Published in final edited form as:

AJR Am J Roentgenol. 2011 August ; 197(2): 325–333. doi:10.2214/AJR.10.5909.

Key Variables for Interpreting ^{99m}Tc-Mercaptoacetyltriglycine Diuretic Scans: Development and Validation of a Predictive Model

Jieqiong Bao¹, Amita Manatunga¹, Jose Nilo G. Binongo¹, and Andrew T. Taylor²

¹Department of Biostatistics and Bioinformatics, Emory University School of Public Health, Atlanta, GA

²Department of Radiology and Imaging Sciences, Emory University Hospital, 1364 Clifton Rd NE, Atlanta, GA 30322

Abstract

OBJECTIVE—The purpose of this study was to facilitate interpretation of ^{99m}Tc-mercaptoacetyltriglycine (MAG3) diuretic scans by identifying key interpretative variables and developing a predictive model for computer-assisted diagnosis.

MATERIALS AND METHODS—Ninety-seven studies were randomly selected from an archived database of MAG3 baseline and furosemide acquisitions and scan interpretations (obstruction, equivocal finding, or no obstruction) derived from a consensus of three experts. Sixty-one studies (120 kidneys) were randomly chosen to build a predictive model for diagnosing or excluding obstruction. The other 36 studies (71 kidneys) composed the validation group. The probability of normal drainage (no obstruction) at the baseline acquisition and the probability of no obstruction, equivocal finding, or obstruction after furosemide administration were determined by logistic regression analysis and proportional odds modeling of MAG3 renographic data.

RESULTS—The single most important baseline variable for excluding obstruction was the ratio of postvoid counts to maximum counts. Renal counts in the last minute of furosemide acquisition divided by the maximum baseline acquisition renal counts and time to half-maximum counts after furosemide administration in a pelvic region of interest were the critical variables for determining obstruction. The area under the receiver operating characteristic curve (AUC) for predicting normal drainage in the validation sample was 0.93 (standard error, 0.02); sensitivity, 85%; specificity, 93%. The AUC for the diagnosis of obstruction after furosemide administration was 0.84 (standard error, 0.06); sensitivity, 82%; specificity, 83%.

CONCLUSION—A predictive system has been developed that provides a promising computer-assisted diagnosis approach to the interpretation of MAG3 diuretic renal scans; this system has also identified the key variables required for scan interpretation.

Keywords

computer-assisted diagnosis; diuretic renography; logistic regression; MAG3 renography; proportional odds model

Innovative approaches are required to address the escalating costs of medical care. In radiology and nuclear medicine, one such approach has been the use of computer-assisted

diagnosis to increase efficiency, improve accuracy, and ultimately reduce costs by helping radiologists interpret studies at a faster rate with higher degree of accuracy. One particularly successful example of computer-assisted diagnosis in nuclear medicine is bulls-eye analysis in SPECT myocardial perfusion imaging [1]. Other areas that can especially benefit from this approach are low-volume studies such as diuretic renography, with which many general radiologists might have had only limited training and experience [2–4]. Limited experience in diuretic renography is a widespread problem. A large percentage of the estimated 590,000 renal scans obtained annually in the United States are interpreted at sites that perform fewer than three studies per week [5]. Even full-time nuclear medicine physicians disagree as much as 20% of the time about whether a kidney is obstructed, indeterminate with respect to obstruction, or not obstructed [6].

To evaluate suspected renal obstruction, an international consensus panel has recommended baseline radionuclide imaging with ^{99m}Tc -mercaptoacetyltriglycine (MAG3) followed by furosemide administration and an additional 15 minutes of imaging [7]. From the baseline and post-furosemide administration acquisitions, quantitative variables, such as time to half-maximum counts after furosemide administration, are derived to help evaluate possible kidney obstruction. Physicians with limited experience or insufficient training may try to compensate with overreliance on a single parameter, such as time to half-maximum counts, but such an approach can lead to erroneous scan interpretation due to failure to consider variables that can be equally or even more important.

To address the problem of limited experience in diuretic renography, we used a statistical approach to develop a predictive model for evaluating suspected obstruction in patients referred for diuresis renography. We had two goals. The first was to develop a statistical decision support system for evaluation of the presence or absence of obstruction with data derived from the MAG3 renogram. The second goal was to determine the most important variables for making or excluding the diagnosis of obstruction so that general radiologists and nuclear medicine physicians could immediately apply this information in their practices to reach an informed interpretation.

Materials and Methods

General Protocol and Patient Selection

Data collection and database use were compliant with the terms of HIPAA and followed institutional review board approval with waiver of the requirement for informed consent. Our diuretic renography protocol is based on the protocol recommended by an international consensus panel [7]. A baseline scan is obtained for 24 minutes after MAG3 administration. In approximately one third of our patients, obstruction can be excluded with the baseline scan, and the study is complete. The other two thirds of patients receive an IV injection of furosemide, and an additional 20-minute acquisition is performed.

The raw and processed data for all MAG3 scans obtained for possible obstruction are archived in a database. The database also includes a subgroup of patients whose scans experts have interpreted as showing obstruction, equivocal findings, or no obstruction. Readers were defined as experts if they had more than 20 years of experience in full-time academic nuclear medicine, multiple publications on renal nuclear medicine, and been invited to present renal nuclear medicine educational sessions at national radiology and nuclear medicine meetings.

Using information available only from the scan, each expert independently scored each kidney on the baseline scan as not obstructed or needing furosemide and scored the baseline and furosemide scans for the presence of obstruction on a 5-point scale: 1, not obstructed; 2,

probably not obstructed; 3, equivocal; 4, probably obstructed; and 5, obstructed. For analysis, the 5-point scale was condensed to a 3-point scale on which not obstructed and probably not obstructed were considered not obstructed and probably obstructed and obstructed were considered obstructed. The consensus interpretation was determined by majority vote unless there was substantial disagreement. In those cases, a conference of the three readers was used to achieve a consensus interpretation. In general, there was very good agreement among the three readers. On only 8 of 191 kidneys (4.2%) did two of the three expert readers disagree regarding the presence or absence of obstruction. This disagreement was resolved at a consensus conference of the three readers.

Studies of 97 patients (54 men [56%], 43 women [44%]; mean age, 54 years; SD, 17 years; range, 20–87 years) were randomly collected from the archived database containing both MAG3 scans and the expert consensus reading; three patients had only one kidney. Patients with renal grafts, ileal conduits, and neobladders were excluded from the database because there were too few patients in these categories to build and test a predictive model. Eight patients underwent two studies, and one patient underwent three studies. The mean time between studies was 12.8 months (range, 1–41 months).

The studies were coded and dated. Expert readers might have recognized that there were two studies of the same patient (same code, different date), but the entry criteria were the same for both studies, and each study was interpreted independently. At the time of the baseline acquisition, 32 of the 97 studies were interpreted as having normal drainage from both kidneys. Because normal drainage excludes obstruction, furosemide acquisition was not performed. The experts classified 121 of the total 191 kidneys (63%) as not obstructed, 30 (16%) as having equivocal findings, and 40 (21%) as obstructed. Because a larger number of patients are generally needed to develop a robust model than to test the model once it is developed, we built the predictive model by randomly allocating 61 patient studies (120 kidneys) to the training set and allocating the other 36 patient studies (71 kidneys) to determining the predictive accuracy of the model.

Acquisition Protocol and Data Processing

After injection of approximately 10 mCi of ^{99m}Tc -MAG3, a baseline acquisition was performed with the patient supine. All studies were processed with an updated in-house version of the QuantEM program (Emory University) developed specifically for use with ^{99m}Tc -MAG3. The program has been validated in a multicenter trial and generates specific quantitative parameters recommended for scan interpretation and calculating camera-based MAG3 clearance [8–11]. Background-subtracted whole-kidney and cortical curves and multiple quantitative curve parameters were generated (Table 1). Age and sex also were recorded.

A standard 40-mg dose of furosemide was administered at the beginning of a separate 20-minute acquisition. The dose of furosemide was occasionally increased to 60 or 80 mg if the MAG3 clearance was reduced or if the patient was known to have an elevated creatinine concentration. Technologists assigned a region of interest over the whole kidney and a region of interest limited to retained activity in the renal collecting system (pelvic region of interest) (Fig. 1). The time to half-maximum counts was based on a linear fit. Quantitative parameters (Fig. 1 and Table 1) were automatically extracted.

Statistical Analysis and Development of the Predictive Model

There were two outcomes of interest: whether experts diagnosed normal kidney drainage on the baseline scan ($Y_1 = 1$ for normal drainage, which excludes obstruction, and $Y_1 = 0$ otherwise) and whether experts diagnosed obstruction after furosemide administration ($Y_2 =$

0 for no obstruction, $Y_2 = 1$ for equivocal finding, and $Y_2 = 2$ for obstruction). The expert consensus served as the reference standard. To predict normal drainage (Y_1) on the baseline scan, we used a logistic regression model with the quantitative variables from baseline scan variables in a vector (X_1). That is,

$$\text{logit}(P(Y_{1i}=1)|x_{1i})=\beta_0+\beta_1^T x_{1i} \quad (1)$$

The regression coefficient vector β_1^T indicates the strength of association between the quantitative scan variables and the outcome Y_1 . Because the diagnosis of obstruction was rated at three levels after furosemide administration, we used a proportional odds model [12] for the baseline quantitative variables (X_1) and furosemide acquisition scan variables (X_2) as the independent variables. However, two important statistical issues had to be addressed. First, crucial second-stage furosemide predictor variables were not available when a patient did not receive furosemide, and the absence of these variables had to be considered in modeling of the probability of kidney obstruction [13]. Second, whenever they diagnosed normal kidney drainage ($Y_1 = 1$), the experts simultaneously inferred that the kidney was not obstructed ($Y_2 = 0$). Because there was no variability in the outcome Y_2 when $Y_1 = 1$, we next modeled the out-come of Y_2 given Y_1 using the proportional odds regression model, as follows:

$$\text{logit}(P(Y_{2i} \geq j|Y_{1i}=0, x_{1i}, x_{2i}, \gamma))=\gamma_0+\gamma_1^T x_{1i}+\gamma_2^T x_{2i}, j=1, 2 \quad (2)$$

When X_2 was missing, we used a multiple imputation approach [14, 15] by imputing missing values based on the conditional distribution of X_2 given X_1 . The overall likelihood is described in Appendix 1. We maximized this likelihood to obtain parameter estimates. For predicting the outcome Y_1 , we used equation 1 to obtain the predicted value $P(Y_1 = 1)$. We used equation 2 to facilitate the prediction probabilities for each category—no obstruction, equivocal finding, and obstruction—and, consequently, the prediction probabilities of any combined categories. To evaluate the predictive power with receiver operating characteristic (ROC) curves [13], we focused on two prediction categories: $P(Y_2 \geq 1)$ and $P(Y_2 = 2)$, which gave the predicted probability for the combined no obstructed–equivocal interpretation and the predicted probability of obstruction.

For the logistic regression model (equation 1) and the proportional odds model (equation 2), we first used a backward regression procedure to determine important predictors needed for the modeling process. To ensure we did not eliminate any important variables, we fitted a series of models to determine whether the model could be improved by adding other variables. We used the Akaike information criterion to rank several competing multivariate models based on the selected important predictors [16]. A smaller value of the Akaike information criterion suggests a more parsimonious model. In addition, Hosmer-Lemeshow tests were performed to assess the goodness of fit [17] and calibration of the model [18].

Results

To develop predictive models for the two outcomes—diagnosis of normal drainage (no obstruction) on the baseline study and diagnosis of obstruction after furosemide administration—we first carried out a variable selection procedure using the training dataset of 61 patients (120 kidneys) using the variables listed in Table 1. Demographic variables such as age and sex were initially considered but were not found significant. In the univariate model, when each variable was considered individually, all variables were found

significantly predictive of normal drainage (values greater than 2.71) (Table 2). Higher values suggest greater importance of a particular variable, but because many of the values are highly correlated, the relative importance can best be evaluated in a multivariate model. In the multivariate model, only three variables—ratio of the counts in the kidney 19–20 minutes on the baseline acquisition to the maximum counts in the kidney (20 min/max count ratio), ratio of the counts in the kidney after voiding to the maximum counts in the kidney (postvoid/max count ratio), and the time in minutes to reach maximum counts in the cortical region of interest (cortical $T_{\max}$)—were identified as useful for predicting the probability of normal drainage on the baseline study. When these three variables were included in the multivariate model, adding other variables did not improve the goodness of fit. Stated differently, all of the remaining variables were associated with $p > 0.05$, indicating that their inclusion in the model did not improve the results (Table 2). Table 2 does not show p for 20 min/max count ratio, cortical $T_{\max}$, or postvoid/max count ratio because these variables had already been chosen in the multivariate model.

The probability of obstruction was based on results of analysis of the variables in Table 3. In the univariate model, when each variable was considered in isolation, approximately two thirds of the variables were predictive of the presence or absence of obstruction. Again, when the multivariate model was considered, only two variables—ratio of kidney counts in the last minute of furosemide acquisition to maximum kidney counts in the baseline acquisition (postfurosemide 20 min/max baseline count ratio) and time in minutes for counts in the renal pelvis to decrease 50% after furosemide administration (pelvic $T_{1/2}$)—were singled out as significant for predicting the probability of diagnosing obstruction. The other variables in Table 3 correlated with the two selected variables, and their inclusion in the model did not improve the goodness of fit (all $p > 0.05$) (Table 3). In Table 3, p is not provided for postfurosemide 20 min/max baseline count ratio or pelvic $T_{1/2}$ because these variables had already been chosen in the multivariate model. In summary, only a few variables were identified as significant for predicting both no obstruction and obstruction.

Table 4 shows the regression coefficients associated with two probabilistic models. The results correspond to the probability of finding normal drainage on the baseline scan (model 1) and the probability of the diagnosis of obstruction after furosemide administration (model 2). The variable postvoid/max ratio was the most significant variable ($p < 0.0001$) in determining normal kidney drainage and, consequently, the most significant variable in excluding the diagnosis of obstruction. The mean postvoid/max count ratio for the kidneys interpreted as having normal drainage was 0.18 (SD, 0.16), compared with 0.78 (SD, 0.28) for kidneys that had abnormal baseline drainage, might have been obstructed, and required a second acquisition after furosemide administration. With our models, it was possible to calculate the effect of critical variables. For example, if we assume that three kidneys had the same average value of 20 min/max count ratio and cortical $T_{\max}$ but had different postvoid/max ratios of 0.18, 0.78, and 1.38, the results obtained with our predictive model suggest that the probabilities of normal drainage (no obstruction) in these three kidneys are 0.89, 0.04, and 0.001.

The ratio of count 20 minutes after furosemide administration to baseline maximum count was the most significant variable ($p < 0.0001$) in determining the diagnosis of obstruction (Table 4, model 2). The mean values of this variable for the no obstruction and obstruction groups were 0.11 (SD, 0.12) and 0.92 (SD, 0.36). The effect of specific values of these variables also can be calculated with this model. For example, if we assume three kidneys have the same average pelvic $T_{1/2}$ value but different postfurosemide 20 min/max baseline count ratios of 0.11, 0.92, and 1.90, the predicted probabilities for the diagnosis of obstruction would be 0.002, 0.268, and 0.994. Moreover, the probabilities of diagnosis of

either obstruction or equivocal evidence of obstruction of a kidney would be 0.012, 0.665, and 0.999. The predictive equations are shown in the footnote to Table 4.

ROC curves were constructed to determine the predictive accuracy of the regression equations. The area under the ROC curve (AUC) for normal kidney drainage on the baseline scan was very high for both the training set (AUC, 0.97; 95% CI, 0.94–0.99) and the validation set (AUC, 0.93; 95% CI, 0.89–0.98) (Fig. 2). Not surprisingly, the diagnosis of obstruction was more difficult than the diagnosis of normal drainage. We constructed two ROC curves for the diagnosis of obstruction: case 1, prediction of the probability of the interpretation of obstruction versus interpretation of no obstruction or equivocal finding, and case 2, prediction of the probability of obstruction or equivocal interpretation versus interpretation of no obstruction. In both cases, the AUCs were large for the training set: 0.88 (95% CI, 0.81–0.94) and 0.89 (95% CI: 0.82–0.96). For the validation data set, performance was still reasonably high: the AUC was 0.84 for case 1 (95% CI, 0.75–0.92) and 0.84 for case 2 (95% CI, 0.76–0.92) (Fig. 3).

The predictive equations result in probability as opposed to a binary (yes or no) clinical decision or a classification rule. To make the results useful to clinicians, we chose cutoff points in the training set for obtaining probabilities with high sensitivity and specificity in the range 0.83–0.95 (Table 5). For predicting normal kidney drainage on the baseline scan, the cutoff point was 0.5; for predicting the probability of renal obstruction versus no obstruction or equivocal finding, the cutoff point was 0.4, and for predicting the probability of combined obstruction and equivocal finding versus no renal obstruction, the cutoff point was 0.6. The sensitivity and specificity of the classification rules were independently evaluated for the validation data set with the same cutoff points, and the results are shown in Table 5. The validation set had high sensitivity and specificity for normal drainage (no obstruction) on the baseline scan (sensitivity, 0.85; specificity, 0.93). As expected, predicting obstruction was more difficult than predicting normal drainage, but the model had reasonably high sensitivity (case 1, 0.82; case 2, 0.78) and specificity (case 1, 0.83; case 2, 0.95) for the validation group.

Discussion

Our statistical model can be used to predict normal drainage on baseline scans and the presence or absence of obstruction on scans obtained after furosemide administration and to determine the critical variables for excluding or making the diagnosis of obstruction. The analysis would have been more straightforward if we had evaluated only scans obtained after furosemide administration, but this design would have eliminated a large segment of patients, would have resulted in a sample not representative of our patient population, might have biased our results, and would have limited the applicability of the predictive equation to future patients.

A contribution of our paper is that our predictive model was designed to apply to the general problem that arises when an experimental design itself introduces missing data. We encountered two nontrivial statistical challenges in analyzing the data. First, the experimental design itself introduced missing data. When it was determined that a patient had normal drainage (no obstruction), a second acquisition with furosemide was not performed, and consequently, the furosemide acquisition data were missing. Our proposed imputation approach of handling missing data led to reasonable results, was further validated with an independent data set, and could be applied to clinical studies in which the experimental design includes decision nodes that introduce missing data. Second, the kidney outcomes of the same individual could be correlated, and hierarchic modeling might have been more appropriate. However, we first fitted a hierarchic model (not reported here) and

found that the odds ratio of the outcomes was almost 1. We therefore focused on a simpler approach, modeling the marginal distribution of kidney-specific outcomes while not accounting for the correlation between outcomes. This method still allowed consistent estimates of regression coefficients [19]. Most important, it resulted in a model that can be easily interpreted in clinical practice.

Based on the univariate model (Table 3), the ratios of furosemide acquisition 20-minute counts to baseline 1- to 2-minute counts, furosemide acquisition 1- to 2-minute counts to maximum baseline counts, and furosemide acquisition 20-minute counts to maximum baseline counts all have comparable likelihood ratios. These variables are all highly correlated with each other, and use of a model containing any one of the variables and the pelvic time to half-maximum count would be reasonable. We chose the model containing the furosemide acquisition 20-minute to baseline maximum count ratio because it yielded the lowest value for the Akaike information criterion, suggesting that it is the most parsimonious model fitting the data. Based on the importance of the postvoid to maximum count ratio on the baseline scan and the importance of the furosemide acquisition 20-minute to baseline maximum counts ratio, the ratio of postvoid kidney counts after furosemide administration to maximum counts at baseline is also likely to be a robust predictor of obstruction or absence of obstruction and should be evaluated in future studies.

The time to half-maximum counts after furosemide administration is commonly used to help make or exclude the diagnosis of obstruction [20], but this time can be affected by dehydration, size and compliance of the renal pelvis, bladder distention, patient position, dose of furosemide, technical factors such as region of interest assignment (whole kidney, cortical parenchyma, pelvis), and the algorithm used to calculate the time to half-maximum counts. In particular, owing to both reduced clearance and prolonged transit time, time to half-maximum counts is prolonged in kidneys with impaired function. As renal function deteriorates, more radiotracer remains in circulation and tracer continuing to enter the kidney compromises measurement of tracer exiting the kidney. This problem can be ameliorated by delaying the administration of furosemide until the entire dilated collecting system is filled with tracer or by assigning a region of interest over the dilated collecting system [21, 22]. In practice, the times to half-maximum counts for the whole kidney and pelvic regions of interest are quite similar because of the dominance of counts in the collecting system, but there can be major differences in selected cases (unpublished data). In summary, the pelvic time to half-maximum counts is a better index of emptying from the collecting system because there is no confounding parenchymal component.

Another set of variables relates the postvoid and prevoid kidney counts to the 1- to 2-minute or 2- to 3-minute baseline counts [23, 24]. The ratios of postvoid to 1- to 2-minute counts and of furosemide acquisition counts at 20 minutes to baseline counts at 1–2 minutes were clearly important variables based on the results of univariate analysis. They did not, however, have quite the same predictive value as ratios based on maximum counts.

Limitations

We evaluated a diuresis renography protocol based on the approach recommended by an international consensus panel [7]; our results cannot be extrapolated to other protocols. Our studies also were confined to adults and patients with native kidneys, although we expect that the interpretative criteria would also apply to renal transplant patients with suspected obstruction. Patients with renal grafts, ileal conduits, and neobladders were not included in the database because their numbers were too few to build and test a predictive model. We assigned 61 studies (120 kidneys) to the training set and 36 studies (71 kidneys) to the validation set because, in general, more studies are required to develop a robust model than to validate it. Our results from the validation data set showed reasonably high estimates for

sensitivity and specificity. The variability and CI of these estimates would have been smaller with a larger validation set.

Ideally, ROC analysis should have a reference standard independent of the method under evaluation. It could be argued that a Whitaker test or clinical outcome would be a better standard, but Whitaker tests are rarely performed, and clinical outcome is a compromised reference standard because the clinical decision to intervene or not to intervene is often contingent on the interpretation of the diuresis renographic study. For our purposes, the absence of an independent reference standard is not an issue because our goal was to develop a predictive statistical model that leads to study interpretations matching those of experts. This goal was based on the widely accepted assumption that radiologists who specialize in specific areas have more expertise in those areas than do general radiologists, who have a much broader and diverse practice.

Conclusion

The statistical predictive system performs well in assessment of normal kidney drainage on baseline scans and in identifying an obstruction or absence of obstruction of a kidney after administration of furosemide. Multivariate analysis showed that the ratio of 20-minute to maximum counts, cortical time to half-maximum counts, and the ratio of postvoid to maximum counts were critical variables for excluding obstruction on baseline scans. For the diagnosis of obstruction, the ratio of furosemide acquisition counts at 20 minutes to baseline maximum counts and pelvic time to half-maximum counts were the critical variables. Both of these variables are easy to measure and can be useful in interpretation of diuretic MAG3 renal scans. In addition to use in identifying the most important variables for interpreting MAG3 scans of patients with suspected obstruction, predictive model systems have potential for educating trainees and for assisting radiologists in clinical practice in interpretation of diuretic MAG3 scans.

Acknowledgments

Supported by National Institutes of Health grant RO1-EB008838, the National Institute of Biomedical Imaging and Bioengineering, the National Institute of Diabetes and Digestive and Kidney Diseases, and a URC grant from Emory University.

References

1. Cerqueira MD, Weissman NJ, Dilsizian V, et al. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart. AHA Writing Group on Myocardial Segmentation and Registration for Cardiac Imaging. *Circulation*. 2002; 105:539–542. [PubMed: 11815441]
2. Garcia EV, Taylor A, Halkar R, Folks R, Krishnan M, Dubovsky E. RENEX: an expert system for the interpretation of ^{99m}Tc -MAG3 scans to detect renal obstruction. *J Nucl Med*. 2006; 47:320–329. [PubMed: 16455639]
3. Taylor A, Garcia E, Binongo J, et al. Diagnostic performance of an expert system for interpretation of ^{99m}Tc MAG3 scans in suspected renal obstruction. *J Nucl Med*. 2008; 49:216–224. [PubMed: 18199609]
4. Binongo JN, Taylor A, Hill AN, et al. Use of classification and regression trees in diuresis renography. *Acad Radiol*. 2007; 14:306–311. [PubMed: 17307663]
5. IMV Medical Information Division. 2003 nuclear medicine census market summary report. Des Plaines, IL: IMV; 2003. p. 7-11.
6. Hunsche, A. Value of quantitative data in the interpretation of diuresis renography for suspected urinary tract obstruction (thesis). Porto Alegre, Rio Grande do Sul, Brazil: Federal University of Rio Grande do Sul; 2006.

7. O'Reilly P, Aurell M, Britton K, Kletter K, Rosenthal L, Testa T. Consensus on diuresis renography for investigating the dilated upper urinary tract. *J Nucl Med.* 1996; 37:1872–1876. [PubMed: 8917195]
8. Taylor A, Corrigan PL, Galt J, et al. Measuring technetium -99m-MAG3 clearance with an improved camera-based method. *J Nucl Med.* 1995; 36:1689–1695. [PubMed: 7658232]
9. Taylor A, Manatunga A, Morton K, et al. Multi-center trial validation of a camera-based method to measure Tc-99m mercaptoacetyltriglycine, or Tc-99m MAG3, clearance. *Radiology.* 1997; 204:47–54. [PubMed: 9205222]
10. Taylor A, Lewis C, Giacometti A, et al. Improved formulas for the estimation of renal depth in adults. *J Nucl Med.* 1993; 34:1766–1769. [PubMed: 8410296]
11. Taylor A. Formulas for estimate renal depth in adults. *J Nucl Med.* 1994; 35:2054–2055.
12. Agresti, A. Categorical data analysis. 2. New York, NY: Wiley; 2002. p. 275-277.
13. Pepe, MS. The statistical evaluation of medical tests for classification and prediction. Vol. 66–67. New York, NY: Oxford University Press; 2003. p. 168-170.
14. Rubin, DB. Multiple imputation for nonresponse in surveys. New York, NY: Wiley; 1987. p. 15-16.
15. Schafer, JL. Analysis of incomplete multivariate data. New York, NY: Chapman and Hall; 1997. p. 200-202.
16. Burnham, KP.; Anderson, DR. Model selection and multimodel inference: a practical information-theoretic approach. 2. New York, NY: Springer; 2002. p. 60-64.
17. Hosmer DW, Lemeshow S. A goodness-of-fit test for the multiple logistic regression model. *Commun Stat Theory Methods.* 1980; 9:1043–1069.
18. Altman SG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ.* 2009; 338:b605. [PubMed: 19477892]
19. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika.* 1986; 73:13–22.
20. Thrall, JH.; Ziessman, HA. Nuclear medicine: the requisites. St. Louis, MO: Mosby; 1995. Genitourinary system; p. 283-320.
21. Shulkin BL, Mandell GA, Cooper JA, et al. Procedure guideline for diuretic renography in children 3.0. *J Nucl Med Technol.* 2008; 36:162–168. [PubMed: 18765635]
22. Connolly LP, Zurakowski D, Peters CA, et al. Variability of diuresis renography interpretation due to method of post-diuretic renal pelvic clearance half-time determination. *J Urol.* 2000; 164:467–471. [PubMed: 10893624]
23. Piepsz A, Tondeur M, Ham HR. NORA: a simple and reliable parameter for estimating renal output with or without frusemide challenge. *Nucl Med Commun.* 2000; 21:317–323. [PubMed: 10845219]
24. Piepsz A, Kuyvenhoven JD, Tondeur M, Ham H. Normalized residual activity: usual values and robustness of the method. *J Nucl Med.* 2002; 43:33–38. [PubMed: 11801700]

APPENDIX 1: Likelihood, Estimation, and Prediction

Let (Y_{1i}, Y_{2i}) denote the outcomes of i^{th} kidney, normal kidney drainage, and diagnosis of obstruction. We modeled the outcomes as follows:

$$\begin{aligned} \text{logit}(P(Y_{1i}=1)|x_{1i}) &= \beta_0 + \beta_1^T x_{1i} \\ \text{logit}(P(Y_{2i} \geq j|Y_{1i}=1, x_{1i}, x_{2i}, \gamma)) &= \gamma_{1j} + \gamma_1^T x_{1i} + \gamma_2^T x_{2i}, j=1, 2 \\ L(\psi_1, \psi_2|Y_{1i}, Y_{2i}, X_{1i}, X_{2i}^{obs}) &= P(Y_{1i}|X_{1i}, \psi_1)P(Y_{2i}|Y_{1i}, X_{1i}, X_{2i}^{obs}, \psi_2) \equiv L_1(\psi_1) \times L_2(\psi_2) \end{aligned}$$

The parameter vectors of interest were $\psi_1 = (\beta_0, \beta_1^T)^T$ and $\psi_2 = (\gamma_0, \gamma_1^T, \gamma_2^T)^T$. For kidneys that had normal drainage, we first imputed the post-furosemide administration covariate X_{2i} by assuming $X_{2i}|X_{1i} \sim MVN(X_{1i}\zeta, \Sigma)$ where MVN is multivariate normal distribution and ζ and

Σ can be estimated from the training sample. We maximized the likelihood to obtain parameters and obtained standard error by inverting the Fisher information.

In prediction of normal kidney drainage, for a new kidney with baseline predictor x_1^* , the probabilities of normal drainage can be predicted as follows:

$$\widehat{P}(Y_1=1|x_1^*)=\frac{\exp(\widehat{\beta}_0+\widehat{\beta}_1^T x_1^*)}{1+\exp(\widehat{\beta}_0+\widehat{\beta}_1^T x_1^*)}$$

where the parameters with a hat (^) are estimates.

In prediction of kidney obstruction, for a new patient who receives furosemide, both x_1^* and x_2^* are observed for the kidneys. The probabilities of kidney obstruction can be predicted as follows:

$$\widehat{P}(Y_2 \geq j|x_1^*, x_2^*)=\widehat{P}(Y_2 \geq j, Y_1=1|x_1^*, x_2^*)+\widehat{P}(Y_2 \geq j, Y_1=0|x_1^*, x_2^*)=\widehat{P}(Y_2 \geq j, Y_1=1|x_1^*, x_2^*), j+1, 2$$

Hence, the predictive equation for kidney obstruction is

$$\widehat{P}(Y_2 \geq j|x_1^*, x_2^*)=\frac{\exp(\widehat{\gamma}_0+\widehat{\gamma}_1^T x_1^*+\widehat{\gamma}_2^T x_2^*)}{1+\exp(\widehat{\gamma}_0+\widehat{\gamma}_1^T x_1^*+\widehat{\gamma}_2^T x_2^*)} \frac{\exp(\widehat{\beta}_0+\widehat{\beta}_1^T x_1^*)}{1+\exp(\widehat{\beta}_0+\widehat{\beta}_1^T x_1^*)}, j+1, 2$$

For a new patient not receiving furosemide, we used the imputation method described in the estimation part and predicted obstruction as follows:

$$\widehat{P}(Y_2 \geq j|x_1^*)=\int \widehat{P}(Y_2 \geq j|x_1^*, x_2^*)f(x_2^*|x_1^*)dx_2^*$$

where $f(x_2|x_1)$ is the conditional probability density function of x_2 given x_1 , which is assumed to be normally distributed. When $j = 2$, the results of the last two equations are predictive of the probability of kidney obstruction; when $j = 1$, the results of the two equations are predictive of the probability that a kidney is obstructed or that the finding is equivocal.

Fig. 1. 88-year-old woman with suspected obstruction showing different regions of interest

A, Technetium-99m–mercaptoacetyl triglycine (MAG3) image obtained 2–3 minutes after injection shows whole-kidney region of interest (ROI).

B, MAG3 image obtained 20 minutes after injection to generate renogram curve shows parenchymal (cortical) defined ROI representing parenchymal function uncontaminated by pelvic or calyceal retention.

C, MAG3 image obtained 2 minutes after furosemide administration shows whole-kidney ROI.

D, MAG3 image obtained 2 minutes after furosemide administration shows ROI over retention in pelvis and calyces.

Fig. 2. Receiver operating characteristic plot shows accuracy of probability of normal kidney drainage predicted with logistic model against expert consensus for training sample ($n = 120$ kidneys) and validation sample ($n = 71$ kidneys).

Fig. 3. Receiver operating characteristics

A, Plot shows accuracy of probability of interpretation of obstruction versus interpretation of no obstruction or equivocal findings predicted with proportional odds model against expert consensus for training sample ($n = 120$ kidneys) and validation sample ($n = 71$ kidneys).

B, Plot shows accuracy of probability of interpretation of obstruction or equivocal finding versus no obstruction predicted by proportional odds model against expert consensus for training sample ($n = 120$ kidneys) and validation sample ($n = 71$ kidneys).

TABLE 1

Definition of Variables

Variable	Definition
Baseline	
MAG3 clearance	Camera-based MAG3 clearance of kidney under evaluation
Relative uptake	Relative percentage uptake of MAG3 in kidney under evaluation
20 min/2–3-min count ratio	Ratio of count in the whole kidney ROI at 19–20 min to count at 2–3 min
20 min/max count ratio	Ratio of count in whole kidney ROI at 19–20 min to maximum count
T _{max}	Time (min) to the maximum count in whole-kidney ROI
Baseline kidney T _{1/2}	Time (min) from maximum count to half-maximum count for whole-kidney ROI on baseline scan
Cortical 20 min/max count ratio	Ratio of count in cortical ROI at 19–20 min to maximum cortical count
Cortical 20 min/2–3 min count ratio	Ratio of count in renal cortex at 19–20 min to count at 2–3 min
Cortical T _{max}	Time (min) to maximum count in cortical ROI
Cortical T _{1/2}	Time (min) from maximum count to half-maximum counts in cortical ROI
Postvoid/prevoid count ratio	Ratio of count in whole-kidney ROI after voiding to prevoid count
Postvoid/max count ratio	Ratio of count in whole-kidney ROI after voiding to maximum count
After furosemide administration	
Time between studies	Time between completion of baseline study and initiation of furosemide acquisition
Whole-kidney T _{1/2}	Time (min) from maximum count to half-maximum count in whole-kidney ROI
Pelvic T _{1/2}	Time (min) from maximum count to half-maximum count in ROI around tracer retained in renal collecting system
Postfurosemide 20 min/baseline 1–2 min count ratio	Ratio of whole-kidney count after furosemide administration at 19–20 min to 1- to 2-min count of baseline acquisition
Postfurosemide 1–2 min/baseline max count ratio	Ratio of 1- to 2-min whole-kidney count after furosemide administration to maximum count of baseline acquisition
Postfurosemide 20 min/baseline max count ratio	Ratio of the whole-kidney 19- to 20-min count to maximum count of baseline acquisition

Note—MAG3 = ^{99m}Tc-mercaptoacetyl triglycine, ROI = region of interest.

TABLE 2

Selection of Variables for Predicting Normal Drainage (120 Kidneys)

Variable	Univariate Model ^a		Multivariate Model ^b	
	Decrease in $-2\log L$	p	Decrease in $-2\log L$	p
20 min/2–3 min count ratio	70.38(1)	< 0.0001	0.53	0.47
MAG3 clearance	33.58	< 0.0001	0.25	0.61
Postvoid/prevoid count ratio	90.65	< 0.0001	0.16	0.69
Cortical 20 min/max count ratio	109.22	< 0.0001	1.25	0.27
Cortical 20 min/2–3 min count ratio	61.08	< 0.0001	0.62	0.43
Relative uptake	9.35	0.0020	0.0004	0.98
Postvoid/1–2 min count ratio	48.73	< 0.0001	2.10	0.15
Cortical $T_{1/2}$	88.40	< 0.0001	0.52	0.47
$T_{\max}$	84.35	< 0.0001	1.89	0.17
Baseline kidney $T_{1/2}$	58.004	< 0.0001	0.23	0.63
20 min/max count ratio	115.07	< 0.0001	—	—
Cortical $T_{\max}$	43.64	< 0.0001	—	—
Postvoid/max count ratio	162.76	< 0.0001	—	—

Note—Dash (—) indicates variable already chosen to create multivariate model. MAG3 = ^{99m}Tc -mercaptoacetyltriglycine.

^a L = likelihood. The decrease in $-2\log L$ reflects the contribution of each variable. A value greater than 2.71 (chi-square with degrees of freedom of 1) is considered important.

^bThe decrease in $-2\log L$ was less than 2.71 for all remaining variables and indicates that none of the remaining variables improved the model when added to the multivariate model consisting of the 20 min/max count ratio, postvoid/max count ratio, and cortical $T_{\max}$.

TABLE 3

Selection of Variables for Diagnosis of Obstruction (120 Kidneys)

Variable	Univariate Model ^a		Multivariate Model ^b	
	Decrease in -2Log L	<i>p</i>	Decrease in -2Log L	<i>p</i>
20 min/2–3 min count ratio	3.30	0.07	0.60	0.44
MAG3 clearance	0.92	0.34	0.22	0.64
Postvoid/prevoid count ratio	13.11	< 0.01	0.62	0.43
Cortical 20 min/max count ratio	7.50	< 0.01	2.33	0.13
Cortical 20 min/2–3 min count ratio	9.79	< 0.01	1.84	0.18
Relative uptake	3.08	0.08	2.29	0.13
Postvoid/1–2 min count ratio	10.89	< 0.01	1.67	0.20
Cortical T _{1/2}	0.29	0.59	0.01	0.91
20/max count ratio	0.31	0.58	0.11	0.75
Cortical T _{max}	11.49	< 0.01	0.16	0.69
T _{max}	11.79	< 0.01	2.57	0.11
Baseline kidney T _{1/2}	0.0244	0.88	0.43	0.51
Postvoid/max count ratio	15.89	< 0.01	2.17	0.14
Whole-kidney T _{1/2}	12.93	< 0.01	1.32	0.25
Time between studies	3.32	0.07	2.39	0.12
Postfurosemide 20 min/baseline 1–2 min count ratio	24.97	< 0.01	0.90	0.34
Postfurosemide 1–2 min/baseline max count ratio	27.37	< 0.01	1.00	0.32
Postfurosemide 20 min/baseline max count ratio	25.75	< 0.01	—	—
Pelvic T _{1/2}	10.71	< 0.01	—	—

Note—Dash (—) indicates variable already chosen to create multivariate model. MAG3 = ^{99m}Tc-mercaptoacetyltriglycine.

^aL = likelihood. The decrease in -2log L reflects the contribution of each variable. A value greater than 2.71 (chi-square with degrees of freedom of 1) is considered important.

^bThe decrease in -2log L was less than 2.71 for all remaining variables and indicates that none of the remaining variables improved the model when added to the multivariate model consisting of post-furosemide administration 20 min/baseline max count ratio and pelvic T_{1/2}.

TABLE 4

Regression Parameters and Predictive Equations for Training Data ($n = 120$)

Model Effect	Estimate	SE	Wald Chi-Square	<i>p</i>
Model 1: Probability of normal drainage (Y_1)				
Intercept (β_0)	4.85	0.82	35.07	< 0.0001
20/max count ratio (β_1)	-3.31	1.77	3.49	0.061
Postvoid/max count ratio (β_2)	-8.81	1.67	27.83	< 0.0001
Cortical $T_{\max}$ (β_3)	0.14	0.06	5.84	0.016
Model 2: Probability of obstruction given abnormal drainage on the baseline scan (Y_2)				
Intercept (obstructed γ_{02})	-3.15	0.73	18.72	< 0.0001
Intercept (obstructed or equivocal γ_{01})	-1.46	0.56	6.69	0.010
Postfurosemide 20 min/baseline max count ratio (γ_1)	6.31	1.56	6.69	< 0.0001
Pelvic $T_{1/2}$ (γ_2)	0.03	0.02	2.71	0.100

Note—SE = standard error. $\text{logit}(\text{Prob of normal drainage}) = 4.85 - 3.31 \times 20/\text{max_ratio} - 8.81 \times \text{postvoid}/\text{max_ratio} + 0.14 \times \text{cortical } T_{\max}$. Given that the kidney has abnormal drainage on the baseline scan, $\text{logit}(\text{Prob of obstruction}) = -3.15 + 6.31 \times \text{postfurosemide 20 min}/\text{baseline max_ratio} - 0.03 \times \text{pelvic } T_{1/2}$. $\text{logit}(\text{Prob of obstruction or equivocal}) = -1.46 + 6.31 \times \text{postfurosemide 20 min}/\text{baseline max_ratio} - 0.03 \times \text{pelvic } T_{1/2}$.

TABLE 5

Accuracy Measures for Training Data ($n = 120$) and Validation Data ($n = 71$)

Statistic	Cutoff	Sensitivity	Specificity	AUC
Probability of normal drainage ($Y_1 = 1$)				
Training data	0.5	0.89 (0.03)	0.93 (0.02)	0.97 (0.01)
Validation data	0.5	0.85 (0.04)	0.93 (0.02)	0.93 (0.02)
Probability of obstructed kidney ($Y_2 = 2$)				
Training data	0.4	0.86 (0.07)	0.83 (0.07)	0.88 (0.05)
Validation data	0.4	0.82 (0.08)	0.83 (0.07)	0.84 (0.06)
Probability of combined obstructed/equivocal ($Y_2 \geq 1$)				
Training data	0.6	0.85 (0.06)	0.95 (0.05)	0.89 (0.05)
Validation data	0.6	0.78 (0.07)	0.95 (0.05)	0.84 (0.06)

Note—Values in parentheses are standard error. AUC = area under the receiver operating characteristic curve.