minn. 2000) (pointing out the potential for inconsistent rulings under abuse of discretion review); see also Brim v. State, 695 So. 2d 268, 274 (Fla. 1997).²³

We acknowledge that several other states apply an abuse of discretion standard of review to reliability determinations under Frye or Daubert. See, e.g., In re Costco Stormwater Discharge Permit, 151 A.3d 320, 331 (Vt. 2016); Rochkind v. Stevenson, 236 A.3d 630, 651 (Md. 2020); Thomas v. Lewis, 289 So. 3d 734, 738 (Miss. 2019); Walsh v. BASF Corp., 234 A.3d 446, 456 (Pa. 2020); Schafersman v. Agland Coop, 631 N.W.2d 862, 868, 871 (Neb. 2001); Commonwealth v. Mathews, 882 N.E.2d 833, 844 (Mass. 2008). The Supreme Court of Vermont, for example, has held that abuse of discretion review is

²³ Florida has since adopted the Daubert standard. See In re Amends. to the Fla. Evidence Code, 278 So. 3d 551, 551-52 (Fla. 2019). The Florida Supreme Court has not addressed the standard-of-review issue under Daubert.

appropriate in this context because a reliability determination “depends heavily on the record made in the trial court and the credibility of the expert witness presenting the disputed evidence.” USGen New England, Inc. v. Town of Rockingham, 862 A.2d 269, 277 (Vt. 2004).

Legal Scholarship

Commentators are divided on the standard of review issue. One commentator, writing in support of an abuse of discretion standard, contends that trial courts are better equipped to make reliability determinations than appellate courts, which are “less experienced in evidentiary determinations, removed from the heat of the moment of trial, and working with a cold trial record.” Ryan, 47 U. Tol. L. Rev. at 370.

Other commentators advocate for a stricter standard of review than abuse of discretion. For example, one scholar asserts that it is “inappropriate to view [the] threshold question of reliability as a matter within each trial judge’s individual discretion” because “the reliability of a scientific technique or process does not vary according to the circumstances of each case.” Paul C. Giannelli, The Admissibility of Novel Scientific Evidence: Frye v. United States, a Half-Century Later, 80 Colum. L. Rev. 1197, 1223 (1980) (quoting Reed v. State, 391 A.2d 364, 367 (Md. 1978)).

Several commentators suggest a hybrid approach, applying de novo review for reliability determinations, but abuse of discretion review for the application of that scientific methodology to the facts of a particular case.²⁴ “When the scientific evidence transcends the particular case, the appellate court should apply a ‘hard-look’ or de novo review to the basis for the expert opinion.” David L. Faigman, Appellate Review of Scientific Evidence Under Daubert and Joiner, 48 Hastings L.J. 969, 976 (1997). However, “[w]hen the scientific evidence involves facts specific to the particular case, the appellate court should defer to the trier of fact.” Ibid.

It has also been observed that appellate courts have more “time and distance to become familiar with . . . complex science” and that “appellate judges sit on panels and thus have the benefit of shared experience and expertise.” Id. at 979; see also Christopher B. Mueller, Daubert Asks the

²⁴ See, e.g., Amy B. Hargis & Joe R. Patranella, Rethinking Review: The Increasing Need for A Practical Standard of Review on Daubert Issues in Place of Joiner, 52 S. Tex. L. Rev. 409, 422, 424 (2011); David L. Faigman, Appellate Review of Scientific Evidence Under Daubert and Joiner, 48 Hastings L.J. 969, 976, 979 (1997); Developments in the Law -- Confronting the New Challenges of Scientific Evidence, 108 Harv. L. Rev. 1481, 1529 (1995); see also Christopher B. Mueller, Daubert Asks the Right Questions: Now Appellate Courts Should Help Find the Right Answers, 33 Seton Hall L. Rev. 987, 1019-22 (2003) (advocating for a more exacting standard that maintains a hybrid element in which “some degree of deference to the decision of the trial judge is in order” if “the admissibility decision actually focuses” on narrower, case-specific questions).

Right Questions: Now Appellate Courts Should Help Find the Right Answers, 33 Seton Hall L. Rev. 987, 1021 (2003) (agreeing that appellate courts are better situated for determining reliability because they generally have more time, more judges assigned to individual cases, and more thorough briefing). Professor Faigman further observes that, unlike other types of evidentiary rulings, reliability does not turn on witness credibility: “Good scientific research simply does not depend on the credibility of individual witness.” 48 Hastings L.J. at 978.

Our Adoption of a Hybrid Review Standard

As we have noted, the State and the prosecution-aligned amici have argued to us that an abuse of discretion standard of review should govern Daubert reliability rulings in New Jersey criminal and quasi-criminal cases, whereas the Public Defender and the defense amici have urged that we adopt de novo review of such reliability decisions.

Having duly considered those competing viewpoints, we unanimously adopt a hybrid standard of review, akin to the approach endorsed by the Alaska Supreme Court in Sharpe. Going forward, we hold that in New Jersey criminal and quasi-criminal cases in which the trial court has admitted or excluded an expert witness based upon Daubert reliability factors, our appellate courts shall review that reliability determination de novo. However, other case-specific

determinations about the expert evidence -- such as whether the witness has sufficient expertise, whether the evidence can assist the trier of fact in that case, and whether the relevant theory or technique can properly be applied to the facts in issue -- should be reviewed for an abuse of discretion.

We adopt the hybrid approach for several reasons. To begin with, it continues the tradition of our case law in criminal matters -- most recently expressed in J.L.G. -- to engage in more rigorous appellate review of the bona fides of an expert's methodology than under a deferential abuse of discretion standard. The shift we announced in February from a Frye general acceptance regime to a Daubert-based multi-factor regime in criminal and quasi-criminal cases does not warrant a departure from that tradition. There is no reason to weaken our appellate courts' oversight of the gatekeeping functions of criminal trial judges because of that shift.

In addition, the abundant reasons explained in Sharpe for adopting a hybrid review standard in criminal and quasi-criminal contexts are persuasive. The permissible methodologies of experts who are allowed to present their opinions in criminal and quasi-criminal prosecutions should not vary from case to case or from trial judge to trial judge.

Many categories of experts who testify frequently in criminal cases -- such as ballistics experts, fingerprint experts, DNA analysts, coroners,

serologists, toxicologists, accident reconstruction experts, cell tower experts, and so on -- use the same methodologies repetitively. They are called upon by prosecutors and defense counsel to testify with regularity. It would be dysfunctional to have the admissibility of their opinions depend upon how individual trial judges assess the reliability of their methodologies under the Daubert factors, based on varying presentations by varied counsel, and require appellate courts to defer to those varying and potentially conflicting rulings. The stability and fairness of the criminal justice system would be undermined by such uneven and unpredictable rulings. These concerns justify a more stringent and less deferential appellate review of the trial court's gatekeeping decision.

Moreover, employing an abuse of discretion standard for reviewing expert reliability determinations in criminal cases would consume time and money, particularly the publicly funded budgets of prosecutors and public defenders, who are the main litigators for such cases. With no assurance of statewide consistency, the reliability of a particular kind of expert methodology under Daubert could be relitigated over and over again in the criminal trial courts. The massive scale of the present record compiled before the Special Master through over forty days of hearings involving multiple lawyers and witnesses illustrates the point. Absent materially new or different

evidence, there is no need, nor the realistic ability, to repeat such a colossal undertaking in the courtrooms of individual trial judges confronted with DRE evidence.

We appreciate the enormity of this record and the efforts by the Special Master, the parties, and the many expert witnesses during the hearing to sift through the documentary exhibits presented. We particularly note the illuminating commentary and context provided at the hearing for the various published studies. In future expert admissibility disputes applying the Daubert factors, the parties -- preferably at the trial level -- should present all relevant scientific and technical evidence and published studies. Such presentations will enable appropriate witnesses to properly contextualize those materials, and testify about their significance or insignificance, for the trial court's and ultimately the appellate court's benefit.

We acknowledge that an appellate court's de novo ruling about the reliability of a certain kind of expert methodology should not be frozen in time. If new scientific research emerges that calls into question the wisdom of such precedent, then prosecutors and criminal defense lawyers should be free to present that new research to the trial courts, with appropriate testimony, and advocate for a change in the law. A hybrid review standard with a de novo component need not perpetuate obsolete scientific principles.

For those reasons, we prospectively adopt for criminal and quasi-criminal cases²⁵ the hybrid review standard used in Alaska and similarly followed in many other states. We now apply that standard to the Special Master's decision here concerning DREs.

V. De Novo Application of the Daubert Factors

Through the prism of de novo review, we proceed to apply the Daubert reliability factors to the record developed before the Special Master. As we noted in Olenowski I, the United States Supreme Court identified in Daubert a list of four factors for assessing reliability of an expert's methodology under Federal Rule of Evidence 702:²⁶

- (1) whether the scientific theory or technique can be, or has been, tested;
- (2) whether it “has been subjected to peer review and publication”;
- (3) “the known or potential rate of error” as well as the existence of standards governing the operation of the particular scientific technique; and

²⁵ Because the question is not before us, we do not address here whether a similar hybrid standard should be adopted for civil cases or whether the abuse of discretion standard we endorsed in Accutane should remain in force in the civil arena.

²⁶ New Jersey's version of Rule 702, which was modeled after the federal rule before it was amended, reads slightly differently. It states: “If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education may testify thereto in the form of an opinion or otherwise.” N.J.R.E. 702.

(4) general acceptance in the relevant scientific community.

[Olenowski I, 253 N.J. at 147 (quoting Daubert, 509 U.S. at 593-94).]

The United States Supreme Court in Daubert made clear that the factors it enumerated are non-exclusive and that the reliability inquiry is “flexible,” signaling that other considerations may also be pertinent. See ibid. (quoting Daubert, 509 U.S. at 594). As the Supreme Court advised, “[m]any factors will bear on the inquiry, and we do not presume to set out a definitive checklist or test.” Daubert, 509 U.S. at 593.

Likewise, our opinions in Accutane and in Olenowski I both cautioned that the Daubert factors should not be applied rigidly. See Accutane, 234 N.J. at 398-99 (describing Daubert’s list of factors as “a helpful -- but not necessary or definitive guide”); Olenowski I, 253 N.J. at 147-49 (emphasizing Daubert’s flexibility). We also made clear that the federal Daubert jurisprudence should not be applied in lockstep fashion and that New Jersey evidence principles ultimately govern admissibility in our state courts. See Accutane, 234 N.J. at 399 (declining “to embrace the full body of Daubert case law”); Olenowski I, 253 N.J. at 154 (adopting that principle for the application of the Daubert standard in criminal cases).

For ease of discussion in this particular case, we reorganize the Supreme Court’s listing of Daubert factors in a few ways. The sequence in which we address the Daubert factors here does not reflect their relative importance; all of them bear upon the analysis.²⁷ The “testability” factor, listed first by the Court conceptually, frequently ties in closely with the “error rate” component of the Court’s third factor, particularly in this case. Given that nexus, we shall discuss testability and error rate together. The other component of the Court’s

²⁷ Federal appellate case law applying Daubert has not rigidly followed the sequence of factors listed in the Daubert opinion. In fact, the Supreme Court itself has not adhered to that sequence in its two opinions applying Daubert. In Kumho Tire Co. v. Carmichael, 526 U.S. 137, 157-58 (1999), the Court first considered the lack of general acceptance of other experts in the field (factor #4), followed by the absence of published articles or papers that validated the expert’s approach (factor #2), and then it moved on to discuss other deficiencies. In Joiner, 522 U.S. at 517-19, the Court first addressed the published studies relied upon by the experts (factor #2), and explained why that reliance was analytically unjustified, without specifically addressing the other factors. The federal courts of appeals also at times have not adhered to a 1-2-3-4 sequence in discussing the factors, and have, in some instances, analytically combined multiple factors. See, e.g., McKiver v. Murphy-Brown, LLC, 980 F.3d 937, 959-60 (4th Cir. 2020) (discussing peer reviewed publication (factor #2) and acceptability (factor #4) before mentioning testability (factor #1)); Lawes v. CSA Architect & Engineer, LLP, 963 F.3d 72, 99-106 (1st Cir. 2020) (first and mainly analyzing published articles in the expert’s field (factor #2), then referring to other factors); UGI Sunberry LLC v. A Permanent Easement for 1.7575 Acres, 949 F.3d 825, 834-35 (3d Cir. 2020) (first discussing peer review (factor #2), then general acceptance (factor #4), then error rate and standards (factor #3), and then testability (factor #1)). That said, we do not prescribe that our trial judges are to follow the sequence we use in the present case, but emphasize that the pertinent factors should all be covered within the analysis.

third listed factor -- the adequacy of standards -- thereby becomes a standalone factor. Because the adequacy of standards logically affects many of the other factors (indeed, a standardless methodology presumably would be unreliable), we choose to address that subject in this case first. Following that, we will proceed to a discussion of the publication/peer review factor, and then move on to the others.

We therefore apply the Daubert factors to this particular record in the following sequence: (A) adequacy of standards; (B) publication and peer review; (C) testability and error rate; and (D) general acceptance. We then conclude with an overall assessment.

A. Adequacy of Standards

The Supreme Court in Daubert underscored the importance of “the existence and maintenance of standards controlling the [expert’s] technique’s operation.” 509 U.S. at 594. The Court illuminated the need for experts to adhere to reliable standards in Kumho Tire. 526 U.S. at 141. In that case, an expert in “tire failure analysis” had developed his own multi-factor test for determining whether a manufacturing or design flaw caused a tire failure. Id. at 143-44. The expert posited that there were four tell-tale visual and tactile signs that a tire failed due to misuse rather than a manufacturing or design flaw. Id. at 144. When the expert found any combination of two or more of

those signs, he would conclude that misuse caused the failure; if one or none was present, he would find a manufacturing or design defect. Ibid.

The trial court in Kumho Tire found that the expert’s methodology satisfied none of Daubert’s prongs. Id. at 145. The Supreme Court agreed. Id. at 158. Among other things, the Court noted that the expert’s four-factor test was administered in an undisciplined, standardless fashion, and that no one else in the field utilized his method. Id. at 154-57.²⁸ Those observations inform our analysis of the “adequacy of the standards” prong.

Here, in his second report, the Special Master recognized the “long process of initiating and developing the DECP and DRE protocol until it reached a level of standardization and developed into a program used in all fifty states, all provinces of Canada, and a number of other countries.” SM Report II at 38-39. In delineating the State’s arguments about the standards, he agreed with them, finding significant “the rigorous training, certification and recertification procedures” for DREs. Ibid. He further acknowledged the DECP program’s continued use of the Technical Advisory Panel, which has

²⁸ In a similar vein, our “net opinion” doctrine under New Jersey evidence law weeds out experts who base their opinions on purely personal standards or “rules of thumb.” See, e.g., State v. Burney, 255 N.J. 1, 23-24 (2023); Pomerantz Paper Corp. v. New Cmty. Corp., 207 N.J. 344, 372-74 (2011).

members from the relevant scientific fields, and the support of administrative and regulatory authorities such as the NHTSA and the IACP. Ibid.

As the Special Master noted, those many factors, according to the State, “assure that the [DECP] program is standardized [and] that it maintains a continuing process, with the advice and input of relevant experts, to continually be aware of new information that might affect the program.” Ibid. Through those means, “evaluations by DREs will be performed in accordance with a standardized procedure.” Ibid.

The Special Master concluded that the program “has established and continually maintains a well-organized structure for the DECP that provides careful and competent supervision and management and assures the reliable implementation of the standardized DRE program generally, and particularly in New Jersey.” Id. at 39. Hence, the Special Master “attribute[d] significant weight to this component” of the Daubert analysis. Ibid.

Although the standards component was not a prominent focus of the Public Defender’s briefing before the Special Master, the Public Defender did levy various criticisms about the skills of the DREs and the operation of the program. Among other things, the Public Defender argues that “[t]he DRE protocol is not a checklist” because DREs making observations at each step of the protocol are not compelled to make a particular finding, as, for example, a

psychiatric diagnosis would require. The Public Defender further submits that “[f]or two defendants exhibiting the exact same clues, DREs could, without violating any of the guidelines surrounding the protocol, conclude that one of the defendants is impaired while the other is not.” Thus, the Public Defender argues that the protocol has inadequate standards.

Applying de novo review, we agree with and adopt the Special Master’s finding on the standards component. The twelve-step DRE process is elaborate and standardized. It is grounded in a program that has been used across the nation and abroad for decades and is periodically modified. The program adheres to a standardized manual and uses a uniform seven-column matrix card and other tools for each DRE’s evaluation. The more than 400 certified DREs in the State who are deployed to perform the evaluations have been extensively trained, and are supervised and recertified. This is in stark contrast to the expert in Kumho Tire, whose idiosyncratic methodology the Supreme Court found lacking in reliable standards. See 526 U.S. at 154-57.

We acknowledge the concerns of the Public Defender and the defense amici that DREs are neither physicians nor medical professionals. The DREs have been trained, however, to ask drivers during the protocol about whether they have medical conditions or about other causes that might impair them or affect their performance on the field sobriety tests. The DREs take note of, but

do not opine about, any such medical information a driver may disclose. They do not render a medical diagnosis. They are not medically trained to obtain a fulsome medical history through follow-up questions. DREs are trained only to be aware of the major non-drug causes of impairment that may mimic signs of drug or alcohol impairment (e.g., head trauma, low blood sugar in diabetics, seizures and neurological disorders, conjunctivitis, some mental health issues, and “physical defects” like injuries that might affect performance of certain steps of the protocol). If the driver needs immediate medical attention, the DRE is trained to halt the examination and obtain medical assistance.

The State presented expert testimony, which the Special Master credited, SM Report II at 21, attesting that it is generally accepted that persons such as DREs who are not licensed medical professionals can be reliably trained to conduct certain medically related tasks such as checking a driver’s pulse and other vitals. At trial, the defense is free to cross-examine and impeach DREs about their limited medical knowledge. In addition, where applicable, the defense may present a medical expert witness to show that the defendant’s behavior and condition have a benign medical explanation, such as the prescribed use of medication or an underlying medical condition.

Additionally, we are unpersuaded by the Public Defender’s argument that the protocol is unreliable because two DREs applying it to the same driver

can reach different opinions. Such potential differences of opinion do not necessarily make a diagnostic standard unsound. In the field of medicine, for example, two physicians applying the same diagnostic standards and relying on the same clinical tests can legitimately disagree about a patient's condition. In fact, that is why patients often will seek a second doctor's opinion before proceeding with a course of treatment. There can be room for interpretation.

We are cognizant that the standards presently used to train and certify DREs might be further enhanced. For example, the program's certification match criteria, which assign a "passing" score to a DRE when the DRE accurately predicts only one out of two toxidrome categories present in a driver, or only two out of three or more categories, arguably might be made more stringent. And, as we noted earlier, the DREs should be obligated to attempt to complete all of the steps of the protocol, unless it is infeasible to do so. But such future enhancement of the training and certification standards is a policy decision for the program administrators. As is, the standards are adequate to reasonably support admissibility, with limitations. We further note that the passing rate and other aspects of the DRE training and certification standards are a fair subject of defense impeachment at trial.

On the whole, we concur with the Special Master that the State has established ample standardization to meet this Daubert factor.

B. Peer Review and Publication

We next discuss the Daubert factor of “whether the theory or technique has been subjected to peer review and publication.” 509 U.S. at 593. As the Supreme Court noted, for scientific experts, “submission to the scrutiny of the scientific community is a component of ‘good science,’ in part because it increases the likelihood that substantive flaws in methodology will be detected.” Ibid. However, the Court cautioned that publication “is not a sine qua non of admissibility; it does not necessarily correlate with reliability, and in some instances well-grounded but innovative theories will not have been published.” Ibid. (citation omitted). Thus, “publication (or lack thereof) in a peer reviewed journal” is “a relevant, though not dispositive, consideration in assessing the scientific validity of a particular . . . methodology on which an opinion is premised.” Id. at 594.

The Court in Kumho Tire observed that the Daubert factors “do not all necessarily apply even in every instance in which the reliability of scientific testimony is challenged,” noting that “[i]t might not be surprising in a particular case . . . that a claim made by a scientific witness has never been the subject of peer review.” 526 U.S. at 151. And case law has been mindful that several non-scientific fields of expertise are not typically studied in peer reviewed academic journals. See, e.g., Bitler, 400 F.3d at 1235 (admitting

expert testimony about the cause of an explosion under Daubert even though the fire investigator's experience-based methodology was "not susceptible to testing or peer review"); United States v. Hankey, 203 F.3d 1160, 1169 (9th Cir. 2000) (admitting under Daubert expert testimony on common gang practices that was not peer reviewed).

The Special Master determined that "the results of the many studies related to the DECP that have been undertaken since 1985 and that were entered into evidence by the parties support the State's position that the DRE protocol has consistently been found to be a reliable method for detecting impairment by drugs." SM Report I 285-96. The Special Master repeated that finding concerning the published studies in his second report. SM Report II 28-29. He found "most relevant and useful" two peer reviewed studies and noted that several of the other studies in the record were peer reviewed. Id. at 26-28.

Those conclusions concerning the various studies were the product of detailed and thoughtful analysis. Throughout sixty-five pages of his first report, the Special Master extensively discussed (1) three early field sobriety studies conducted by the Southern California Research Institute (SCRI) and funded by the NHTSA between 1977 and 1986 addressing driver alcohol

abuse, which formed the basis of the SFSTs;²⁹ (2) three more field validation studies funded by NHTSA between 1995 and 1998 concerning the accuracy of the SFSTs;³⁰ (3) three more studies from 2002, 2007, and 2011 examining the relationship between the SFSTs and alcohol impairment, two of which were peer reviewed, and two of which the NHTSA funded;³¹ (4) three studies from 2005, 2014, and 2020 evaluating a relationship between the SFST and drug-induced impairment, two of which were peer reviewed;³² (5) three foundational

²⁹ Marcelline Burns & Herbert Moskowitz, Psychophysical Tests for DWI Arrests (1977); Van Tharp et al., Development and Field Test of Psychological Tests for DWI Arrests (1981); Theodore E. Anderson et al., Field Evaluation of a Behavioral Test Battery for DWI (1983).

³⁰ Marcelline Burns & Theodore E. Anderson, Colo. Dep't of Transp., A Colorado Validation Study of the Standardized Field Sobriety Test (SFST) Battery (1995); Marcelline Burns, Fla. Dep't of Transp., A Florida Validation Study of the Standardized Field Sobriety Test (SFST) Battery (1997); Marcelline Burns & Jack Stuster, Validation of the Standardized Field Sobriety Test Battery at BACs Below 0.10 Percent (1998).

³¹ James McKnight et al., Sobriety Tests for Low Blood Alcohol Concentrations, Accident Analysis and Prevention, 34 Accident Analysis & Prevention 305 (2002) (NHTSA-funded and peer reviewed); Marcelline Burns, The Robustness of the Horizontal Gaze Nystagmus Test (2007) (NHTSA funded); and Karl Citek et al., Sleep Deprivation Does Not Mimic Alcohol Intoxication on Field Sobriety Testing, 56 J. Forensic Sci. 1170 (2011) (peer reviewed). We note that Dr. Citek was one of the expert witnesses who testified for the State before the Special Master.

³² K. Papafotiou et al., An Evaluation of the Sensitivity of the Standardized Field Sobriety tests (SFSTs) to Detect Impairment Due to Marijuana Intoxication, 180 Psychopharmacology 107 (2005) (peer reviewed); Amy J.

studies from 1985, 1986, and 1994 relating to the DECP protocol referenced in the DRE training manual, each of which were government-sponsored;³³ (6) eight other field and retrospective studies conducted in the United States and Canada examining the reliability of the protocol, some of which were peer reviewed and some of which were government-sponsored;³⁴ and (7) three

Porath & Douglas Beirness, An Examination of the Validity of the Standardized Field Sobriety Test in Detecting Drug Impairment Using Data from the Drug Evaluation and Classification Program, 15 Traffic Injury Prevention 125 (2014) (peer reviewed); Dary Fiorentino et al., The Usefulness of SFSTs in Detecting Drugs Other than Alcohol (2020) (not peer reviewed). Dr. Fiorentino was a witness for the State at the Special Master hearings.

³³ These include the Bigelow study and the Compton study, cited above, see supra n.12, and relied upon by the LAPD in the expansion of the DECP. See also Eugene V. Adler & Marcelline Burns, Ariz. Dep't of Pub. Safety, Drug Recognition Expert (DRE) Validation Study (1994).

³⁴ These include the 1993 Hardin study (government sponsored), the 2009 Beirness/Canada study (peer reviewed), and the 2021 Vaillancourt study (peer reviewed), cited above, see supra, n.17, and relied on by the Special Master as the three most useful studies. These will be discussed in detail below. See also D.F. Preusser et al., Evaluation of the Impact of the Drug Evaluation and Classification Program on Enforcement and Adjudication (1992) (funded by the NHTSA); Amy J. Porath et al., Toward a More Parsimonious Approach to Drug Recognition Expert Evaluations, 10 Traffic Injury Prevention 513 (2009) (peer reviewed); Amy J. Porath & Douglass Beirness, Simplifying the Process for Identifying Drug Combinations by Drug, 11 Traffic Injury Prevention 453 (2010) (peer reviewed); Amy J. Porath & Douglass Beirness, Predicting Categories of Drugs Used by Suspected Drug-Impaired Drivers Using the Drug Evaluation and Classification Program Tests, 20 Traffic Injury Prevention 255 (2019) (peer reviewed); Rebecca L. Hartman et al., Drug Recognition Expert (DRE) Examination Characteristics of Cannabis Impairment, 92 Accident Analysis & Prevention 219 (2016) (peer reviewed).

laboratory studies from 1996 and 1998 concerning the protocol, all of which were peer reviewed and two of which were government-sponsored.³⁵ SM Report I at 222-286.

We need not elaborate in this opinion on the details of those twenty-six studies, which the Special Master aptly described at length. We note that the Special Master particularly found significant the Hardin, Beirness/Canada, and Vaillancourt field studies, two of which were peer reviewed, “because they actually assessed the overall reliability of DREs evaluating subjects in the field.” Id. at 272. That helps to ensure a higher degree of correlation between peer review publication and reliability. See Daubert, 509 U.S. at 593 (noting peer review “does not necessarily correlate with reliability”).

The 1993 Hardin study, which was conducted in Minnesota and was government-sponsored, examined 71 field cases in which a DRE opined that a subject was under the influence of a drug and for which a urine sample was provided. SM Report I at 260. The Hardin study authors found an overall

³⁵ Stephen J. Heishman et al., Laboratory Validation Study of Drug Evaluation and Classification Program: Ethanol, Cocaine, and Marijuana, 20 J. Analytical Toxicology 468 (1996) (peer reviewed and NHTSA funded); Stephen J. Heishman et al., Laboratory Validation Study of Drug Evaluation and Classification Program: Alprazolam, d-Amphetamine, Codeine, and Marijuana, 22 J. Analytical Toxicology 503 (1998) (peer reviewed and NHTSA funded); David Shinar & Edna Schechtman, Drug Identification Performance on the Basis of Observable Signs and Symptoms, 37 Accident Analysis & Prevention 843 (2005) (peer reviewed only).

“corroboration rate” of 84.5%, applying impairment match criteria.³⁶ Ibid. It concluded “[t]he DRE protocol, if followed properly, appears to be a useful screening tool for predicting whether a subject is under the influence of drugs.” Ibid. (alteration in original) (quoting Hardin et al., Minnesota Corroboration Study 2). The Special Master found it was “the least helpful” of the three studies because of its small sample size. Id. at 272.

The 2009 Beirness/Canada study, which was published in a Canadian forensic science journal, analyzed 1,349 evaluations performed by DREs in Canada. Id. at 261. That study determined that in 92.1% of cases, the DRE’s opinion matched the drug class identified by a toxicological analysis. Ibid. In only nine cases did the DRE indicate a drug to be present and no drug was found. Ibid. The authors concluded that overall the drug evaluations conducted by the DREs were over 95% accurate,³⁷ which “provides confidence in the use of the DEC procedure to detect persons impaired by substances other

³⁶ The impairment match criteria, in contrast to the certification match criteria described above, require only that the DRE opine the presence of any impairing drug and that the toxicological analysis confirm the presence of any impairing drug. SM Report I at 142. Drug categories are irrelevant to the impairment match standard. Although the certification match standard is used for DRE certification, the impairment match standard is used in some studies evaluating DRE performance. Ibid.

³⁷ For reasons we explain in Part V.3, accuracy rates may be difficult to calculate reliably.

than alcohol.” Id. at 262 (quoting Beirness et al., 42 Can. Soc. Forensic Sci. J. at 79).

The Vaillancourt study published in 2021 was a retrospective study of 2,982 DECP cases in Quebec between 2014 and 2018. Id. at 269. The study encompassed all alleged drugged drivers arrested with signs of impairment following a DECP investigation in which a toxicological sample was available. Ibid. The study revealed that at least one drug with impairing potential was found in 98% of the cases, with at least one drug matching the DRE’s identified categories in 89% of the cases. Id. at 270. In only 9% of the cases, the DRE opined a drug category and the toxicology did not corroborate any drug in that category. Ibid.

That said, the Special Master correctly recognized limitations with those studies, which we likewise recognize. Most significantly, the studies assessed populations with an extremely high prevalence of drug-positivity and a low prevalence of drug-negativity. Id. at 272-73. As we will discuss in the next section, such an inherently skewed composition of samples means that a reliable error rate, particularly a false positive rate, might not be ascertained. Additionally, toxicology in those studies was typically not performed in cases in which the DRE opined no impairment by the driver. Id. at 273. Hence,

“whether those cases are true negatives or false negatives remains undetermined.” Ibid.

Further, the Special Master found that the lab studies relied upon by the Public Defender indicating lower accuracy rates (both Heishman studies and Shinar’s re-analysis of the same data) had “only marginal usefulness to this proceeding” and that the field studies were more “meaningful.” Id. at 282, 285. In particular, he noted the State’s experts testified that the conditions of the Heishman studies could have misled the DREs. Id. at 283-85. In particular, the researchers possibly used lower dosing levels than seen in the field, and they also allowed the test subjects to practice and improve upon their performance of the psychomotor tests. Ibid.³⁸

Despite their recognized limitations, we hold that the two dozen studies presented in the record and considered by the Special Master are sufficient to meet the Daubert factor of publication and peer review. The Special Master appropriately considered not only the existence of those studies but also their substantive content and conclusions. Many of the studies appeared in peer

³⁸ The Nebraska Supreme Court has also critiqued these studies, observing that, because the DREs did not question subjects about recent drug use and did not examine evidence that would be found at an arrest, the study inappropriately “examined an abbreviated evaluation that is different from the standardized protocol that is actually used.” State v. Daly, 775 N.W.2d 47, 59-60 (Neb. 2009).

reviewed publications, enabling other researchers to comment on the findings and to undertake their own studies.

Although some of the studies were sponsored by government agencies such as the NHTSA and not peer reviewed by academics, that does not undermine those studies' relevance or evidential weight. See Contini v. Bd. of Educ. of Newark, 286 N.J. Super. 106, 124-25 (App. Div. 1995) (finding statistical reports on school performance compiled by the State Board of Education to be reliable and admissible). Indeed, our evidence rules recognize that statistical findings in government reports presumptively have sufficient reliability to qualify for admission under the hearsay exception for public records, N.J.R.E. 803(c)(8). See Biunno, Weissbard & Zegas, Current N.J. Rules of Evidence, cmt. 1 on N.J.R.E. 803(c)(8) (2023-2024) (noting “the special trustworthiness of official written statements”).

To be sure, there has not yet been a published study that specifically examines the New Jersey DRE program. Instead, experts in the present case endeavored to analyze the “retrospective data” collected between 2017 and 2018. We discuss that data analysis in the following section.

C. Testability and Error Rate

Testability and error rate present more difficult issues in this case. Our extensive discussion of them follows, aided by the context we have already presented concerning standards and publication.

The United States Supreme Court stated in Daubert that “[o]rdinarily, a key question to be answered in determining whether a theory or technique is scientific knowledge that will assist the trier of fact will be whether it can be (and has been) tested.” 509 U.S. at 593. The Court repeated the qualifying term “ordinarily” in announcing the Daubert factor of error rate, stating that “in the case of a particular scientific technique, the court ordinarily should consider the known or potential rate of error.” Id. at 594. The term “ordinarily” conveys that a judge’s findings of testability and reasonably low error rates from test results are expected -- but not always required -- elements of a proponent’s reliability showing.

Testability, sometimes called “falsifiability” or “refutability,” is meant to help “separat[e] science from metaphysics” and thus the knowable and factual from the unknowable and speculative. D.H. Kaye, On “Falsification” and “Falsifiability”: The First Daubert Factor and the Philosophy of Science, 45 Jurimetrics J. 473, 476 (2005). Admissible evidence must consist of “knowable fact[s]” relevant to the determination of the question before the

trier of fact. 1 Wigmore on Evidence § 1 (Tillers rev. 1983). If an expert's testimony conveys only a "conclusion" or an "assumption," Volk v. DeMeerleer, 386 P.3d 254, 277 (Wash. 2016), or if it is mere "speculation or conjecture," Townsend v. Pierre, 221 N.J. 36, 55 (2015) (quoting Davidson v. Slater, 189 N.J. 166, 185 (2007)), it is not factual and not helpful to the trier of fact.

The difference between fact and speculation, however, is often unclear. That is particularly so in the domain of what are considered "soft" sciences such as social sciences, as opposed to "hard" sciences such as physical sciences. Case law applying Daubert in other jurisdictions has generally been less demanding concerning testability and error rates for experts in soft sciences. See, e.g., Commonwealth v. Hinds, 166 N.E.3d 441, 453-56 (Mass. 2021) (sociology); Morris v. State, 361 S.W.3d 649, 654 (Tex. Crim. App. 2011) (child psychology). The DRE program is a mix of both soft social sciences, such as psychology and human behavior, and hard sciences, such as toxicology.

As the Special Master recognized, there are inherent practical limitations within the DRE program that complicate efforts to test the program results empirically and to obtain meaningful error rates. See SM Report I at 216-17,

(noting constitutional, ethical, and practical constraints); SM Report II at 25-27 (same). Those practical limitations have numerous dimensions.

First, as we have already noted, the sample of drivers who are stopped for suspected DWI or DUID because of observed “erratic and dangerous driving” do not represent the general population. See Bealor, 187 N.J. at 590. The sample is heavily skewed towards persons who are likely to be impaired because of their usage of alcohol, drugs, or both. As the Special Master correctly recognized, our constitutions and laws do not allow DRE researchers to stop every driver and infringe on their liberty to perform the DRE protocol without probable cause to arrest the person for driving under the influence of drugs. SM Report I at 272; SM Report II at 27; see also Delaware v. Prouse, 440 U.S. 648, 663 (1979); State v. Williams, 254 N.J. 8, 44-45 (2023).

Second, laboratory simulations cannot replicate all twelve steps of the DRE protocol. Key portions of the protocol (particularly Steps 2, 3, and 10) ask whether the subject made admissions of drug use to the arresting officer or DRE. Such admissions could not be made in a “double blind” study in which the test subjects would be unaware of whether they had ingested an actual impairment-causing drug or a placebo. Further, a DRE’s observations of an injection site for signs of drug use cannot be simulated.

Third, as the Special Master rightly noted, there are ethical and legal constraints, as well as medical risks, in subjecting humans to high doses of mind-altering drugs. That is especially true of new pharmacological substances (NPS) and polydrug combinations, about which there is little scientific knowledge and which might even be harmful or lethal to the test subject.

Fourth, as the record also shows, toxicology has several limitations. The experts and the parties agree that toxicology alone can only reveal the presence of a drug in a person's body. It does not measure the actual effect of a substance on the test subject, much less impairment beyond a legally cognizable threshold.

Fifth, toxicology uses drug-specific cutoff levels to detect the presence of substances in the extracted urine or blood samples. As several of the experts noted, some drugs and combinations of drugs may be impairing but below the cutoff levels.

Sixth, the concentration of drugs in a body can dissipate over time. This is especially of concern with respect to a blood sample. As the Special Master noted, drivers have a legal right to refuse to consent to the extraction of a blood sample from their bodies. Cf. Schmerber v. California, 384 U.S. 757, 767-70 (1966). If a driver invokes that right and refuses to consent to a blood

draw, the police must obtain a warrant from a judge or the warrantless blood draw must be justified by exigent circumstances. See id. at 770-71; Missouri v. McNeely, 569 U.S. 141, 164-65 (2013); see also State v. Zalcborg, 232 N.J. 335, 351-52 (2018). During the time expended in obtaining a warrant -- assuming there is probable cause to support one -- the drug levels in the driver's blood may diminish before a blood sample is collected.

Seventh, urine samples, although easier to obtain from a driver, are less informative than blood toxicology because urine metabolites may remain in a person's body for days or weeks. Hence, a positive toxicology result derived from a urine sample does not signify that the test subject was impaired at the time of driving or had recently ingested the drug(s).

Those and other inherent constraints make the DRE program less "testable" and the error rate less "knowable" than the ideal. With that in mind, we proceed to discuss the New Jersey retrospective data in the record, and the opinions presented about that data in the testimony of the statisticians and the other experts.

As we noted above, the New Jersey retrospective data collected from the DRE program in 2017 and 2018 encompassed 5,855 DRE reports. Of that total, 2,551 were non-training cases that included a toxicology report for corroboration of the DRE conclusion. In about 27% of the 5,855 cases, there

was no toxicology report obtained. That can occur because the driver refused to provide a urine sample; because the DRE concluded that the driver was not impaired by drugs and so did not request a urine sample; or for some other reason.

Sensitivity, Specificity, and Accuracy

The experts who testified before the Special Master discussed this data by using several core concepts within the field of statistics, chiefly “sensitivity,” “specificity,” and “accuracy.”

Sensitivity refers to the detection of true positives. In this context, it calculates the percentage of times a DRE correctly opined the presence of specific drug categories (under the certification match criteria described above) out of the total number of instances where the drivers had drugs in their systems. SM Report I at 188. Mathematically, that entails dividing the number of true positives by the sum of true positives and false negatives.³⁹
Ibid.

Specificity refers to the detection of true negatives. In this context, it means how often the DRE will opine that persons have no drugs in their system, if they indeed have no drugs in their systems. Id. at 188-89. The

³⁹ Sensitivity = True Positives/(True Positives + False Negatives)

specificity is calculated by dividing the number of true negatives by the sum of true negatives and false positives.⁴⁰ Ibid. The false positive rate shows in this case how often DREs opine that drivers have a drug or drugs in their system when, according to a toxicology report, they do not. It can be calculated by dividing the number of false positives by the sum of false positives and true negatives,⁴¹ or by subtracting the specificity rate from 100%.

Accuracy “summarizes the ability of the test being able to truly discriminate between true positives and true negatives.” Id. at 189. It considers when the subject condition is present and when it is not. Ibid. One of the State’s statistical experts, Dr. Brian D. Martin, testified that it is “the most commonly valued statistic associated with a test.” Ibid. As expressed in a mathematical formula, it means taking the sum of true negatives and true positives and dividing that figure by the sum of all four potential outcomes -- true positive, false positive, true negative, and false negative.⁴² Ibid.

The high sensitivity rate in the New Jersey data, ranging from 82.5% to 92.6%, is an important, albeit not dispositive, starting indicator of reliability.

⁴⁰ Specificity = True Negatives/(False Positives + True Negatives)

⁴¹ False Positive Rate = False Positives/(False Positives + True Negatives)

⁴² Accuracy = (True Negatives + True Positives)/(True Positives + False Positives + True Negatives + False Negatives)

Based on the available data, it appears that when a DRE yields a positive result, indicating that a person is displaying signs consistent with a specific category of drug, the result is very often correct as corroborated by toxicology. There were relatively few “false negative” instances in which the DRE examination failed to detect the presence of at least one out of two (or two out of three or more) drug categories detected by toxicology. Had the sensitivity rate been low, it would cast doubt on the protocol’s reliability.

Notably, one of the State’s key witnesses, Dr. Enrique Schisterman, who chairs the University of Pennsylvania Medical School’s Department of Epidemiology, testified that the sensitivity rate within the New Jersey data was “quite robust.” Dr. Schisterman noted that he was “confident” in sensitivity as an estimator for this data, in part because the DRE test was designed to be utilized in evaluating suspected intoxicated drivers, not the general driving population. As the Special Master summarized it, “DREs are excellent at identifying true positive cases.” Id. at 215.

The Public Defender essentially contends the DRE methodology is worthless because a high sensitivity rate would also be attained by assuming that all drivers who are subjected to the protocol are drug impaired. That argument overlooks the DRE’s informative role in narrowing down the possible sources of drug use within the matrix’s seven categories of

toxidromes. After completing the protocol, the DRE designates which of the seven categories, if any, match the driver's presentation. The methodology is a nuanced multi-step protocol, not a crude "guess them all" exercise.

Calculation of the specificity rate, however, presents substantial obstacles. As we have already noted, there are many practical reasons -- such as delays in obtaining a warrant; time otherwise consumed in getting a sample without a warrant; lab testing cutoffs; the non-testability of NPS substances and polydrug combinations; the differences between blood and urine analysis; and so on -- that can explain why a driver might have actually been impaired at the time of the DRE's assessment despite a negative toxicology report. Thus, one cannot assume that the instances in which the DRE made a positive finding that was not corroborated by a later toxicology exam are necessarily "false" positives.

Regardless of the actual composition of the toxicology instances, we accept Dr. Schisterman's assessment that the retrospective New Jersey data is inadequate to enable a fair calculation of actual false positives and specificity. We therefore decline to adopt the argument of the Public Defender that the false positive rate must be 78% because in 105 total instances where a toxicology report revealed no drugs in the driver's system, a DRE nonetheless opined that the subject was impaired by drugs in 82 of them. Nor, however, do

we adopt the 3% error rate the State ascribed to false positives; that rate was incorrectly calculated by using all cases, rather than all negative cases, as the denominator in the formula.

As the Special Master found, that data shows a high accuracy rate of between 91% and 95%. Such a high rate is to be expected because the sample of persons subjected to the DRE protocol are drivers suspected of being impaired and who were, in most instances, observed by an officer to have driven their vehicles erratically and did not have a BAC at or above the legal limit. But because accuracy is a function of sensitivity and specificity, it cannot be reliably calculated from the retrospective dataset, as Dr. Schisterman acknowledged in his testimony. The same dataset constraints exist in the other published retrospective studies.

In sum, the testability and false-positive error rate aspects of the Daubert analysis are largely inconclusive, due to variables that are neither controllable nor known, and thus must be understood in the context of the datasets from which they were calculated. The field studies that support reliability for the DRE are limited in authoritativeness by the skewed sample of motorists subjected to the DRE protocol. On the other hand, the double-blind studies that question the reliability of parts of the DRE protocol give the methodology short shrift because of the inherent ethical and study-design limitations.

We reject our dissenting colleagues’ assertion that testability and error rates are categorically the most important Daubert factors. Post at ____ (slip op. at 9, 16-17). Case law does not support according those factors such pre-eminent or dispositive status. Despite the important role testability often may play in assessing reliability, “testability is not a prerequisite to admission.” Seifert v. Balink, 888 N.W.2d 816, 841 (Wis. 2017) (allowing expert medical testimony on the standard of reasonable care for obstetricians based on the expert’s untestable personal experiences). “While the testability and error rates of a scientific theory are factors a trial court may consider in assessing reliability, the trial court may give these factors less weight or disregard them altogether if the case so requires.” Estate of Ford v. Eicher, 250 P.3d 262, 269 (Colo. 2011) (en banc) (finding and recognizing that “ethics prevent testing the [expert’s] intrauterine contraction theory”).

“In certain fields, experience is the predominant, if not sole, basis for a great deal of reliable expert testimony.” Seifert, 888 N.W.2d at 841 (quoting Fed. R. Evid. 702, Advisory Comm. Note (2000)). The inability to calculate a methodology’s error rate with precision can be a realistic constraint in situations where, as here, the testing would involve human subjects. See, e.g., John’s Heating Serv. v. Lamb, 46 P.3d 1024, 1035-36 (Alaska 2002) (upholding the admissibility of toxicology experts who evaluated the leakage

of carbon monoxide into homes causing neurological illnesses, noting that “testing on humans [to determine dangerousness levels] simply cannot be ethically undertaken”); United States v. Pollard, 128 F. Supp. 2d 1104, 1120, 1123 (E.D. Tenn. 2001) (finding reliable a doctor’s estimation of the age of a child within an illicit video, despite reliance on a scientific scale with questionable error rates because the scale was one of several factors relied upon, including the expert’s over twenty years of professional experience as a pediatrician).

The absence of a definitive rate of error in the present case should not be a dispositive basis to exclude all DRE testimony. In essence, the defense is demanding that the State prove a “null hypothesis” that the DRE protocol will not produce an intolerable percentage of false positives. But, as the Appellate Division recognized in Carl v. Johnson & Johnson, “no set of statistical results is capable of establishing that [a] null hypothesis is actually true or false.” 464 N.J. Super. 446, 456 (App. Div. 2020). That is why “absolute scientific certainty is not the standard for the admissibility of expert testimony.” Paolino v. Ferreira, 153 A.3d 505, 523 (R.I. 2017) (quoting State v. Abdullah, 967 A.2d 469, 478 (R.I. 2009)).

If, as the Public Defender and the defense amici argue, testing to validate the DRE protocol must be more robust, expanded testing would entail

stopping and administering the DRE protocol to a large sample of drivers who had only committed a motor vehicle violation, thereby detaining people in violation of their liberties and constitutional rights. The appellants and amici surely would not favor such infringements in a quest to accumulate more reliable data about DRE error rates. In short, the inconclusiveness of the error rate here should not categorically bar the admission of this useful evidentiary source.

D. General Acceptance

We noted in Olenowski I that the previously dispositive Frye admissibility standard, which hinged upon the “general acceptance” of an expert’s methodology, has now been folded in as a single factor within the multi-factor Daubert test. See 253 N.J. at 147 (citing Daubert, 509 U.S. at 593-94). As the Supreme Court instructed, “[a] ‘reliability assessment does not require, although it does permit, explicit identification of a relevant scientific community and an express determination of a particular degree of acceptance within that community.’” Daubert, 509 U.S. at 594 (quoting United States v. Downing, 753 F.2d 1224, 1238 (3d Cir. 1985)). “Widespread acceptance can be an important factor in ruling particular evidence admissible, and ‘a known technique which has been able to attract only minimal support

within the community’ may properly be viewed with skepticism.” Ibid.
(quoting Downing, 753 F.2d at 1238).

As the Special Master correctly found, the record here amply establishes such “[w]idespread acceptance” and support of the DRE protocol. See *ibid.*
At the end of his comprehensive initial report applying the Frye standard, the Special Master wrote:

I conclude for all of the reasons stated in this report that DRE testimony is reliable. The reliability is established by the expert testimony presented by the State, which establishes that the DRE protocol replicates generally accepted medical practices for identifying the presence of impairing drugs and their likely identity through a toxidrome recognition process. This testimony has also established that the DRE matrix comports with matrices designed for this purpose and generally accepted and used in the medical field. This testimony has also established that the training DREs receive is comparable to that received by medical technicians and that DREs are thus enabled to reliably apply the protocol. Therefore, by implication, the DRE protocol as a whole and its individual components are generally accepted in the scientific communities to which they belong, namely medicine and toxicology.

As with all evidence, and as I have stated repeatedly regarding each individual step, DRE evidence and the DRE opinion will be tested by cross-examination and the factfinder will ascribe to it such credibility assessments and weight allocations as he or she deems appropriate.

[SM Report I at 331 (emphases added).]

Upon our de novo review of the record, we concur with the Special Master's conclusions, subject to caveats we will detail in the next portion of this opinion.⁴³

For many years, the DRE protocol has been widely and regularly used across this country and abroad. No state has discontinued it, and no state's highest court has nullified it. The protocol has been studied multiple times and periodically revised and enhanced. When DRE evidence is presented in courts far and wide, defense attorneys have had repeated opportunities to impeach it on cross-examination and to counter it with competing expert opinion that may be critical of the methodology. Although it has imperfections, the protocol has stood the test of time in its widespread acceptance.

Our case law has instructed that there need not be complete agreement within the scientific community to satisfy the general acceptance test. “[P]ractically every new scientific discovery has its detractors and unbelievers, but neither unanimity of opinion nor universal infallibility is required for judicial acceptance of generally recognized matters.” Chun, 194 N.J. at 92 (quoting Johnson, 42 N.J. at 171). The test does not require the “exclusion of the possibility of error.” Ibid. (quoting Harvey, 151 N.J. at 171).

⁴³ The dissent acknowledges that the Daubert factor of general acceptance has been demonstrated. Post at ____ (slip op. at 17-19).

The twelve expert witnesses who testified for the State in support of the DRE protocol's reliability were highly credentialed. They collectively explained in depth why the protocol is reliable and widely used. And the Public Defender, in arguing before us, did not rely on the testimony of the four defense experts who expressed an opposing viewpoint. The record provides a solid foundation for the Special Master's conclusion of general acceptance.

Case law in other jurisdictions has generally upheld the admissibility and reliability of the DRE protocol, which supports the protocol's general acceptance. See State v. Kelly, 97 N.J. 178, 210 (1984) (noting case law is one indicator of general acceptance). The cases can be divided into three groupings.

First, several courts have ruled that DRE evidence is admissible as expert testimony based upon specialized, not scientific, knowledge. See State v. Aleman, 194 P.3d 110, 112, 117 (N.M. Ct. App. 2008); Williams v. State, 710 So. 2d 24, 25, 28 (Fla. Dist. Ct. App. 1998); State v. Layman, 953 P.2d 782, 786 (Utah Ct. App. 1998); Mace v. State, 944 S.W.2d 830, 834 (Ark. 1997); United States v. Everett, 972 F. Supp. 1313, 1319-21 (D. Nev. 1997); State v. Klawitter, 518 N.W.2d 577, 579, 584-85 (Minn. 1994).

Second, the Washington Supreme Court and the New York County Court, Suffolk County, have held that DRE testimony is admissible under the

Frye standard. State v. Baity, 991 P.2d 1151, 1157-61 (Wash. 2000) (en banc); People v. Quinn, 580 N.Y.S.2d 818, 826 (Dist. Ct. 1991), rev'd on other grounds, 607 N.Y.S.2d 534 (App. Div. 1993).

Third, the Wisconsin Court of Appeals, the Nebraska Supreme Court, and the Oregon Court of Appeals have deemed DRE evidence admissible under the Daubert standard. State v. Chitwood, 879 N.W.2d 786, 793, 796-801 (Wis. Ct. App. 2016); Daly, 775 N.W.2d at 62; State v. Rambo, 279 P.3d 361, 366-67 (Or. Ct. App. 2012).

Only a handful of courts -- none of which are a state's highest court -- have held that DRE evidence is inadmissible. The Public Defender and defense-aligned amici have cited to some unpublished opinions that have done so, none of which warrant our citation or reliance. See R. 1:36-3.

The Public Defender urges that we take note of last year's 2-1 published majority opinion of the Michigan Court of Appeals in People v. Bowden, which concluded that a DRE's testimony was inadmissible under Michigan Rule of Evidence 702. ___ N.W.2d ___, ___ (Mich. Ct. App. 2022), appeal denied, 994 N.W.2d 776 (Mich. 2023). The Bowden majority found that the State failed to meet its burden to establish the reliability of DRE testimony under the Daubert standard. Ibid. The majority agreed with the defendant that the studies relied on by the State "validated the DRE protocol's accuracy in

determining the presence of a substance in a subject's blood but did not validate the DRE protocol for determining a subject's degree of impairment." Id. at ___ (slip op. at 9) (emphases added). The record in Bowden failed to support "the purpose for which the prosecution intended to use the results of the protocol in this case -- to provide evidence of defendant's level of impairment and impaired driving ability." Id. at ___ (slip op. at 9-10) (emphasis omitted).

The dissent in Bowden concluded the trial court had not erred in admitting the DRE's testimony. Id. at ___ (slip op. at 1) (Redford, J., dissenting). The dissent acknowledged that the DRE protocol cannot definitively establish a person's degree of impairment, but deemed it reliable enough to assist the trier of fact. Id. at ___ (slip op. at 8).

Notably, the testimony offered in support of the protocol in Bowden came solely from a single DRE officer, id. at ___ (slip op. at 1-4) (majority opinion), in contrast with the forty-two days of hearings developed here before the Special Master with sixteen expert witnesses. Moreover, as we will explain in the next section of this opinion, we accept the Bowden court's premise that DRE testimony does not, in and of itself, establish impairment. But we further hold that such testimony is sufficiently reliable to be admitted for a less ambitious purpose, and with critical safeguards.

VI. Analysis

Applying a de novo standard to the Special Master's determination, and applying the Daubert factors in combination, does the record developed before the Special Master reflect that DRE testimony is sufficiently reliable to be admitted in our courts? The answer is yes, but subject to stringent limitations we now set forth.

For the reasons we have explored in this opinion, there are many facets of the DRE protocol that weigh in favor of its reliability, but the protocol has several weaknesses as well. The protocol is elaborate and widely utilized, has been studied and scrutinized, and is the subject of extensive training and oversight within New Jersey. Although the adequacy of the data can be debated, it frequently predicts correctly that a driver who has been stopped but who does not have an illegal BAC level has ingested one or more identified categories of drugs.

The protocol does not, however, establish that a driver is actually impaired, or that the drug categories identified by the DRE are definitively the cause of any such impairment. A toxicology report, particularly one based on a blood sample instead of a urine sample, can help corroborate the presence of such drugs in the driver's system. But even that toxicology cannot prove that the driver was actually impaired by drugs while behind the wheel because

there are no per se DUID violations in our statutes. Further, no studies in this record identify a drug level that establishes impairment per se.

We also recognize that, as the defense argues, there are palpable risks of confirmation bias when a DRE officer administers the protocol, particularly in the more subjective aspects of the examination, such as the SFSTs and the eye tests. Such bias may consciously or subconsciously affect the DRE's opinion concerning a driver, despite an officer's good faith and training to remain objective. In many instances, drivers admit to the arresting officer or DRE that they have been using drugs, which potentially influences how the DRE evaluates other steps of the protocol. DREs are called only when there is a suspected drugged driver, as we have underscored.

Because of those concerns, which to some extent undercut but do not refute the reliability of the DRE's methodology, we adopt several limitations on the admissibility and probative use of a DRE's opinion in criminal and quasi-criminal cases. Several of these limitations have been recommended by the Public Defender as alternatives to its preferred outcome of total exclusion.

The boundaries of reliability we now delineate are not unusual. Some fields of expertise are only sufficiently reliable to be admitted with appropriate

restrictions and limitations.⁴⁴ The fact that an expert’s methodology cannot reliably prove everything a proponent would like it to prove does not mean that it cannot be a reliable and useful tool for a more limited purpose. The scope of the proffer is critical.

A. The “Consistency Only” Limitation

First and foremost, a DRE’s opinion must not be allowed to prove too much. We reject the notion that the DRE’s opinion at Step 11 establishes causation, i.e., that particular drugs or categories of drugs were ingested by the driver and caused the driver to be impaired. Impairment instead must be

⁴⁴ See, e.g., J.L.G., 234 N.J. at 272, 302-03 (permitting as “reliable” limited expert testimony that victims of child sexual abuse will often delay disclosure of the incident, but barring as not “sufficiently reliable” “any reference to ‘[Child Sexual Abuse Accommodation Syndrome],’ an abuse ‘syndrome,’ other CSAAS ‘behaviors’ aside from delayed disclosure, or causes for delayed disclosure” under the Frye standard); In re Bair Hugger Forced Air Warming Devices Prod. Liab. Litig., 9 F.4th 768, 782-83 (8th Cir. 2021) (overturning the lower court’s “categorical exclusion” of an expert and his forced-air warming model, and instead limiting the scope of his permissible testimony to the narrow hypothesis his model was able to test), cert. denied, 142 S. Ct. 2731 (2022); United States v. Hill, 818 F.3d 289, 298-99 (7th Cir. 2016) (holding that “[h]istorical cell-site analysis can show with sufficient reliability that a phone was in a general area,” but not the phone’s specific location, and “caution[ing]” that expert testimony “that overpromises on the technique’s precision -- or fails to account adequately for its potential flaws -- may well be an abuse of discretion”); United States v. Willock, 696 F. Supp. 2d 536, 569-72 (D. Md. 2010) (holding “that firearms toolmark identification evidence is only relevant, reliable, and helpful to a jury if it is offered with the proper qualifications regarding its accuracy” and outlining the relevant “safeguards”), aff’d sub nom. United States v. Mouzone, 687 F.3d 207 (4th Cir. 2012).

proven by the State with independent evidence, as we held in Bealor. See 187 N.J. at 577. That evidence can include, for example, specific factual observations of impaired behavior by the arresting officer or the DRE, a driver’s admissions, information from a passenger or other observer about the driver’s recent drug use, or drugs or paraphernalia found in the vehicle. Id. at 590-91 (explaining that a factfinder may draw inferences connecting the simultaneous presence of “objective facts of intoxication” and proof of the “presence of a cause of intoxication” to “conclude that [a] defendant drove while intoxicated”). And a DRE’s opinions tying such factual observations and the protocol results to specific drug categories must be more restricted.

For reasons we now explain, we hold that a DRE is only allowed to opine in court that the protocol has presented indicia that are “consistent with” the driver’s usage of certain categories of drugs. The DRE’s expert opinion testimony must not go further than that.

It is axiomatic that correlation (sometimes termed “consistency” or “association”) does not equal causation. See State v. Loftin, 157 N.J. 253, 379 n.3 (1999) (Handler, J., dissenting) (“A positive correlation may be the product of mathematical randomness rather than actual cause and effect.”); Landrigan, 127 N.J. at 415 (“Statistical associations, however, do not necessarily imply causation.” (internal quotation omitted)); State in Int. of A.B., 109 N.J. 195,

200 (1988) (observing no causal relationship between learning disabilities and juvenile delinquency because “[w]hile there may be some correlation between slow learning and the commission of crimes, we suspect that the correlation is as much a result of background factors as of any direct link between the two”).

Sometimes a positive correlation can be mere coincidence. A simple example illustrates the point: the fact that, between 1908 and 2020, the same political party’s candidate won the presidential election in two-thirds of the years in which a National League team won the World Series does not mean that the World Series outcome caused the presidential election result. The correlation between those events is obviously coincidental. By contrast here, a toxicology match with a DRE’s opinion is not an unexplained coincidence. The match is supported by principles explained by the medical and other experts who testified before the Special Master detailing the rationales for the various steps within the protocol.

Proof of consistency can be pertinent as one component within the totality of the evidence to support an inference that drugs caused a driver’s impairment. See Bealor, 187 N.J. at 590-91 (permitting the factfinder to infer a driver’s impairment was caused by drugs when there are “objective facts of intoxication” and “the proven presence of a cause of intoxication,” where a

consistency opinion is evidence of the latter). A number of courts have recognized such general principles about causation when applying Daubert.⁴⁵

Several of the testifying experts and the Special Master used this nomenclature. They discussed whether certain findings generated through the DRE protocol were “consistent with” certain inferences, or quoted from studies and DRE materials that identified such consistency, including, significantly, the DRE matrix that assists the DREs in reaching their

⁴⁵ See, e.g., United States v. Valencia, 600 F.3d 389, 425 (5th Cir. 2010) (“[W]here evidence of correlation itself is potentially relevant and unlikely to mislead the jury, an expert who reliably discerns this relationship can present such conclusions to the jury.”); Etherton v. Owners Ins. Co., 829 F.3d 1209, 1220-21 (10th Cir. 2016) (“Although correlation alone may be insufficient to establish causation . . . it is nonetheless relevant to identifying causal relationships. Indeed, it may be ‘a necessary but not sufficient condition for causation.’” (citations omitted) (quoting Joseph F. Healey, The Essentials of Statistics 350 (4th ed. 2015))); In re Bair Hugger, 9 F.4th at 779 (noting that “epidemiology enables experts to find associations, which by themselves do not entail causation,” but that such studies can nonetheless “be brought to bear on the question of causation, and can be very useful to answering that question” (citations and internal quotations omitted)); Milward v. Acuity Specialty Prods. Grp., Inc., 639 F.3d 11, 17-19, 23 (1st Cir. 2011) (citing the research of Sir Arthur Hill and observing that an association between an exposure and a disease may bear upon the plausible explanations for that disease, considering the “weight of the evidence” and a holistic evaluation of data and scientific evidence); Hendrix ex rel. G.P. v. Evenflo Co., Inc., 255 F.R.D. 568, 592 (N.D. Fla. 2009) (permitting under Fed. R. Evid. 702 an expert engineer to opine “whether, from a biomechanics standpoint, [the plaintiff]’s injuries are consistent with those expected from an exploding airbag,” although the expert was “not qualified to offer medical causation testimony” regarding the specific cause of the plaintiff’s injuries), aff’d, 609 F.3d 1183 (11th Cir. 2010).

conclusion in Step 11. See, e.g., SM Report I at 11, 185, 277, 279, 288. And as part of his analysis in his first report, see id. at 287-88, the Special Master quoted from the Washington Supreme Court’s opinion in Baity, which held that “[t]he DRE officer, properly qualified, may express an opinion that a suspect’s behavior and physical attributes are or are not consistent with the behavioral and physical signs associated with certain categories of drugs.” 991 P.2d at 1160-61 (emphasis added). We likewise conclude that a DRE may opine that the protocol results are consistent with a driver’s use of drugs in the specific matrix categories. However, we emphasize that such a consistency opinion is the outer limit of reliability with which a DRE can offer admissible expert testimony.

B. The Absence of a Toxicology Report (“Step 12”)

The parties and amici have sharply differed about whether a corroborating toxicology report must be a precondition to admitting any opinion testimony from a DRE, assuming such testimony is to be admitted at all.

We agree with the Public Defender and the supporting amici that a toxicology report corroborating a DRE’s opinion is important evidence. The toxicology report can strengthen the State’s case or, alternatively, undermine it. However, as we have already noted, a toxicology report can detect only

drug presence; it cannot establish the amount or timing of the driver's drug usage. And, for a variety of reasons we have already discussed, the toxicology report may not detect some combinations of drugs or newer "designer drugs" that are resistant to detection.

Because toxicology can be relevant and helpful to a trier of fact, we encourage that it be performed. Indeed, toxicology of a driver's blood is routinely conducted after fatal traffic accidents in our state. See SM Report I at 319-20; N.J.S.A. 26:2B-24 (requiring testing for alcohol of victims and drivers involved in fatal accidents).

On the other hand, as we have noted, there are many practical reasons why toxicology may not be feasible, such as a driver's lack of consent to provide a sample or the delays or obstacles to obtaining a warrant. Moreover, the record shows that Steps 1-10 of the DRE protocol can reveal information useful to the trier of fact, even without a corroborating toxicology report. Of the jurisdictions that have addressed the admissibility of DRE evidence, most have not required that a toxicology report be obtained for the DRE to provide expert testimony.⁴⁶

⁴⁶ Three states that admit DRE testimony have required toxicology be completed: Oregon, New Mexico, and Washington. See Rambo, 279 P.3d at 366-67; Aleman, 194 P.3d at 121; Baity, 991 P.2d at 1160.

Bearing in mind those considerations, we hold that DRE officers must make a reasonable attempt to obtain a toxicology report when it is feasible to do so -- and preferably to obtain a blood sample rather than a urine sample -- when their protocol indicates at Step 11 an opinion of consistency with drug use.⁴⁷ If the court finds no reasonable attempt was made, despite its feasibility, the DRE evidence shall be excluded. However, if the State establishes a reasonable justification for the lack of a toxicology report, then the DRE evidence is admissible, subject to defense impeachment and counterproofs.

C. Fair Opportunity for Defense Impeachment and Counterproofs

As just noted, if the trial court admits DRE evidence for the State -- with the limitations we have prescribed -- the defense shall have a fair opportunity to impeach or rebut it through cross-examination of the DRE and with counterproofs. The adversarial process can then explore the probative strengths and weaknesses of the DRE evidence. See, e.g., Hisenaj v. Kuehner, 194 N.J. 6, 23-24 (2008) (noting that parties are “free to pursue” on cross-examination specific weaknesses in an expert’s admissible methodology).

For example, defense attorneys can explore any doubts and inconsistencies within the DRE findings, such as discrete indicators that the

⁴⁷ A DRE’s opinion corroborated by toxicology based on a urine sample will still be admissible, but toxicology based on a blood sample would be evidentially stronger.

DRE found or did not find and whether they could be consistent or inconsistent with several categories about which the DRE did or did not opine. The defense may also show that there are benign medical or other reasons why a driver may appear impaired. Additionally, the defense may call qualified experts who can opine about flaws within the DRE process and urge that the trier of fact ascribe little or no weight to the DRE's testimony. These impeachment techniques are not exclusive; counsel may pursue other avenues to undermine the DRE's opinion within the usual boundaries of the rules of evidence.

D. Jury Instructions

Most of the time, DRE officers will be testifying before municipal or Superior Court judges in non-jury proceedings. For some cases, however, such as vehicular homicide prosecutions, the State may call a DRE to support its case in a jury trial. In such jury trials, it may be beneficial for the court to provide jurors with an explanatory instruction about the DRE evidence, such as the consistency limitation. We refer this subject to the Model Criminal Jury Charges Committee for its consideration of a model charge on this subject.

We respectfully decline to adopt the other safeguards advocated by the Public Defender in its fallback argument. We do so without foreclosing future

cases in which such proposals can be re-evaluated after the measures prescribed by this opinion have been implemented.⁴⁸

VII. Impact on the State's Burden of Proof

Having set forth those general principles guiding the admissibility of DRE evidence, we briefly address how they can affect the State's burden of proof in DUID cases. In doing so, we reaffirm and extend the guidance provided in Bealor, 187 N.J. at 590-91.

When assessing the proofs, trial judges must consider the evidential ramifications of the presence or absence of a toxicology report under Step 12. They must also consider whether such a report, if one exists, corroborates or conflicts with the DRE's consistency opinion under Step 11.

A positive DRE opinion at Step 11, though admissible under N.J.R.E. 702 subject to the strictures prescribed today, is not dispositive of a driver's guilt of DUID. Unlike a BAC reading of .08% or more in a drunk driving case, the DRE's opinion is not used as a per se test of guilt. Instead, the DRE

⁴⁸ We note our opinion broadly addresses the general admissibility of DRE expert testimony, subject to delineated limitations. We do not address the admissibility of a DRE opinion for every type of drug or polydrug combination, which may be subject to independent scientific study. For example, inhalants, hallucinogens, dissociative analgesics (such as PCP) may not reliably be detected by the indicators listed on the DRE matrix because they have only been rarely found in the retrospective data that has been studied. See SM Report I at 259, 262. Those drug-specific issues are not before us, and so we reserve those discrete questions for another day.

testimony is just one part of the evidence as a whole, and it can be amplified or rebutted.

We note the State would have a much steeper burden to prove a driver's guilt when it lacks corroborating proof from a toxicology report. Although we do not require the completion of Step 12 when it is not feasible, we anticipate that prosecutors will have considerable incentives to obtain corroborating toxicology evidence before they pursue these cases.

VIII. Conclusion

This is a complicated appeal, which has generated over several years an enormous record, with well-presented arguments by counsel and viewpoints on the protocol from experts on both sides. The reliability of DRE evidence is surely a controversial and difficult subject.

We conclude from our de novo review that such DRE evidence is sufficiently reliable under an analysis of the Daubert factors and can be admitted for certain purposes. But we also have imposed important limitations that recognize legitimate concerns about such evidence.

The DRE protocol's function in identifying categories of drugs that a driver may have ingested has significant, albeit impeachable, evidential value. Although imperfect, the DRE protocol is a useful tool that can be helpful to the trier of fact in the search for truth. See N.J.R.E. 702. The record here does not

justify discarding it, as the dissent would mandate. The total exclusion of all DRE expert testimony advocated by the Public Defender and defense amici, and reliance instead on non-standardized lay observations of a driver and a toxicology report, could produce less reliable, rather than more reliable, outcomes.

We presume that researchers will continue to study the efficacy of the DRE methodology, and we do not foreclose future litigation with appropriate testimony to re-examine it. Further, under our Rule 702 jurisprudence, trial judges still have a gatekeeping responsibility in ensuring that DRE expert witnesses demonstrate that they have reliably applied the methodology under the framework we have laid out today.

For these reasons, the reports and findings of the Special Master concerning the admissibility of DRE evidence are adopted as modified. Because Olenowski's convictions were based upon DRE testimony that did not adhere to the guidelines we have set forth today, we posthumously vacate the judgments entered against him.

We close by reiterating our appreciation to Judge Lisa, who patiently presided over these hearings for forty-two days and issued two thoughtful and detailed reports. Although we have modified some of his conclusions de novo, we agree with many of his observations and analyses. The parties and the

public at large have benefitted immeasurably from his dedicated service on this case.

JUSTICES PATTERSON, SOLOMON, WAINER APTER, and FASCIALE join in JUDGE SABATINO's opinion. JUSTICE PIERRE-LOUIS filed a dissent, in which CHIEF JUSTICE RABNER joins.

State of New Jersey,
Plaintiff-Respondent,

v.

Michael Olenowski,
Defendant-Appellant.

JUSTICE PIERRE-LOUIS, dissenting.

Last term, this Court adopted the principles outlined in Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993), to guide the admission of expert testimony under N.J.R.E. 702 in criminal cases. See State v. Olenowski (Olenowski I), 253 N.J. 133, 139 (2023). By shifting away from the test established in Frye v. United States, 293 F. 1013 (D.C. Cir. 1923), towards the Daubert standard in criminal cases, we embraced “an approach that focuses directly on reliability by evaluating the methodology and reasoning underlying proposed expert testimony.” Olenowski I, 253 N.J. at 138. We acknowledged the criticisms that Frye’s emphasis on a technique’s general acceptance allowed for the admission of evidence that may be scientifically unreliable as long as that methodology is generally accepted. Id. at 150. And we noted that the Frye test does not ensure that “[e]xpert

techniques and modes of analysis . . . ‘have a sufficient scientific basis to produce uniform and reasonably reliable results.’” Id. at 150 (quoting State v. Kelly, 97 N.J. 178, 210 (1984)). We therefore adopted the Daubert standard in criminal cases as a means to ensure reliability through concentration on “the soundness of the methodology used to validate a scientific theory or technique, the strength of the reasoning underlying it, and the accuracy of the theory or technique in practice.” Ibid. (emphasis added). This Court determined that the Daubert analysis’s focus “on testing, peer review, [and] error rates” was the better format to allow judges to gauge the reliability of the technique or theory in question in criminal cases. Id. at 151-52. In short, a desire to ensure the reliability of evidence -- what we viewed as “the heart of the issue” -- motivated our change in standards. Id. at 150.

This case presents the Court’s first opportunity to apply the Daubert standard to a criminal matter and asks us to determine the admissibility of expert testimony related to the standardized 12-step Drug Recognition Expert (DRE) protocol used to determine whether a driver is impaired. Under Olenowski I, today’s decision should therefore focus on the objective testability and soundness of the methodology in question. Yet the majority opinion discounts legitimate concerns about the reliability and accuracy of the DRE protocol and upholds the admission of DRE evidence despite

acknowledging that “the factors of testability and false positive error rate are largely inconclusive” and that “DRE testimony does not, in and of itself, establish impairment.” Ante at ___ (slip op. at 6, 95). The bench and bar will undoubtedly wonder why the Court adopted a standard to promote reliability only to downplay reliability in its first application of that standard, and the precedent the majority opinion sets for restyling and adjusting the Daubert factors to fit particular circumstances is concerning in light of the divergent trial court results that may follow.

Here, because the State is the party seeking to admit DRE evidence, it must carry the burden to “clearly establish” that the testimony is sufficiently reliable under N.J.R.E. 702. See State v. Cassidy, 235 N.J. 482, 492 (2018). The State has not met its burden. Most importantly, the DRE protocol unacceptably fails under the first and third Daubert factors -- testability and error rate -- the two elements of Daubert that truly distinguish the standard from the Frye test. Because I find that DRE testimony cannot meet the scrutiny demanded pursuant to the Daubert test, I respectfully dissent.

I.

Daubert instructs courts to conduct a “preliminary assessment of whether the reasoning or methodology underlying the [expert] testimony is scientifically valid and of whether that reasoning or methodology properly can

be applied to the facts in issue.” 509 U.S. at 592-93. Daubert provides the following non-exclusive list of four factors to guide judges in their admissibility analysis:

- (1) whether the scientific theory or technique can be, or has been, tested;
- (2) whether it “has been subjected to peer review and publication”;
- (3) “the known or potential rate of error” as well as the existence of standards governing the operation of the particular scientific technique; and
- (4) general acceptance in the relevant scientific community.

[Olenowski I, 253 N.J. at 147 (quoting Daubert, 509 U.S. at 593-94).]

This Court has affirmed that this framework “applies not only to testimony based on ‘scientific’ knowledge, but also to testimony based on ‘technical’ and ‘other specialized’ knowledge.” Id. at 148 (quoting Kumho Tire Co., Ltd. v. Carmichael, 526 U.S. 137, 141 (1999)).

The Daubert factors “provide a helpful -- but not necessary or definitive -- guide for our courts to consider when performing their gatekeeper role concerning the admission of expert testimony.” In re Accutane Litig., 234 N.J. 340, 398-99 (2018). But the fact that the Daubert factors are not dispositive does not mean that their helpfulness does not derive from their structure and

the rigor they collectively require. The Daubert test rightfully places the emphasis on important factors to be considered in assessing the reliability of expert evidence through the lens of validating the theory or technique by testing its accuracy, among other considerations. Compressing those important factors undercuts their efficacy and transforms the test we adopted precisely for its rigor into an approval based on judicial acceptance rather than reliability.

Unlike the majority, I decline to reorganize and reformulate the Daubert factors in applying them to the DRE protocol in this case. I instead proceed by analyzing the factors in the same manner in which they have been routinely analyzed in the three decades since the Supreme Court decided Daubert.

A.

The first Daubert factor asks whether a scientific theory or technique “can be (and has been) tested.” 509 U.S. at 593. It is fitting that testability is the first factor to consider in this analysis because testability goes to the heart of determining whether the scientific or other specialized knowledge at issue is reliable. As the Supreme Court noted, testability is ordinarily a “key question” in determining reliability. Ibid. Indeed, I view the testability factor as the starting point upon which many of the other factors and considerations regarding reliability of the methodology rest. Studies based upon an

inadequately tested methodology or technique, whether peer reviewed or not, would be a curious foundation upon which a determination of reliability is made. The same applies to the error rate and standards query in factor three of Daubert. If a method does not lend itself to accurate testing, how can any reliance be placed upon the error rates that the flawed testing produced? Additionally, standards are certainly essential for ensuring consistency in any technique or methodology, but if those standards are based on a theory that is potentially unreliable, even the most stringent standards would fail to ensure reliability.

It is undisputed that the State presented no studies assessing the ability of DREs to identify “impairment” -- that is, whether a driver was impaired by a particular class of drugs or combination of drugs at the time of testing. At best, the studies cited by the State and the majority compare DRE opinions to toxicology, but toxicology does not establish impairment, let alone impairment caused by drugs. The State further acknowledges the lack of any “gold standard” against which studies can compare DRE results with scientific testing. The State’s own witness, Bridget D. Verdino, M.S., acknowledged that a positive toxicology report from a blood or urine sample “infers use but not necessarily recent use or impairment.” Verdino explained that data from a blood sample can be hard to interpret because of differences in individuals’

metabolization rates and tolerance, and because different ingestion methods lead to different levels and rates of impairment.

Testing urine does not provide accurate testing capabilities either. Despite being employed in 90% of drug tests for suspected drug-impaired-driving in New Jersey, urine samples are even less indicative of impairment than blood samples. According to State witness Dr. Lewis Nelson, M.D., urine “tends not to reflect the clinical conditions at the time” because “urine concentration, urine volume, all change all the time, and because . . . urine tends to concentrate drug.” Although toxicology is used in the studies as a proxy for impairment, Dr. Nelson testified that “most drugs last in the urine . . . for about three days after drug use.” Given such a broad timeframe, toxicology can hardly be viewed as an accurate indicator of impairment at any given moment. Since many of the symptoms detected during the DRE examination may be the result of non-drug causes of impairment (e.g., bipolar disorder, diabetes, head trauma, seizures), reliance on studies that do not even measure “impairment” risks criminalizing non-drug symptoms that are purportedly confirmed via toxicology testing.

Furthermore, the majority, the State, and the Special Master, Judge Lisa, all agree that all studies of DRE protocol have additional “inherent shortcomings and limitations” beyond the lack of a “gold standard,” including

problems gathering a representative sample, an inability to conduct double-blind testing, and the impossibility of replicating dosing levels and multi-drug use. As the majority notes, “the studies assessed populations with an extremely high prevalence of drug-positivity and a low prevalence of drug-negativity” -- such “an inherently skewed composition of samples means that a reliable error rate, particularly a false positive rate, might not be ascertained.” Ante at ___ (slip op. at 75). Those admitted limitations make it impossible to confidently say that the studies conducted on the DRE protocol present an accurate picture of whether the methodology can be tested and, if so, that the methodology is accurate in determining driver impairment.

Because Daubert makes no distinction between scientific evidence and expert evidence based upon specialized knowledge, there remains an expectation under Daubert that the validity of the DRE protocol will stand up to scrutiny via testing. Although it is true that some of the limitations of testing the DRE protocol are the result of constitutional and practical considerations beyond the State’s control, that does not justify resignation to the fiction that an untestable methodology is apparently good enough and should be deemed reliable despite evidence to the contrary.

In light of those concerns, I find that the first Daubert factor of testability, which I view as the heart of the Daubert analysis, weighs heavily against admitting DRE evidence.

B.

The next factor is whether the DRE protocol “has been subjected to peer review and publication.” Daubert, 509 U.S. at 593.

In the first and second Special Master reports, Judge Lisa noted that the two laboratory studies in the record were only marginally useful because they suffered from serious limitations in their designs that “rendered the data gleaned unhelpful or distorted if used as a measure for the accuracy of the portions of the truncated DRE protocol that was administered.” Special Master’s Report of Findings and Conclusions of Law 283 (Aug. 22, 2022) (SM Report I). In his second Special Master report, Judge Lisa noted that he “found the Beirness/Canada and Vaillancourt studies most relevant and useful,” although he conceded that the studies have “inherent limitations that cannot be avoided in actual law enforcement scenarios.” Special Master’s Supplemental Report of Findings and Conclusions of Law 27-28 (April 13, 2023) (SM Report II).

Despite the conceded concerns from Judge Lisa and all parties regarding the limitations of the studies, the majority concludes that the peer reviewed

studies presented on the effectiveness of DRE protocol -- specifically the Vaillancourt, Beirness/Canada, and Hardin studies -- support the reliability and admission of DRE evidence. Ante at ____ (slip op. at 73, 76). The studies, however, scarcely inspire confidence that the testing conducted presents a complete picture of the accuracy of the DRE protocol in practice. For instance, the Vaillancourt study, which used urine testing to measure the DREs' accuracy, candidly stated that it can draw "no direct link between [the study's] analyses results and impairment." Lucie Vaillancourt et al., Drugs and Driving Prior to Cannabis Legalization: A 5-Year Review from DECP (DRE) Cases in the Province of Quebec, Canada, 149 Accident Analysis & Prevention, Jan. 2021, at 1, 4. In the Hardin study, subjects were only tested after the DREs "felt that the subject was under the influence," meaning the sample size was not representative of the general public. Glenn G. Hardin et al., Minn. Dep't of Pub. Safety, Minnesota Corroboration Study: DRE Opinions and Toxicology Evaluations 1 (1993). The Beirness/Canada study suffers from similar flaws. Douglas Beirness et al., The Accuracy of Evaluations By Drug Recognition Experts in Canada, 42 Can. Soc. Forensic Sci. J. 75 (2009).

Further, the cited studies and articles only compared DRE assessments with toxicological results. Although toxicology testing may provide some

insight, these articles do not inform whether DREs can reliably identify drug-impaired drivers -- nor do they claim to do so. The majority concludes that “despite their recognized limitations . . . the two dozen studies presented in the record and considered by the Special Master are sufficient to meet the Daubert factor of publication and peer review.” Ante at ____ (slip op. at 76).

In light of the studies’ shortcomings, the majority’s conclusion that this second factor weighs in favor of admitting DRE evidence parallels the logic from one jurisdiction that seemingly interpreted the second factor to inquire only about the existence of specialized literature, not the credibility of the conclusions reached by that literature. See State v. Sampson, 6 P.3d 543, 549-50, 556 (Or. Ct. App. 2000) (“The difficulty with defendant’s argument is that it attacks the credibility of the literature bolstering the reliability of the DRE protocol, not its existence. Furthermore, [the defendant] does not cite peer reviewed articles that have effectively ‘discredited the underlying theory’ of the DRE protocol.” (quoting State v. Lyons, 924 P.2d 802, 814 (Or. 1996))).¹

¹ The Oregon Court of Appeals has since held that where only the first eleven steps of the DRE protocol were conducted and no toxicological sample was taken, a DRE could not testify regarding the DRE protocol, but could testify as a specialized expert as to whether an individual was under the influence of alcohol or a controlled substance. State v. Rambo, 279 P.3d 361, 366-67 (Or. Ct. App. 2012).

Our Court, however, should not take the position that Daubert's second factor is satisfied simply by the existence of peer reviewed publications, notwithstanding the admitted issues with the studies contained therein. To do so would diminish the purpose of this factor. It cannot be that Daubert sought to simply identify whether peer review articles exist regardless of the content and conclusions of those publications. The studies relied on here have significant limitations, as identified by all parties. If the purpose of the second Daubert factor is not to carefully examine the studies for their validity in scrutinizing the technique at issue, simply being satisfied with the existence of peer reviewed publications informs an assessing court of nothing regarding the reliability of the evidence sought to be admitted. As the Supreme Court noted in Daubert, the existence of peer review alone "does not necessarily correlate with reliability." 509 U.S. at 593. It is the substance of that peer review that matters, and here, peer review evidences flaws in the ability to accurately scrutinize DRE evidence.

Because the studies and publications in this case suffer from significant shortcomings in the ability to test the DRE protocol and do not study the accuracy of the protocol regarding impairment from drug use, the second factor also weighs against the admission of DRE evidence.

C.

The third Daubert factor instructs the Court to consider “the known or potential rate of error” as well as the existence of standards governing the operation of the particular scientific technique. 509 U.S. at 594. The majority concedes the error rate analyses are “largely inconclusive.” Ante at ___ (slip op. at 87). Because the error rate here is, at best, unknown and, at worst, unacceptably high, this factor strongly weighs against admission of DRE evidence. The majority concedes as much in declining to adopt either the State’s or the Office of the Public Defender’s (OPD) asserted error rates in examining the 2017-2018 retrospective New Jersey DRE data.

Judge Lisa’s calculation of the error rate is flawed simply based on the mathematical calculation employed. By dividing the total number of false positives from the retrospective data by the overall number of positive toxicology tests -- a questionable denominator -- the State and Judge Lisa calculate the error rate to be 3.2%. SM Report I 201. That is an unhelpful data point because it does not indicate how often DREs incorrectly identify subjects as drug-positive, which is the entire purpose of the third Daubert factor. The OPD takes the same data set and considers the sample of negative toxicology tests, noting that 78.1% of those who tested negative for drugs were nonetheless identified as drug-positive through the DRE protocol. If the true

false positive rate is in fact 78.1%, it would obviously be unacceptably and alarmingly high.

Both proposed error rates, however, should be assessed with caution for several reasons. First, as Judge Lisa stressed in the second Special Master report, some unknown number of the negative toxicology results can be attributed to the limitations of toxicology testing since some substances, such as fentanyl, can go undetected. SM Report II 30-31. Second and inversely, as the OPD stresses, some unknown number of the toxicology positives may be entirely irrelevant in analyzing current impairment because, as previously noted, a person can test positive for a drug days after ingesting it.

Third, the data presents issues of confirmation bias given that suspects admitted to drug use in 87% of the non-training evaluations. As expert State witness Dr. Enrique Schisterman, Ph.D., M.A., explained, the DREs were over 1,800% more likely than chance to correctly predict the toxicology result in the event of an admission. Without a suspect admission, however, the DRE was only 90% more likely to correctly predict the toxicology result. Fourth, the prevalence of urine testing in New Jersey raises legitimate doubt as to the accuracy of the figures because “the quantitation or the amount of drug in urine does not reflect . . . anything about what’s in the brain,” according to State expert Verdino.

Lastly, as the State conceded at oral argument, not all 12 steps of the protocol are administered during each DRE evaluation. The data thus comprises an endless number of conceivable fact patterns and combinations of steps administered that may differ in important yet unknown ways. In short, the DRE protocol is an imprecise and non-uniform mix of medical-based assessments and regular police practices, and it would be a mistake to draw empirical conclusions from a data set encompassing such a wide range of scenarios. For all those reasons, it is impossible to calculate the true error rate with confidence.

As we made clear in Olenowski I, Daubert's focus on the soundness of the methodology at issue "matter[s] in this and other cases -- for example, when it soon comes time to directly evaluate error rates associated with DRE evidence." 253 N.J. at 154. The majority concedes that "the testability and false-positive error rate aspects of the Daubert analysis are largely inconclusive" for the DRE protocol "due to variables that are either not controllable or not known." Ante at ___ (slip op. at 87). That fact should be given significant weight in the Court's analysis of reliability.

With regard to the existence-of-standards component of the third Daubert factor, I agree that DREs are subject to extensive training and certification procedures and that the 12-step protocol is a method significantly

more detailed than the one found to be insufficient in Kuhmo Tire. The State argued that the DRE standards themselves are rigorous, but conceded at oral argument that DREs do not necessarily follow the 12-step protocol in every instance and that there is no uniformity among DREs or in the protocol for determining when or why certain steps of the protocol will be disregarded. Even if detailed standards are in place, those standards cannot be expected to produce reasonably uniform or consistent results if DREs can exclude or include steps in the course of their assessment without limitation or explanation.

And, once again, testability is the cornerstone of the Daubert analysis. It may be that DREs are subject to rigorous training and certification procedures, but it begs the question whether all that training is for naught if the methodology upon which DREs are trained is unsound. DREs can train and learn the protocol extensively, but if the protocol that they are learning to administer is flawed or, in this instance, may have an unacceptably high error rate, all the training in the world cannot fix an untested or untestable methodology.

Accordingly, the third factor also weighs against the admission of DRE evidence.

D.

The fourth and final Daubert factor asks the Court to consider whether the DRE protocol has been generally accepted by the relevant scientific community. 509 U.S. at 594. This question was the crux of the Frye test that we abandoned last term in Olenowski I. As such, although general acceptance is one factor to consider under the Daubert analysis, the factor should not outweigh the others, particularly the factors that focus on testing and error rates. Furthermore, this Court should not allow the historic use of DRE evidence to dominate the new Daubert analysis.

The majority notes that despite its imperfections, the “DRE protocol has been widely and regularly used across this country and abroad” and gives consideration to the fact that “[n]o state has discontinued it, and no state’s highest court has nullified it.” Ante at ____ (slip op. at 92). I agree that the DRE protocol has been utilized for many years and is generally accepted, so this factor weighs in favor of admitting DRE testimony. But the mere fact that similar police procedures have been used for years does not mean the DRE protocol is reliable under Daubert. In fact, even under Frye, this Court did not hesitate to invalidate unreliable methods after years of use by State and law enforcement officials. See State v. J.L.G., 234 N.J. 265, 272, 288 (2018) (holding that, after “decades” of widespread acceptance and use by 40 states

for some purpose, “it is no longer possible to conclude that [“Child Sexual Abuse Accommodation Syndrome”] has a sufficiently reliable basis in science to be the subject of expert testimony” in light of developments in psychological research); Cassidy, 235 N.J. at 486-87, 498 (holding that the use of Alcotest devices calibrated without the use of a NIST-traceable digital thermometer in the calibration process undermines the reliability of the Alcotest); Windmere, Inc. v. Int’l Ins. Co., 105 N.J. 373, 375, 386 (1987) (questioning the reliability of voiceprint evidence).

Any deference to historical practices is even less relevant now in criminal cases under the new Daubert framework. Judge Lisa herself acknowledged that under Daubert, courts “now directly assess reliability” and ask whether “experts in the relevant field would accept the DRE protocol as reliable if they were aware of” all relevant facts. SM Report II 41-42. As the discussion of the first three factors reveals, there are flaws inherent in testing the protocol and determining whether DREs can accurately detect drug impairment. Although the majority concludes that this factor weighs in favor of admission, and I agree that is true, this Court should hardly overlook the significant weaknesses of this evidence as borne out by the other three Daubert factors.

II.

Considering the above factors, I find that the State has failed to establish the reliability of the DRE protocol under Daubert given that three of the four factors weigh against admissibility. The lack of scientifically rigorous testing and the unknown error rate -- the two pillars of the Daubert framework -- make it impossible for me to reach any other conclusion.

The majority opinion reads, however, as if it is simply one factor that is inconclusive and weighs against admissibility, having “reorganize[d]” and combined certain factors. Ante at ___ (slip op. at 62). The majority’s reorganization of the factors -- the combining of factors one and three, and the separating out of a subfactor -- not only obscures the distinct factors put in place by Daubert 30 years ago, but also minimizes the importance of the testability and error-rate analyses. Instead of making clear that the two factors that go to the heart of the scientific query are inconclusive, the majority opinion buries that fact by combining those important factors into one.

The Supreme Court explained that while Daubert’s reliability inquiry is “a flexible one,” the Daubert framework very intentionally placed emphasis on testability and error rates. 509 U.S. at 594. The majority’s application of the factors subverts that emphasis and fundamentally alters Daubert in the process. And this is the first case in which this Court analyzes the admission of expert

testimony in a criminal case pursuant to Daubert. I query whether the majority opinion's reorganization of the Daubert factors in this matter signals that trial and appellate courts should do the same going forward. Endorsing and eroding a standard in a single opinion is not the clear guidance the bench and the bar require. I believe our adoption of Daubert in criminal cases was intended to be just that -- adoption of the four well-established, non-exclusive factors without obscuring one or more in a way that fits with the facts or argument in a particular case.

When this Court makes an important statement of law, that statement should be clear and consistent to help guide courts throughout the Judiciary. Accordingly, our recent adoption of Daubert in criminal cases must be accompanied by a thorough and accurate Daubert analysis; otherwise, the purpose of our holding in Olenowski I is thwarted.

The majority opinion thoughtfully attempts to remedy the apparent unreliability and weaknesses of DRE evidence by imposing “major safeguards” and guardrails on the use of such evidence.

It is not clear, first, that the majority's guardrails could “fix” the problems they reflect. At Step 11, for example, the DRE forms a “final opinion based on the totality of the examination.” Ante at ___ (slip op. at 17). The majority finds that “the DRE's opinion at Step 11” does not “establish

causation, i.e., that particular drugs or categories of drugs were ingested by the driver and caused the driver to be impaired.” Ante at ____ (slip op. at 98). The majority holds that a better practice is to allow a DRE to testify at Step 11 only that the results of the examination are “‘consistent with’ the driver’s usage of certain categories of drugs” because consistency does not equal causation. Ante at ____ (slip op. at 99). The reality, however, is that such testimony, coming from a witness qualified before a judge or jury as an “expert,” will undoubtedly carry a significant amount of gravitas, notwithstanding the terminology used. If the DRE testifies that, based on the examination conducted, results are consistent with a defendant’s drug usage, that is, practically speaking, all that would be needed for a judge or jury to find causation.

But more fundamental than any shortcomings of the proposed guardrails is the simple fact that the Court should not be fashioning such guardrails in the first place. The majority carefully attempts to remedy the unreliable nature of the DRE protocol, but it is not within the Court’s province to fix a scientific technique or method to attempt to make it better when the answer to the reliability question is anything other than “yes.” The query under the Daubert framework is whether a scientific technique or methodology is reliable. If, after applying the Daubert factors and any other factors that may be relevant, a

particular methodology is found to be unreliable, neither this nor any court can fix that flaw and attempt to make the methodology reliable. No amount of gymnastics will make an unreliable scientific theory or specialized knowledge technique reliable simply by imposing guardrails.²

The majority emphasizes the constitutional and ethical barriers to creating a study that rigorously examines the reliability of the DRE protocol. Those obstacles are true and legitimate, but that alone is not reason to ignore the important facets of the Daubert test and approve a technique that has not been -- and apparently cannot be -- tested or shown to have an acceptably low error rate.

Furthermore, despite the concerns expressed by the State, there is no reason to believe that admitting unreliable DRE evidence is so essential to convicting impaired drivers that we should resign ourselves to approving a protocol that has not been effectively tested. The majority is undoubtedly

² In People v. Bowden, the Court of Appeals of Michigan, by a 2-1 majority, held that DRE testimony was inadmissible under Michigan Rule of Evidence 702 because the State failed to establish reliability under Daubert. ___ N.W.2d ___ (Mich. Ct. App. 2022) (slip op. at 10), appeal denied, 994 N.W.2d 776 (Mich. 2023). The majority notes that it agrees with the premise of Bowden that “DRE testimony does not, in and of itself, establish impairment,” but holds that the testimony is reliably admissible “for a less ambitious purpose, and with critical safeguards.” Id. at ___ (slip op. at 95). As discussed, I disagree.

correct that drug-impaired driving is an important public safety issue to be taken seriously by law enforcement. Ante at ____ (slip op. at 4). Without DRE evidence, the State would still have the ability to present other evidence of the arrest and encounter with the defendant. As the majority explains, a driver's impairment must be proven with evidence independent of DRE testimony, such as "specific factual observation of impaired behavior by the arresting officer or the DRE, a driver's admissions, information from a passenger or other observer about the driver's recent drug use, or drugs or paraphernalia found in the vehicle." Ante at ____ (slip op. at 98-99). The prosecution could also include other evidence of the defendant driving erratically, including dashcam video footage. Even without a DRE expert's testimony, that evidence would provide law enforcement with the tools needed to ensure New Jersey roads are safe for drivers.

III.

Under Daubert, it is not this Court's charge to create safeguards to try to preserve the use of techniques that cannot withstand rigorous scrutiny. Our task instead is to ensure that if evidence is given the weight of an expert's endorsement, that evidence has "a sufficient scientific basis to produce uniform and reasonably reliable results" because "[a]n expert opinion that is not reliable is of no assistance to anyone." Olenowski I, 253 N.J. at 150

(alteration in original) (quoting Kelly, 97 N.J. at 209-10). The means to ensure that is to faithfully apply, not reconfigure and reduce, the Daubert factors that emphasize reliability. By altering the Daubert factors here, the majority not only reaches a determination of reliability that is not supported by the test, it also upends the clear guidance this Court set out to provide in Olenowski I regarding placing the focus of these expert reliability determinations on testing, peer review, and error rates.

Because, in my view, the State did not meet its burden of clearly establishing under Daubert that DRE evidence is reliable and many questions remain about the reliability of DRE evidence, I would hold that DRE evidence is not admissible under N.J.R.E. 702.

For those reasons, I respectfully dissent.