Original article

Nuclear Medicine Communications

An electronic technetium-99m-diethylenetriaminepentaacetic acid glomerular filtration rate spreadsheet with novel embedded quality assurance features

Ran Klein[a], Simin Razavi[b], Rayhan Memon[c] and Lionel S. Zuckier[a]

Background **Critical clinical decisions are made on the basis of the glomerular filtration rate (GFR) measured using technetium-99m-diethylenetriaminepentaacetic acid (DTPA) administration, followed by multiple time-point plasma sampling. As GFR studies rely on few data points and produce a single result, they are prone to technical errors that may remain inconspicuous.**

Objective **We describe a data analysis worksheet that provides real-time quality control (QC) indicators and evaluate our initial clinical experience.**

Methods **Two hundred and forty-six consecutive GFR studies carried out at our clinics were included. Our protocol used plasma samples at 2, 3, and 4 h after injection of technetium-99m-DTPA. Duplicate plasma samples, background samples, and aliquots of an activity dilution standard were counted. Times were logged for injection and dilution standard preparation, blood sampling, and counting. Data were entered into a custom GFR analysis spreadsheet that flagged QC in real time at warning and error levels, including QC of the expected ratio between dilution standard counts–activity ratio (CARs) measurements, which was newly introduced to our clinic. The prevalence of QC events was analyzed in three phases: baseline, training, and evaluation ($n = 31$, 69, and 146, respectively).**

Results **From the baseline and training phases ($n = 100$), CAR reference values were determined for each of two sites. In the absence of the CAR QC indicator, errors were present in 5/31 (16%) examinations, but with QC indication decreased to 7/146 (5%) ($P < 0.05$), suggesting that the real-time QC information guided the technologists to ensure proper standard preparation and sample handling, as intended. Improvements in other QC measures were also noted, resulting in an overall error rate reduction from 23 to 8%.**

Conclusion **Real-time analysis of redundant information as a component of the GFR worksheet ensures quality results, but training of technologists and interpreting physicians is essential for optimal utilization of these QC indicators.** ***Nucl Med Commun*** **40:30–40**

Nuclear Medicine Communications 2019, **40**:30–40

Keywords: glomerular filtration rate, quality assurance, renal function, technetium-99m-diethylenetriaminepentaacetic acid

[a]Department of Medicine, Division of Nuclear Medicine, University of Ottawa, [b]Ottawa Hospital Research Institute and [c]Department of Systems and Computer Engineering, Carleton University, Ottawa, Ontario, Canada

Correspondence to Ran Klein, PhD, The Ottawa Hospital, PO Box 232, 1053 Carling Ave, Ottawa, ON, Canada K1Y 4E9
Tel: +1 613 761 4072; fax: +1 613 761 5306; e-mail: rklein@toh.ca

Received 14 May 2018 **Revised** 4 September 2018 **Accepted** 1 October 2018

Introduction

Glomerular filtration rate (GFR) is a well-established metric of renal function [1] used to guide vital clinical decisions [2–5]. Although several methods are available for quantifying renal function [6,7], serial plasma sampling following the administration of technetium-99m (^{99m}Tc)-diethylenetriaminepentaacetic acid (DTPA) has been shown to be a practical and accurate technique [8,9]. This GFR method can also be combined with gamma-camera imaging to derive differential function indices, thereby enabling evaluation of individual kidneys [10].

Administration of a known amount of DTPA intravenously is followed by blood sampling at serial time points. Plasma is separated from blood and the concentration of activity is measured and compared with the initially administered dosage. The rate at which DTPA is cleared from the plasma, or GFR, can then be calculated. A 'corrected GFR' value is also customarily reported by normalizing the patient's body-surface area (BSA) to that of a standard-sized individual [4].

Although methods of obtaining precise GFR measurements have utilized as many as 10 blood samples to capture multi-exponential clearance, a more feasible 'slope-intercept' method incorporates only two to four blood samples starting at two hours following injection. The slope-intercept method has been shown to characterize the terminal DTPA clearance phase with sufficient accuracy and precision to be used in clinical practice, although additional corrections have been proposed to further improve accuracy [11–13]. Simpler single-blood-sample methods have also been described [14–17], and

Preliminary data for this article were presented at the Society of Nuclear Medicine and Molecular Imaging Annual Meeting 2016 and the abstract is published as: Klein R, Memon R, Zuckier L. Quality control of 99mTc-DTPA-derived glomerular filtration rate studies using multiple plasma samples. J Nucl Med 2016; 57(Suppl 2):2719. Incremental data and more robust data analysis clearly distinguish these works.

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website, www.nuclearmedicinecomm.com.

0143-3636

DOI: 10.1097/MNM.0000000000000929

a recent review has shown that the Fleming method [17] can achieve equivalent performance to the slope-intercept method [18], especially in pediatrics [19]. However, single-blood-sample GFR may be inaccurate below 30 ml/min/1.73 m^2 [18,19], making it ill-suited for a general population consisting of renal failure cases, and it has not yet been endorsed ubiquitously [8].

Although these DTPA clearance examinations have been characterized and validated in specialized laboratories, the reliability of any individual test needs to be ensured when ported to routine clinical practice. It is of concern that tests that rely on discrete measurements, such as the DTPA GFR method, are subject to technical errors or blunders that corrupt accuracy, and that the results may not clearly indicating the presence of these errors. In a recent study, McMeekin *et al.* [20] evaluated the effectiveness of previously published quality control (QC) methods for the detection of clinically significant errors. They concluded that QC methods based on expected volume of distribution and agreement between different model results (e.g. single-sample vs. slope-intercept) were not useful error predictors. The correlation between multiple time blood samples was a moderately useful QC indicator, especially at high GFR values, but insufficient on its own. They emphasize that there is no substitution for careful measurements and their recording, along with the use of well-established physics-based QC.

Duplicate sampling and measurement can be useful for detecting random errors and blunders by exploiting physics-based relationships between data; these strategies have been utilized previously to improve the quality of GFR examinations [12,14,20]. In this paper, we have extended this concept by leveraging other known interdependencies in the measured data that are used to flag inherent discrepancies indicative of probable technical errors. An integrated approach, encompassing all exam variables including dilution and measurement of standard, has, to our knowledge, not been described previously. A further novelty of our technique, intended to improve the ease and accuracy of performing this examination in our routine clinical practice, is that we have embedded the QC flags into the technologist's electronic spreadsheet, so that QC checks are incorporated automatically into the clinical workflow with minimum added effort. In this communication, we describe our GFR measurement protocol, related QC methods, and describe our experience during introduction and subsequent routine use of the method in a busy clinical environment.

Methods

This work encompasses all GFR studies carried out between 13 May 2015 and 13 February 2018, at either of two affiliated clinics (Table 1). Studies carried out before 1 September 2015 were used to establish baseline QC performance metrics. Studies carried out between 1 September 2015 and 15 March 2016 (inclusive) were used to phase in the finalized GFR calculation spreadsheet and train the staff on the QC indicators. Training included the following: (a) updating of the study protocol (including intended meaning of QC indicators and appropriate staff response actions), (b) presentation of the revised protocol to technologists at the in-service meeting at each site, (c) presentation of the revised protocol at departmental grand rounds, (d) e-mail communication to all technologists and physicians, and (e) one-on-one consultation of staff members with the department physicist. Studies carried out after 15 March 2016 were used to evaluate the performance of the method with QC fully implemented. In this paper, the three periods are referred to as the baseline, training, and evaluation phases, respectively.

Table 1 Scintillator counters and parameters

	Site 1	Site 2
Counter manufacturer/ model	PerkinElmer Wizard 3 (PerkinElmer, Waltham, Massachusetts, USA)	LTI Multi-Wiper (LTI, Elburn, Illinois, USA)
Type	Single detector, automated feed	10 detectors, manual feed
Count time per sample (s)	120	120
Energy windows (keV)	128–149	128–149
Reference dose calibrator model	Capintec CRC-25R (Capintec, Florham Park, New Jersey, USA)	Capintec CRC-15R (Capintec, Florham Park, New Jersey, USA)
CAR (counts/kBq)[a]	9.84 ± 0.25	15.24 ± 0.56
CAR precision (%)[b]	2.50	3.67*

CAR, counts–activity ratio.
[a]CAR measures are reported as median ± interquartile range/1.35, to account for outlier values in the data.
[b]Precision = interquartile range/1.35/median.
*$P < 0.05$ between sites.

As a clinical quality assurance improvement initiative, this project was exempt from requiring research ethics board approval, as confirmed by our local research ethics board.

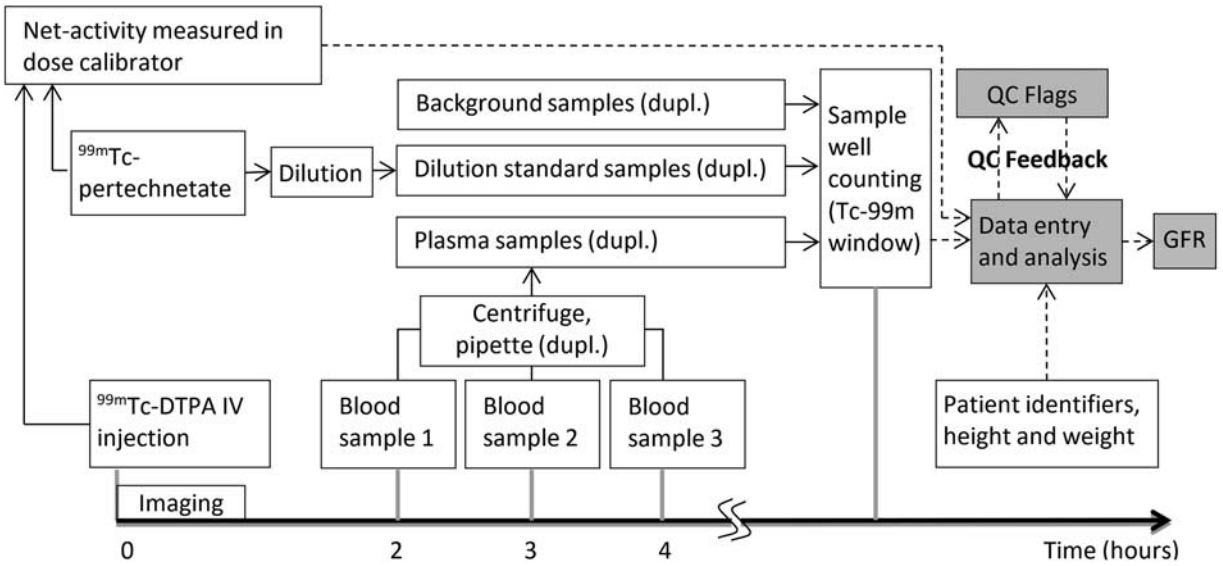
Clinical workflow

All aspects of the studies, including both in-vivo and in-vitro laboratory work, were carried out by one of the certified nuclear medicine technologists on duty on the day of the exam, as per regular work assignment. The clinical workflow is highlighted in Fig. 1 and follows guidelines by the British Nuclear Medicine Society [8]. The times of all radiopharmaceutical administration, blood draws, and counting of activity were initially hand-recorded from a single clock on a printed worksheet in real time; data were transcribed subsequently into the electronic spreadsheet as detailed below.

Glomerular filtration rate protocol

Dose calibrators and well counters at each of our clinical sites (Table 1) are subjected to routine daily QC to ensure consistency with respect to reference sources. Patients were reviewed for counter indications including previous administration of radiopharmaceuticals, expanded body space (e.g. ascites, oedema), and hydration therapy. Patient height and weight were measured and recorded at the start of the exam. Patients were instructed to drink 500 ml of

Fig. 1



GFR workflow. "dupl." indicates that samples were taken in two duplicates. Imaging is an optional component of this workflow. Solid arrows indicate physical manipulations, dashed arrows indicate data flow and gray lines indicate timed events. QC, quality control; ^{99m}Tc-DTPA, technetium-99m-diethylenetriaminepentaacetic acid.

water and to void before commencement of the study. Activity in the injection syringe was assayed in the dose calibrator using the lowest possible range setting both before and after administration of radiopharmaceutical to account for residual activity in the syringe and IV access setup. Radiopharmaceutical injection setup was a butterfly IV access with a four-way stopcock enabling a 10 ml 0.9% saline flush immediately after radiopharmaceutical injection. Imaging of the injection site at ~45 min after injection was included to rule out extravasation. In our practice, the clear majority of studies follow IV administration of ~370 MBq (10 mCi) of ^{99m}Tc-DTPA and incorporate imaging over the subsequent 45 min. A minority of patients were studied without imaging and were administered a reduced ~37 MBq (1.0 mCi) dosage of radiopharmaceutical.

In all cases, venous blood samples for the calculation of GFR were withdrawn at 2, 3, and 4 h after injection using a preinserted, saline-flushed, angiocatheter (20 or 22 G) in the opposite arm used for radiopharmaceutical injection. Each withdrawal consisted of a saline flush of the catheter, draw of ~3 ml blood to be discarded, draw of ~6 ml blood sample into an evacuated heparinized vial, and postdraw catheter flush with 10–20 ml saline. All vials were labeled immediately with patient identifiers and time of blood draw.

Dilution standard

To relate the injected activity measured in the dose calibrator to the sample activity measured in the well-counter, a dilution standard was prepared for cross-calibration. Initially, a 500 ml volumetric flask was filled with ~400 ml of tap water. A syringe of ~12 MBq of ^{99m}Tc-pertechnetate was measured in the dose calibrator both before and after discharge into the flask. The flask was topped up with tap water to the 500 ml meniscus line, capped, and agitated rigorously. 100 µl aliquots of the dilution standard were pipetted in duplicate into well-counter vials after a second, vigorous mixing.

Sample preparation and counting

At the completion of the study, the venous blood samples were centrifuged at 2000*g* for 10 min or longer as needed. 100 µl aliquots of plasma were pipetted in duplicate from each of the venous blood samples into well-counter vials. Two additional blank vials were used to measure the background counts. All 10 vials were then counted according to the settings delineated in Table 1.

Data entry and analysis

During the course of each patient's study, all data, including time, activity, and demographics, were recorded on a temporary worksheet. Following counting of the 10 samples, all data were transcribed from the temporary worksheet into an electronic spreadsheet that was implemented in Microsoft Excel 2010/2016 (Microsoft, Redmond, Washington). Details of the embedded QC checks and calculation of GFR are described below (Fig. 2 and Table 2).

Glomerular filtration rate calculation

All dose calibrator and well-counter measurements were decay corrected to the time of DTPA injection on the basis

of the recorded times and the 6.01 h physical half-life of ^{99m}Tc. Methods of GFR calculation have been described previously [4] and are summarized here briefly as they relate to our QC process. A mono-exponential washout function [$N(t) = N_0 e^{-\lambda t}$] was fitted to the plasma count data, $N(t)$, where N_0 represents the extrapolated plasma counts at the time of injection ($t = 0$) and λ represents the rate of DTPA washout from the plasma (min^{-1}). GFR was then calculated using the following equation:

$$\mathrm{GFR} = \frac{500\,\mathrm{ml} \times N_{\mathrm{std}} \times \lambda \times A}{A_{\mathrm{std}} \times N_0}, \quad (1)$$

where A is the activity administered to the patient (in Bq), A_{std} is the activity of the dilution standard (in Bq), and N_{std} is the count of the 100 µl dilution sample. The dilution standard serves to relate the N_0 sample counts into an activity concentration (Bq/ml). GFR corrected for standardized BSA [21] was calculated as $\mathrm{GFR} \times 1.73\,\mathrm{m}^2/\mathrm{BSA}$ using the Du Bois and Du Bois equation [22]:

$$\mathrm{BSA}\,(\mathrm{m}^2) = 0.20247 \times [\mathrm{height}\,(\mathrm{cm}) / 100]^{0.725} \times \mathrm{weight}\,(\mathrm{kg}) \times 0.425. \quad (2)$$

Quality control testing

QC embedded into the spreadsheet was designed to evaluate all entered data (Table 2) using predefined threshold values to indicate three outcome levels: pass, warning, and error (Table 3). Warnings directed staff to re-evaluate their procedure, but the values were considered tolerable after review; a study with an error indication was considered not acceptable under any circumstance. Warning and error levels were selected to be sensitive and specific to QC issues with intended $\sim 1:20$ and $\sim 1:1000$ false-positive rates of alarm, respectively. Warning levels were therefore more stringent than error levels. Values falling within the low and high warning thresholds passed QC. Four specific analyses were carried out:

(1) Repeat count agreement

Significant differences between paired sample background counts (N_1 and N_2) were tested using a z-score assuming a Poisson distribution (i.e. variance equal to the mean, $\bar{N}$) of counts:

$$z = \frac{|N_2 - N_1|}{\sqrt{\bar{N}}}, \quad (3)$$

$$\bar{N} = \frac{N_1 + N_2}{2}. \quad (4)$$

Similarly, for standard and plasma sample pairs, a z-score incorporating $\sigma_{\mathrm{P}} = 1\%$ pipetting variance (measured in-house by a novice user using an analytical balance as a reference), was used:

$$z = \frac{|N_2 - N_1|}{\sqrt{2\bar{N} + 2\bar{N}^2 \sigma_{\mathrm{p}}}}. \quad (5)$$

To account for four independent tests (standard and three plasma), a Bonferroni correction factor of 3 was applied to reduce the overall study false-positive rate.

(2) Sequential plasma sample agreement

Agreement of the mean counts of the three consecutive plasma samples with mono-exponential washout was evaluated using the Pearson correlation (r) goodness-of-fit parameter.

(3) Counts–activity ratio (CAR) testing

The CAR (counts/kBq) was calculated as the ratio between dilution standard mean number of sample counts [$N_{\mathrm{std}} = (N_{\mathrm{std1}} + N_{\mathrm{std2}})/2$] and the net activity used to prepare the standard (kBq), A_{std}, after decay correction. This formulation is not to be confused with the efficiency of the well counter (counts/s/Bq) as CAR is intentionally designed to be influenced (and therefore reflect) by multiple potential sources of operator error and instrumentation faults associated with the well counter, dose calibrator, pipette, and volume flask. Because standard dilution volume (500 ml), pipetting volume (100 µl), and counting time were used (Table 1), a simple ratio could be used to calculate CAR when using decay-corrected values:

$$C = \frac{N_{\mathrm{std}}}{A_{\mathrm{std}}}\,(\mathrm{counts} / \mathrm{Bq}). \quad (6)$$

Deviation of CAR from the reference truth, μ_C, was reported as a percent error.

$$\Delta C = \frac{C - \mu_C}{\mu_C} \times 100\,\%. \quad (7)$$

The reference CAR, μ_C, and SD, SD_C, were determined for each site independently using the composite baseline and training (i.e. pre-evaluation) data and are listed in Table 1. To account for outlier values, μ_C was determined as the median CAR, C, and SD_C was estimated using SD_{np} = interquartile range/1.35. The evaluation phase data were used to validate our estimate of μ_C and SD_C.

(4) Range checking

To eliminate errors of data entry, several miscellaneous checks were performed to ensure that the recorded data were within expected ranges, including patient BSA, injected activity according to protocol, residual activities in syringes, and blood withdrawal timing. In addition,

Fig. 2

The Ottawa Hospital | L'Hôpital d'Ottawa
Nuclear Imaging
Imagerie Nucléaire

GLOMERULAR FILTRATION RATE WORKSHEET

PATIENT	John, Doe		**HEIGHT(cm)**	182	**Data Quality Assurance**
ID	12345678		**WEIGHT(kg)**	70	
DATE	31-Aug-18		**BSA (m^2)**	1.90	✓ BSA in range
PROTOCOL	GFR with Imaging		**SITE**	Civic	

Set assay times only if different from injection time.	**Patient Dose** Activity (kBq)	Assay Time	**Standard** Activity (kBq)	Assay Time	
Pre assay	391000	09:15	14000	09:15	✓ Injected activity match protocol
Residual assay	6000	09:21	1230	09:15	✓ Standard activity match protocol
Decayed to Injection	384930		12770		✓ Injection residual activity
					✓ Standard prep. residual activity

Time of Injection	09:15	Injection site:		
Time of Plasma 1 (2 hr)	11:10	115		✓
Time of Plasma 2 (3 hr)	12:21	186	Min. after injection	✓ Sample time match protocol
Time of Plasma 3 (4 hr)	13:17	242		✓
Sample Counting	14:34	☐ Following day		✓ Counting after samples prepared

SAMPLE	**Sample 1** (counts)	**Sample 2** (counts)	MEAN-BKGD (counts)	% Diff	z-score	
Background	34	41	38		0.3	✓
Standard	68790	69141	68928	0.51%	0.1	✓
Plasma 1	35704	35761	35695	0.16%	0.0	✓ Sample 1 and 2 agreement
Plasma 2	27478	27817	27610	1.23%	0.2	✓
Plasma 3	22236	22427	22294	0.86%	0.1	✓

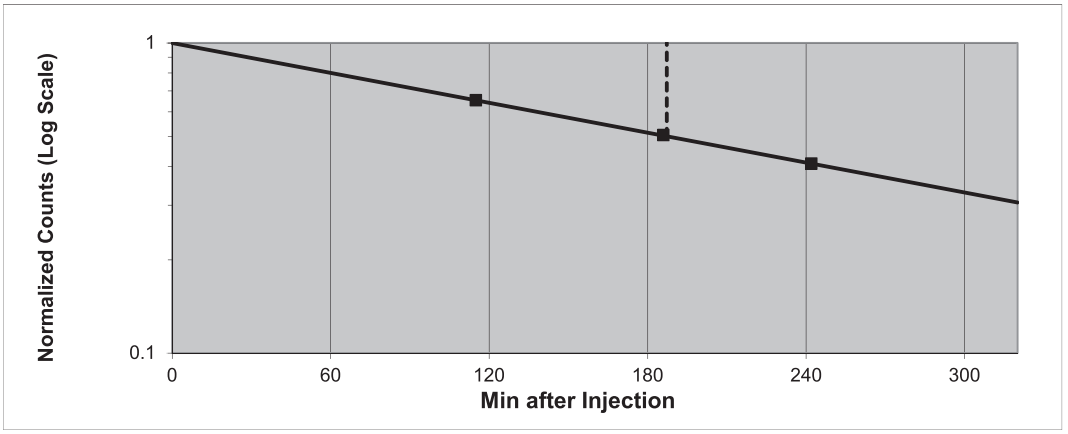
Counts at To from graph	54742	
Time of $T_{1/2}$ from graph (min)	187	

✓ **Counts-Activity Ratio**
9.97 cts/kBq of std.
1.59% Deviation

✓ Curve fit
0.9998 r^2

$$\text{GFR (mL/min)} = \frac{\text{Std counts x 500 mL x ln(2) x patient dose (kBq)}}{\text{Plasma cts at } T_o \text{ x } T_{1/2} \text{ x std dose (kBq)}}$$

Corrected GFR (mL/min/1.73 m^2) = GFR x 1.73 / BSA



GFR = 70 ml/min	**QC tests passed.**	Entered by
Corrected GFR = 64 ml/min/1.73 m^2	✓	Checked by
		Spreadsheet date: 2016-Sep-07

Example spreadsheet for the calculation of glomerular filtration rate (GFR), body surface area-corrected GFR, and quality control checks. Shaded boxes indicate input data. The last two columns consist of a visual indicator of quality control (checkmark, x, or exclamation marks) and an explanation thereof. BSA, body surface area GFR, glomerular filtration rate.

Table 2 Glomerular filtration rate spreadsheet data input, dependent calculations, and related quality control checks

Input	Dependent calculations	Related quality control
Patient name	–	–
Patient identification number	–	Correct number of digits
Exam date	–	–
Patient height (cm)	BSA, GFR correction	Height and BSA in range
Patient weight (kg)	BSA, GFR correction	Weight and BSA in range
Protocol (with or without imaging)	Injected activity QC limits	–
Injection site	–	–
Patient injection activity preinjection (kBq) and time	Injected activity, decay corrections	Injected activity in range, event sequence in order
Patient injection residual activity (kBq) and time		
Standard activity preinjection (kBq) and time	Standard activity concentration	Dilution standard activity in range, event sequence in order
Standard residual activity (kBq) and time		
Time of injection	GFR	Timing of events follows protocol, event sequence in order
Time of blood sample 1 (2 h)		
Time of blood sample 2 (3 h)		
Time of blood sample 3 (4 h)		
Blank sample counts (in duplicate)	Background activity correction	Agreement between background
Standard sample counts (in duplicate)	GFR, CAR	Agreement between standard samples, standard preparation
Plasma sample counts (in duplicate)	Washout rate, GFR	Agreement between samples, exponential washout pattern

Injections, blood draws, and measurement of activity are all associated with the corresponding time stamps.
BSA, body surface area; CAR, counts–activity ratio; GFR, glomerular filtration rate; QC, quality control.

Table 3 Glomerular filtration rate spreadsheet quality control checks, nominal value, and warning and error limits

		Limits			
Test	Nominal value	Low error	Low warning	High warning	High error
BSA	–	0.25	0.5	3.0	3.5
Injected activity match protocol (%)		50	70	150	200
With imaging (MBq)	370				
Without imaging (MBq)	37				
Standard activity match protocol (MBq) (%)	12.5	40	60	300	600
Injection residual activity (%)	0	0	0	25	50
Standard residual activity (%)	0	0	0	80	90
Blood sample times (min)	0	−45	−15	+15	+45
Sample agreement error (*z*-score)	–	–	–	1.96	3.29
CAR (%deviation from reference)	0	−15	−10	10	15
Washout curve fit (*r*)	1	0.95	0.985	–	–
Injection time	–	Injection assay time	–	–	Injection residual assay time
Standard assay time	–	–	–	–	Standard residual assay time
Counting after samples	–	4 h blood sample time	–	–	–

BSA, body surface area; CAR, counts–activity ratio.

timing of patient injection, blood withdrawals, and sample counting were tested for proper sequence.

Quality control indicators

QC test limits incorporated into the GFR calculation spreadsheet are summarized in Table 3. During the evaluation phase, a QC summary indicator was added beside the final GFR metrics (as shown in Fig. 2) that reported the most severe status of all the above-described tests with the following accompanying wording:

(1) Pass – 'QC test passed'.
(2) Warning – 'QC test at tolerance – Results may be inaccurate'.
(3) Error – 'QC test failed – Results are likely inaccurate'.

On the basis of the direct QA feedback embedded in the spreadsheet, when warnings or errors were encountered, the technologists were prompted to recheck their data, often resulting in repeated pipetting or self-correction of other errors, without consulting with the physicist. In these cases, the interim spreadsheet was replaced by the subsequent correction. In other cases, especially when errors were refractory, the technologist staff consulted with the department physicist in real time, which always resulted in successful remediation of the examination. As a general rule, the physicist did not review study results in real time during study accrual unless consulted.

Retrospective review of quality control

At the conclusion of the evaluation phase, all GFR studies carried out in our department were reviewed by the department physicist to assess changes in quality. The results presented here therefore represent an evaluation of routine technologist-driven and physician-driven performance. Incidents of warning and error indicators were recollected from the clinically reported GFR spreadsheets. Rates of warning and error indications were

tabulated retrospectively for each study phase. Incidences of interim errors, which were self-corrected, were not captured.

Statistical analysis

Event rates were compared using a two-proportion z-test. Except for CAR, continuous variables were summarized as mean ± SD, with precision expressed as the standard error (SE = SD/mean × 100%). Comparisons between the means of continuous variables were performed using the Student t-test, and variances were compared using the f-test. P values less than 0.05 were considered statistically significant. Data analysis was carried out using Matlab 2017b (MathWorks, Natick, Massachusetts, USA).

Results

Two hundred and forty-six GFR exams were performed in 207 patients; 39 patients underwent two examinations. The first 31 studies were carried out during the baseline phase, 69 studies were carried out during the training phase, and the remaining 146 studies were carried out during the evaluation phase. Studies were roughly evenly split between the two performance sites (110 vs. 136 studies).

Patient demographics and GFR measurements are summarized in Table 4. All patients were adults (>18 years old). For all 246 studies, GFR measured 99.6 ± 38.4 (9.6–209) ml/min [mean ± SD [minimum–maximum)], with BSA-corrected GFR of 88.5 ± 31.7 (10.1–218) ml/min/1.73 m^2. There were no significant differences in the means between the three phases of the study ($P > 0.24$) and the referral patterns also did not change; thus, we assumed that the patient population remained similar. 92% of the studies incorporated imaging, whereas 8% did not. Injected activities were 367 ± 89 (233–442) MBq for studies with imaging and 89 ± 111 (20–392) MBq without imaging. The activity concentration of the dilution standard measured 16.0 ± 5.9 (1.5–47.6) kBq/ml at the time of sample counting.

Counter–activity ratios

Reference CAR and precisions, as reported in Table 1, were determined from the pre-evaluation (baseline and training) phase data. Standard error estimates (precision) between the two sites varied significantly even after excluding outlier values ($P = 0.013$); hence, we decided to use ± 10 and ± 15% deviations in CAR measurements for warning and error levels, respectively, instead of normal distribution-derived z-score limits. CAR deviations from reference values are shown as percent error over the course of the study in Fig. 3.

In the final evaluation phase of the study, CARs were 9.84 ± 0.20 and 15.06 ± 0.49 counts/kBq for the two respective sites. There was no statistically significant change in average or precision (median or SD_{np}) values between the pre-evaluation phases and the evaluation phase for either site. Precision remained superior at site 1 than at site 2 (2.01 vs. 3.24%, respectively, $P < 0.001$).

Table 4 Patient demographics

Parameters ($N = 246$)	Mean ± SD (minimum–maximum)
Age (years)	56 ± 13 (23–85)
Female (%)	115/246 (47)
Weight (kg)	83.8 ± 19.1 (44–140)
Height (cm)	169 ± 11 (117–203)
BSA (m^2)	1.94 ± 0.25 (1.44–2.58)
GFR (ml/min)	99.6 ± 38.4 (9.6–208.8)
GFR (ml/min/1.73 m^2)	88.5 ± 31.7 (10.1–218.2)

BSA, body surface area; GFR, glomerular filtration rate.

Quality indicator rates

The resulting warning and error rates at each phase of the study are reported in Table 5. For most indicators (adherence to protocoled blood sampling times, agreement between duplicate counting samples, and curve quality of fit), the incidences of errors and warnings were too small to test for significant change between study phases. CAR error rates were higher in the pre-evaluation phases than in the evaluation phase (14 vs. 5%, $P = 0.01$), as were the combined warning and error rates (18 vs. 5%, $P = 0.002$), showing improved adherence to quality standards during the evaluation phase. The summary QC indicator similarly indicated higher error rates in the pre-evaluation phases than in the evaluation phase (17 vs. 8%, $P = 0.02$). Rates of combined error and warning also decreased between these corresponding phases (33 vs. 22%, $P < 0.05$).

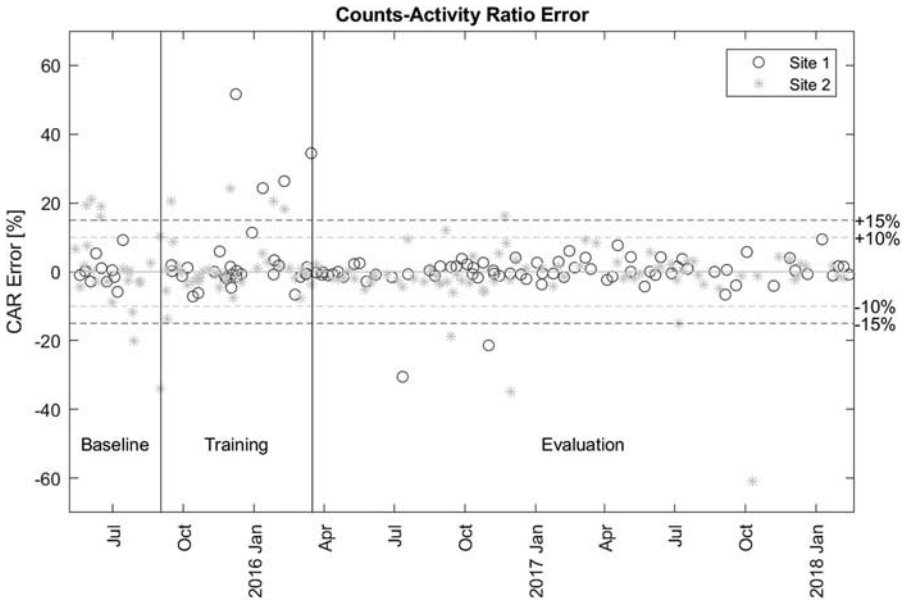
Discussion

This work describes implementation of a rigorous QC program for nuclear GFR measurements studies, and evaluates its adoption in a routine clinical practice. The GFR protocol involves several manual steps that are prone to human error that may not be apparent in the final results, which has motivated our attempt to incorporate automated quality testing into the clinical worksheet. These tests are therefore intended to identify potentially inaccurate data, prompting real-time review and correction, with minimal incremental workload for the staff. Our experience has shown that our GFR protocol and QC efforts can be implemented seamlessly and effectively in a busy clinical setting.

A key novelty of our methodology is testing of the dilution standard against an expected count-rate dependent on the well-counter and dose-calibrator efficiencies, which should remain constant over the life of the devices [8]. The incremental work required by the technologists to perform these QC measures is negligible, primarily requiring recording of the time of sample counting.

Retrospective review of data from the baseline phase of this study showed that in the absence of a QC indicator, CAR errors were common (16% of studies), justifying the

Fig. 3



Count–activity ratio (CAR) errors over the course of the study. The vertical lines indicate the boundaries of the three study phases and the horizontal lines represent the warning and error indicator threshold levels.

Table 5 Quality control test results from the clinical glomerular filtration rate worksheet review

	Number of events (%prevalence) [*n* (%)]					
	Baseline set ($n=31$)		Training phase ($n=69$)		Evaluation phase ($n=146$)	
Test	Warning	Error	Warning	Error	Warning	Error
Protocol timing	2 (6)	1 (3)	7 (10)	0 (0)	9 (6)	1 (1)
Duplicate samples agreement	0	1 (3)	4 (6)	1 (1)	13 (9)	2 (1)
Curve quality of fit	4 (13)	0 (0)	4 (6%)	0 (0)	5 (3)	2 (1)
CAR agreement	1 (3)	5 (16)	3 (4)	9 (13)	1 (1)	7 (5)*
Summary quality indicator	4 (13)	7 (23)	12 (17)	10 (14)	21 (14)	11 (8)*

CAR, counts–activity ratio.
*$P<0.05$ for error rates between pre-evaluation phases and the evaluation phase.

need for more rigorous QC. With QC feedback, in the final evaluation phase of our study, errors that would have otherwise gone unnoticed were identified and corrected before interpretation of the GFR measures, as indicated by lower error rates (5%). CAR errors were driven by dilution standard preparation and/or pipetting errors – both of which proportionately affect GFR results.

Our experience has shown that training of technologists and interpreting physicians was required to facilitate understanding of QC indicators and how they should be utilized. On the basis of these lessons, we iteratively adjusted the spreadsheet to provide more intuitive feedback, with the final version including a summary QC indicator. Change management requires ongoing education and development of a quality-oriented culture. Stringent QC is paramount for accurate and precise results, and improves clinician confidence.

Data redundancy

QC testing is possible through redundant information such as duplicate samples and utilization of the goodness-of-fit (Pearson *r*) parameter of the mono-exponential washout model. CAR leverages redundancy between the activity of the dilution standard and the counts of its samples by taking advantage of fixed counter and dose calibrator efficiencies, and protocoled pipetting and standard volumes. In many GFR protocols, the dilution standard traditionally serves a dual purpose: (a) converting counts into activity concentration, (b) performing decay correction between the time of patient injection and the time of sample counting. By assuming that the counter and dose calibrator efficiencies are stable over time and applying accurate time-based decay correction of the radiopharmaceutical activity, it is possible to carry out an accurate GFR study in the absence of a standard altogether. We opted to continue the use of a dilution

standard, and maintain redundancy, to enable quality testing that encompasses pipetting and counter accuracy.

Counts–activity ratio error limits

Originally, we sought to determine CAR warning and error limits on the basis of empirical measurements of the natural variability from the pre-evaluation phases of the study and assuming a Gaussian distribution. On the basis of these results (Table 1), the respective CAR error and warning limits (associated with 5 and 0.1% false-positive rates) would have been 5.0 and 8.2% for site 1 and 7.3 and 12.1% for site 2. In practice, we encountered much higher warning and error rates, prompting us to use arbitrarily selected 10 and 15% warnings and errors limits that we believed were acceptable within the medical context [20]. We speculate that the main sources of CAR errors were variations in pipetting techniques amongst operators and/or standard preparation as the instrumentation (i.e. dose calibrator, counter, pipets, volumetric flask) precisions far exceed these measures. Nevertheless, we are hopeful that with ongoing efforts, the magnitude of the CAR errors will continue to decrease, enabling us to apply more stringent error limits in the future.

Additional quality control measures

Sanity checks on entered and calculated data can be effective safeguards for detecting errors such as use of erroneous units or transcription errors. In our spreadsheet, in addition to the tests listed in Table 3, we also ensured that patient identification numbers contained eight numerical digits and that time stamps followed the sequence shown in Fig. 1.

Other physiologic parameters could be considered for QC testing, including volume of distribution, extra-cellular volume, and time to 50% washout ($T_{1/2}$) [4,23]. We did not include these error measures as variabilities, even in normal populations of potential donors, were relatively high compared with the physics-based quality metrics described in this work, and certainly were outside of 'normal' values in patients with renal disease. QC on the basis of physiologic metrics is difficult in a heterogenous clinic such of ours, consisting of patient populations ranging from healthy, potential kidney donors to patients with severe renal dysfunction. As an example, errors between volume of distribution estimated using BSA and Brochner–Mortensen corrected GFR and BSA alone have been proposed for QC [23], but had 24.0% variability, which is too large to be of clinical utility as a QC measure [14]. This conclusion is consistent with those of McMeekin *et al.* [20].

The British Nuclear Medicine Society guidelines recommend that sample counts exceed 10 000 counts [8]. In this work, 40/246 exams had one or more sample that did not conform to this requirement. Of these, 18 tests did not include imaging and therefore had a lower injected activity of ~37 MBq. Of the remaining 22 exams, 20 exams had duplicate counts that summed to more than 10 000 counts. The two remaining cases had sample counting performed on the following day (using 10 min/sample) because of technical issues. In one case, duplicate sample counts agreed to within the 95% confidence interval. The single remaining case indicated proper injected activity (374 MBq) and no evidence of subcutaneous injection (by imaging of the injection site), but did have a warning for low agreement between duplicate 3 h plasma counts – nevertheless, correlation between time samples was high ($r=0.998$), indicating a good model fit. The use of a broader energy window in the well-counter settings (Table 1) would serve to increase count statistics, but would necessitate derivation of new reference CAR values.

The British Nuclear Medicine Society guidelines [8] recommend $r=0.985$ as a lower limit for the correlation coefficient between multiple sample count rates. We applied this value as a warning threshold and used $r=0.95$ as the error threshold, recognizing that lower r values are expected in low GFR cases, in which the gradient of the fit is low [20]. Future work could evaluate the use of a GFR-dependent lower limit for r or another metric for consistency between multiple samples.

Persistent errors

Although we sought to eliminate all errors in the GFR exams performed by our clinic, some remaining errors did persist even into the evaluation phase (Fig. 3 and Table 5). Five of the seven studies with indications of CAR errors were in the first months of the evaluation phase, following which the prevalence of this error decreased markedly. One of the remaining two errors was associated with an intermittent counter system hardware issue that triggered servicing of the device, showing the utility of the CAR test in detecting equipment errors. The last CAR error (−61%) was not handled in a timely manner and therefore no corrective action was undertaken – an error margin was included in the patient report with recommendation for a repeat study.

In a single study an error was indicated for late withdrawal of a blood sample which resulted from tardy return of the patient to the department. This study also had a low correlation between the plasma sample counts error indicated ($r=0.911$). The physicist was consulted on this case in real time and it was determined that there were no other technical issues with the study and that the low correlation was associated with the patient having very low plasma clearance (GFR =10 ml/min), consistent with a clinical history of chronic kidney disease. An additional blood sample was recommended for validation, but the patient could not comply with this request.

Of all the examinations in the evaluation phase, 22% had one or more warnings, which was an improvement over the pre-evaluation phases (33%). Roughly 30% of the warnings were associated with blood withdrawal timing

deviating from the protocol. Because the actual time of the blood draw is to a large degree accounted by the curve fitting, this type of deviation from the protocol may not result in GFR inaccuracy [18], unless the error was because of an error in time keeping [8,20]. Through ongoing quality initiatives, we will endeavor to further reduce these and other warning rates.

Our clinic is staffed by over 25 technologists, eight physicians, and multiple trainees, requiring training of over 35 individuals across two practice locations. Most staff are shift workers, further complicating face-to-face communication with all individuals. Furthermore, because of the relatively small volume of GFR studies in our clinic, it is likely that this study is carried out infrequently by some individuals. Thus, some individuals have had limited opportunity for hands-on experience with our new QC measures. One potential means to achieve more rapid improvement in quality is to improve the proficiency of select individuals in the group by restricting the performance of GFR studies to a subset of technologists and physicians. Improved communication and training techniques may also have been able to achieve better compliance with quality assurance directives, and are the subject of future efforts.

Translation to other sites

Translation of our QC methods to other institutions requires the configuration of site-specific parameters, including CAR, size of dilution volume, pipetting volume, and protocolled injection activity, not to mention warning and error thresholds. An editable GFR calculation spreadsheet is included in Supplementary Material 1 (Supplemental digital content 1, *http://links.lww.com/NMC/A138*), and includes a settings page where these parameters may be configured.

Measurement variability is expected because of physical, instrumentation, and human factors and can be estimated from prospective, ad-hoc measurement experiments (e.g. measurement of activity, dilution volume, pipetting volume, and counting statistics). Furthermore, these evaluations should also incorporate multiple technologists to account for interoperator variability. An alternative approach, such as that described in this work for the CAR metric, is to analyze retrospective data that inherently reflect measurement variability resulting from all the above-mentioned sources.

Alternative protocols

Our clinical protocol consists of three blood samples at 2, 3, and 4 h after a radiopharmaceutical injection. Alternative protocols consisting of different numbers of blood samples, different sampling times, and different modeling techniques have been proposed previously [8,18]. The QC measures described in this work could be adapted easily by a qualified physicist for alternative protocols.

Limitations

A limitation of this work is the relatively small number of GFR exams, which is because of the infrequent performance of this exam at our institution (~100 studies/year). The use of previous data to increase the exam pool was not possible as the protocol used for earlier studies did not include recording of exact times of sample counting, precluding CAR calculations. Nevertheless, appropriate statistical testing could identify significant differences that support our conclusions.

Conclusion

We have developed, validated, and reported on a rigorous quality assurance workflow for GFR studies using the ^{99m}Tc-DTPA methodology and based on a dedicated calculator spreadsheet that is readily implementable in a routine clinical practice. This real-time QC reduced the frequency and magnitude of errors, which increased confidence in GFR measurements. Efforts to provide education and develop a culture of quality were required to ensure its optimal use.

Acknowledgements

The authors thank all the technologists at the Department of Nuclear Medicine at The Ottawa Hospital who collaborated in implementing these changes in the clinic and provided invaluable feedback.

Conflicts of interest

Ran Klein is consultant with Jubilant DRAXimage and has received grant funding from industry partnership programs including GE Healthcare, Jubilant DRAXimage, Shelley Medical Solutions, and Hermes Medical Solutions. Ran Klein receives revenues from rubidium generator technology licensed to Jubilant DRAXimage and revenue shares from the sale of FlowQuant. For the remaining authors there are no conflicts of interest.

References

1 Stevens LA, Coresh J, Greene T, Levey AS. Assessing kidney function–measured and estimated glomerular filtration rate. *N Engl J Med* 2006; **354**:2473–2483.
2 Fehrman-Ekholm I, Skeppholm L. Renal function in the elderly (>70 years old) measured by means of iohexol clearance, serum creatinine, serum urea and estimated clearance. *Scand J Urol Nephrol* 2004; **38**:73–77.
3 Jeon YK, Kim MR, Huh JE, Mok JY, Song SH, Kim SS, *et al.* Cystatin C as an early biomarker of nephropathy in patients with type 2 diabetes. *J Korean Med Sci* 2011; **26**:258–263.
4 Murray AW, Barnfield MC, Waller ML, Telford T, Peters AM. Assessment of glomerular filtration rate measurement with plasma sampling: a technical review. *J Nucl Med Technol* 2013; **41**:67–75.
5 Song W, Sung HH, Han DH, Jeong BC, Seo SI, Jeon SS, *et al.* The effect of contralateral kidney volume on renal function after radical nephroureterectomy: implications for eligibility for neoadjuvant chemotherapy for upper tract urothelial cancer. *Urol Oncol Semin Orig Investig* 2017; **35**:114.e1–114.e7.
6 Knight EL, Verhave JC, Spiegelman D, Hillege HL, de Zeeuw D, Curhan GC, *et al.* Factors influencing serum cystatin C levels other than renal function and the impact on renal function measurement. *Kidney Int* 2004; **65**:1416–1421.
7 White CA, Akbari A. The estimation, measurement, and relevance of the glomerular filtration rate in stage 5 chronic kidney disease. *Semin Dial* 2011; **24**:540–549.
8 Fleming JS, Zivanovic MA, Blake GM, Burniston M, Cosgriff PS, British Nuclear Medicine S. British Nuclear Medicine Society. Guidelines for the measurement of glomerular filtration rate using plasma sampling. *Nucl Med Commun* 2004; **25**:759–769.
9 Levey AS, Inker LA. Assessment of glomerular filtration rate in health and disease: a state of the art review. *Clin Pharmacol Ther* 2017; **102**:405–419.

10 Bokacheva L, Rusinek H, Zhang JL, Lee VS. Assessment of renal function with dynamic contrast-enhanced MR imaging. *Magn Reson Imaging Clin N Am* 2008; **16**:597–611.
11 Bröchner-Mortensen J. A simple method for the determination of glomerular filtration rate. *Scand J Clin Lab Invest* 1972; **30**:271–274.
12 Wanasundara SN, Wesolowski MJ, Barnfield MC, Waller ML, Murray AW, Burniston MT, *et al*. Accurate and precise plasma clearance measurement using four 99mTc-DTPA plasma samples over 4 h. *Nucl Med Commun* 2016; **37**:79–86.
13 Fleming JS. An improved equation for correcting slope-intercept measurements of glomerular filtration rate for the single exponential approximation. *Nucl Med Commun* 2007; **28**:315–320.
14 Fleming JS, Persaud L, Zivanovic MA. Quality control of slope-intercept measurements of glomerular filtration rate using single-sample estimates. *Nucl Med Commun* 2005; **26**:737–742.
15 De Sadeleer C, Piepsz A, Ham HR. Post-test quality control for the single blood sample technique in glomerular filtration rate measurement in children. *Nucl Med Commun* 2006; **27**:255–260.
16 Christensen AB, Groth S. Determination of 99mTc-DTPA clearance by a single plasma sample method. *Clin Physiol* 1986; **6**:579–588.
17 Fleming JS, Persaud L, Zivanovic MA. A general equation for estimating glomerular filtration rate from a single plasma sample. *Nucl Med Commun* 2005; **26**:743–748.
18 McMeekin H, Wickham F, Barnfield M, Burniston M. A systematic review of single-sample glomerular filtration rate measurement techniques and demonstration of equal accuracy to slope-intercept methods. *Nucl Med Commun* 2016; **37**:743–755.
19 Blaufox MD, Aurell M, Bubeck B, Fommei E, Piepsz A, Russell C, *et al*. Report of the radionuclides in nephrourology committee on renal clearance. *J Nucl Med* 1996; **37**:1883–1890.
20 McMeekin H, Wickham F, Barnfield M, Burniston M. Effectiveness of quality control methods for glomerular filtration rate calculation. *Nucl Med Commun* 2016; **37**:756–766.
21 Larsson A, Malm J, Grubb A, Hansson LO. Calculation of glomerular filtration rate expressed in mL/min from plasma cystatin C values in mg/L. *Scand J Clin Lab Invest* 2004; **64**:25–30.
22 Du Bois D, Du Bois E. A formula to estimate the approximate surface area if height and weight be known. *Arch Intern Med* 1916; **17**:863–871.
23 Holness JL, Fleming JS, Malaroda AL, Warwick JM. (99m)Tc-DTPA volume of distribution, half-life and glomerular filtration rate in normal adults. *Nucl Med Commun* 2013; **34**:1005–1014.