(18)
RECEIVED
10/22

BEFORE THE
DEPARTMENT OF HEALTH AND HUMAN SERVICES

Substance Abuse and Mental Health Services Administration	)	
In the Matter of:	)	
MANDATORY GUIDELINES FOR WORPLACE DRUG TESTING PROGRAMS	)	FR Doc. 01-20945

COMMENTS OF THE AIR LINE PILOTS ASSOCIATION AND THE TRANSPORTATION TRADES DEPARTMENT OF THE AFL-CIO

CAPTAIN DUANE E. WOERTH, President
JONATHAN A. COHEN, Director, Legal Dept.
SUZANNE L. KALFUS, Senior Attorney, Legal Dept.
AIR LINE PILOTS ASSOCIATION
535 Herndon Parkway
Herndon, VA 20170
Telephone: (703) 689-4324

EDWARD WYTKIND, Executive Director
TRANSPORTATION TRADES DEPARTMENT, AFL-CIO
888 16th Street, N.W., Suite 650
Washington, D.C. 20006
Telephone: (202) 628-9262

October 22, 2001

TABLE OF CONTENTS

Introduction and Summary 1

I. VALIDITY TESTING 2

A. The Accuracy Of Any Required Or Permitted Validity Test Must Be Guaranteed. 4

1. Any Screening Tests To Detect Adulterants or Measure Urine Dilution Should Be Permitted Only After The Accuracy And Reliability Of Such Tests Have Been Determined, A Process Accomplished By FDA Clearance Or Approval. 4

2. It Is Essential That Any Required Validity Test Be Confirmed With A Second Test That Utilizes A Different Physical Or Chemical Property Than The Initial Test. 6

3. The Non-Specific Testing Proposed For Oxidizing Adulterants Fails To Meet Scientific Standards. 10

4. Quality Controls Are A Vital Part Of Any Testing Scheme. 12

B. The Claim That Urine Reported With Creatinine Of Less Than Or Equal To 5 mg/dL And Specific Gravity Of Less Than Or Equal To 1.001 Cannot Be Considered Human Urine Is Contrary To Actual Experience And Study Results. 12

1. It Is Necessary To Account For Laboratory And Other Errors 13

2. Humans Do Produce Urine With Creatinine And Specific Gravity Below The Proposed Regulatory Cutoffs 14

C. Additional Procedural Protections Should Be Added To Support The Integrity Of Any Validity Tests. 22

BEFORE THE
DEPARTMENT OF HEALTH AND HUMAN SERVICES

Substance Abuse and Mental Health Services Administration	
In the Matter of:	
MANDATORY GUIDELINES FOR WORPLACE DRUG TESTING PROGRAMS	**FR Doc. 01-20945**

COMMENTS OF THE AIR LINE PILOTS ASSOCIATION AND THE TRANSPORTATION TRADES DEPARTMENT OF THE AFL-CIO

Introduction and Summary

The Air Line Pilots Association ("ALPA") is the principal labor union representing the nation's commercial pilots. It represents more than 67,000 pilots at 47 airlines in the United States and Canada. The Transportation Trades Department of the AFL-CIO ("TTD") is an organization of the AFL-CIO comprised of 33 unions that represent millions of employees in transportation industries.[1] Some of the unions in TTD also represent federal employees. These affiliated unions represent employees subject to DOT regulated drug testing (governed by 49 CFR Part 40) as well as federal agency drug testing, all of whom are affected by the proposed regulations. ALPA submits these comments on its own behalf and on behalf of TTD in response to the above-captioned

[1] The unions represented by TTD are listed in attachment 1 to these Comments.

Notice of Proposed Rulemaking ("NPRM"). Mandatory Guidelines for Federal Workplace Drug Testing Programs, 66 Fed. Reg. 43,876 (August 21, 2001).

ALPA and TTD maintain their opposition to mandatory "validity" testing in the manner in which HHS and DOT seek to implement it. While we appreciate the improvements that have been made to the procedures, we are still concerned about the lack of certain fundamental safeguards, the failure to meet acceptable scientific standards and the continued risk of innocent employees being improperly deemed to be rule violators.

Experience over the past few years has shown that innocent employees have been falsely reported to have adulterated or substituted their urine samples, and have been terminated from their jobs as a result. Such reports have resulted both from laboratory error – including egregious misconduct and failing to comply with applicable quality control standards – and from a small number of otherwise healthy individuals who produce ultra-dilute urine and who have had creatinine and specific gravity reported below the regulatory cutoff levels. The dire consequences to an employee of being reported as having tampered with his or her urine specimen necessitates that any validity testing protect such innocent individuals as well as meet the highest forensic and due process standards.

I. VALIDITY TESTING

Although HHS cites figures of failed validity tests over the past few years, it fails to distinguish between tests of applicants versus those of actual employees. There is a significant difference in test results between the two categories. From what we have

seen, and what experienced collectors have reported, the overwhelming number of problematic test results has been on preemployment tests – not those of current employees.

Moreover, the percentage of specimens reportedly "adulterated" or "substituted" is negligible, at worst. See 66 Fed. Reg. at 43,877 (allegedly 6,440 adulterated specimens and 2,821 "substituted" specimens out of 13 million specimens). If these numbers are adjusted to consider tests of only current employees, the problematic results would be virtually nil.

It should also be noted that "Bayes Theorem" postulates that when the prevalence of a condition being tested is quite low, the probability that a particular result is true is greatly decreased. See R.S. Galen and S.R. Gambino, Beyond Normality: The Predictive Value and Efficiency of Medical Diagnoses (1975). This respected theorem has shown the statistical risks of increased false results in a population, recognizing that all testing procedures are less than 100% perfect. Based on the HHS figures, the extremely low prevalence of "adulterated" results (.049%) and the even lower prevalence of "substituted" results (.021%), would cause the number of false positives to be much greater than the number of true positive readings. Id.

We also submit that there is the least justification and the greatest concern about the so-called "substitution" testing. Even considering reported results that include applicants, the number of "substituted" specimens is far, far less than reported "adulterants." Id. Moreover, as we show below, the "substitution" cut-offs include otherwise innocent employees, as shown by actual experience and the existing scientific studies.

And finally, while improvements have been made to the testing for adulterants – such as some requirement for "true" confirmation tests – such protections remain wholly lacking for creatinine and specific gravity tests. In sum, while we believe the government has failed to produce evidence justifying any mandatory validity testing, the deficiencies involved in the required testing for urine dilution overwhelmingly mandate a serious reconsideration of this aspect of the proposed regulations.

A. The Accuracy Of Any Required Or Permitted Validity Test Must Be Guaranteed.

1. Any Screening Tests To Detect Adulterants Or Measure Urine Dilution Should Be Permitted Only After The Accuracy And Reliability Of Such Tests Have Been Determined, A Process Accomplished By FDA Clearance Or Approval.

The proposed screening tests are gross tools that have been approved or cleared largely for clinical purposes but do not have the same degree of precision as is necessary for workplace employee testing. Whether a patient's urine has specific gravity of 1.001 or 1.002 is of little consequence in the context of patient care. Likewise, a difference between 5 mg/dL and 6 mg/dL of creatinine is not clinically significant. Under the mandatory validity testing program, however, such differences determine whether an individual's livelihood is terminated. Thus, while these tests may be sufficiently accurate or reliable for clinical care of patients, they do not meet the requisite forensic standards for a government imposed employee testing program.

FDA review can address some of these concerns. Such review is especially important for validity tests for a number of reasons. First, the proposed regulations allow final results – results upon which an employee's career will depend – to be based on

screening tests alone. Accordingly, it is all the more important to identify the error rate of the screening tests. Second, there are no restrictions or limitations on the particular screens or assays that may be used. Unlike NHTSA approval of alcohol testing devices or HHS certification of laboratories, there is no such government certification of these tests. Nor is information about the content or composition of the reagents used in the screening tests necessarily available to the public without FDA review. Such information is essential in order to evaluate and eliminate alternative explanations or causes of results reported as positive. Such information is also necessary to enable meaningful MRO review.

The FDA plays a vital role in reviewing commercial tests involving health, food and drugs, as well as commercial tests sold for diagnostic purposes. The process by which the FDA clears or approves such tests involves having the manufacturer prove that the test accurately and reliably does what it purports to do. Such review identifies the diagnostic sensitivity (true positive rate), diagnostic specificity (true negative rate), and the predictive value of the detection of the compound (or property) being tested. It also looks at the chemical or scientific basis of the test, identifies the reagents, and inspects the manufacturing process to verify the components of the test. These review processes identify potential causes of false positive and false negative test results. The immunoassays used to screen for drugs and drug metabolites under the DOT drug testing program have long been subject to such FDA oversight.[2]

---

[2] This requirement was previously contained in 14 C.F.R. Section 40.29(e). See 54 Fed. Reg. 49,854, 49,872 (Dec. 1, 1989). It was removed when the regulations were revised in December 2000. See 14 CFR Section 40.87(a), replacing 40.29(e). See also Section 2.4(e)(1) adopted in the HHS Guidelines. 53 Fed. Reg. 11,970, 11,983 (Apr. 11, 1988).

The FDA oversight process identifies the error rate of a test or device, when it is used for the purpose for which the FDA is reviewing it. While some of the screening tests used for validity testing may have been cleared or approved for other purposes – such as to identify bacterial infections, or to test renal function – they have not been assessed for their ability to identify adulterants, validate urine at the cutoff levels, or even measure creatinine in urine. Such FDA review would result in the identification of the predictive probabilities of the tests for the variable being tested for, and ultimately determine the error rate of the test being used. In sum, FDA review is an extremely valuable tool for assuring a high standard for employee testing, and a vital means of obtaining key information about the applicable tests. Such review should be mandated by the regulations.

2. It Is Essential That Any Required Validity Test Be Confirmed With A Second Test That Utilizes A Different Physical Or Chemical Property Than The Initial Test.

As ALPA expressed in its Comments to Docket No. 0ST-99-6578 (Part 40 NPRM), it is a fundamental principle in forensic toxicology that at least two different analytical techniques must be used in order to assure an accurate test result. See attachment 2 at 6, citing J. J. Reese, American Academy of Forensic Sciences Policies on Confirmation; R.H. Cravey and R.C. Baselt, Introduction to Forensic Toxicology (Biomedical Publications, Davis CA 1981); B. Levine, Principles of Forensic Toxicology (1999); American Academy of Forensic Science/Society of Forensic Toxicologists Forensic Toxicology Guidelines. Underlying this tenet is recognition that no test is 100% accurate; therefore the probability of a correct result is increased when the result is confirmed using a different methodology. Or, put another way, the consistent,

corroborative findings of independent tests of the same value or fact increases the probability that the finding is true. The failure to require a confirmation test that utilizes a different testing methodology is a serious procedural flaw, which can cause grossly inaccurate test results.

This significant failing is not cured by repeating the same test on two different aliquots because the second test is not independent and thus would be subject to the same errors and interferences as the first. For example, if there is interference with the screening test – such as from medication or menstrual blood in the urine – such interference would be repeated if the same procedure, method, and/or instrument is used again, regardless of the nature of the testing method, and the number of times it is used.

For this reason, to increase the predictive value of a test to acceptable levels, at least two independent procedures must agree on the result. To be independent, they must be carried out on two different aliquots and they must be based on different physical or chemical properties of the analyte. See e.g., V.R. Spiehler et al., Confirmation and certainty intoxicological screening, Clin. Chem. 34:1535-39 (1988); M. Zweig et al., NCCLS GP 10P, Assessment of clinical sensitivity and specificity of laboratory tests, National Committee for Clinical Laboratory Standards, Villanova, PA (1987); M.H. Zweig et al., Receiver-Operating Characteristic (ROC) plots; a fundamental evaluation tool in clinical medicine, Clin. Chem. 39:561-77 (1993).

The proposed regulations state that such confirmation testing shall be used for adulterants, but then follow that statement with an exception that eviscerates the protection. The two sections of the proposed regulations governing testing for adulterants state: "A confirmatory test . . . shall use a different analytical principle or chemical

reaction than that used for the initial test unless a recognized reference method is used for both the initial and confirmatory test" (emphasis added). See proposed sections 2.5(h)(2) and (j)(2). 66 Fed. Reg. at 43,881-82. This loophole eliminates a true confirmation test if a "recognized reference method" is used.

The term "recognized reference method" is not defined in the regulations and could be subject to various interpretations. But regardless of what testing method is used and how accurate and reliable it is supposed to be, repeating the same test cannot afford the protection that a truly independent test can. A confirmation test is not required because the initial tests are unreliable; all tests used by NLCP laboratories are, or should be, reliable. A confirmation test is required to increase the probability that the result is a true positive or true negative to a level of probability sufficiently certain to stake a person's career upon that reported result.

The proposed rule also states "In some cases both initial and confirmatory validity tests may use the same procedure, instrument or method." See Subpart B, ¶10. (d)(1), 66 Fed. Reg. 43,881. Like the rigorous standards required by law when testing employees' urine samples for the presence of illegal drugs, so too should such equally stringent requirements apply to validity tests which can be similarly career-ending. The increased assurance of true results that comes from confirmation tests should apply to any tests used to measure urine dilution. Accordingly, we strongly urge that these exceptions and loopholes be removed from the rules.

The same need for confirmation exists with respect to tests of creatinine and specific gravity. DOT sought to distinguish these tests claiming there was no need to confirm such results with a test using a different physical or chemical property because

creatinine and specific gravity are unlike the "chemically complex substances" that comprise illegal drugs. See 65 Fed. Reg. 79,462, 79,480 (Dec. 19, 2000).[3] DOT also claimed that because creatinine is expected to be found in urine, that fact somehow makes confirmation testing superfluous. Id. These facts have no relevance to the rationale and need for confirmation testing.

The pertinent issue is not the chemical complexity of creatinine in relation to the chemical complexity of other substances expected to be found in urine, as DOT seems to be saying. What is significant is that simple colorimetric tests, such as those for creatinine, may falsely report a positive result due to various reasons (such as a machine pipette failure, a reagent reaction problem, etc.). Any such problems would not be cured by repeating the same test.

It is also incorrect, as DOT claimed, that an initial creatinine validity test is analogous to a confirmation drug test. Id. Neither from a chemical, scientific, analytical or probability standpoint, does an initial creatinine test accomplish the same goals as a drug confirmation test. Nor does the mere fact that a screening test produces a "quantified" result increase the probability that the results are true – as does an independent confirmation test. How "chemically complex" creatinine is, and the fact that some levels are normally expected to be in urine, have no relation to the function and

[3] In fact, creatinine is more chemically complex than methamphetamine.

value of an independent creatinine confirmation test. The same is true for specific gravity tests.[4]

Nor is there any practical reason to exempt these validity tests from higher scientific standards than other validity tests. Creatinine can be confirmed by chromatographic tests, which are commonly performed by laboratories and readily available. If initial specific gravity tests are done using refractomers, confirmation tests can be done either with a hydrometer or with a balance. These methods are easily done and likewise, readily available. There is simply no justification for exempting any mandated testing, with potentially career-ending consequences, from standards less protective than those applicable to drug testing.

3. The Non-Specific Testing Proposed For Oxidizing Adulterants Fails To Meet Scientific Standards.

In contrast to the testing proposed for specific adulterants, creatinine and specific gravity, HHS is seeking to allow tests for other unidentified "oxidizing" agents that, if present in a sample at any detectable level, requires the test to be reported as an "invalid specimen" – a category that implies employee misconduct and one which subjects that person to directly observed testing. While we understand that HHS is concerned about

[4] Following the creatinine test with a test for specific gravity does not constitute a "true" confirmation test either. Each test is measuring a different property and is subject to its own variations in precision and accuracy. Moreover, specific gravity is neither a reliable confirmation of a creatinine level nor a suitably independent measure of urine diluteness to meet forensic standards. See Spiehler Report (Attachment 1 at attachment 1, p.9) citing W.V.R. Vieweg et al., Psychogenic Polydipsia and Water Intoxication-Concepts That Have Failed. Biol. Psychiatry 20:1308-20 (1985); S.B. Needleman et al., Creatinine Analysis In Single Collection Urine Specimens. J. For. Sci. 37:1125-33 (1992). The variable correlation between specific gravity and creatinine in individual patients has been found to range from 0.618 to 0.935. Id.

identifying new adulterants, the approach suggested has serious problems and should not be adopted.

First, there are no standards governing the types of tests that may (and may not) be used for such general adulterant testing. It is essential to know the chemical composition of any such tests because that will determine what other compounds will react positively to the testing reagent and could cause "false positives." F. Urry et al., Nitrite adulteration of workplace urine drug-testing specimens. Sources and associated concentrations of nitrite in urine and distinction between natural source and adulteration. Journal of Analytic Toxicology 22:89-95 (1998).

It is also important to know the chemical composition of these tests in order to know what properties of the specimen might interfere with the tests. For example, nitrites are detected through a color-producing reagent. Blood in the urine can cause that test to report false positives since the color of the blood can falsely indicate a "positive" result. There are also many oxidants that occur naturally in the environment and in humans, such as blood, feces, bacteria, iron, etc. Knowledge of the basis of a test's oxidizing reaction is essential so that MROs, affected individuals and their representatives can identify the true causes of any such reported positive.

Any such tests should also be subject to the same rigorous standards as those for other validity and drug tests. The accuracy and reliability of the testing methodology should be identified and reviewed before employees suffer any actions as a result of those test results. Additionally, before any test reaction is treated as a true "positive" there must be some showing that the actual quantity of the compound in the urine that caused the reaction is a true reading.

The gross screening tests which, alone, would be a basis for finding a sample "invalid" are not sufficiently precise and accurate to justify this conclusion. We submit that canceling any such test result is the more reasonable approach until testing meets appropriate scientific standards.

4. Quality Controls Are A Vital Part Of Any Testing Scheme.

We are pleased to see that the proposed regulations have included and enhanced the quality control requirements enumerated in NLCP Program Document #37, Notice to HHS Certified Laboratories and Inspectors, Subject: Specimen Validity Testing (July 28, 1999) ("PD 37"). Experience has shown the large number of incorrectly reported results, and the significant number of employees adversely affected when such protocol is not followed. It is absolutely vital that such protections be included in any testing scheme, and we are glad to see them included here.

B. The Claim That Urine Reported With Creatinine Of Less Than Or Equal To 5 mg/dL And Specific Gravity Of Less Than Or Equal To 1.001 Cannot Be Considered Human Urine Is Contrary To Actual Experience And Study Results.

There has been no showing by either HHS or DOT that a urine sample with creatinine and specific gravity below the regulated cutoffs cannot be human urine. For this reason, it is essential that any such cutoffs be used, at most, as only a trigger for the need to obtain additional information, and not as a presumption of individual wrongdoing with the burden placed on the employee to prove his or her "innocence." As recent experience (limited as it has been) has shown, there are several factors that may contribute to a sample being reported as below the cutoffs.

1. It Is Necessary To Account For Laboratory And Other Errors.

It is important to recognize that the levels reported may not accurately reflect the actual levels of creatinine and the specific gravity of the urine. As was shown in the high profile Delta pilot case handled by ALPA, reported levels can be unreliable and inaccurate due to laboratory error. The structure of the DOT rules and the HHS proposed guidelines fails to provide sufficient mechanisms by which a reported result can be overcome where, as there, the reported result is from faulty laboratory work as opposed to "a legitimate medical explanation."

We have previously addressed (and incorporate by reference) the need for employees and their unions to have ready access to information necessary to uncover such laboratory error, both to demonstrate employee innocence and to protect the integrity of the system. (See attachment 3, incorporated by reference).[5] However, it should also be recognized that it is extremely time consuming and exorbitantly expensive to investigate and uncover such laboratory error. The magnitude of that undertaking will likely deter innocent individuals (especially non-represented workers), or result in futile attempts that are unsuccessful due to lack of resources and specialized expertise. That burden should not fall on individual employees, or even their labor unions.

---

[5] Union access to such information should include laboratory oversight and quality assurance documents, such as laboratory proficiency checks, inspection reports and critiques, etc. See e.g., material referenced in Section 3.2(b). 66 Fed. Reg. at 43882. Allowing unions and other interested parties to serve as "watch dogs" is a useful function that fosters the integrity of the program. ALPA's identification of the problems at LabOne that led to a special HHS investigation causing the cancellation of 300 test results, illustrates this point.

(18) p. 16

2. Humans Do Produce Urine With Creatinine And Specific Gravity Below The Proposed Regulatory Cutoffs.

Another reason for a so-called "substituted" result may be that the individual is one of the small minority of those whose body produces ultra-dilute urine. While we agree that the overwhelming majority of individuals will not have urine with measures at these levels, a small percentage of otherwise healthy people (due to their physiology, diet or other factors) will produce readings at or close to the proposed cut offs. And in a program of this magnitude covering millions of workers,[6] even a fraction of a percentage of the covered employees yields a significant number of individuals at risk of losing their careers.

Contrary to the government's claims that specimens below the proposed cutoffs must be treated as non-human urine, experience has shown that individuals producing such urine exist, and have come forward with such samples obtained under direct observation. We have also seen that an individual may have a sample above the cutoffs on one occasion, and then produce urine below the cutoffs (again, under direct observation) on another.

Such cases have likewise been confirmed by the MROs.

> [t]here have been a few cases where MROs have personally observed the production of specimens that have subsequently been reported as substituted. Thus, it does appear that the proposition that it is impossible to physiologically produce substituted urine is not a sustainable one.

Theodore F. Shults, MS, JD, The MRO's Oversight of the Referral Physician, MRO ALERT at 4, Vol. XII, No. 3 (April 2001) ("MRO ALERT").

---

[6] DOT says that its testing program covers 8.34 million employees. 64 Fed. Reg. at 69,093. Millions of federal employees are subject to compulsory testing as well.

The literature reviewed by HHS and DOT (66 Fed. Reg. at 43,877), also fails to prove that individuals cannot produce urine with readings below these cutoffs. The extremely limited value of the studies cited in NLCP: State Of The Science: Update #1, Subject: Urine Specimen Validity Testing: Evaluation of the Data Used to Define a Urine Specimen as Substituted (Feb. 14, 2000)("NLCP #1") (attachment 4) was discussed in detail in ALPA's Comments to the Part 40 NPRM. (Attachment 3 at 25-31). In sum, a review of the data revealed the appalling dearth of "paired data" – data from specimens where both the creatinine level and the specific gravity was measured. Notwithstanding the 45 different papers cited,[7] individual paired data was presented from only eight men and two women, described in four papers. Unquestioningly, so few studies with such limited data are scientifically and statistically inadequate to confirm the proposition that urine with creatinine and specific gravity at or below the cutoff cannot be human urine.

Moreover, many of the remaining studies – none of which actually measured specific gravity and creatinine in the same urine sample – do show that particular individuals' urine had one variable that measured at or below the "substitution" cutoffs. For example, in nine different studies the specific gravity of 20 subjects' urine was at or below the cutoff.[8] Subjects' urine had creatinine at or below the cutoff in other studies.[9]

---

[7] The bibliography references 48 numbered studies but 3 of them were listed twice.

[8] Specific gravity at 1.000 was reported in 12 subjects and 12 specimens in the following studies: E.J. Cone et al., In Vivo Adulteration: Excess Fluid Ingestion Causes False-Negative Marijuana and Cocaine Urine Test Results, Journal of Analytic Toxicology, 22:460-73 (1998); G.M. Homer et al., A Discussion of Creatinine Analysis in Single Collection Urine Specimens, Journal of Forensic Sciences, 38:501-02, (1993); W.V.R. Vieweg et al., Psychogenic Polydipsia and Water Intoxication – Concepts that Have Failed, Biol. Psychiatry, 20:1308-20 (1985); G. Rinard, Water Intoxication, American Journal of Nursing, 89:1635-38 (1989); R.T. Frizzell et al., Hyponatremia and Ultra-marathon Running, JAMA, No. 6, 265:772-74 (1986). Specific gravity of 1.001 was

The more recent paper cited in the NPRM, The Characterization of Human Urine for Specimen Validity Determination in Workplace Drug Testing: A Review, Journal of Analytical Toxicology 24:579-88 (2000), revealed similarly few relevant scientific studies and no conclusive support for the contention that urine with creatinine and specific gravity below the cutoffs cannot be human urine. As the authors noted, "A review of the literature revealed no studies that specifically posed and answered the question 'How can one determine with certainty whether a specimen is urine or not.'" Id. at 582. Accordingly, that article referenced the same random urine clinical studies, medical overhydration studies, and water loading studies identified in NLCP #1. The only new data it referenced is from some additional medical overhydration studies, none of which included paired data. It should be noted, however, that many of the specimens in those new studies reported specific gravity at or below the regulatory cutoff.[10]

Nor does the limited study of creatinine and specific gravity readings of selected employees, conducted by DOT, prove that an individual cannot produce urine below the cutoffs. To the contrary, the data suggests that individuals – particularly women – will

---

reported in 8 subjects and 25 specimens in the following studies: R.A. Braithwaite, An investigation into the extent of possible dilution of specimens received for urinary drugs of abuse screening, Addiction, 90:967-70 (1995); S.L. Nickman et al., Further Experiences with Water Intoxication, Pediatrics, 41:149-51 (1968); and M. Okura, M. and S. Morii, Polydipsia, Polyuria and Water Intoxication Observed in Psychiatric Inpatients, Tokushima Journal Exp. Med., 33:1-5 (1986).

[9] See e.g., S. George and R.A. Braithwaite, An investigation into the extent of possible dilution of specimens received for urinary drugs of abuse screening, Addiction, 90:967-70 (1995); E.J. Cone et al., In Vivo Adulteration: Excess Fluid Ingestion Causes False-Negative Marijuana and Cocaine Urine Test Results, Journal of Analytic Toxicology, 22:460-73 (1998); W.V.R. Vieweg et al., Psychogenic Polydipsia and Water Intoxication-Concepts That Have Failed. Biol. Psychiatry 20:1308-20 (1985).

[10] Id. at 583, Table VI., references 48-71.

produce urine with readings at these levels. That study looked at paired measurements in 54 volunteers, 41 of whom were women. It should be noted, however, that there were serious flaws in the study protocol, which could have skewed the results.

First, cigarette smoking or the use of products or medications containing nicotine was not controlled. Nicotine is well recognized as inhibiting dilution of urine by water loading. As a result, creatinine values on urine dilution from populations that contain smokers will be much higher than such levels in non-smoking populations. See J. Cates and O. Garrod, The effect of nicotine on urinary flow in diabetes insipidus, Clin. Sci. 10:145 (1951); J. Walker, The effect of smoking on water diuresis in man, Q. J. Med. 18:51 (1949); W. Chin et al., Water intoxication caused by smoking in a compulsive water drinker, Clin. Res. 24:625A (1976).

Second, the study included the participants' first urine specimen after awakening, on both the first and second morning of the test collections. First morning urine is far more concentrated than urine randomly collected during the day. See M. Krieg and K. Gunser, Quantitative analysis of clinical and chemical parameters in the 24 hour urine and in the morning. J. Clin. Chem. Clin. Biochem. 24:863 (1986); W. Ottinger, A discussion of creatinine analysis in single collection urine specimens, Journal Forensic Sciences 38:501 (1993); C. Ricos et al., Biological variation in urine samples used for analyte measurements, Clin. Chem. 40:472-77 (1994). Accordingly, including such data in the analysis would skew the results, making the creatinine levels higher than those from urine produced for random drug tests. For example, the study abstract reports that 113 of the 500 specimens, or 22.6%, were "dilute." However, if the 112 first morning

urine specimens are subtracted from the data as not representative of random specimens, this would make 113 of 388, or 30%, dilute.

Third, the specimens were not "blinded" to the laboratory, and the study lacked controls. Such controls should have included non-urine specimens, as well as nonhuman urine, to test whether laboratory testing can distinguish between nonhuman and human urine, and non-urine liquids and human urine. Ordinarily such controls would be part of any scientific study subject to peer review. Moreover, neither the agency authorizing the study or the researchers conducting it, were disinterested parties. The researchers included individuals who had previously endorsed the regulatory creatinine and specific gravity cutoffs and had a professional interest in having them validated. Unwittingly, that bias could have affected the data reporting and analysis.

Fourth, the study encompassed a relatively small number of people not statistically significant enough to be representative of the population. Nor is the number of women subjects - 41- a sufficient quantity from which to draw conclusions representative of all females. It is also significant that the study did not track or identify the female subjects' place in their menstrual cycle. Creatinine output in women varies during the menstrual cycle. See M. Gault et al., Mid-menstrual cycle decline in creatinine and urea clearances, Nephron 67:158-66 (1994). Thus, it is not known what impact the failure to track or control for that variable had on the data.

Fifth, the study also failed to include any significant number of low weight and small body mass participants – especially women, who are at the greatest risk for producing low levels of creatinine. Only nine of the 54 participants were women

weighing 115 pounds or less. As such, it failed to test any representative number of the type of individuals most likely to have low creatinine levels.[11]

Yet, despite these flaws, **three of the subjects – or nearly six percent of the group – produced specimens at the "substituted" level within the margin of error of the method and the laboratory procedures.**[12] Although, as we have discussed (supra at 5-6) the precision of the test is not actually known, if we conservatively assume the test has a standard deviation of ±.1 mg/dL (a coefficient of variation of 2%),[13] then a specimen with a true value of 5.1 mg/dL would test between 5.0 and 5.2 68% of the time, and between 4.9 and 5.2 mg/dL 95% of the time. N. Tietz, Textbook of Clinical Chemistry 48-51 (Philadelpia, W.B. Saunders Co. 1986). **This means that a person with creatinine at 5.1 mg/dL would likely report a result of 5.0 mg/dL or lower -- 33% of the time.**

If, however, the precision is actually less, a greater number of specimens will read lower than their actual levels a greater percentage of the time. So, for example, if the

---

[11] It is noteworthy that the study showed that even women with sizable body mass could consistently produce dilute urine. For example, subject "S1" – a 26 year old women weighing 165 pounds and 67 inches tall, consistently produced dilute specimens.

[12] Subject "E17" (an Asian female) produced urine with 5.2 mg/dL of creatinine and 1.002 specific gravity after drinking 1004 ml in three hours. Subject "E20" (a white, female) produced five dilute specimens, the lowest of which were 5.1 mg/dL of creatinine and 1.001 specific gravity; and 5.6 mg/dL of creatinine and 1.001 specific gravity, apparently, after drinking 5202 ml of water over eight hours. Subject "S1" (a white, female) produced eight dilute specimens, the lowest of which were 5.2 mg/dL and 1.001, and 5.8 mg/dL creatinine and 1.001 specific gravity, apparently, after drinking 2740 ml over five hours.

[13] This coefficient of variation is based on data provided by the manufacturer of the Olympus Creatinine Reagent when it is used to measure creatinine in blood. (Attachment 5). Of course, there has been no showing that the same precision applies when the test is used to measure urine.

(18) p. 22

standard deviation is ±.25 (a coefficient of variation of 5%), then a specimen with a true value of 5.1 would test between 4.8 and 5.3 mg/dL 68% of the time, and between 4.6 and 5.6 mg/dL 95% of the time. **This means that a person with creatinine at that level - like the woman in the study – would likely report a result of less than 5.0 mg/dL - 50% of the time.**

At this juncture, however, the amount of error that is tolerated in these tests is unknown. Whether the coefficient of variation or the "precision" of a particular assay or screening test is 1%, 2%, 5% or even more, will determine the range of values within which a known quantity is expected to measure.

There is also the question of how much variation is allowed at a laboratory between the different runs on known controls. This too must be taken into account in assessing whether a reported result is truly within the "normal" range for creatinine levels.

These findings do mean, however, that if a larger group of subjects had been used, it is likely that additional data at or below these levels would have been obtained. Accordingly, contrary to the conclusion reached by DOT and HHS, the study suggests that individuals can produce urine with creatinine and specific gravity levels measuring at or below the cutoffs.

Others reviewing the study have reached the same conclusion. Noting the relatively low "*n*-" numbers, Theodore F. Shults of the American Association of Medical Review Officers, observed

> The limited data that has been compiled shows (or at least statistically suggests) that it is not physiologically impossible to produce substituted urine. What is revealing about the study is that one volunteer had urine that was right on the border of being identified as a substituted specimen.

(18) p. 43

> This borderline data point is significant. Although it is not "technically" substituted urine, one should ask what is the "technical" standard deviation measurement of creatinine here? It is certainly not zero; thus the measurement could easily have been below the 5.0 cutoff level for creatinine. So at best there is a zero margin of safety. Statistics would also indicate that in a similar study with larger *n*- numbers and a normal distribution of data, some incidence of data points would be identified as meeting and exceeding the substituted criteria.

MRO ALERT at 4.

The results of the DOT study also contradict the "theoretical dilution limits" of urine as identified in the literature review cited by HHS. See The Characterization of Human Urine for Specimen Validity Determination in Workplace Drug Testing: A Review, Table IV, Journal of Analytical Toxicology 24:579-88 (2000). The scientific paper stated that with a daily urine output of one liter (identified in the text as the "normal" excretion volume), the lowest possible value of creatinine is 50 mg/dL, and the lowest possible specific gravity is 1.019. The article indicates that with a level of fluid consumption of 10 liters or more per day, the lowest possible value of creatinine is 5.0 mg/dL and the lowest specific gravity is 1.002. Id.

In DOT's limited study with its inherent flaws, the subjects produced much lower levels of creatinine and specific gravity, based on much less fluid consumption, and over a shorter period of time than the theoretical limits. Three subjects and five or more specimens reached the theoretical limits of 5 mg/dL and 1.001 specific gravity (even assuming the precision of the tests is 2%) after drinking only one to five liters of water over three to eight hours.[14]

This data and our actual experience shows that the "absolute" limits of creatinine and specific gravity levels in human urine differ from the theoretical concepts. It also

[14] Supra at 19, n.12.

shows that otherwise healthy individuals – without any serious or unusual medical conditions, and without having tampered with their specimens – can and do produce urine below the creatinine and specific gravity cutoffs. Such data must be seriously considered and the limits eliminated as proposed.

C. Additional Procedural Protections Should Be Added To Support The Integrity Of Any Validity Tests.

It is not acceptable to leave the degree of error in a testing device to the test manufacturer, or to allow laboratories to establish the margin of error for their procedures as they deem fit. Any such mandated testing, with mandatory employee sanctions, should ensure that the applicable testing methods and laboratory methodology comply with strict standards with respect to precision and accuracy.

Once the margin of error of the validity tests is established - including the precision, accuracy, false positive rate, etc. - it is essential that this range be added to the reported values of test results for creatinine and specific gravity, to ensure that employees are not penalized as a result of variable screening results. Thus, if the margin of error of a testing method for creatinine is ± .1, than a 4.9 reading should be recognized as capable of actually being 5.0. This approach has been required of other forensic laboratories in other countries. See International Standard Organization ("ISO") Standard 17025 (1999).

Additionally, the other means by which reported results can be distorted should likewise be recognized in the process. MRO's should be given instruction about testing limitations and the uncertainty of the measurements. Recognition of such test limitations should be taken into account in the regulatory scheme. MROs should also be trained on

the physiological (and other) reasons why otherwise healthy individuals may have readings below any regulatory cut-offs.

Finally, we urge that any regulatory standards recognize that innocent people can produce urine with creatinine and specific gravity readings at or near the currently proposed cut-offs. Individuals should not suffer adverse consequences merely for producing ultra-dilute urine by having their samples deemed "substituted" or "invalid." If drug testing on such samples cannot be accomplished within the bounds of scientific certainty, then those tests should be cancelled. The focus should remain on testing for the detection of illegal drugs as opposed to penalizing employees for ultra-dilute urine, or uncorroborated results on non-specific oxidation tests.

CAPT. DUANE E. WOERTH, President
JONATHAN A. COHEN, Director, Legal Dept.
SUZANNE L. KALFUS, Sr. Attorney, Legal Dept.

AIR LINE PILOTS ASSOCIATION, INT'L
535 Herndon Parkway
Herndon, VA 20170
Phone: (703) 689-4324

EDWARD WYTKIND, Executive Director

TRANSPORTATION TRADES DEPARTMENT, ALF-CIO
888 16TH Street, N.W., Suite 650
Washington, D.C. 20006
Phone: (202) 628-9262

Dated: October 22, 2001

(18) p. 24

TTD AFFILIATES

The following labor organizations are members of and represented by the TTD:

Air Line Pilots Association
Amalgamated Transit Union
American Federation of State, County and Municipal Employees
American Federation of Teachers
Association of Flight Attendants
American Train Dispatchers Department
Brotherhood of Locomotive Engineers
Brotherhood of Maintenance of Way Employes
Brotherhood of Railroad Signalmen
Communications Workers of America
Hotel Employees and Restaurant Employees Union
International Association of Fire Fighters
International Association of Machinists and Aerospace Workers
International Brotherhood of Boilermakers, Blacksmiths, Forgers and Helpers
International Brotherhood of Electrical Workers
International Brotherhood of Teamsters
International Longshoremen's Association
International Longshoremen's and Warehousemen's Union
International Organization of Masters, Mates & Pilots, ILA
International Union of Operating Engineers
Marine Engineers Beneficial Association
National Air Traffic Controllers Association
National Association of Letter Carriers
National Federation of Public and Private Employees
Office and Professional Employees International Union
Professional Airways Systems Specialists
Retail, Wholesale and Department Store Union
Service Employees International Union
Sheet Metal Workers International Association
Transportation • Communications International Union
Transport Workers Union of America
United Mine Workers of America
United Steelworkers of America

October 2001

(18)
P. 37

NOTE

THE COMMENTS SUBMITTED BY THE AIR LINE PILOTS ASSOCIATION AND THE TRANSPORTATION TRADES DEPARTMENT OF THE AFL-CIO INCLUDED

- **The comments of the Airline Pilots Association submitted to DOT during the public comment period when Part 40 was being revised.**

The APA comments submitted to DOT have not been scanned and placed on this website.

The APA comments are available on the Federal Register website under Docket No. OST-6578