Reference air kerma and kerma-area product as estimators of peak skin dose for fluoroscopically guided interventions

Deukwoo Kwon and Mark P. Little^{a)}

Radiation Epidemiology Branch, National Cancer Institute, Division of Cancer Epidemiology and Genetics, Bethesda, Maryland 20852-7238

Donald L. Miller^{b)}

Radiation Epidemiology Branch, National Cancer Institute, Division of Cancer Epidemiology and Genetics, Bethesda, Maryland 20852-7238 and Department of Radiology and Radiological Sciences, F. Edward Hébert School of Medicine, Uniformed Services University, Bethesda, Maryland 20814

(Received 14 February 2011; revised 9 April 2011; accepted for publication 18 April 2011; published 28 June 2011)

Purpose: To determine more accurate regression formulas for estimating peak skin dose (PSD) from reference air kerma (RAK) or kerma-area product (KAP).

Methods: After grouping of the data from 21 procedures into 13 clinically similar groups, assessments were made of optimal clustering using the Bayesian information criterion to obtain the optimal linear regressions of (log-transformed) PSD vs RAK, PSD vs KAP, and PSD vs RAK and KAP.

Results: Three clusters of clinical groups were optimal in regression of PSD vs RAK, seven clusters of clinical groups were optimal in regression of PSD vs KAP, and six clusters of clinical groups were optimal in regression of PSD vs RAK and KAP. Prediction of PSD using both RAK and KAP is significantly better than prediction of PSD with either RAK or KAP alone. The regression of PSD vs RAK provided better predictions of PSD than the regression of PSD vs KAP. The partial-pooling (clustered) method yields smaller mean squared errors compared with the complete-pooling method.

Conclusion: PSD distributions for interventional radiology procedures are log-normal. Estimates of PSD derived from RAK and KAP jointly are most accurate, followed closely by estimates derived from RAK alone. Estimates of PSD derived from KAP alone are the least accurate. Using a stochastic search approach, it is possible to cluster together certain dissimilar types of procedures to minimize the total error sum of squares. © 2011 American Association of Physicists in Medicine. [DOI: 10.1118/1.3590358]

Key words: peak skin dose, reference air kerma, kerma-area product

I. INTRODUCTION

The International Commission on Radiological Protection (ICRP), the National Council on Radiation Protection and Measurements (NCRP), and professional societies have emphasized the importance of estimating skin dose during and after fluoroscopically guided interventions (FGI), as a means of predicting the likelihood and severity of deterministic skin effects and as a means of indicating which patients require follow-up for detection of possible skin injuries.¹⁻⁴ The Joint Commission considers a skin dose that exceeds 15 Gy to be a sentinel event.⁵ It is thus important to be able to estimate peak skin dose (PSD)—the greatest absorbed skin dose at any point on the patient's skin—during and after a FGI.

PSD can be measured using film, thermoluminescent or optically stimulated luminescence dosimeters placed on the patient, or similar means.³ However, these require prior planning and must be placed accurately at the site of PSD, which is difficult to do.⁶ Since mid-2006, the U.S. Food and Drug Administration has required that all fluoroscopes sold in the U.S. be capable of displaying the total air kerma at the interventional reference point, a quantity termed reference

air kerma (RAK) by the International Electrotechnical Commission.⁷ Most fluoroscopes can also display kerma-area product (KAP).

It is useful to be able to estimate PSD, and this can be done using either RAK or KAP. In the U.S., there is a preference for RAK.^{2,4,8,9} In the European Union, KAP is the most commonly used dose metric, as it has been recommended by the International Commission on Radiation Units (ICRU) to measure patient doses in IR.^{10,11} An earlier study (the RAD-IR study) suggested that RAK was superior to KAP as a predictor of PSD but looked only at a pooled sample containing cases of all FGI in the database.⁸

In this study, we performed a secondary analysis of the data from the RAD-IR study to determine more precise regression formulas for estimating PSD from RAK and KAP. In particular we sought to determine whether RAK, KAP, or a combination of the two is the best predictor of PSD. We also sought to determine the optimal grouping of FGI procedures for the purpose of estimating PSD from RAK and KAP. We estimate the likely errors in PSD at specific values of RAK and DAP and compare the predictive values of RAK and KAP.

II. MATERIALS AND METHODS

II.A. RAD-IR study

The RAD-IR study documented radiation doses from certain interventional radiology procedures. During 1999–2002, seven academic medical centers in the U.S. collected radiation dose data for each of 21 different interventional radiology procedures, as well as several procedure subsets defined *a priori*. The results were reported in three parts. Part I provided overall dose data, identified procedures associated with higher radiation doses, analyzed the effect of operator training level on dose, and provided recommendations for recording overall dose.¹² Part II provided skin dose data for a subset of cases.⁸ Part III presented the physics data that validated the radiation dose data.¹³

The RAD-IR study protocol was reviewed and approved by the Institutional Review Board (IRB) at each participating institution. The data used here have been analyzed previously to determine reference levels.^{2,4,8,9} The present secondary analysis was also reviewed and approved by our IRB and was HIPAA compliant.

II.B. Dose metrics

Four radiation dose metrics (RAK in Gy, KAP in $\text{Gy} \cdot \text{cm}^2$, fluoroscopy time in decimal minutes, and number of radiographic images) were collected for all 2142 cases in the RAD-IR study.

KAP, also known as dose-area product or DAP, is the integral of air kerma (the energy extracted from an x-ray beam per unit mass of air in a small irradiated air volume; for diagnostic x-rays, the dose delivered to that volume of air) across the entire x-ray beam emitted from the x-ray tube. It is a surrogate measure of the amount of energy delivered to the patient.¹²

RAK (also known as reference dose, cumulative dose, or cumulative dose at a reference point) is the air kerma accumulated at a specific point in space (the patient entrance reference point) relative to the fluoroscopic gantry.¹² For C-arm fluoroscopic systems, the patient entrance reference point is a point along the central ray of the x-ray beam, 15 cm back from the isocenter toward the focal spot.⁷

Some fluoroscopic units in the RAD-IR study were equipped with an additional dose measurement system (CAREGRAPH; Siemens Medical Systems, Malvern, PA). Data on measured KAP, collimator field size, tabletop position, and C-arm angulation are transferred to the CAREGRAPH software in real time, along with additional data inputs of patient height, weight, and location on the procedure table. The software uses a mathematical model of the skin as an oblate cylinder with an average circumference of 100 cm, modified for the patient's height and weight to model the surface shape. This is used, along with the patient's location on the procedure table, to calculate x-ray entrance field size, location, and air kerma incident on the skin.¹⁴ These calculations are performed every 500 ms for areas of skin 0.5×0.5 cm.^{14–16} The maximum air kerma level without backscatter (i.e., maximum entrance skin exposure) and spatial distribution of

the dose on the skin (skin dose map) are displayed on a graphical representation of the skin surface.¹⁷ For this analysis, skin dose values recorded by the CAREGRAPH software were corrected to account for systematic errors in the software calculations and to incorporate the effect of back scatter, using conversion factors established previously.¹⁸

II.C. Procedures

Skin dose data were available for 800 of the 2142 cases in the RAD-IR data set. To minimize the effects of large variations in subject body habitus on the results, we excluded subjects less than 18 years old. There were 53 such cases. Of these, 51/53 (96%) were neuroembolization procedures. The remaining two pediatric cases were one nephrostomy performed for obstruction and one pelvic vein embolization performed for varicocele. We eliminated 14 procedures of the 35 procedure types in the RAD-IR dataset, because they had insufficient numbers (less than four) of patients; a total of 39 patients were excluded for this reason.

After excluding pediatric patients and taking account of the limited number of data points available for certain procedures in the RAD-IR study, a dataset with 708 records and 21 procedures remained (Table I). One statistical outlier case (embolization for arteriovenous malformation of the brain) was removed from the analysis, leaving an analysis dataset of 707 cases.

Clinically similar procedures (e.g., chemoembolization of the liver and embolization of organs in the abdomen and pelvis; embolization of all lesions in the head; embolization of all lesions in the spine) were grouped together for this analysis to achieve larger sample sizes (Table II). Descriptions of these procedures are provided elsewhere.¹²

II.D. Statistical methods

Our interest is in estimating the linear relationship between PSD (in Gy) and two other variables, RAK (Gy), and KAP (in $\text{Gy} \cdot \text{cm}^2$) for 21 interventional radiology procedures. All analysis is restricted to the 707 observations with PSD measurements (from a total of 2142 sets of observations). This is a reduced subset because of the limited availability of equipment for making skin dose measurements and the exclusion of 93 observations as noted above. For each individual i ($1 \leq i \leq 707$) there are measurements of PSD (PSD_i), RAK (RAK_i), and KAP (KAP_i).

The sample size for each procedure ranges from 4 to 142. Although not strictly necessary for the regression, we evaluated the normality of the dependent variable, PSD, using the Shapiro–Wilk test.¹⁹ Before testing for normality and to increase the number of data points in each group, we grouped the 21 procedures into 13 groups based on clinical similarities among the procedures. These procedures and groups are listed in Table I. The Shapiro–Wilk test is used to test for departures from normality (the assumed null hypothesis), so that conventionally a p -value below 5% ($p \leq 0.05$) is grounds for rejection of the assumption of normality.

TABLE I. List of procedures and groups.

Procedure	Sample size	Clinical group	Cluster number in analysis of ln[PSD] vs ln[RAK]	Cluster number in analysis of ln[PSD] vs ln[KAP]	Cluster number in analysis of ln[PSD] vs ln[RAK] and ln[KAP]
Transjugular intrahepatic portosystemic shunt (TIPS) creation	19	1	1	1	1
Biliary drainage	23	2	2	2	1
Nephrostomy for obstruction	22	3	3	1	2
Nephrostomy for stone access	4	3	3	1	2
Pulmonary angiography	13	4	3	3	3
Pulmonary angiography with inferior vena cava filter placement	73	5	2	4	4
Renal or visceral angioplasty without stent	18	6	1	2	1
Renal or visceral angioplasty with stent	21	6	1	2	1
Iliac angioplasty without stent	9	7	2	5	2
Iliac angioplasty with stent	22	7	2	5	2
Bronchial artery embolization	6	8	1	2	5
Hepatic chemoembolization	26	9	1	1	4
Fibroid embolization	14	9	1	1	4
Other tumor embolization	23	9	1	1	4
Pelvic artery embolization trauma/tumor	6	9	1	1	4
Gastrointestinal hemorrhage localization and treatment	25	10	1	4	4
Embolization in the head for arteriovenous malformation	125	11	2	6	2
Embolization in the head for aneurysm	142	11	2	6	2
Embolization in the head for tumor	39	11	2	6	2
Vertebroplasty	61	12	3	6	2
Embolization in the spine for AVM or tumor	16	13	1	7	6

We used mean square error (MSE) to assess the overall error of each procedure and in total. We explored a variety of further clusters of groups, using a penalized least squares approach to amalgamate clinical groups and so protect against overfitting. We consider three models: (a) a no-pooling model, in which a separate regression was fitted to each of the $L = 13$ groups of clinically related procedures; (b) complete-pooling model, in which a single regression was used for the combined data; and (c) a partial-pooling or

cluster model, in which the $L = 13$ groups were collapsed into a smaller number of clusters. Cases (a) and (b) are of course special cases of the partial-pooling model (c). We used a stochastic search approach to find the optimum combination of clusters. This algorithm minimizes the Bayesian information criterion (BIC), which adds a penalty term (depending on the number of parameters being estimated) to the residual sum of squares (RSS).²⁰ For purposes of summarizing goodness of fit we also employed Akaike's

TABLE II. Shapiro–Wilk test (Ref. 19) of normality of peak skin dose (PSD).

Procedure	Clinical group number	Sample size	p -value (original scale)	p -value (ln[PSD])
Transjugular intrahepatic portosystemic shunt creation	1	19	0.163	0.290
Biliary drainage	2	23	<0.001	0.771
Nephrostomy for obstruction/nephrostomy for stone access	3	26	<0.001	0.641
Pulmonary angiography	4	13	0.019	0.069
Pulmonary angiography with inferior vena cava filter placement	5	73	<0.001	0.201
Renal or visceral angioplasty with or without stent	6	39	0.003	0.461
Iliac angioplasty with or without stent	7	31	<0.001	0.229
Bronchial artery embolization	8	6	0.042	0.515
Hepatic chemoembolization/fibroid embolization/other tumor embolization/pelvic artery embolization trauma/tumor	9	69	<0.001	0.115
Gastrointestinal hemorrhage localization and treatment	10	25	0.235	0.158
Embolization in the head for arteriovenous malformation/embolization in the head for aneurysm/embolization in the head for tumor	11	306	<0.001	0.375
Vertebroplasty	12	61	<0.001	0.397
Embolization in the spine for AVM or tumor	13	16	0.932	0.090
Total	—	—	<0.001	<0.001

TABLE III. Analysis of variance of models of $\ln[\text{PSD (Gy)}]$ vs $\ln[\text{RAK (Gy)}]$ and $\ln[\text{KAP (Gy cm}^2\text{)}]$.

Model	Residual sum of squares	Degrees of freedom	<i>F</i> statistic	<i>p</i> -value	Akaike information criterion	Bayesian information criterion
$\ln[\text{PSD}]$ vs $\ln[\text{RAK}]$	91.61	705	155.68	$<10^{-10a}$	567.6	581.3
$\ln[\text{PSD}]$ vs $\ln[\text{KAP}]$	157.42	705	773.26	$<10^{-10a}$	950.4	964.1
$\ln[\text{PSD}]$ vs $\ln[\text{RAK}]$ and $\ln[\text{KAP}]$	75.02	704	—	—	428.4	446.6

^aTest of improvement of fit compared with model of $\ln[\text{PSD}]$ vs $\ln[\text{RAK}]$ and $\ln[\text{KAP}]$, computed using an *F*-test.

information criterion (AIC), which adds a similar penalty term to the RSS.²¹ Further details are given in Appendix A.

III. RESULTS

III.A. Test for normality

We examined PSD for all 13 groups. On the original (nontransformed) scale only three procedures were consistent with normality ($p > 0.05$) (transjugular intrahepatic portosystemic shunt creation, gastrointestinal hemorrhage localization and treatment, and embolization in the spine for AVM or tumor). After logarithmic transformation, all 13 groups were consistent with a normal distribution of PSD ($p > 0.05$) (Table II). For example, the Shapiro–Wilk test *p*-value for the original (nontransformed) measurements for biliary drainage procedures was $p < 0.001$, but after logarithmic transformation this became $p = 0.771$ (Table II). Although PSDs for individual groups were consistent with a lognormal distribution, the combined distribution of the log-transformed PSDs over all 13 groups was not consistent with normality due to heterogeneity between the groups (result not shown).

III.B. Analysis of variance

Table III shows the result of analysis of variance. It is clear from this that a model of PSD with both KAP and RAK as explanatory variables provides a statistically signifi-

cantly better fit ($p < 10^{-10}$) than a model with either of these explanatory variables separately. Also, the model $\ln[\text{PSD}]$ vs $\ln[\text{RAK}]$ yields a much better fit than does the $\ln[\text{PSD}]$ vs $\ln[\text{KAP}]$, as implied by the substantially lower BIC (581.3 vs 964.1) and AIC (567.6 vs 950.4) values for this regression (Table III).

III.C. Estimation of optimal clustering

We found three clusters of clinical groups when we used $\ln[\text{PSD}]$ vs $\ln[\text{RAK}]$ (Table IV, Fig. 1) and seven clusters of clinical groups with $\ln[\text{PSD}]$ vs $\ln[\text{KAP}]$ (Table V, Fig. 2). We also observed six clusters of clinical groups in joint modeling of $\ln[\text{PSD}]$ vs $\ln[\text{RAK}]$ and $\ln[\text{KAP}]$. In Tables IV–VI we show the regression coefficients and the MSEs for the three pooling methods. These demonstrate, as also do Figs. 1 and 2, that RAK is more predictive of PSD than is KAP, although as demonstrated by Tables III and VII the best prediction is given by a model of $\ln[\text{PSD}]$ vs $\ln[\text{RAK}]$ and $\ln[\text{KAP}]$. The model of $\ln[\text{PSD}]$ vs $\ln[\text{RAK}]$ and $\ln[\text{KAP}]$ has the lowest BIC (91.8) (Table VII); although it is not much smaller than that for the model of $\ln[\text{PSD}]$ vs $\ln[\text{RAK}]$ (BIC 116.0), the fit of either model is appreciably better than that of $\ln[\text{PSD}]$ vs $\ln[\text{KAP}]$ (BIC 225.4) (Table VII). The AIC statistics suggest the same ranking of the three models (Table VII). [Note: by definition, the optimal model is chosen as that which minimizes BIC. One trivial “pooling” model is that with no-pooling, so this must have higher BIC

TABLE IV. Optimal clustering of $\ln[\text{PSD (Gy)}]$ vs $\ln[\text{RAK (Gy)}]$ (three clusters).

Clinical group	No-pooling			Partial-pooling			Complete-pooling		
	Intercept $[\ln(\text{Gy})]$	Slope	MSE ^a	Intercept $[\ln(\text{Gy})]$	Slope	MSE ^a	Intercept $[\ln(\text{Gy})]$	Slope	MSE ^a
1	−0.081	0.874	0.030	−0.135	0.873	0.033	−0.407	0.861	0.156
6	−0.145	0.918	0.074			0.074			0.158
8	−0.047	0.917	0.052			0.063			0.240
9	−0.120	0.906	0.066			0.068			0.162
10	−0.248	0.863	0.116			0.132			0.144
13	0.170	0.665	0.194			0.207			0.278
2	−0.406	0.846	0.095	−0.457	0.834	0.101			0.096
5	−0.332	0.879	0.079			0.087			0.081
7	−0.421	0.999	0.128			0.141			0.135
11	−0.627	0.954	0.124			0.141			0.151
3	−0.595	0.862	0.079	−0.617	0.838	0.080			0.118
4	−0.777	0.731	0.115			0.121			0.176
12	−0.608	0.816	0.146			0.146			0.188

^aMean square error.

FIG. 1. Partial clustering analysis of $\ln[\text{peak skin dose (PSD)}]$ vs $\ln[\text{reference air kerma (RAK)}]$.

than the optimal pooling-model in each case. The mean square error must be lower for the no-pooling model (because the model is an extension of the pooling model), but this is more than offset by the additional complexity penalty.] When we used clusters identified using the partial-pooling approach described above, MSEs were generally only slightly larger than those for the no-pooling approach but much smaller than those for complete-pooling.

For RAK, for KAP, and for RAK and KAP together, Fig. 3 shows the percentage error in model predictions for estimates derived using the pooled or clustered models as compared to clinical groups. The clustered (partial-pooling) model generally provides a lower error than the pooled model, regardless of the dose metric used to estimate PSD.

When the median values of the dose metrics for each clinical group were used, percentage errors in PSD ranged from:

- 6 to 12% for the clustered model and –37 to 28% for the pooled model of $\ln[\text{PSD}]$ vs $\ln[\text{RAK}]$ and $\ln[\text{KAP}]$;
- from –4 to 8% for the clustered model and –24 to 29% for the pooled model of $\ln[\text{PSD}]$ vs $\ln[\text{RAK}]$; and
- from –6 to 10% for the clustered model and –101 to 36% for the pooled model of $\ln[\text{PSD}]$ vs $\ln[\text{KAP}]$ (Fig. 3).

IV. EXAMPLE CALCULATIONS

Suppose that an RAK of 10 Gy and a KAP of 500 Gy cm^2 were recorded for a transjugular intrahepatic portosystemic shunt creation. This procedure is in clinical group one (Table I). From Table IV we estimate the PSD using the model of $\ln[\text{PSD}]$ vs $\ln[\text{RAK}]$ without pooling as $\exp[-0.081 + 0.874 \times \ln[10]] = 6.90$ Gy; for the partial-pooling and complete-pooling models the analogous figures are $\exp[-0.135 + 0.873 \times \ln[10]] = 6.52$ Gy and $\exp[-0.407 + 0.861 \times \ln[10]] = 4.83$ Gy. From the coefficients in Table V we see that the model of $\ln[\text{PSD}]$ vs $\ln[\text{KAP}]$ predicts a PSD using the no-pooling, partial-pooling, and complete-pooling models of $\exp[-5.028 + 0.957 \times \ln[500]] = 2.51$ Gy, $\exp[-4.627 + 0.889 \times \ln[500]] = 2.45$ Gy, and $\exp[-4.936 + 0.959 \times \ln[500]] = 2.78$ Gy, respectively. From the coefficients in Table VI, we see that the model of $\ln[\text{PSD}]$ vs $\ln[\text{RAK}]$ and $\ln[\text{KAP}]$ predicts a PSD using the no-pooling, partial-pooling, and complete-pooling models of $\exp[-1.767 + 0.587 \times \ln[10] + 0.325 \times \ln[500]] = 4.97$ Gy, $\exp[-2.088 + 0.540 \times \ln[10] + 0.395 \times \ln[500]] = 5.00$ Gy, and $\exp[-1.938 + 0.614 \times \ln[10] + 0.320 \times \ln[500]] = 4.33$ Gy, respectively. The partial-pooling model of $\ln[\text{PSD}]$ vs $\ln[\text{RAK}]$ and $\ln[\text{KAP}]$ is (by the results of Table VII) optimal (in the sense of minimizing BIC) so that the value of 5.00 Gy is the “best” estimate in this case.

TABLE V. Optimal clustering of $\ln[\text{PSD (Gy)}]$ vs $\ln[\text{KAP (Gy cm}^2\text{)}]$ (seven clusters).

Clinical group	No-pooling			Partial-pooling			Complete-pooling		
	Intercept $[\ln(\text{Gy})]$	Slope	MSE ^a	Intercept $[\ln(\text{Gy})]$	Slope	MSE ^a	Intercept $[\ln(\text{Gy})]$	Slope	MSE ^a
1	–5.028	0.957	0.038	–4.627	0.889	0.040	–4.936	0.959	0.050
3	–4.805	0.963	0.156			0.164			0.178
9	–4.872	0.929	0.244			0.245			0.254
2	–4.722	0.983	0.068	–4.581	0.932	0.075			0.176
6	–4.649	0.940	0.115			0.116			0.153
8	–4.841	0.974	0.147			0.152			0.190
4	–4.642	0.720	0.145	–4.642	0.720	0.145			0.736
5	–5.595	0.983	0.056	–6.076	1.121	0.067			0.387
10	–5.740	1.081	0.176			0.189			0.191
7	–4.883	0.873	0.127	–4.883	0.873	0.127			0.295
11	–3.944	0.796	0.164	–3.706	0.755	0.170			0.169
12	–3.549	0.723	0.193			0.208			0.435
13	–1.893	0.520	0.168	–1.893	0.520	0.168			0.435

^aMean square error.

FIG. 2. Partial clustering analysis of $\ln[\text{peak skin dose (PSD)}]$ vs $\ln[\text{kerma-area product (KAP)}]$.

V. DISCUSSION

It is generally recognized that FGI procedures are characterized by a log-normal dose distribution, but skin dose distributions for interventional radiology procedures have not been specifically evaluated. All of the RAD-IR procedures evaluated in this study demonstrated a log-normal distribution. This has implications for using reference levels. Cases with high skin doses are located in the high-dose “tail” portion of the dose distribution.¹⁷ The high dose tail cannot be fully evaluated by methods such as diagnostic reference levels that utilize measures of central tendency. Alternative methods are required.²

Estimates of skin dose are important for patients undergoing fluoroscopically guided interventions, because these patients may receive radiation doses high enough to cause deterministic effects.¹ When deterministic effects are of concern, the most desirable dose estimate is PSD.² ICRP Publication 85 recommends that “records of exposure should be

kept if the estimated maximum cumulative dose to skin is 3 Gy or above (1 Gy or above for procedures likely to be repeated),” and that “all patients with estimated skin doses of 3 Gy or above should be followed up 10 to 14 days after exposure.”¹ Clearly, this is not possible if PSD cannot be estimated with reasonable accuracy.

Because PSD depends on a large number of factors, including gantry angle and position, field size and shape, and patient size and position, it cannot be calculated directly from overall measures of dose such as RAK and KAP.¹⁷ At present, PSD is rarely measured in routine clinical practice, because current methods are cumbersome, expensive or both.⁸ Automated methods are under development, but are not generally available as of 2011.^{2,22}

Since PSD is not usually measured directly, overall measures of dose are typically used as surrogates. The American College of Radiology recommends recording RAK and KAP for all fluoroscopy procedures.²³ The NCRP, the Society of Interventional Radiology (SIR) and the Cardiovascular and Interventional Radiological Society of Europe (CIRSE) recommend using RAK as the best clinical approximation of skin dose.^{2,4} KAP is considered less useful, because KAP is the product of x-ray fluence and field size. A small radiation dose delivered to a large area can yield the same KAP as a large radiation dose delivered to a small area.^{24,25} Other studies, principally performed in Europe, have used KAP as a surrogate for PSD or to estimate PSD, either because RAK data were not available or because the ICRU has recommended using KAP to measure patient dose for interventional radiology procedures.^{10,26–28}

While the use of both RAK and KAP together yields the best estimates of PSD, the use of RAK alone is nearly as good and is appreciably better than using KAP alone. If RAK data are available, we recommend against using KAP alone.

The NCRP and SIR-CIRSE recommendations provide formulas for estimation of PSD from RAK and from KAP, but the formulas are derived from the pooled sample of 800 cases in the RAD-IR database with PSD data.^{2,4} In the present study, we have shown that combining procedures into clinically similar groups and then further combining these groups into clusters using a stochastic search approach yields better estimates of PSD than using pooled data from all procedures.

FIG. 3. Percentage error in model predictions at the median dose per procedure (for clustered and pooled models).

TABLE VI. Optimal clustering of ln[PSD (Gy)] vs ln[RAK (Gy)] and ln[KAP (Gy cm²)] (six clusters).

Clinical group	No-pooling				Partial-pooling				Complete-pooling			
	Intercept (ln(Gy))	Slope (RAK)	Slope (KAP)	MSE ^a	Intercept (ln(Gy))	Slope (RAK)	Slope (KAP)	MSE ^a	Intercept (ln(Gy))	Slope (RAK)	Slope (KAP)	MSE ^a
1	−1.767	0.587	0.325	0.027	−2.088	0.540	0.395	0.031	−1.938	0.614	0.320	0.062
2	−3.308	0.290	0.663	0.058				0.067				0.093
6	−1.335	0.707	0.246	0.069				0.074				0.137
3	−1.591	0.679	0.237	0.072	−1.612	0.668	0.238	0.077				0.073
7	−2.845	0.532	0.473	0.102				0.110				0.121
11	−1.891	0.683	0.283	0.117				0.124				0.131
12	−1.284	0.655	0.167	0.142				0.156				0.155
4	−1.887	0.555	0.217	0.110	−1.887	0.555	0.217	0.110				0.232
5	−3.922	0.298	0.679	0.049	−1.978	0.669	0.345	0.061				0.062
9	−1.392	0.750	0.243	0.052				0.058				0.090
10	−2.588	0.625	0.442	0.095				0.104				0.102
8	3.175	1.427	−0.658	0.036	3.175	1.427	−0.658	0.036				0.220
13	−2.361	−0.198	0.651	0.166	−2.361	−0.198	0.651	0.166				0.279

^aMean square error.

We provide more accurate regressions for estimation of PSD from specific interventional radiology procedures.

In deriving the optimal clusters of clinical groups, we minimized the Bayesian information criterion, which is the total error sum of squares with a penalty term, added to guard against overfitting.²⁰ Thirteen separate regression lines would minimize the total error sum of squares, but this would result in overfitting, so that the resulting regression would be poorly predictive when applied to new data. In aggregating studies there is inevitably a tradeoff between bias and variance—however, this has little to do with the rationale for using BIC to determine the optimal clusters.

Given a more extensive patient dose database one could refine the models used here, and in particular could obviate the need to cluster clinical groups. The method described here is most useful when relatively small data sets are available. In the future, regional or national datasets should become available, and the statistical methods described here will not be as necessary, except for groups of procedures that are not commonly performed. Ideally, methods for estimation of PSD from Radiation Dose Structured Reports (RDSRs) will become available and can be used to provide PSD estimates in real time based on information on radiation field size and location in individual cases.²² The regressions provided here would still be useful in several settings: for

retrospective dose reconstruction, in those facilities with older equipment that cannot be upgraded to generate RDSRs, and for comparison to older studies in the literature.

ACKNOWLEDGMENTS

The Chief, Navy Bureau of Medicine and Surgery, Washington, D.C., Clinical Investigation Program sponsored this study, Grant No. NNMCMC.2008.0044. This research was partially supported by the Intramural Research Program of the National Institutes of Health, the National Cancer Institute, Division of Cancer Epidemiology and Genetics.

The views expressed in this paper are those of the authors and do not necessarily reflect the official policy or position of the U.S. Food and Drug Administration, the National Cancer Institute, the Department of Health and Human Services or the U.S. Government. The authors are grateful for the detailed and helpful comments of the referee. The authors have no conflicts of interest.

APPENDIX: STATISTICAL AND MODELING DETAILS

Some of the heterogeneity in the PSDs can be explained by relationships between PSDs and other variables (RAK, KAP). Thus, we evaluated approaches for predicting PSD from either RAK or KAP. In each case, we used linear regression with ln[PSD] as the dependent variable and either ln[RAK] or ln[KAP] as the independent variable. For example, the model for PSD vs RAK and KAP assumes that for patient j in procedure group g :

$$\ln[\text{PSD}_{g,i}] = \beta_{0,k} + \beta_{1,k} \ln[\text{RAK}_{g,i}] + \beta_{2,k} \ln[\text{KAP}_{g,i}] + \varepsilon_{g,i} \quad (\text{A1})$$

where the error term, $\varepsilon_{g,i}$, is assumed to be normal with mean 0 and variance σ_g^2 i.e., $\varepsilon_{g,i} \sim N(0, \sigma_g^2)$.

If we assume a partitioning of $M_1 \cup \dots \cup M_K = \{1, 2, \dots, L\}$ into K disjoint subsets, the deviance ($-2 \times \ln[\text{likelihood}]$) corresponding to model (A1) is therefore

TABLE VII. Bayesian information criterion (BIC), Akaike information criterion (AIC) and Residual Sum of Squares (RSS) of optimal clustering models (Tables IV–VI).

Partial-pooling model	Akaike information criterion	Bayesian information criterion	Residual sum of squares	R ²
ln[PSD] vs ln[RAK] (three clusters)	111.0	116.0	81.19	0.91
ln[PSD] vs ln[KAP] (seven clusters)	213.7	225.4	108.25	0.88
ln[PSD] vs ln[RAK] and ln[KAP] (six clusters)	79.2	91.8	69.82	0.93

$$-2 \ln L = \sum_{k=1}^K \sum_{g=1}^L 1_{g \in M_k} \left\{ \frac{\sum_{i=1}^{n_g} (\ln[\text{PSD}_{g,i}] - \beta_{0,k} - \beta_{1,k} \ln[\text{RAK}_{g,i}] - \beta_{2,k} \ln[\text{KAP}_{g,i}])^2}{\sigma^2} \right. \\ \left. + n_g (\ln[\sigma^2] + \ln[2\pi]) \right\}, \quad (\text{A2})$$

where n_g is the number of patients in procedure group g . This has minimum when

$$-2 \ln \hat{L} = C + N \ln \left[\frac{1}{N} \sum_{k=1}^K \sum_{g=1}^L 1_{g \in M_k} \sum_{i=1}^{n_g} (\ln[\text{PSD}_{g,i}] - \hat{\beta}_{0,k} - \hat{\beta}_{1,k} \ln[\text{RAK}_{g,i}] - \hat{\beta}_{2,k} \ln[\text{KAP}_{g,i}])^2 \right] \quad (\text{A3})$$

for some constant C . We consider three models: (a) a no-pooling model, in which a separate regression was fitted to each of the $L = 13$ groups of clinically related procedures; (b) a complete-pooling model, in which a single regression was used for the combined data; and (c) a partial-pooling model, in which the $L = 13$ groups were collapsed into a smaller number of clusters. Cases (a) and (b) are of course special cases of the partial-pooling model (c). We use a stochastic search approach to find the optimum combination of clusters. This algorithm minimizes the BIC,²⁰ which for the analysis of $\ln[\text{PSD}]$ vs $\ln[\text{RAK}]$ and $\ln[\text{KAP}]$ is given by

$$\text{BIC} = -2 \ln \hat{L} + 3K \ln[N] \\ = N \ln \left[\frac{1}{N} \sum_{k=1}^K \sum_{g=1}^L 1_{g \in M_k} \sum_{i=1}^{n_g} \left(\ln[\text{PSD}_{g,i}] - \hat{\beta}_{0,k} - \hat{\beta}_{1,k} \ln[\text{RAK}_{g,i}] - \hat{\beta}_{2,k} \ln[\text{KAP}_{g,i}] \right)^2 \right] + 3K \ln[N]. \quad (\text{A4})$$

A similar expression is minimized in the analysis of $\ln[\text{PSD}]$ vs $\ln[\text{KAP}]$ or of $\ln[\text{PSD}]$ vs $\ln[\text{RAK}]$, with in each case the number of parameters, $2K$, replacing $3K$ in the last term. The minimization is taken over all possible partitioning subsets

$$M_1 \cup \dots \cup M_K = \{1, 2, \dots, L\}$$

where $N = n_1 + \dots + n_L$ is the total number of observations. We used mean square error (MSE) to assess the overall error of each procedure and in total. For purposes of summarizing goodness of fit we also employ Akaike's information criterion (AIC)²¹ given by

$$\text{AIC} = -2 \ln[\hat{L}] + 2 \times 3K \\ = N \ln \left[\frac{1}{N} \sum_{k=1}^K \sum_{g=1}^L 1_{g \in M_k} \sum_{i=1}^{n_g} \left(\ln[\text{PSD}_{g,i}] - \hat{\beta}_{0,k} - \hat{\beta}_{1,k} \ln[\text{RAK}_{g,i}] - \hat{\beta}_{2,k} \ln[\text{KAP}_{g,i}] \right)^2 \right] + 2 \times 3K. \quad (\text{A5})$$

Again, there is a similar expression in the analysis of $\ln[\text{PSD}]$ vs $\ln[\text{KAP}]$ or of $\ln[\text{PSD}]$ vs $\ln[\text{RAK}]$, with in each case the number of parameters, $2K$, replacing $3K$ in the last term.

^{a)}Author to whom correspondence should be addressed. Electronic mail: mark.little@nih.gov; Telephone: +1 301 875 3413; Fax: +1 301 402 0207.

^{b)}Current address: Food and Drug Administration, Center for Devices and Radiological Health, Silver Spring, Maryland 20993.

¹International Commission on Radiological Protection, "Avoidance of radiation injuries from medical interventional procedures. ICRP Publication 85," Ann. ICRP **30**, 7–67 (2000).

²NCRP, "Radiation dose management for fluoroscopically guided interventional medical procedures," Report No. 168. (National Council on Radiation Protection and Measurements, Bethesda, MD, 2010).

³D. L. Miller, S. Balter, L. K. Wagner, J. F. Cardella, T. W. I. Clark, C. D. Neithamer Jr., M. S. Schwartzberg, T. L. Swan, R. B. Towbin, K. S. Rholl, and D. Sacks, "Quality improvement guidelines for recording patient radiation dose in the medical record," *J. Vasc. Interv. Radiol.* **15**, 423–429 (2004).

⁴M. S. Stecker, S. Balter, R. B. Towbin, D. L. Miller, E. Vañó, G. Bartal, J. F. Angle, C. P. Chao, A. M. Cohen, R. G. Dixon, K. Gross, G. G. Hartnell, B. Schueler, J. D. Statler, T. de Baère and J. F. Cardella, "Guidelines for patient radiation dose management," *J. Vasc. Interv. Radiol.* **20**, S263–S273 (2009).

⁵The Joint Commission, *Radiation Overdose as a Reviewable Sentinel Event*, Update March 7, 2006. Available at: http://www.jointcommission.org/assets/1/18/Radiation_Overdose.pdf. Accessed May 18, 2011.

⁶R. Nicholson, F. Tuffee, and M. C. Uthappa, "Skin sparing in interventional radiology: The effect of copper filtration," *Br. J. Radiol.* **73**, 36–42 (2000).

⁷International Electrotechnical Commission, Report No. 60601–2–43, 2010.

- ⁸D. L. Miller, S. Balter, P. E. Cole, H. T. Lu, A. Berenstein, R. Albert, B. A. Schueler, J. D. Georgia, P. T. Noonan, E. Russell, T. Malisch, R. L. Vogelzang, M. Geisinger, J. St George, J. F. Cardella, G. Miller, and J. Anderson, "Radiation doses in interventional radiology procedures: The RAD-IR study Part II. Skin dose," *J. Vasc. Interv. Radiol.* **14**, 977–990 (2003).
- ⁹D. L. Miller, D. Kwon, and G. H. Bonavia, "Reference levels for patient radiation doses in interventional radiology: Proposed initial values for U.S. practice," *Radiology* **253**, 753–764 (2009).
- ¹⁰E. Vano, R. Sanchez, J. M. Fernandez, J. J. Gallego, J. F. Verdu, M. G. de Garay, A. Azpiazu, A. Segarra, M. T. Hernandez, M. Canis, F. Diaz, F. Moreno, and J. Palmero, "Patient dose reference levels for interventional radiology: A national approach," *Cardiovasc. Intervent. Radiol.* **32**, 19–24 (2009).
- ¹¹International Commission on Radiation Units and Measurements, "Patient dosimetry for X rays used in medical imaging," ICRU Report No. 74 (J. ICRU 5, 1–113, 2005).
- ¹²D. L. Miller, S. Balter, P. E. Cole, H. T. Lu, B. A. Schueler, M. Geisinger, A. Berenstein, R. Albert, J. D. Georgia, P. T. Noonan, J. St George, J. F. Cardella, E. Russell, T. Malisch, R. L. Vogelzang, G. Miller, and J. Anderson, "Radiation doses in interventional radiology procedures: The RAD-IR study Part I. Overall measures of dose," *J. Vasc. Interv. Radiol.* **14**, 711–727 (2003).
- ¹³S. Balter, D. L. Miller, B. A. Schueler, P. E. Cole, A. Berenstein, R. Albert, J. D. Georgia, P. T. Noonan, E. J. Russell, T. W. Malisch, R. L. Vogelzang, M. Geisinger, J. F. Cardella, J. St George, G. L. Miller III, and J. Anderson, "Radiation doses in interventional radiology procedures: The RAD-IR study. Part III. Dosimetric performance of the fluoroscopes," *J. Vasc. Interv. Radiol.* **15**, 919–926 (2004).
- ¹⁴A. den Boer, P. J. de Feijter, P. W. Serruys, and J. R. T. C. Roelandt, "Real-time quantification and display of skin radiation during coronary angiography and intervention," *Circulation* **104**, 1779–1784 (2001).
- ¹⁵S. Balter, T. B. Shope Jr., D. W. Fletcher, H. M. Kuan, and R. C. Chan, "Techniques to estimate radiation dose to skin during fluoroscopically guided procedures," in *Intravascular Brachytherapy. Fluoroscopically Guided Interventions* (Medical Physics Publishing, Madison, WI, 2002), pp. 731–764.
- ¹⁶D. W. Fletcher, D. L. Miller, S. Balter, and M. A. Taylor, "Comparison of four techniques to estimate radiation dose to skin during angiographic and interventional radiology procedures," *J. Vasc. Interv. Radiol.* **13**, 391–397 (2002).
- ¹⁷D. L. Miller, S. Balter, P. T. Noonan, and J. D. Georgia, "Minimizing radiation-induced skin injury in interventional radiology procedures," *Radiology* **225**, 329–336 (2002).
- ¹⁸M. A. Ozeroglu, Verification of Caregraph® peak skin dose data using radiochromic film. M.P.H. thesis, Department of Preventive Medicine and Biometrics, Bethesda, MD: Uniformed Services University, 2005. Available at: <http://www.lrc.usuhs.mil/dissertations/pdf/Ozeroglu2005.pdf>. Accessed May 18, 2011
- ¹⁹S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika* **52**, 591–611 (1965).
- ²⁰G. Schwarz, "Estimating the dimensions of a model," *Ann. Stat.* **6**, 461–464 (1978).
- ²¹H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Second International Symposium on Information Theory*, edited by B. N. Petrov and F. Czaki (Akadémia Kiadó, Budapest, 1973), pp. 267–281.
- ²²Y. Khodadadegan, M. Zhang, W. Pavlicek, R. G. Paden, B. Chong, B. A. Schueler, K. A. Fetterly, S. G. Langer and T. C. Wu, "Automatic monitoring of localized skin dose with fluoroscopic and interventional procedures," *J. Digit Imaging* (2010).
- ²³American College of Radiology, "ACR technical standard for management of the use of radiation in fluoroscopic procedures," in *Practice Guidelines and Technical Standards 2008* (American College of Radiology, Reston, VA, 2008), pp. 1143–1149.
- ²⁴L. T. Dauer, R. Thornton, Y. Erdi, H. Ching, K. Hamacher, D. C. Boylan, M. J. Williamson, S. Balter, and J. St Germain, "Estimating radiation doses to the skin from interventional radiology procedures for a patient population with cancer," *J. Vasc. Interv. Radiol.* **20**, 782–788 (2009).
- ²⁵A. Trianni, G. Chizzola, H. Toh, E. Quai, E. Cragolini, G. Bernardi, A. Proclemer, and R. Padovani, "Patient skin dosimetry in haemodynamic and electrophysiology interventional cardiology," *Radiat. Prot. Dosim.* **117**, 241–246 (2005).
- ²⁶E. Bogaert, K. Bacher, K. Lemmens, M. Carlier, W. Desmet, X. De Wagter, D. Djian, C. Hanet, G. Heyndrickx, V. Legrand, Y. Taeymans, and H. Thierens, "A large-scale multicentre study of patient skin doses in interventional cardiology: Dose-area product action levels and dose reference levels," *Br. J. Radiol.* **82**, 303–312 (2009).
- ²⁷D. Bor, T. Olğar, T. Toklu, A. Çağlan, E. Önal and R. Padovani, "Patient doses and dosimetric evaluations in interventional cardiology," *Phys. Med.* **25**, 31–42 (2009).
- ²⁸K. Chida, H. Saito, H. Otani, M. Kohzuki, S. Takahashi, S. Yamada, K. Shirato, and M. Zuguchi, "Relationship between fluoroscopic time, dose-area product, body weight, and maximum radiation skin dose in cardiac interventional procedures," *Am. J. Roentgenol.* **186**, 774–778 (2006).