Measurement of Urinary Metanephrines to Screen for Pheochromocytoma in an Unselected Hospital Referral Population

KEITH L. BRAIN,1* JONATHAN KAY,2 and BRIAN SHINE2

**Background:** Despite the rarity of pheochromocytoma, diagnosis is important because of the dangers of uncontrolled severe hypertension and the availability of very effective surgical treatment. Urinary or plasma catecholamines or catecholamine derivatives are commonly used to screen for pheochromocytomas before imaging, but data from 24-h urinary metanephrine results, patient age, and sex may better predict tumors in populations with a low pretest probability.

**Methods:** We retrospectively studied outcomes of an unselected population (1819 patients) referred to a tertiary hospital laboratory for urinary metanephrine testing and investigated the usefulness of some simple derivative measures for detecting pheochromocytoma. We normalized values for urinary 24-h excretion of metanephrine, normetanephrine, and 3-methoxytyramine by dividing by an age- and sex-specific reference range. We then compared pheochromocytoma prediction by the use of products of these normalized measures with the gold standard of biopsy-confirmed tumor.

**Results:** The product of the excretion of normalized metanephrine (nMAD) and normalized normetanephrine (nNMT) (nMAD/nNMT) was a highly sensitive (100%) and specific (99.1%) measure, yielding a positive predictive value of 82%. ROC curves were not improved by including the normalized 3-methoxytyramine concentration in the product. The test for nMAD-nNMT gave higher sensitivity and specificity than the tests for either substance alone.

**Conclusion:** The test for nMAD-nNMT is a useful measure for identifying pheochromocytoma in a population with a low pretest probability.

© 2006 American Association for Clinical Chemistry

Although the best available screening tests for pheochromocytoma have good sensitivity, their positive predictive value is low because the rarity of pheochromocytoma, even in people with suggestive symptoms such as hypertension, makes the pretest probability low. The most common biochemical approaches used to detect pheochromocytoma are measurements of metanephrines (MADs) or catecholamines in plasma or urine, although some studies have advocated the use of chromogranin (1) or platelet catecholamines (2).

Recent studies have advocated the measurement of plasma metanephrines (3), a method supported in relevant reviews (4, 5). However, Sawka et al. (6) reported that in a screening population, the probability likelihoods for raised plasma metanephrines and urinary total metanephrines were 6.3 and 58.9, respectively. This high probability for urinary metanephrines indicates that this screening test is more suitable for patients with a low pretest probability, although the high upper limits of detection applied in this study may have led to unacceptable sensitivity. This controversial issue is perhaps best resolved by comparing ROC curves when evaluating which test might be most useful for screening purposes.

The measurement of fractionated urinary metanephrines allows detection of tumors that predominantly produce 1 of the 3 O-methylated metabolites and thus has been widely accepted as preferable to the measurement of total urinary metanephrines (7, 8). A common approach to fractionated urinary MAD measurement is to consider a result to be positive if either normetanephrine (NMT) or

---

1 Department of Pharmacology, University of Oxford, Oxford, United Kingdom.
2 Department of Clinical Biochemistry, John Radcliffe Hospital, Oxford, United Kingdom.
*Address correspondence to this author at: Department of Pharmacology, University of Oxford, Mansfield Rd., Oxford OX1 3QT, United Kingdom. Fax 44-1865-271853; e-mail keith.brain@pharm.ox.ac.uk.
Received April 1, 2006; accepted August 29, 2006.
Previously published online at DOI: 10.1373/clinchem.2006.070805

Nonstandard abbreviations: MAD, metanephrine; NMT, normetanephrine; n3MT, normalized 3-methoxytyramine; nMAD, normalized metanephrine; nNMT, normalized normetanephrine.
MAD is high [for example, see reference (6)]. Although the sensitivity of urinary fractionated metanephrines for detection of pheochromocytoma is high, specificity is low, at only 45% in patients undergoing testing because of suspicion of sporadic forms of the tumor (3). To improve methods for interpretation of urinary fractionated metanephrines, we investigated the outcomes of a cohort of patients referred to a tertiary hospital laboratory for pheochromocytoma screening by measurement of fractionated urinary metanephrines over a 28-month period and compared the usefulness of derived measures of urinary fractionated metanephrines.

Materials and Methods
We performed a retrospective analysis of all adult 24-h urinary MAD tests referred to the John Radcliffe Hospital (Oxford, United Kingdom) between January 1, 2003, and May 4, 2005. At this referral center, such urinary MAD tests are the preferred testing strategy for the biochemical diagnosis of pheochromocytoma. Samples from each acidified 24-h specimen were hydrolyzed in boiling acidic conditions to release conjugated metanephrines. Ion-exchange chromatography with first a cationic and then an anionic exchange column was performed before isocratic reversed-phase HPLC with Bio-Rad Analytical HPLC Cartridges (catalog no. 195-6088). Electrochemical detection was with an ESA COULOCHROM II coulometric electrochemical detector (ESA Analytical; 1st electrode at 350 mV, 2nd at 350 mV). Regular control solutions were Bio-Rad Urine Standard and Lyphocheck Urine controls (catalog nos. C-390-10 and C-395-10). The performance of the method was assessed in the United Kingdom National External Quality Assessment Service (http://www. ukneqas.org.uk/) scheme for urinary metanephrines.

For all specimens in which either NMT, MAD, or 3-methoxytyramine concentration was above the age- and sex-specific upper reference limits and the patient was seen at the John Radcliffe Hospital or a general practice within the referral area, a review of all biochemistry, hematology, immunology, and online radiology results was carried out (including an assessment of clinical information provided at the time of the request). In the case of atypical test results, if the diagnosis was unclear, the hospital notes were reviewed.

Each measured value of NMT and MAD was normalized by dividing by the age- and sex-dependent upper limits of the reference interval (9), in the 1st test after January 1, 2003, for each patient. The 24-h laboratory reference ranges were as follows: for MAD, 1.9 μmol for men and 1.4 μmol for women; for NMT, 3.6 μmol for men and 3.0 μmol for women (ages 18–40 years), 4.25 μmol for men and 3.45 μmol for women (ages 40–60 years), and 4.5 μmol for men and 3.65 μmol for women (ages >60 years); and for 3-methoxytyramine, 2.75 μmol for both men and women (ages 18–40 years), 2.55 μmol (ages 40–60 years), and 2.3 μmol (ages >60 years).

The John Radcliffe Hospital is a regional referral center for the treatment of pheochromocytomas, so we expected the hospital to have reports on most pheochromocytomas occurring in the laboratory referral area. Therefore, we retrospectively searched histopathology reports for the terms “pheochromocytoma”, “pheochromocytoma”, “paraganglioma”, and “paraganglionomia” to identify all pheochromocytomas histologically diagnosed in the referral population.

All searches were performed with in-house laboratory database software. ROC curves were plotted with Excel for Mac, version 11 (Microsoft); areas under the ROC curve were calculated numerically with Excel. We used GraphPad Prism (version 4.0b; GraphPad Software) with a method described by Hanley and McNeil (10) to calculate the SE for calculating the area under ROC curves. When comparing tests, we used the following equation to calculate the z value:

\[
z = \frac{|Area_1 - Area_2|}{\sqrt{SE_1^2 + SE_2^2 - 2rSE_1 \cdot SE_2}}
\]

Area and are the areas being compared, SE and SE are their corresponding SEs, and is the estimated correlation coefficient. Because we considered multiple measures based on the same underlying data, these measures were likely to be correlated. One-tailed P values were obtained by comparing this z value with a standard gaussian distribution.

Results
In the referral population, 53% were women and the mean age was 51 years; 23% of tests were requested by local general practitioners, 21% were from hospital outpatient clinics, and 55% were from hospital wards (or through referral hospital laboratories).

For a total of 1819 patients, 2470 24-h urine MAD tests were performed during the study period, of which 95 patients had 149 tests with results above the upper reference limit for either MAD or NMT. A mean of 1.36 24-h urine collections per patient was obtained. Hereafter, only the 1st test after January 1, 2003, for each patient is considered. Of the 95 patients with MAD or NMT over the upper reference limit, 30 patients had no further follow-up because negative results were obtained for repeat testing of either urinary fractionated metanephrines or urinary free catecholamines. Imaging analysis was performed on 28 patients (abdominal computerized tomographic scan or magnetic resonance imaging; 1 patient had abdominal ultrasound only), with resulting diagnoses of pheochromocytoma in 14 patients; other pathologies (metastatic carcinoma, adrenal hemorrhage, adrenal myelolipoma, pelvic fibromatosis, or neuromyotonia, which was diagnosed after neurophysiology tests) in 9 patients, and no atypical findings in 5 patients.
In 22 study patients, an intercurrent condition was present during testing (conditions included intracranial hemorrhage, congestive cardiac failure, pancreatitis, and postoperative sepsis from nonabdominal surgery or pregnancy); in 3 of these patients, subsequent repeat tests were within reference intervals and no additional follow-up was performed; the remaining patients either died of their other pathology or had symptom resolution with no further follow-up. In 7 patients, no further follow-up of the atypical result could be identified, and in 5 patients, the atypical results were attributed to drug interference or effects (monoamine oxidase inhibitors, imipramine, or polydrug overdose).

Of the 14 patients in whom pheochromocytoma was detected, 11 were referred from hospital wards, 2 were referred from hospital clinics, and only 1 was referred by a general practitioner.

A search of local histopathology records identified 22 pheochromocytoma patients, of whom 14 underwent presurgery urinary metanephrines measurement during the study period. In all 14 of these patients, urinary metanephrines were atypical. In 2 of these 14 cases, multiple endocrine neoplasia was subsequently diagnosed. In addition, 4 paragangliomas were identified (a carotid body tumor, a left renal hilum paraganglioma in the context of Carney triad, a urinary bladder paraganglioma, and a jugulotympanic paraganglioma). Urinary metanephrines were measured preoperatively in only 1 of these cases; NMT and MAD were within the reference interval, but 3-methoxytyramine was increased.

When we used our current laboratory thresholds and either MAD or NMT were increased, the sensitivity for detecting pheochromocytoma was 100%, with a specificity of 95%. However, because the pretest probability in this population was very low, the positive predictive value was only 16%. To investigate better ways of interpreting the existing laboratory data, we examined the distribution of metanephrines in the presence and absence of pheochromocytomas in those patients in whom no previous pheochromocytoma or paraganglioma had been detected (Fig. 1). For most patients with pheochromocytoma, both MAD and NMT were increased. However, if the criterion for an atypical result had been both high MAD and high NMT, then 2 of 14 cases of pheochromocytoma would have been missed (because 1 pheochromocytoma patient had MAD and another had NMT within reference intervals).

A simple measurement that incorporates information about both MAD and NMT is the product of their normalized (to correct for age and sex) values (nMAD\_nNMT). In our population, the smallest nMAD\_nNMT value in which pheochromocytoma was histologically diagnosed was 3.1. The use of nMAD\_nNMT $\geq$3 as the threshold for pheochromocytoma yielded a test with a sensitivity of 100%, a specificity of 99.1%, and a positive predictive value of 82%. To compare alternative measurement for the detection of pheochromocytoma in our referral population and to see whether the normalized 3-methoxytyramine (n3MT) concentration could add additional discriminative information when combined in a simple product, we compared the ROC curves (Fig. 2). The areas under the ROC curves for the tests were as follows: nMAD, 0.9970; nNMT, 0.9881; n3MT, 0.76; nMAD\_nNMT, 0.9995; and nMAD\_nNMT\_n3MT, 0.9968. A common practice is to report a test positive if either nMAD or nNMT is assessed to be atypical. We assessed this approach by calculating $k = \max(n\text{MAD}, n\text{NMT})$ (i.e., the maximum of nMAD and nNMT) and testing the variation in sensitivity and specificity as the threshold value $k$ varies [a modification of the technique used by Lenders et al. (11)]. With our results, this calculation gave an area under the ROC curve of 0.9983. Finally, the sum of fractionated metanephrines (nMAD + nNMT) yielded an area under the ROC curve of 0.9989. Of all tests investigated, the highest area under the ROC curve was for nMAD\_nNMT (0.9995), which was statistically significantly different from each of the other 6 tests described above ($P < 0.05$).

In all 24-h urinary MAD tests, we also measured 3-methoxytyramine and found isolated increases in 52 patients in the target population. Of these patients, 22 had typical results for repeat tests of 3-methoxytyramine or free urinary dopamine and were not further investigated, and 15 had no identifiable follow-up. Of the patients who underwent additional follow-up, 5 underwent imaging studies. Of these 5 studies, 3 showed no pathologies and 2 showed adrenal masses, both of which were excised; one mass was a benign neurofibroma and the other was not adrenal but was a left renal hilum paraganglioma.
Four patients who underwent follow-up were tested in the presence of known pathology (recurrent chemodectomas, a previous pheochromocytoma, or iodine-131-meta-iodobenzylguanidine-hot multiple liver lesions), 4 were tested during a severe intercurrent illness (in 2 of these, 3-methoxytyramine was within the reference interval on repeat testing), and 2 were being treated with Levodopa.

**Discussion**

Detecting rare problems is difficult, particularly in a diverse and unselected population. By examining the outcomes of those patients referred to our service for the testing of urinary metanephrines, we hoped to aid clinical decision-making by determining the causes commonly associated with atypical test results and to optimize the use of available laboratory data. Of those patients with atypical urinary metanephrines, approximately one third were considered, on the basis of repeat testing, to have had initial false-positive results, one third went on to imaging studies, and one third had atypical results that were attributed to an acute intercurrent condition or the effects of drugs. Even with atypical findings for urinary metanephrines and adrenal imaging, several other pathologies were discovered, including metastatic carcinoma, adrenal hemorrhage, and adrenal myelolipoma. Hence, the combination of positive results for urinary metanephrines and adrenal imaging is not an appropriate gold standard for detecting pheochromocytomas.

The present study has demonstrated that the normalized product of urinary metanephrine and normetanephrine (nMAD/nNMT) provides a useful measurement for the detection of pheochromocytoma in the unselected referral population of a tertiary laboratory. Several other approaches have been used to increase the sensitivity or specificity of urinary metanephrines, including careful analysis of the patient’s drug history (taking particular note of phenoxybenzamine, tricyclic antidepressants, and beta blockers) (12), absence of clonidine suppression of plasma NMT (12–14), use of overnight (12 h) testing to avoid the variability introduced by daytime physical activity and to increase specificity because of lower resting sympathetic activity overnight (15–17), special reporting for samples referred from intensive therapy units (18), and use of immunoassays for metanephrines to decrease drug interference (19). Each of these approaches has been and will be found useful in different situations depending on the pretest probability and the mix of pathologic conditions within the referral population. The practical advantages of nMAD/nNMT are that it requires no additional testing and its calculation can be readily automated.

Perhaps the strongest study supporting the use of plasma free metanephrines instead of urinary fractionate metanephrines is that of Lenders et al. (3). This predominantly prospective study was of a group of 1003 patients (over a 7-year period from 4 referral centers). Given that in this population pheochromocytoma was found in 214 patients, the mean pretest probability of >20% makes this population very different from our hospital referral base. Indeed, the frequency of occurrence of pheochromocytoma is 0.5% in a screened population with hypertension (20), 5% in a population with adrenal incidentaloma (21), and 42% in a population presenting with multiple endocrine neoplasia type 2A and medullary thyroid cancer (22). In our unselected hospital referral population the mean pretest probability for pheochromocytoma was 0.8% (14 of 1819), more than an order of magnitude lower than that described by Lenders et al. (3). In such low pretest probability populations, testing for nMAD/nNMT provides excellent specificity while maintaining sensitivity, suggesting that this derivative measure could be used more widely as a screening test for pheochromocytoma.

**Study Limitations**

The reference test chosen for detecting pheochromocytomas was biopsy or excision-biopsy histopathology. Although this method is very specific, it is likely that small pheochromocytomas may have been missed, and hence...
the sensitivity of the assays in the present work may be overestimated, particularly because a strong correlation exists between urinary metanephrines (unfractionated) and tumor size in patients with catecholamine-secreting tumors (23). Even more important is the fact that in this center, the results of urinary fractionated metanephrines were used as part of the work-up for tumor excision, a protocol that biases our analysis toward higher sensitivity values for our analyses, particularly for the test “nMAD or nNMT abnormal”, which generates an abnormal report from our laboratory. Nevertheless, the nMAD-nNMT measurement was better at predicting subsequent pheochromocytoma. In this study, the alternative gold standard of adrenal mass has been demonstrated to lack specificity, and it seemed impractical, if not unethical, to subject all 1819 patients to a computerized tomographic scan or magnetic resonance imaging. A longer period of follow-up would be useful in future studies. The sample size in this study (1819 patients) was relatively small compared with some multicenter trials, and the follow-up time of 28 months may not have been sufficient to allow the detection of all pheochromocytomas, a shortcoming that may have led to an overestimate of sensitivity. A prospective study of the use of nMAD-nNMT in the diagnosis of pheochromocytoma is now required.

We gratefully acknowledge the support of Dr. S. Manek for allowing access to the histopathology database and the staff of the Department of Clinical Biochemistry who performed the analyses. The Wellcome Trust supports K.L.B.

References