Evaluation of agreement between two instruments in measurements of radiation dose rates in cats that underwent iodine I 131 treatment

Heather J. Chalmers, DVM; Peter V. Scrivani, DVM; Nathan L. Dykes, DVM; Lorna J. Hubble; Jeff M. Hobbs; Hollis N. Erb, DVM, PhD

Objective—To assess agreement between a commercially available Geiger-Meuller (GM) survey meter and millirem tissue-equivalent (TE) meter for measuring radioactivity in cats treated with sodium iodine I 131 (^{131}I).

Animals—15 cats with hyperthyroidism and undergoing ^{131}I treatment.

Procedures—Duplicate measurements were obtained at a distance of 30 cm from the thyroid region of each cat's neck by 2 observers who used both meters on day 3 or 5 after ^{131}I administration. Comparisons of measurements between meters and observers were made, with limits of agreement defined as the mean difference ± 2 SDs of the differences.

Results—For observer 1, the mean of the differences in the 2 meters' measurements in all cats was 0.012 mSv/h (SD, 0.011 mSv/h). For observer 2, the mean of the differences in measurements was 0.012 mSv/h (SD, 0.010 mSv/h). For the GM meter, the mean of the differences of the 2 observers for all cats was 0.003 mSv/h (SD, 0.011 mSv/h). For the TE meter, the mean of the differences of the 2 observers for all cats was 0.003 mSv/h (SD, 0.007 mSv/h).

Conclusions and Clinical Relevance—Results indicate that there was considerable agreement between meters and observers in measurements of radioactivity in cats treated with ^{131}I. Measurements obtained by use of the GM meter may be approximately 0.01 mSv/h less than or 0.03 mSv/h higher than those obtained with the TE meter. If this range is acceptable for an institution's release criteria, the 2 meters should be considered interchangeable and acceptable for clinical use. (*Am J Vet Res* 2007;68:354–357)

Use of instruments to measure radioactivity is common in veterinary nuclear medicine. Much of the daily operation of veterinary nuclear medicine facilities depends on appropriate and correct use of radiation detection instruments for detection of benchtop contamination, assessment of patient radioiodine uptake after treatment, and postprocedural monitoring of residual radioactivity in patients until release criteria are satisfied. Radiation measurements obtained to determine whether cats treated with ^{131}I have met release criteria must be accurate to ensure safe release practices. Although various types of instruments are available, GM survey meters are used for all these applications in many veterinary nuclear medicine facilities. In general, this type of instrument is simple to use, portable, less expensive than other machines, and durable.[1]

Although use of a GM survey meter is appropriate for detecting contamination in the laboratory, use of this instrument to estimate residual radioactivity in treated cats is less than ideal because of limitations inherent in the detection system.[2] The GM counter is known to over-read radiation exposure. The meter consists of an ion chamber in which a single photon (eg, a gamma ray) produces > 1 ion, causing the radiation exposure reading to be higher than actual radiation exposure. As such, the count rate yielded by this type of instrument is not related to the energies of the detected gamma rays or to the gamma-ray exposure rate.[2] Even with calibration of the instrument's scale to display exposure rate units, these readings may be in error by 2- to 3-fold in some instances.[2] Additionally, GM survey meters calibrated to cesium 137m dose rates are known to over-respond to lower-energy gamma emitters.[3]

In contrast, organic or TE scintillation detectors are designed to estimate dose and generate a TE dose in units of millirems per hour. These instruments are intended for use in situations in which accurate measurements are desired at lower-energy radiation levels and each photon produces a single scintillation event (a light photon) that is converted to an electron, which is recorded. These instruments have a flatter energy response at all levels and thus do not overrespond to the same degree as GM meters.[2] Moreover, TE meters are portable, durable, and easy to use. The cost of such an instrument, however, is substantially higher than that of a GM survey meter.

Abbreviations

GM	Geiger-Mueller
TE	Tissue-equivalent
^{131}I	Sodium radioiodine

Received June 27, 2006.
Accepted October 18, 2006.
From the Departments of Clinical Sciences (Chalmers, Scrivani, Dykes, Hobbs), Environmental Health & Safety (Hubble), and Population Medicine and Diagnostic Sciences (Erb), College of Veterinary Medicine, Cornell University, Ithaca, NY 14853.
Address correspondence to Dr. Scrivani.

The advantages and limitations of the GM survey meter versus those of the TE scintillation detector in determining radiation dose are such that the use of a GM survey meter to satisfy release criteria warrants evaluation in a clinically relevant situation. Our objective was to use 2 observers with different levels of expertise to evaluate agreement in measurements between the 2 types of instruments of radiation dose rates from cats that had received ^{131}I.

Materials and Methods

The study population consisted of all cats admitted to the Cornell University Hospital for Animals that underwent ^{131}I treatment for hyperthyroidism from February 1, 2005, to June 30, 2005. Fractious cats (n = 2) were excluded. Additionally, readings were not obtained if both observers were not available (n = 4 cats). All cats received ^{131}I (between 148 and 222 MBq [4 and 6 mCi], SC) in the caudal portion of the thoracic or lumbar region. The project was approved by the Cornell Institutional Animal Care and Use Committee.

Radiation dose rates were measured by use of a commercially available GM survey meter[a] and a millirem TE meter[b] 3 or 5 days after ^{131}I administration. Recorded measurements (mrem/h) were converted to SI units (mSv/h) for publication (100 mrem = 1 mSv). Both instruments were calibrated to a National Institute of Standards and Technology traceable cesium 137 source annually according to institutional requirements.

Cats were transferred from the ^{131}I ward to a shielded area for obtaining measurements. Measurements were made 30 cm from the area in proximity to the thyroid gland on the ventral aspect of the neck of each cat by a board-certified veterinary radiologist (NLD) and a veterinary imaging resident (HJC). Each measurement was obtained twice to yield replicates for each observer for each meter. The observers were not blinded to cats' identities, but each was unaware of the other's measurements. For determining agreement between the 2 meters and between observers, the mean value of the replicate measures for each observer was used.

Instrument comparison—Results obtained by the veterinary radiologist and resident were plotted separately; the pairs of measurements from each meter for each cat were plotted in a scatter diagram, and the linear association was quantified via the Spearman rank correlation. Perfect within-pair agreement in measurements would yield a graph with all observations on the line of equality and *r* would equal 1.0; disagreement in measurements but perfect agreement in rankings of measurements between meters would yield observations that were not on the line of equality, but *r* would still equal 1.0.

As another way of describing agreement between instruments, a separate Bland-Altman bias graph was drawn for each observer. The x-axis was the cat-specific mean of the pair of measurements for each instrument ([GM + TE]/2), and the y-axis was the cat-specific difference (GM – TE) between the instruments' measurements. With perfect agreement between instruments, all observations would fall on the horizontal line at y = 0; if there was a simple additive bias (difference) between instruments but the bias was highly repeatable, the measurements would fall with little vertical scatter near a nonzero horizontal line. To calculate the limits of agreement for each Bland-Altman graph, the mean and SD of the 15 between-instrument, cat-specific differences (the y-axis data) were calculated; the limits of agreement were the means of differences ± 2 SDs of the difference.[4]

Use of repeated measures by each of the 2 methods may yield an estimate of the SD of the differences that is too small because some of the effect of repeated-measurement error has been removed.[4] For this reason, a correction was made.[4] The limits of agreement are estimates of values that apply to the whole population, and a second sample would yield different results.[4] As such, SEs were calculated to determine how precise our estimates were.[4]

Interobserver agreement—Bland-Altman analysis similar to that described for the observer-specific graphs was performed to describe the comparison of the 2 instruments. Interobserver variation for each instrument was described separately.

Repeatability—A method may be considered to be repeatable when 95% of (the absolute values of) differences between duplicate measures are < 2 SDs of the differences.[4] Therefore, assuming a mean difference of 0, repeatability was assessed by obtaining the SD of the differences between replicate measures per cat and comparing them by inspection to the differences between replicate measures for each instrument and observer.[4]

Results

Fifteen cats met all selection criteria and were enrolled in the study. Scatter diagrams of the mean of the replicate measures per cat for the 2 instruments and 2 observers were created. For both observers, there was a strong linear association ($r \geq 0.80$), but use of the GM meter generally yielded slightly higher values **(Figure 1)**.

Instrument comparison—The mean of the 2 instruments' measurements versus the difference in the 2 instruments' measurements were plotted for each observer **(Figure 2)**. There were considerable agreement between the 2 instruments for both observers, with the GM yielding slightly higher measurements. The upper and lower limits of agreement were simi-

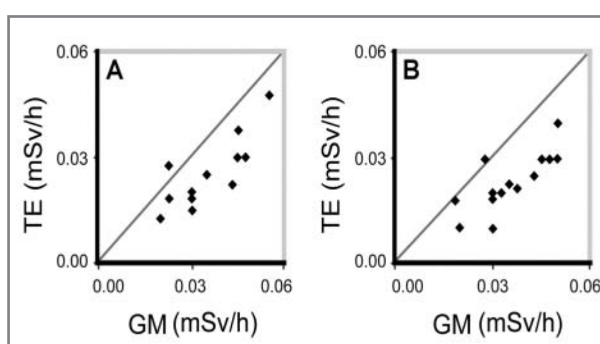

Figure 1—Scatter diagrams with lines of equality for radiation dose rates measured with a GM[a] and TE meter[b] by a board-certified radiologist (observer 1; A) and a veterinary imaging resident (observer 2; B) in 15 cats treated with ^{131}I. Notice that although there is a strong linear association between the 2 instruments for observer 1 ($r = 0.89$) and observer 2 ($r = 0.80$), values yielded by the GM meter were higher. Only 1 data point is depicted for measurements shared by > 1 cat.

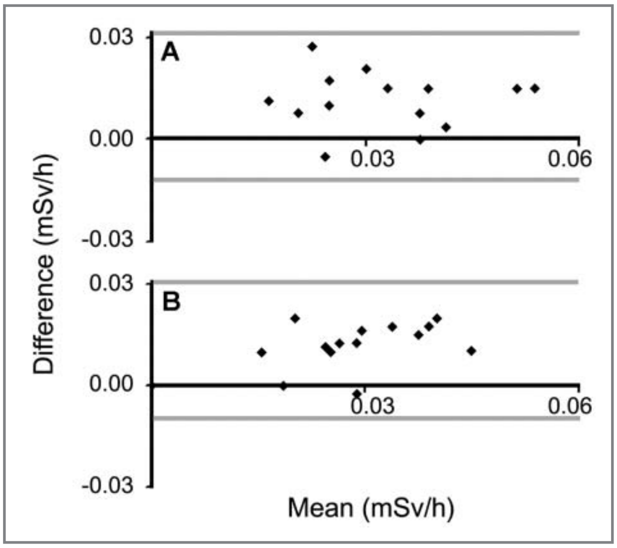

Figure 2—Bias graphs depicting the within-cat means of the 2 meters' measurements plotted against the within-cat differences obtained by the same 2 observers who used the same 2 instruments as in Figure 1. Notice that there was good agreement between the 2 meters for both observers. Gray lines represent limits of agreement. A = Observer 1. B = Observer 2. *See* Figure 1 for key.

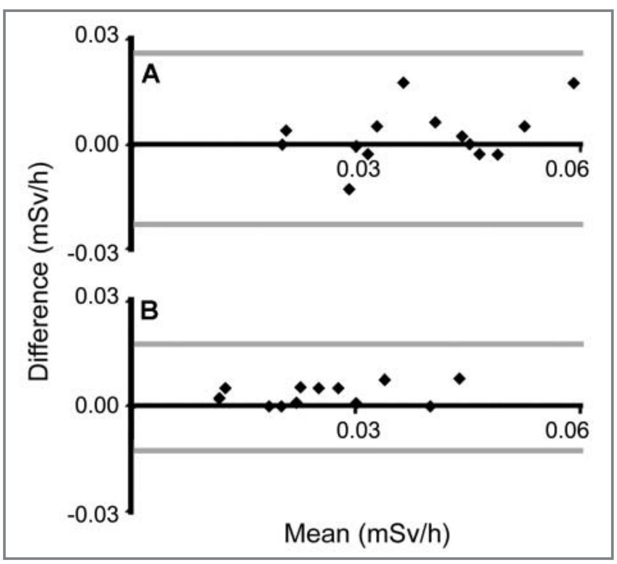

Figure 3—Bias graphs depicting the within-cat means of the 2 observers' measurements plotted against the within-cat differences for each meter. Notice that there was slightly less agreement in values with the GM meter. Gray lines represent limits of agreement. A = GM meter. B = TE meter. *See* Figure 1 for key.

Table 1—Mean differences (mrem/h) in measurements from a GM survey meter and a TE meter when used by a board-certified radiologist (observer 1) and a veterinary imaging resident (observer 2). Limits of agreement were defined as ± 2 times the SD.

Comparison	Mean difference	SD	Limits of agreement
Meter			
Observer 1	1.2	1.1	–0 to 3.4
Observer 2	1.2	1.0	–0.8 to 3.2
Interobserver			
GM	0.3	1.1	–1.9 to 2.5
TE	0.3	0.7	–1.1 to 1.7

Table 2—Mean (SD) differences between duplicate measurements of radiation dose in 15 cats that underwent treatment with ^{131}I. Mean and SD values were used to assess repeatability for the same 2 meters used by the same 2 observers as in Table 1. Data are given as millisieverts per hour.

Observer	GM survey meter	TE meter
1	0.039 (0.003)	0.027 (0.002)
2	0.036 (0.002)	0.024 (0.002)

lar. For the board-certified radiologist, the mean of the differences of the 2 instruments for all cats was 0.012 mSv/h (SD, 0.011 mSv/h) and the limits of agreement were –0.010 mSv/h and 0.034 mSv/h (SE, 0.004 mSv/h). For the resident, the mean of the differences of the 2 instruments for all cats was 0.012 mSv/h (SD, 0.010 mSv/h) and the limits of agreement were –0.008 mSv/h and 0.032 mSv/h (SE, 0.004 mSv/h). The differences did not appear to vary in a systematic way over the range of measurements.

Interobserver agreement—Plots of the mean of the 2 observers' measurements versus the differences in the 2 observers' measurements were created for each instrument **(Figure 3)**. There was substantial agreement between the 2 instruments for both observers. Slightly less agreement, however, was observed with the GM than with the TE meter. For the GM survey meter, the mean of the differences of the observers for all cats was 0.003 mSv/h (SD, 0.011 mSv/h) and the limits of agreement were –0.019 mSv/h and 0.025 mSv/h (SE, 0.004 mSv/h). For the TE, the mean of the differences of the observers for all cats was 0.003 mSv/h (SD, 0.007 mSv/h) and the limits of agreement were –0.011 mSv/h and 0.017 mSv/h (SE, 0.001 mSv/h). The differences did not appear to vary in a systematic way over the range of the measurements.

To make the results of this study generalize more effectively to clinical use, we converted some results to units of millirems per hour because that is the unit displayed on the instruments, and personnel working with radiation in the United States are often more familiar with units of millirems than millisieverts **(Table 1)**.

Repeatability—The SD of the differences between replicate measures per cat was < 0.003 mSv/h for each instrument and observer. At least 95% of differences (ie, all differences in each plot) between replicate measures were within ± 2 SDs of the differences for each instrument and observer **(Table 2)**.

Discussion

Our aims were to assess whether the GM and TE meters are interchangeable and whether use of a GM survey meter is acceptable for clinical use in estimating residual radioactivity in cats treated with ^{131}I. Accordingly, we evaluated the similarity of results yielded by the 2 instruments for the same sample (agreement). The superiority of 1 instrument over the other was not evaluated in terms of validity (ie, how well each method estimated the true measurement) or repeatability (ie, how much variability each method had when the same measurement was repeated on the same sample), although our results indicate that measurements obtained with both instruments were repeatable. Although results obtained with the TE meter are considered to be more valid for the reasons described, on the basis of the agreement between the 2 instruments, it appears that the GM survey meter is a suitable replacement and is acceptable for clinical use.

Overall, there was considerable agreement between instruments and for both observers. Between instruments, the GM meter tended to yield higher values, which was likely attributable to 2 characteristics of that instrument. First, some high-energy beta radiation on the surface of the cat might be detected by the GM meter at the distance used in this study. Second, the instrument overresponds to low-energy gamma radiation.[3] For clinical use, this overestimation is acceptable in terms of evaluating cats for release because it yields a more conservative reading. Between observers, there was slightly less agreement for the GM meter, which was likely a result of the tendency of the needle display on that instrument to have slight motion even at a fixed distance. Thus, observers must determine the most appropriate single reading from a slightly dynamic needle display, which is less desirable for clinical use because a small and exact value may be needed for determining release. In contrast, the TE meter tends to have a fairly stable needle display, making interpretation less variable.

Most cats were measured 3 days after ^{131}I treatment to simulate an early release. As such, the exposure rate was approximately 0.03 mSv/h per cat. We did not observe any bias over the range of measurements observed (approx 0.01 to 0.05 mSv/h), but we did not compare the meters at more extreme ranges of radioactivity. Some of the variability between measurements may not have been related to the accuracy of the instruments but rather to the difficulty in maintaining a uniform distance between the cat and the instrument given patient movement and the process of aligning the probe with the thyroid gland at a distance. It is possible that the high degree of repeatability may have resulted in part from examiners not being blinded to their own measurements.

Measurements obtained with the GM meter may be approximately 1 mrem/h less or 3 mrem/h greater than those obtained with the TE meter. Measurements between examiners for the GM meter varied by approximately ± 2 mrem/h. This observation was close to the variation between observers for the TE meter, which was approximately 1 mrem/h less and 2 mrem/h greater than the other observer. If these ranges are acceptable for an institution's release criteria, the 2 meters should be considered interchangeable and acceptable for clinical use. The overestimation of the dose rate and subsequent more conservative measurements yielded by the GM meter may be an additional factor making that meter suitable for clinical use.

a. Ludlum Model 3 Survey Meter with Model 44-9 Pancake G-M Detector, Ludlum Measurements Inc, Sweetwater, Tex.
b. Bicron Micro Rem tissue-equivalent survey meter with extended detector, Bicron Corp, Newbury, Ohio.

References

1. Sorenson JA, Phelps ME. *Physics in nuclear medicine*. 2nd ed. Orlando, Fla: Grune & Stratton, 1987;68.
2. Knoll GF. *Radiation detection and measurements*. 3rd ed. Hoboken, NJ: John Wiley & Sons, 2000;215.
3. Shleien B, Slaback LA, Birky BK. *The health physics and radiological health handbook*. 3rd ed. Baltimore: The Williams & Wilkins Co, 1998;2120–2436.
4. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–310.