Author's response to reviews

Title: Decision Theory Applied to Image Quality Control in Radiology.

Authors:

Patrícia S Lessa (patlessa@hotmail.com)
Cristofer A Caous (cacaous@gmail.com)
Paula R Arantes (parantes@usp.br)
Edson Amaro Jr (edsonjr@einstein.br)
Fernando M Campello de Souza (fmcs@hotlink.com.br)

Version: 2 Date: 9 June 2008

Author's response to reviews: see over
Manuscript title: “Decision Making Theory Applied to Image Quality Control in Radiology”

# Reviewer 1

I. “Abstract: Generally fine, although comments about later sections may affect this one.”

The abstract was modified supported on the later corrected sections.

II. “… page 4 para 1 line 3: Exactly what is not standardized? In radiology practice there are standard views for nearly every examination that the technologist acquiring the images knows and follows. There are also general dose limits and guidelines for every examination that in general are followed, although these very much depend on the exam, the acquisition device, the size of the patient etc. There are also certain types of film used for certain exams – high-speed, high-contrast, high-dynamic range for example for mammography, that are fairly standard. You must be more precise here about what is not standardized. It should also be noted that with digital radiography, many of the underexposure and overexposure issues are gone and the only major problem left is poor/improper positioning of the patient leading to poor quality. This can be dealt with to some extent by technologist technique, but in many cases the patient simply cannot be positioned properly because they are too ill. Also there are issues with patient motion (breathing when they’re not supposed, shifting position etc.) that cannot be helped or standardized either.”

Some issues are difficult to standardize, although the major of technical procedures were monitored for a proper quality assurance (Papp et al., 1998). The manuscript works on the standardization lack in radiology field relating a statistical evaluation method on decision making theory. As consequence, an inaccurate diagnostic can occur increasing the costs of the patient’s treatment. The authors thank the reviewer for this comment and additional precise information was added to the background of the manuscript.

III. “Page 4 paragraph 2 line 6: When you refer to image detection, edge detection etc. Are you referring to the point in the imaging chain of image acquisition (the actual device capturing the image), the developing of the image once captured (for film the chemical process of developing, fixing etc.), the display on the alternator or view box, or the detection of these things by the human observer? The imaging chain in radiology is quite complex and you are referring to it at a very general level that is not really appropriate.”

The authors referred to inadequate use, inadequate sensibility and or inadequate spatial resolution related to the radiological chain process. The present work applied the decision making theory on acquired images, not identifying other parameters before the image detection. The observation was added to the background section.
IV. “Page 4 paragraph 2 line 8: It is not true that people have not looked at image quality from a more mathematical level such as the one you are proposing. There are lots of studies that have done just that: see for example Art Burgess, Harry Barrett, Craig Abbey, Kyle Myers, Miguel Eckstein, David Wilson, Jeff Johnson, Jeff Lubin, Subok Park and a host of other researchers in radiology who have used a variety of mathematical and decision-based theories to model and investigate image quality.”

The mistake was corrected.

V. “Page 4 paragraph 4 line 1: The problem is – who is your individual? In practical radiology, it is the technician who acquires the image and has the first look at it. It is often up to them – without ever asking the radiologist – whether or not another image must be taken because of poor quality. It should also be noted that the technologists often will acquire an image with certain parameters because they know which radiologists will view it and what their preferences are. For example, some radiologists like the images a little darker and some a little lighter so the technologist will process the film to match that preference – basing image quality on the subjective preference of a single person. In other words there is a range of what is considered good image quality beyond just dose etc. Further, if the individual is the radiologist, as I’ve just said, there are ranges of what is acceptable and preferred as well in terms of image quality and the range still falls within the recommended dose limits etc. for acquisition quality. You also need to review the image quality in radiology literature because there are lots of studies that show that image quality can actually be quite poor yet radiologists have no problem reaching the proper diagnostic decision – they may not like the look of the image and there may even be artifacts, but they can still render the correct decision.”

The classification of this “individual” depends on direct or indirect aspects. He could be the administrator of the radiology service department, the radiologist, the technician and even the patient who could refuse to perform a second or third roll to repeat the acquisition. This part was clarified in the background section.

VI. “Page 6 paragraph 1 line 9: Part of the problem here seems to be a fairly heavy reliance having a posteriori information. In other words, it’s not clear how you would use this model prospectively to make decisions about image quality. You have a database (that you never describe by the way) and presumably decisions about that database in reference to a gold standard of truth to know that those diagnostic decisions are indeed correct. It is not clear at all how you ever could get patient satisfaction information. You have set your model up based on this set of data but then what. In the future you then have a newly acquired image and you need to decide if the image quality is good - the initial point of accept or reject based on image quality is the technologist acquiring the image so you have no input of diagnostic decision from the radiologist & nothing about patient satisfaction. How does the technologist use your model to make a judgment about the quality of the image they have just acquired? This step from retrospective to prospective seems to be missing here.”
The authors thank the reviewers comment. This step was added to the text in order to clarify the study application method.

VII. “Page 6 paragraph 2 line 4: You need to define correct decision here. In radiology there are lots of decisions that the radiologists needs to make and it differs for each type of exam. In general, the radiologist first needs to detect something – is there something abnormal in the image or not. This decision can be right or wrong. Then they need to classify what they detect – is it a nodule or is it a pneumothorax. This decision can be right or wrong. Then there is an action decision – patient needs another exam, patient needs a tube inserted, patient needs surgery or whatever. This decision can be wrong or right. Exactly which decision are you talking about in your model and how are you defining the gold standard of truth when deciding whether or not the decision is right or wrong?”

The decision course action that was considered to this work involves rejecting or accepting a film or image related to the state of nature in decision making theory. A gold standard of truth maybe is a very complex consequence of all studies involving mathematical theories in radiology. This work explored some possibilities according to the dependency values regarding to the image quality assurance.

VIII. “Page 6 paragraph 2 line 9: Again – patient satisfaction is generally not something that is measured in this context. Generally the patient does not even know if the number of images acquired is because they are necessary or if they are repeats due to poor quality. They certainly never know about the dose they just received unless they ask (which is rare indeed) and even if they asked it is questionable whether a technologist could or would tell them. What sort of data then did you have to build your model on?”

The patient satisfaction level was mathematically considered as the probability of “yes” or “no”, “1” or “zero” measure. In addition this personal satisfaction level was not measured in the present study with specific questionnaires.

IX. “Methods: Details are generally missing. As already noted you never even describe the database that you used to build your model on – type of images, number of images, types of abnormalities, whether they were confirmed with a gold standard and how, who provided the diagnostic decisions, where did you get the patient satisfaction data etc. None of this is provided.”

The referred database was supported by the two tables of probabilistic results combining the dependency values. The dependency values include the inadequate or adequate parameters of application, sensibility and resolution, as well as, diagnostic, cost and satisfaction (not personal satisfaction levels).

X. “In general you never explain how the model is validated. In other words – you have some database that you used to design the model, but was it ever tested and how? Typically with these types of models you then need to get a whole new dataset and use the model (and compare against some gold standard or other accepted measure of image quality) to make judgments of image quality on the new set prospectively and see how good you do. It does not
seem as if any of this sort of validation was carried out so how can the reader have faith that it indeed works and can be used in the clinic to make prospective judgments about image quality?"

All values of probable observations were devised using hypothetical prediction based on previous information from real datasets. The present work is not a validation method process.

XI. “Page 7 para 2 line 3: Again – you never described your database. Also – why was a phantom used? Image quality varies from exam to exam and is based on lots of factors including patient size, age etc. What type of phantom was used – what body part were you simulating, what patient size, what age, what structure (soft tissue, bone etc.)? None of these details are provided yet are required.”

There is not a specific dataset or database. The model used herein is hypothetical therefore it is not correct to have a new method validated through this work. In addition, no matter which acquisition data system, there will always be a likelihood function to the considered mathematical model.

XII. “Page 7 para 3: Again – what datasets?”

There is not a specific dataset or database.

XIII. “Page 7 para 3 & Figures 2 & 3: These figures are not ROC figures. The axes in ROC go from 0 to 1 both x and y. You have limited ranges on both your axes. These are transformations of some sort of ROC curves. ROC curves go up from the bottom left to upper right corner & yours go upper left to lower right – without ever even hitting the origin points. The data are basically uninterpretable as well – it looks like a sets of random circles. Is it everything below your line is good, everything above or what? How far above & below are considered good/bad? What are the ranges? These curves and the whole explanation is vague and difficult to figure out.”

The mistake was corrected. Instead of figures should be tables 1 and 2. The ROC curve in terms of P (positive test| positive for disease) versus P (positive test| negative for disease) or sensibility versus false positive, P (altered exam| disease diagnostic) versus (altered exam| not disease diagnostic) could be one of the different ways to get a ROC curve. At the present, risks settings were used (good film| bad film). Better explanations were added to the text.

XIV. “Results: As already noted your two main figures are vague and difficult to interpret and need much better explanation to make the results meaningful and useful.”

The ROC curve axis depict risks not probabilities involving the preference concept of the decision making ( von Neumann-Morgenstern function). The present work aims about decision and not about inference status.
XV. “Page 8 para 1 line 2: You seem to be mixing terms here and not talking about films as they are used in the clinic. You talk about “lots” of film seemingly referring to boxes of film or large sets that would be used with a number of patients. Then you talk about “units” which are presumably individual films. It is rare that an entire lot/box of film would be bad and would have to be thrown away. Typically a single film/unit is bad because of how it is used. It is not clear how and why you are talking about lots and units as if one bad unit would result in throwing away the whole lot.”

Lots of films refer to boxes. Each box contains units of films.

XVI. “Page 8 para 3 line 1: What films? Again you never describe your database.”

The authors referred to radiographic films. The present model is hypothetical and predictive.

XVII. “Page 9 top: You reference mammography here but it is not the only case where different imaging parameters are needed. It simply is not clear anywhere how the model works when each type of exam has different requirements that you mostly do not address.”

The reference to mammography was only an application example. It was not the objective of the present work to test the model for each type of exam, but show that it could be an effective tool for image quality evaluation in other type of exams.

XVIII. “Page 9 top: Your terms have problems. 1) There is a range of what is acceptable for use – it is not a black & white single point decision. As already noted, there is quite a bit in the literature on radiologists judging image quality quite poorly 76 but doing just fine diagnostically. So – how do you define “inadequate for use”?"

The term “inadequate use” reflects the type of the film incorrect applied to different exams.

XIX. “Page 9 top: What is “sensibility” – define and give criteria for judging it.”

The term sensibility is related to the time of film exposure when a best optical density point is reached to the image acquisition.

XX. “Page 11 para 4 line 2: How do you know the patient would be more risk aversive? If the patient knew that a little higher dose would give a higher quality image (which is exactly what higher dose does before noise takes over) then why would they not say “I’ll go with the higher dose to get a better image for my radiologist so they can make a better decision”? You are making assumptions about patient attitudes that may or may not be true because you have no real data to base your assumptions on."

This was not the aim of the present study.
XXI. “Page 11 para 4 line 5: You never describe the imaging equipment so how do you know what the parameters are?”

The signal/noise ratio considered to this study was provided by the manufacture and is a well known parameter in imaging equipments.

XXII. “Page 13 para 1 & Figures 2 & 3: As already noted, these are not ROC curves and it is impossible to make any sense of them without a better explanation as to exactly what you are plotting.”

The present work does not propose a method to improve sensibility or specificity of the data acquisition system. This issue has a technological solution, for that reason, it does not make sense to worry about the traced ROC curve in terms of $P$ (positive test| positive for disease) versus $P$ (positive test| negative for disease) or sensibility versus false positive, $P$ (altered exam| disease diagnostic) versus (altered exam| not disease diagnostic). A better explanation was added to the text.

XXIII. “The tables have values that you really need to explain better. Very little here is easy to interpret because you don’t really explain what anything means.”

Tables 1 and 2 are likelihood function and consequence function, respectively. Table 3 depicts a result of a calculation.

XXIV. “Page 13: Again – it is not clear if this model has been validated in any way with an independent dataset from the one that was used to build the model. It does not seem so and thus there is no way to judge the results as being good or bad or useful or not.”

There is not a specific dataset or database. The model used herein is hypothetical as already clarified.

XXV. “… page 14 para 1 line 5: Again – there are quite a few mathematical & decision-based models about image quality in the literature that you do not seem to have reviewed.”

These models about image quality did not apply the decision making theory.

XXVI. “Much of the discussion is redundant to the introduction – condense. You should discuss things like the limitations of your approach, whether/how it was validated etc. Page 17 para 4 line 1: You did not do a cost-utility analysis at all and you certainly did not really take into account patient preferences because you never had patient data – you made assumptions about what a patient would prefer that may or may not be right. Throughout it is not clear that your assumptions are valid and whether the model itself was ever really validated.”

The discussion section was corrected according to the reviewer’s comments.
“References: Need to do a better review of the decision and image quality literature in radiology. The tables and figures are not very clear as noted above and need much better explanation in the text.”

Other references regarding to image quality were added to the reference section.

“Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

These mistakes were corrected.

# Reviewer 2

I. “The abstract is a little vague and probably misleading.”

The abstract was corrected.

II. “In the Background it is stated that this work is applied in “routine” use?”

It is not stated that the work is applied in routine use. The authors propose that it could possibly be applied in a routine.

III. “With “film lot” do you mean a film batch with a different lot number from the film type already used in a radiology department, as well as, other film types from the same or different manufacturers?”

Yes, that’s correct.

IV. “In the Methods you write: “based on routine films …”. Is this what you did and how? (See later comments concerning the Methods Section).”

The term “based on routine” reflects the observation probability values considering a hypothetical dataset related to real information in radiology parameters.

V. “The value of r (arbitrary selected as 2 or 4) is the least important thing you did compared to the work described in this manuscript.”

Yes, although it is relevant to say that in any risks levels according to the mathematical statement of the present work, the model with decision making theory application is still reliable.

VI. “… should note, that while patient’s opinion could indirectly influence the administration in the choice of the film type used, however patients can not decide!”
Yes, that’s correct. The authors referred to the fact that a patient can decide to not perform a second or third acquisition repetition. This misunderstanding was clarified throughout the text.

VII. “According to your conclusion, I thought that your study contains all the necessary information that I could implement for making some kind of risk assessment (e.g. using various types of films available for mammography with different characteristics), something that is quite far away from the truth.”

The present work is not a validation method. No one (decision maker) had their utility function educated. To reach a stage of risk assessment the model must pass through other studies in a real world setting. Work at all of these levels, represent major efforts toward a quantitative science of medical decision making.

VIII. “… page 4, 2nd paragraph: Film rejection is not due to “lack of standardization” but to a number of causes relating to the X-ray unit operator (e.g. wrong patient positioning), X-ray equipment (e.g. AEC system failure) and film processor (e.g. processor artefacts).”

The manuscript works on the standardization lack in radiology field relating a statistical evaluation method on decision making theory. The film rejection when correct, assures the quality of the image and an accurate diagnostic evaluation.

IX. “Page 5, 1st paragraph: the description of the four main sets, needs revising and re-ordering to follow the block diagram of figure 1, i.e. observations come first, then the states of nature are determined from the analysis of observations, then the possible actions are discussed in view of the consequent payoffs which should also be determined and considered. The same should be done for the probabilistic mechanisms analyzed in the next paragraph.”

This issue was corrected.

X. “Page 5, last paragraph: Instead of “Other three functions” you should write “Three additional functions”, but the major problem is that only two functions are defined in this paragraph: the “utility function” and the “decision rule”. The third rule is most probably the “randomized decision rule” or its alternative “behavioral rule” mentioned only in the next page. These three rules have first to be stated and then clearly explained.”

The authors attempt to the kind observation. The rules were clearly explained.

XI. “Page 6, 1st sentence: the ROC analysis briefly mentioned, relates to the interpretations of observations using a scale associated with a varying level of confidence, referring to the detection of certain structures or serious revision and a more detailed connection between mathematics and what these numbers represent in actual practice.”

The ROC analysis was explained in detail in order to associate the mathematical statements and a practical appliance.
XII. “From the second half of page 12 and up to the Discussion, I could not follow any more and I could not understand the figures 2 and 3 quoted nor the table 3. Some of my questions are given below: what is the Rd that suddenly appears (it has not been defined before), what is its meaning and why is the negative sign used?”

The Rd is the decision rule of the risk parameter function. The negative sign means that is a negative mathematical equation and it is part of it referring to a lost function. The figures and table were now properly explained, as well as Rd that was previously defined in the text.

XIII. “In figures 2 and 3 you give Rd (?1) versus Rd (?o) and not Rd (?) × Rd (?1). These figures (if rotated by 90 degrees to the right) look like ROC curves (True Positive Fraction versus False Positive Fraction), but with negative values they do not make sense to me and I can not see any connection with ROC. What are the circle symbols? The possible decisions or the observations? Please explain in more detail within the text and in the figure legends the meaning of these figures.”

The figures legends were corrected and the possible decisions were explained in the discussion section.

XIV. “What ROC analysis was carried out and in what data? Who were the observers? Where these fixed values of Rd came from? What are the d17, d192 and d16 (decisions?). In table heading it writes “Regra...”. How are they defined, what is their meaning and why you selected these particular values?”

The ROC analysis depicts the decision rules sets of the considered issue. The table heading was corrected. The defined selected values were disposed at the vertex of the ROC curve for each given decision value (d17, d16, ...).

XV. “What happened with the risk estimation mentioned in the abstract?”

The risk estimation was cited throughout the results section, specifically on page 11, paragraph 2 and in the function equations of this model. Please see Rd (θo) and Rd (θ1) defining the risk function.

XVI. “Discussion and Conclusion: I can not make any detailed comments concerning the last two sections since I was long lost, when I read them. I am almost certain that revisions would be required in these sections too, to link the issues discussed with the results presented.”

The authors thank the reviewer comment. The discussion and conclusion were properly revised.
XVII. “I have noticed however that in the 3rd paragraph of page 15 and the 2nd paragraph of page 16 an interesting reference to the real clinical practice is attempted. However, in page 16 it is written “… if the real state of nature was an initial malignant tumor…” but until now the states of nature defined were good film and bad film, the parameters considered were correct and wrong diagnosis, while the patient actual condition (with or without the suspected pathology) were never considered in the model.”

The phrase was only a practical application example. This inaccuracy was corrected.