

6-2010

Optimizing Radiology Peer Review: A Mathematical Model for Selecting Future Cases Based on Prior Errors

Yun Robert Sheu

University of Pittsburgh Medical Center

Elie Feder

CUNY Kingsborough Community College

Igor Balsim

CUNY Kingsborough Community College

Victor F. Levin

University of Pittsburgh Medical Center

Andrew G. Bleicher

University of Pittsburgh Medical Center

See next page for additional authors

[How does access to this work benefit you? Let us know!](#)

Follow this and additional works at: http://academicworks.cuny.edu/kb_pubs

 Part of the [Applied Mathematics Commons](#), and the [Radiology Commons](#)

Recommended Citation

Sheu, Y. R., Feder, E., Balsim, I., Levin, V. F., Bleicher, A. G., and Branstetter, B. F., IV, "Optimizing Radiology Peer Review: A Mathematical Model for Selecting Future Cases Based on Prior Errors", *Journal of the American College of Radiology*, 7(6) (June 2010), 431-438.

This Article is brought to you for free and open access by the Kingsborough Community College at CUNY Academic Works. It has been accepted for inclusion in Publications and Research by an authorized administrator of CUNY Academic Works. For more information, please contact AcademicWorks@cuny.edu.

Authors

Yun Robert Sheu, Elie Feder, Igor Balsim, Victor F. Levin, Andrew G. Bleicher, and Barton F. Branstetter IV

Optimizing Radiology Peer Review: A Mathematical Model for Selecting Future Cases Based on Prior Errors

Yun Robert Sheu, MD, MS^a, Elie Feder, PhD^c, Igor Balsim, PhD^c,
Victor F. Levin, PhD^a, Andrew G. Bleicher, MD^a, Barton F. Branstetter IV, MD^{a,b}

AQ: 1

Background: Peer review is an essential process for physicians because it facilitates improved quality of patient care and continuing physician learning and improvement. However, peer review often is not well received by radiologists, who note that it is time intensive, is subjective, and lacks a demonstrable impact on patient care. Current advances in peer review include the RADPEER[®] system, with its standardization of discrepancies and incorporation of the peer-review process into the PACS itself. The purpose of this study was to build on RADPEER[®] and similar systems by using a mathematical model to optimally select the types of cases to be reviewed, for each radiologist undergoing review, on the basis of the past frequency of interpretive error, the likelihood of morbidity from an error, the financial cost of an error, and the time required for the reviewing radiologist to interpret the study.

Methods: The investigators compiled 612,890 preliminary radiology reports authored by residents and attending radiologists at a large tertiary care medical center from 1999 to 2004. Discrepancies between preliminary and final interpretations were classified by severity and validated by repeat review of major discrepancies. A mathematical model was then used to calculate, for each author of a preliminary report, the combined morbidity and financial costs of expected errors across 3 modalities (MRI, CT, and conventional radiography) and 4 departmental divisions (neuroradiology, abdominal imaging, musculoskeletal imaging, and thoracic imaging).

Results: A customized report was generated for each on-call radiologist that determined the category (modality and body part) with the highest total cost function. A universal total cost based on probability data from all radiologists was also compiled.

Conclusion: The use of mathematical models to guide case selection could optimize the efficiency and effectiveness of physician time spent on peer review and produce more concrete and meaningful feedback to radiologists undergoing peer review.

Key Words: Peer review, cost analysis, quality assurance, diagnostic errors

J Am Coll Radiol 2010;xx:xxx. Copyright © 2010 American College of Radiology

INTRODUCTION

Peer review is a necessary feedback mechanism for any physician, as it allows for quality assessment of patient

care, as well as continued learning and self-improvement for the physician. Both the Joint Commission (formerly the Joint Commission on Accreditation of Healthcare Organizations) and the ACR now formally require a peer-review system for accreditation. RADPEER[®], or an equivalent peer-review program such as peerVUE, is required to qualify for accreditation by the ACR [1]. Similarly, the Joint Commission requires that all staff members participate in peer review, with a continuous random review of 5% of cases for ongoing credentialing [2].

Unfortunately, many radiologists are wary of peer review. Common complaints include a high time commit-

^aDepartment of Radiology, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania.

^bDepartment of Biomedical Informatics, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania.

^cDepartment of Mathematics, Kingsborough Community College, Brooklyn, New York.

Corresponding author and reprints: Barton F. Branstetter IV, MD, University of Pittsburgh Medical Center, Department of Radiology, 200 Lothrop Street, PUH Room D132, Pittsburgh, PA 15213; e-mail: bfb1@pitt.edu.

ment, difficulties in implementation, subjective interpretation, and a lack of clear, actionable results. This leads to poor participation in the peer-review process and inattention to detail from some radiologist participants. Many radiologists view the peer-review process as a requirement that is best bypassed rather than an opportunity for professional growth.

The systems used for peer review have advanced over the past several years, becoming more streamlined, standardized, and unobtrusive. Systems such as RADPEER[®] can now be integrated into PACS, which lessens the disruption to radiologists' work flow. The RADPEER[®] system allows radiologists who are in the process of reviewing comparison studies to rate the accuracy of their colleagues' interpretations. Standardized scoring of discrepancies has been developed for use with the system. Recent updates have included the addition of the clinical significance of errors. Discrepancies are then reconciled in a predefined manner on the basis of the institution [3].

Although RADPEER[®] allows for studies to be rated on their accuracy in a standardized manner and stored for future reference, there are still several elements of the peer-review process that need clarification, namely, the criteria for selecting cases for review and the number of cases to select for adequate representation [4]. With the continuing exponential growth in the number of radiologic studies, the time afforded a radiologist for peer review continues to decrease. It should follow that a peer-review process needs to be as efficient and effective as possible to optimally use radiologists' time.

Current peer-review systems described in the literature range from a random selection of cases, such as the RADPEER[®] system, to voluntary submission of cases, as described by the University of Cincinnati's pediatric radiology department [5]. We believe that improvements can be made in the selection process with the use of a mathematical model. Such a model would be integrated into the electronic peer-review system itself, with no additional input needed by radiologists or by department administration.

The purpose of our study was to create a model that took into account discordances in preliminary and final reports and weighed them in terms of morbidity, financial costs, and time costs, along with their probability of occurrence, to determine optimal case selection for future peer review. We tested the model on an existing database of >60,000 cases to determine the feasibility of these calculations.

METHODS

Data Collection

Between 1999 and 2004, our tertiary care academic center used a preliminary report system that was directly

incorporated into our PACS [6]. Between 5 and 11:00 PM, specialty attending radiologists on call were required to dictate studies in their fields of expertise and provide preliminary reports on studies that were outside their areas of expertise. For the time period between 11 PM and 8 AM, preliminary reports were authored by residents on call. When final interpretations were rendered the next day, dictating radiologists would enter their final impressions alongside the preliminary interpretations. Dictating radiologists would then adjudicate the degrees of agreement between their interpretations and the preliminary reports. The 3 potential categories were "agree," "mild discordance," and "significant discordance." "Agree" would indicate no discrepancies between the preliminary report and the final interpretation. "Mild discordance" would indicate a discrepancy that does not immediately influence patient care. "Significant discordance" would indicate an error that could have an effect on patient care in the time between the preliminary report and the final interpretation (regardless of whether an adverse event actually occurred). Radiology faculty members were trained in the appropriate application of the adjudication terms.

A total of 612,890 preliminary reports were recorded. For the purpose of our study, we included only those preliminary reports that came from our main hospital's emergency department (255,814 studies). The remaining examinations had been performed on hospital inpatients, at outpatient clinics, or at community hospitals. We then excluded those reports in which the preliminary and final authors were the same, which could indicate that the preliminary report system was being used only for the rapid communication of results to referring clinicians. This accounted for 168,570 of the 255,814 studies. Finally, although it was department policy for the final interpreter to provide an adjudication, this was not always the case. Cases without final adjudications were also excluded, accounting for 21,464 cases. This resulted in a final database of 65,870 reports. A senior staff radiologist and an emergency room physician then reviewed each "significant discordance," without knowledge of who authored the preliminary report, to determine whether the differences in interpretation truly merited the classification of "significant discordance." The validation process did not attempt to determine whether the preliminary or final interpretations were accurate; it served only to prevent the misuse of the adjudication terms.

Of the 65,780 preliminary studies, 1,449 (2.2%) had validated significant discordances. Broken down into modalities, there were 0.8% and 2.0% discordance rates for conventional radiography (CR) and CT, respectively, among residents and 0.5% and 2.5% discordance rates for staff radiologists. These discrepant studies were com-

110 piled and represent the data used for the development of
111 our model. They include information regarding the au-
112 thor of the preliminary report, the preliminary findings,
113 and the final findings. The significant discordances were
114 categorized by modality (CT, MR, and CR) as well as by
115 final interpreting division (neuroradiology, abdominal
116 imaging, thoracic imaging, and musculoskeletal imag-
117 ing).

118 This data collection was determined to be exempt
119 from review after consideration by our institutional re-
120 view board. The original data collection is more com-
121 pletely described in a previous report [6].

122 Attributes of the Mathematical Model

123 Our model seeks to calculate a “cost” for each of 12
124 categories that can be used to target areas of weakness.
125 The cost can be defined as the liability addressed per unit
126 of peer-review time. In other words, given a standardized
127 unit of peer-review time, the cost function represents the
128 expected cost (both financial and medically) to the hos-
129 pital and patient if the error is not fixed. Consequently, it
130 would be most efficient to select future peer-review cases
131 from the highest cost categories. We have identified 4
132 attributes that we believe should be taken into account by
133 a mathematical model when choosing a study for peer
134 review:
135
136

- 137 1. *Morbidity cost*: Recent literature indicates that in ad-
138 dition to the frequency of an error, its clinical signif-
139 icance must also be taken into account. This has been
140 addressed in a recent addendum to the RADPEER[®]
141 system that proposes rating clinical significance when
142 there is a discrepancy [3]. An error that could result in
143 patient death is worse than one that has no potential
144 effect on clinical management. A peer-review system
145 should be weighted toward the correction of errors
146 that have the most detrimental effect on patient care.
- 147 2. *Financial cost*: With the current financial environ-
148 ment and ever lowering reimbursements, it has be-
149 come increasingly important to reduce errors for fi-
150 nancial reasons as well as for patient well-being.
151 Insurers have already stopped payment of a variety of
152 presumed iatrogenic illnesses, including catheter in-
153 fections and pressure ulcers [7]. Additionally, new
154 initiatives may result in pay-for-performance reim-
155 bursement, with no payment for poor or inappropriate
156 care [8]. An error that does not cause a patient
157 permanent harm but still incurs large nonreimbursed
158 costs for a hospital is a substantial error, and that
159 effect is reflected in the financial cost of the error.
- 160 3. *Probability*: By using past records to determine the
161 frequency of a particular error for a given doctor, the
162 mathematical model can focus on areas of weakness
163 for a specific radiologist and tailor selection of cases to

110 achieve larger sample sizes within weak areas while
111 avoiding redundant evaluation in areas of known
112 competency. This also allows a radiologist in whom a
113 rare error was serendipitously discovered in peer re-
114 view to be redeemed by repeat measures in the same
115 area.

- 116 4. *Time required to perform the peer review*: The time
117 necessary to review a study should be considered.
118 Would it be more beneficial to peer review a lengthy
119 multisequence MRI study or several traditional CT
120 studies? A balance needs to be struck to maximize the
121 number of relevant studies that a radiologist reviews
122 within the time allotted to peer review. Although it
123 may seem appealing to simply add more cases into the
124 peer-review pool, most radiologists recognize the op-
125 portunity cost of lengthy, complex peer-review cases.

126 The mathematical model that we propose takes into
127 account all 4 of these factors and determines the modality
128 and body part for each radiologist that, on the basis of
129 past errors, have the potential for the greatest morbidity,
130 mortality, and financial cost. This information would
131 allow a peer-review committee or RADPEER[®] to selec-
132 tively pick review cases for a given radiologist to achieve a
133 more efficient review, maximizing the statistical likeli-
134 hood of discovering a true area of weakness for a radiol-
135 ogist. Rather than a new random assortment of cases,
136 there would be definitive proof that a weak area has been
137 addressed and an electronic record made to which future
138 audits could be compared.

139 In the event that a new radiologist does not have
140 enough personal data to specify a targeted area of weak-
141 ness, the model can provide a generic cost function as-
142 suming the universal probability of a given error on the
143 basis of data from all other users or from users within the
144 same subspecialty.
145

146 Quantification of the Cost Function

147 To determine what modality and body part should be
148 targeted for a given radiologist, we created 12 categories
149 corresponding to 3 modalities (CR, CT, and MRI) and 4
150 departmental divisions (neuroradiology, abdominal im-
151 aging, thoracic imaging, and musculoskeletal imaging).
152 We assigned a cost function to each of the 12 categories.
153 The category that yields the highest cost function is the
154 category that should be preferentially reviewed. The cost
155 function is defined as follows:
156

$$157 \text{cost} = \frac{M \times F \times P}{T},$$

158 where M is morbidity cost, F is financial cost, P is prob-
159 ability, and T is the time required to peer review. The
160 elements of this function are quantified as follows.
161
162
163

Morbidity Cost and Financial Cost. The preliminary and final diagnoses were categorized into approximately 20 broad categories per body part (Table 1). Each error type consists of an ordered pair of diagnoses (preliminary diagnosis, final diagnosis), for which the preliminary diagnosis and final diagnosis are different. Each of these error types was then assigned a numerical financial cost value and a morbidity cost value. The criteria for defining these values are as follows:

- Morbidity cost
 - 1 = minor error in which there is no significant change in morbidity or management
 - 10 = moderate error with the possibility of permanent disability without significant deterioration in the future quality of life
 - 100 = major error in which there is the possibility of significant disability that affects the future quality of life or there is a substantial chance of death
- Financial cost
 - 1 = minor error with no significant change in treatment
 - 10 = moderate error resulting in an extended hospital stay or an unnecessary procedure
 - 100 = major error resulting in an additional unnecessary procedure and an extended hospital stay

The specific values of 1, 10, and 100 for these events are used only to document the feasibility of the model; the values can be adjusted by users of the system.

The following example illustrates the difference between morbidity cost and financial cost. Many errors are false-positive errors (an “abnormal” preliminary interpretation and a “normal” final interpretation). Most of these errors pose no threat to patient outcomes but may result in unnecessary hospital stays. They were therefore assigned a low morbidity cost value (1) but a moderate financial cost value (10).

Once we assigned morbidity and cost values for each error type within a given body part and modality, we computed weighted averages for these values. To do this, we found the percentage of each type of significant error within our data for a given modality and body part. We used these percentages to compute a weighted average for the morbidity cost value and for the financial cost value for each of the 12 categories. These weighted averages are the values M for morbidity cost and F for financial cost.

For example, if 30% of neuroradiologic CT errors involved a missed stroke that had major financial and major morbidity cost values, while the remaining 70% of neuroradiologic CT errors were errors with minor morbidity and moderate financial costs, then $M(\text{neuroradiology, CT}) = 0.3 \times 100 + 0.7 \times 1 = 30.7$, while $F(\text{neuroradiology, CT}) = 0.3 \times 100 + 0.7 \times 10 = 37$.

We performed these calculations for all body parts and modalities, for a total of 12 morbidity costs and 12 financial costs.

Probability. We computed the weighted probability, P , that a given doctor would make an error for a given modality and body part as

$$P = \frac{100s + m}{n},$$

where n is the total number of times the doctor performed a given test, s is the number of significant discordances the doctor had for this test, and m is the number of minor discordances the doctor had for this test.

We are assuming that the occurrence of a significant discordance is 100 times more indicative of a doctor’s tendency for error in a specific area than a minor discordance. (This value was chosen only to test feasibility of the model and can be adjusted by users of the system.)

Time. Assuming that a CT scan takes approximately 3 times as long as a conventional radiographic image to review, and an MRI study takes approximately 4 times as long as a conventional radiographic image to review, we let $T = 1$ for CR, $T = 3$ for CT, and $T = 4$ for MRI. Thus, when deciding if a doctor should be reviewed for MRI or CR, although the morbidity and financial costs will generally be higher for an error in MRI, we divide the cost function for MRI by 4 because the review of the MRI takes longer; 4 conventional radiographs can be reviewed in the time it takes to review 1 MRI study. It is clear that this is an oversimplification of modern radiology practice, in which isotropic CT data sets may require far more interpretation time than a simple MRI examination. However, it serves as a simplified model for this analysis and can be easily modified with more data regarding the complexity of imaging.

Once we computed the 12 cost functions for each doctor, we indicated the type of examination for which the cost function is highest. As we collect more data for each particular doctor and for all doctors in the system, the elements of our function will change in real time. The system is programmed to constantly update the cost functions as new data are received and suggest the best test to review whenever an opportunity for review arises.

Note that the type of examination with the highest cost is not the only type of examination that should be reviewed for an individual radiologist. However, high-cost examinations should get preferential weighting within the pseudorandom selection process.

RESULTS

The total cost, as determined by our model, was compiled for 246 on-call attending radiologists and residents in each of the 12 categories (Table 2). The category that

Table 1. Diagnostic categories for error classification

Neuroradiology	CR	CT	MRI	Abdomen	CR	CT
Brain—stroke	0.00	19.28	20.83	Solid viscera—mass	3.70	7.81
Brain—blood	0.00	18.88	4.17	Solid viscera—stone	14.81	6.25
Brain—edema	0.00	1.61	0.00	Solid viscera—lac	0.00	8.13
Brain—mass	0.00	6.43	16.67	Solid viscera—infarct	0.00	0.31
Brain—vascular	1.56	6.83	12.50	Solid viscera—cyst	0.00	3.13
Brain—other	0.00	11.65	12.50	Solid viscera—organomegally	3.70	2.50
Spine—cord	0.00	0.00	16.67	Solid viscera—other	0.93	5.00
Spine—disc	4.69	0.40	0.00	Hollow viscera—obstruction	26.85	8.44
Spine—vertebra	1.56	0.80	4.17	Hollow viscera—ileus	10.19	1.88
Spine—other	0.00	0.80	12.50	Hollow viscera—leak	0.93	5.31
Neck—airway	31.25	0.80	0.00	Hollow viscera—infarct	1.85	2.19
Neck—mass	7.81	2.81	0.00	Hollow viscera—inflammation	5.56	25.00
Neck—gland	0.00	1.20	0.00	Hollow viscera—other	6.48	7.50
Neck—other	4.69	2.81	0.00	Gall bladder	5.56	2.19
Bone—skull Fx	7.81	4.82	0.00	Vessels	1.85	6.56
Bone—skull-base Fx	1.56	2.41	0.00	Nodes	0.00	1.88
Bone—facial Fx	17.19	11.65	0.00	Ascites	0.93	1.25
Bone—other	3.13	3.21	0.00	Air	4.63	4.69
Sinuses	9.38	1.20	0.00	Foreign body	1.85	0.00
Foreign body	0.00	2.01	0.00	Bad tube placement	6.48	0.00
Incomplete study	9.38	0.40	0.00	Incomplete study	3.70	0.00
Chest	CR	CT		Musculoskeletal	CR	CT
Lung—pneumonia	44.72	13.16		Fracture—spine	5.33	44.23
Lung—mass	11.45	7.89		Fracture—rib	2.18	7.69
Lung—PTX	2.86	7.02		Fracture—shoulder	2.67	6.73
Lung—interstitial	3.22	5.26		Fracture—arm	11.15	2.88
Lung—COPD	0.36	0.00		Fracture—hand	12.97	0.00
Lung—other	5.37	4.39		Fracture—pelvis	2.79	10.58
Vascular—CHF/effusion	14.85	4.39		Fracture—hip	2.42	2.88
Vascular—PE	0.54	33.33		Fracture—leg	11.76	1.92
Vascular—cardiomegaly	5.01	0.88		Fracture—foot	9.58	0.00
Vascular—veins	1.07	2.63		Fracture—additional	2.18	9.62
Mediastinum—widened	0.54	0.00		Joints—dislocation	3.39	0.96
Mediastinum—pneumo	0.00	3.51		Joints—OA	0.12	0.00
Mediastinum—mass	1.79	5.26		Joints—inflammation	1.09	1.92
Mediastinum—other	0.89	3.51		Joints—effusion	2.55	0.00
Foreign body	1.25	0.00		Joints—mass	0.12	0.00
Bad tube placement	5.01	0.00		Joints—soft tissue injury	4.36	0.96
Incomplete study	1.07	8.77		Joints—other	0.48	0.00
				Bone—infection	1.45	0.96
				Bone—mass	1.21	1.92
				Bone—other	5.45	0.96
				Spine—alignment	3.39	1.92
				Spine—DJD	0.24	1.92
				Spine—infection	0.00	0.00
				Spine—mass	0.12	0.00
				Spine—other	0.73	0.96
				Foreign body	2.55	0.00
				Incomplete study	9.70	0.96

Note: Each error is considered to be an ordered pair of these diagnoses (preliminary diagnosis, final diagnosis) for which the preliminary and final diagnoses are different. The frequency of each of the errors as a percentage is provided for each of the modalities. There were no MRI discrepancies for chest, musculoskeletal, and abdomen. COPD = chronic obstructive pulmonary disease; CR = conventional radiography; DJD = degenerative joint disease; Fx = fracture; OA = osteoarthritis; PE = pulmonary embolism; PTX = pneumothorax.

Table 2. Total cost of errors in the 12 categories for 5 sample physicians

Physician	CR				CT				MRI			
	Abdomen	Chest	MSK	Neurologic	Abdomen	Chest	MSK	Neurologic	Abdomen	Chest	MSK	Neurologic
1	0.00	0.45	0.00	0.00	49.62	27.05	57.81*	38.20			0.00	
2	0.00	155.18	300.90*	100.33	2.68	0.73	3.06	296.74	0.00		0.00	12.06
3	383.53*	171.83	357.05	0.00	5.47	0.00	6.88	83.43	0.00		0.00	253.86
4	5.27	1.10	87.95	0.00	94.77*	0.00	3.67	8.09				
5	0.56	45.87	135.23	2,533.26*	91.12	44.08	4.01	138.60				

Note: Blank spaces indicate categories with too few cases for statistically meaningful analysis. A value of zero denotes a category with an adequate total number of cases but without any errors. CR = conventional radiography; MSK = musculoskeletal.
*Category with the highest cost.

had the highest cost was then selected as the one that should be preferentially examined in future peer reviews. The total number of cases read for each category was also determined for each attending radiologist and resident. The universal probability for a category was compiled using data from all radiologists, and the total cost was calculated (Table 3). Last, the average cost per category as well as the range of costs per category were tabulated (Table 4).

DISCUSSION

A large sample of significant discordances on the basis of overnight preliminary reports adjudicated by specialty

trained radiologists was compiled. This allowed us to construct a mathematical model that factored in the estimated morbidity and financial cost of each error, along with the probability that an error would be repeated and the time required for peer review. This final aggregated cost was computed for each of 12 categories, with each radiologist’s “highest cost” category used as a selection criterion in future peer-review cases. The goal of this research was to demonstrate the feasibility of using a mathematical cost model to guide selection of radiology examinations for peer review. Additionally, the study creates a baseline “evaluation” for all of the participating attending radiologists, against which future errors can be compared.

There are several limitations to our current, preliminary mathematical model. The weighting factors used for the calculation of the various morbidity, financial, and time costs are generalizations that were selected using broad categories. These weighting factors will need to be refined with future research. For example, the time factor could be tied to the physician work component of the Resource-Based Relative Value Scale used by CMS. This is, however, a suboptimal model, because the physician work component of the Resource-Based Relative Value Scale also takes into account the mental effort and the stress imposed on the physician. Review of the literature does reveal a single study in which average completion times for different modalities and body parts among a large group of radiologists were measured [9]. This information is unfortunately limited by the age of the paper. Published in 1990, in the early stages of MR, the data are likely not relevant to today’s radiology environment. More precise measurements of radiologists’ time commitments are necessary to refine our mathematical model. The same problem exists with our estimates of the financial and morbidity costs. For the purposes of our model, we used a grossly simplified stratification system based on 3 possible outcomes. This obviously is not representative of the complexity of a real hospital environment. Review of the literature yielded no comprehensive publications on the morbidity and financial costs of

Table 3. Universal cost for each category using the universal probability of an error on the basis of data from all radiologists

Study	Total Cost Value	Total Number of Cases
CR		
Abdomen	40.31	4,831
Chest	60.53	16,806
MSK	119.10	15,443
Neurologic	246.11	2,313
CT		
Abdomen	78.72	7,310
Chest	26.62	2,224
MSK	69.40	2,812
Neurologic	238.08	10,340
MRI		
Abdomen	0.00	16
Chest	0.00	9
MSK	12.54	228
Neurologic	117.26	1,166

Note: A grand total of the number of cases interpreted in each category is also provided. Some of the original 65,780 cases did not fit into these categories (such as ultrasound or nuclear medicine) but were too small in number to form their own category, so they are excluded from this table. CR = conventional radiography; MSK = musculoskeletal.

Table 4. Average total cost along with minimum and maximum values for each of the different body parts and modalities

	CR				CT				MRI	
	Abdomen	Chest	MSK	Neurologic	Abdomen	Chest	MSK	Neurologic	MSK	Neurologic
Average	39.73	44.50	106.03	235.88	68.99	31.51	43.52	139.05	28.47	140.68
Minimum	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Maximum	383.53	180.81	537.36	1,558.93	232.37	152.99	252.77	918.95	79.41	680.32

Note: There were no MRI discrepancies for chest, MSK, and abdomen, because of a small number of cases. Calculated costs for physicians who performed <10 studies were excluded in the calculation of a maximum value. CR = conventional radiography; MSK = musculoskeletal.

radiologic errors. This problem too needs to be addressed with future research. Optimally, costs for the extra procedures performed between the time of the preliminary read and the final discrepancy would be tabulated and a system of defining morbidity created. Understandably, this would be a technically challenging as well as time-consuming effort. As the initial purpose of our study was to test the initial feasibility of a mathematical model, we did not delve deeply into a precise calculation of these costs. However, the advantage of the model is that it is extremely flexible, with easy adjustment of these values, and additional gradations can be added as new data become available.

Errors may also occur in our model if the sample size for a given physician is too small (eg, part-time faculty members, new attending radiologists, residents). All the results should be correlated with the total number of cases a radiologist performed in any given category. In the event a sample size is too small, several options are available. The probability could be substituted with the universal probability of a given error. A second option is to treat the error as real despite a small sample size and target that category, allowing a larger sample size to be used in the next peer review. This would determine if the area of concern is valid or the result of a statistical aberrancy.

We initially hoped to reduce the total number of peer-review events that were needed for each physician or to determine a minimum number of cases required for a statistically significant peer review. Unfortunately, we were not able to mathematically achieve this with our current data set. We plan to continue our data collection and hope to address this issue in the future.

Despite these shortcomings, there are already many potential applications for this mathematical model. Collecting data is only the first part a peer-review process. It must then be processed into a relevant and practical form. Many examples of peer review have been described. However, none have specifically addressed selection criteria for peer-review cases. Clinically, such information is useful in many ways. It is useful for radiologists and residents by directing the focus of future study, and it is useful for radiology departments by allowing the

administration to monitor the performance of all staff members relative to their peers. Global mistakes can be detected and communicated to staff members to allow for more care with high-risk studies. It also serves to provide a quantifiable way to ensure that there is improvement in weak areas.

For a residency program director, the model can detect global deficiencies among the residents that can be remedied with targeted teaching. Cases for daily case conferences could also be selected on the basis of the category with the greatest average cost. From residents' standpoint, it provides information on what mistakes they are making, what impact the mistakes had, how many other people in their peer group made the same mistake, and what to work on in the future. Last, with targeted case selection, we hope to reduce the number of cases needed for a successful peer review, saving radiologists' time as well as providing concrete evidence that peer review is having a positive effect.

We envision a system that is seamlessly integrated into the peer-review process, tapping into the information collected and providing a periodic report to every radiologist. Identified areas of weakness could then be addressed by radiologists or submitted to a peer-review committee. Cases could then be selected to comprehensively evaluate an area of weakness, and continued data collection would verify improvements.

In conclusion, peer review is an essential component of radiologists' practice, ensuring continuing improvement in patient care. The rise in the number of studies interpreted per radiologist and other radiologist time commitments, along with a generally unfavorable attitude toward peer review, requires that the peer-review process be as efficient and effective as possible. By identifying trends in errors as well as their costs for patients and hospitals, our proposed model of peer-review case selection will allow a more relevant selection of future cases for review and allow a statistically accurate way of monitoring physician improvement and resolution of areas of weakness. Our model could be implemented as an extension of RADPEER[®] or a similar system, flagging relevant cases for review and providing continually evolving data with minimal manual input.

REFERENCES

380					
381					
382					
383					
384					
385					
386					
387					
388					
389					
390					
391					
392					
393					
394					
395					
396					
397					
398					
399					
400					
401					
402					
403					
404					
405					
406					
407					
408					
409					
410					
411					
412					
413					
414					
415					
416					
417					
418					
419					
420					
421					
422					
423					
424					
425					
426					
427					
428					
429					
430					
431					
432					
433					

UNCORRECTED PROOF

**000 Optimizing Radiology Peer Review: A Mathematical Model for
Selecting Future Cases Based on Prior Errors**

*Yun Robert Sheu, MD, MS, Elie Feder, PhD, Igor Balsim, PhD,
Victor F. Levin, PhD, Andrew G. Bleicher, MD,
Barton F. Branstetter IV, MD*

The use of mathematical models to guide case selection could optimize the efficiency and effectiveness of physician time spent on peer review and produce more concrete and meaningful feedback to radiologists undergoing peer review.

434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487

UNCORRECTED PROOF

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

1

AQ1— Please indicate Dr Levin's institutional affiliation, if any. If he is unaffiliated, an affiliation line is still required stating his place of residence or private practice (city and US state or city and country only).

AQ2— This calculation (assuming a strict left-to-right order of operations) results in a value of 307, not 37. Please clarify.
