MIPS SPECIAL FEATURE

Quantifying the potential for dose reduction with visual grading regression

^{1,2}Ö SMEDBY, MD, Dr Med Sci, ³M FREDRIKSON, PhD, ^{1,2}J DE GEER, MD, ^{4,5}L BORGEN, MD and ^{2,6}M SANDBORG, PhD

¹Radiology (IMH), Linköping University, Linköping, Sweden, ²Center for Medical Image Science and Visualization (CMIV), Linköping University, Linköping, Sweden, ³Occupational Medicine (IKE), Linköping University, Linköping, Sweden, ⁴Radiology, Drammen Hospital, Drammen, Norway, ⁵Radiology, Buskerud University College, Drammen, Norway, and ⁶Radiation Physics (IMH), Linköping University, Linköping, Sweden

Objectives: To propose a method to study the effect of exposure settings on image quality and to estimate the potential for dose reduction when introducing dose-reducing measures.

Methods: Using the framework of visual grading regression (VGR), a log(mAs) term is included in the ordinal logistic regression equation, so that the effect of reducing the dose can be quantitatively related to the effect of adding post-processing. In the ordinal logistic regression, patient and observer identity are treated as random effects using generalised linear latent and mixed models. The potential dose reduction is then estimated from the regression coefficients. The method was applied in a single-image study of coronary CT angiography (CTA) to evaluate two-dimensional (2D) adaptive filters, and in an image-pair study of abdominal CT to evaluate 2D and three-dimensional (3D) adaptive filters.

Results: For five image quality criteria in coronary CTA, dose reductions of 16–26% were predicted when adding 2D filtering. Using five image quality criteria for abdominal CT, it was estimated that 2D filtering permits doses were reduced by 32–41%, and 3D filtering by 42–51%.

Conclusions: VGR including a log(mAs) term can be used for predictions of potential dose reduction that may be useful for guiding researchers in designing subsequent studies evaluating diagnostic value. With appropriate statistical analysis, it is possible to obtain direct numerical estimates of the dose-reducing potential of novel acquisition, reconstruction or post-processing techniques.

Received 24 October 2011
Revised 14 January 2012
Accepted 31 January 2012

DOI: 10.1259/bjr/31197714

© 2013 The British Institute of
Radiology

The ultimate diagnostic value of an imaging procedure can be assessed only in relation to an accepted reference standard, using concepts such as sensitivity, specificity or receiver operating characteristic (ROC). Such studies require considerable resources in terms of image material and competent reviewers, and are typically performed at a late stage in the development of new procedures. However, in the earlier phases, when protocols are being formulated, evaluated and compared, there is often a need to perform repeated evaluations of protocols that differ only slightly (*e.g.* to optimise the diagnostic value in relation to ionising radiation dose in CT or examination time in MRI). In these situations, an attractive alternative is visual grading experiments (*i.e.* experiments in which diagnosticians are asked to assess some aspect of the quality of an image on an ordinal scale, or to compare a pair of images using a similar scale) [1]. The image quality

criteria often relate to the visibility of certain anatomical structures, as is the case with the European Union criteria for CT image quality [2].

In particular, several methods have been proposed to reduce the dose of ionising radiation in CT. As the obvious price of a reduced dose is an increase in image noise, most of these have aimed at suppressing the amount of noise in the images, in order to preserve diagnostic image quality while reducing the dose. Of particular interest is the use of alternative reconstruction techniques (iterative—algebraic or statistical—algorithms rather than filtered back projection), which have a long history [3–5] but have only recently become commercially available [6, 7], as well as the application of post-processing in the form of non-linear filters [8]. For evaluating these methods, visual grading experiments may be useful.

The basic design of visual grading studies is rather straightforward, comprising either evaluation of one image (or stack) at a time or comparison of two images (stacks) simultaneously [1]. However, when it comes to the statistical analysis of the rating data, there is less

Address correspondence to: Professor Örjan Smedby, Radiology (IMH), University Hospital, SE-58185 Linköping, Sweden. E-mail: orjan.smedby@liu.se

consensus on the choice of methods. The use of parametric statistical methods based on least-square approximation, such as analysis of variance, may be tempting, but is not statistically appropriate, as the data to be analysed are defined on an ordinal scale. To solve this problem, Båth and Månsson [9] have proposed the non-parametric visual grading characteristic (VGC) method. In simple situations where two imaging protocols are to be compared, this approach appears to work well. The same is true of an alternative non-parametric method introduced by Svensson [10], which has also been applied to the evaluation of image quality [11]. For more complex situations, however, the visual grading regression (VGR) method [12] is easier to apply if one wants to separately evaluate the effects of acquisition settings, reconstruction algorithms and post-processing.

With the traditional design of visual grading studies evaluating dose reduction measures, images with and without the intervention studied, and acquired at a limited number of different dose levels, are compared with each other, either directly in image-pair experiments or indirectly by studying the grading scores in single-image experiments. It thus seems possible to draw conclusions only on actually tested doses.

An alternative approach to the evaluation of dose-reducing techniques is to try and quantify separately the effects of the intervention and of variations in the dose, and then relate these effects to each other. The goal of the analysis would then be to obtain direct numerical estimates of the potential for dose reduction brought about by the intervention.

The aim of this article is to demonstrate how a parametric statistical model can be used to obtain quantitative estimates of the potential for dose reduction when using post-processing such as adaptive filtering.

Methods and materials

Theory

The framework for the analysis will be VGR, proposed in a previous publication [12]. This is a rather straightforward application of ordinal logistic regression. In order to mathematically model the probability for a certain grading score, the probability p is transformed with the logit function:

$$\text{logit}(p) = \log[p/(1-p)] \quad (1)$$

That is, the (natural) logarithm of the odds $p/(1-p)$. The probability of obtaining a visual grading score not greater than n is then given by

$$\text{logit}[P(y \leq n)] = z \quad (2)$$

where $P(A)$ denotes the probability of the event A , y is the score assigned by the observers, and z is a linear combination of factors assumed to affect the perceived image quality.

Now consider a situation where images are acquired with radiography or CT using several levels of the mAs setting, which is proportional to the effective dose. Then the influence of the mAs setting should also be included in

the model. We have chosen to assume that, in order to achieve a given effect on the perceived image quality, one should multiply the mAs setting by a certain factor rather than adding a constant term. Thus, within a certain interval, the effect of doubling the mAs setting, for example, will always be the same. This is accomplished by introducing a $\log(\text{mAs})$ term in the regression equation.

In a single-image experiment with post-processing applied to some of the images being graded by the observers, the regression equation will then be:

$$\text{logit}[P(y \leq n)] = a \log(\text{mAs}) + b \text{PP} + D_P + E_R - C_n \quad (3)$$

where a and b are coefficients describing the dependency on $\log(\text{mAs})$ and post-processing, respectively, PP is a binary variable (1 for post-processing applied, 0 for not applied), D_P and E_R are terms representing individual patients and reviewers, respectively, and C_n is a constant specific for each scoring level n . With simple algebra, it is possible to estimate the change in $\log(\text{mAs})$ that would yield the same change in score probabilities as the addition of post-processing:

$$\Delta \log(\text{mAs}) = b/a \quad (4)$$

The relative reduction in mAs value can then be estimated by

$$\text{RR}_{\text{mAs}} = 1 - \exp(-b/a) \quad (5)$$

In an image-pair experiment, where pairs of images are being compared with respect to a given image quality criterion, the regression equation will take the form:

$$\text{logit}[P(y \leq n)] = a[\log(\text{mAs}_R) - \log(\text{mAs}_L)] + b(\text{PP}_R - \text{PP}_L) + D_P + E_O - C_n \quad (6)$$

where the subscripts R and L denote the right and left image in the image pair, respectively. If two post-processing procedures PP1 and PP2 are studied simultaneously, then Equation (6) will be expanded to:

$$\text{logit}[P(y \leq n)] = a[\log(\text{mAs}_R) - \log(\text{mAs}_L)] + b_1(\text{PP1}_R - \text{PP1}_L) + b_2(\text{PP2}_R - \text{PP2}_L) + D_P + E_O - C_n \quad (7)$$

The effect of each of the post-processing procedures PP1 and PP2 is given by the relative mAs reduction

$$\text{RR}_{\text{mAs},i} = 1 - \exp(-b_i/a) \quad i = 1, 2 \quad (8)$$

As individual patients and observers are not interesting in themselves, but merely represent samples from larger populations, they should be treated as random effects in the statistical analysis [13, 14]. Standard statistical software does not allow random effects in logistic regression, but using generalised linear latent and mixed models (GLLMM) [15, 16], it is possible to include random effects in a logistic regression model.

The GLLMM algorithm provides standard errors, and thus confidence intervals, for the coefficients in the regression equation. However, these cannot be directly

converted into standard errors or confidence intervals for the dose reductions computed by the non-linear Equations (5) and (8). This problem can be solved by applying the “delta method”, or method of statistical differentials, where an arbitrary random variable is Taylor-expanded around its mean [17].

Evaluation of two-dimensional adaptive filters in coronary CT angiography

In a previously published study [18], coronary CT angiography (CTA) images were acquired in 24 patients using standard settings with a quality reference mAs setting of 310 mAs, and reduced dose of 62 mAs, depending on the cardiac phase (cf. Figure 1). (The quality reference mAs is selected by the user and corresponds to a desired image quality for a reference adult with a body mass of 75 kg.) Single slices from the examinations at the level of the left main coronary artery (LMCA) were used. The reduced-dose images were viewed native and after post-processing with a two-dimensional (2D) adaptive filter (SharpView; Contextvision AB, Linköping, Sweden) [18, 19]. Nine blinded reviewers independently graded the images with respect to the following image quality criteria:

1. visually sharp reproduction of the thoracic aorta
2. visually sharp reproduction of the wall of the thoracic aorta
3. visually sharp reproduction of the heart
4. visually sharp reproduction of the LMCA
5. the image noise in relevant regions is sufficiently low for diagnosis.

The scale used for grading was as follows:

1. criterion is fulfilled
2. criterion is probably fulfilled
3. indecisive
4. criterion is probably not fulfilled
5. criterion is not fulfilled.

The VGR model defined by Equation (3) was applied to the data, and the analysis was performed in Stata v. 10.1 (Stata Corporation, College Station, TX), using GLLAMM, and treating patient and observer as random effects and log(mAs) and post-processing as fixed effects. This was achieved (for Criterion 1) with the following Stata command:

```
gllamm criterion1 logmas filter, link(ologit) i(observer_id patient_id)
```

Figure 2. Design of study of two-dimensional and three-dimensional filters in abdominal CT. Images of four kinds from 12 patients (P1–P12) were pair-wise evaluated by six radiologists (R1–R6), resulting in five comparison scores (S1–S5).

are negative, in accordance with the inverted direction of the scoring scale (with lower scores representing higher image quality). For all image quality criteria, a rather similar dependence on the mAs setting was found. However, the effect of the filtering varied, and tended to be smaller for sharp reproduction of the thoracic aorta and the LMCA than for the other criteria, although the confidence intervals were clearly overlapping.

Estimates of the potential reduction in mAs are also found in Table 1. The estimated reduction thus ranges from 16% for sharp reproduction of the thoracic aorta to 26% for sharp reproduction of the aortic wall. Again, the width of the confidence intervals informs the researcher of the statistical uncertainty of the results.

Using the estimated regression coefficients and Equation (3), the probability of obtaining a certain value of the assessment scores can be predicted also for values not actually studied. Figure 3 illustrates how the probability of obtaining a score of 1 or 2 for Criterion 4 (visually sharp reproduction of the LMCA) varies with the mAs setting, under the assumptions of the model. If an 80% probability of obtaining a score of 1 or 2 is accepted as a reasonable goal for image quality, then an mAs setting of at least 209 mAs would be needed without filtering and at least 165 mAs with filtering, in agreement with the reduction estimate of 21% in Table 1.

Evaluation of two-dimensional and three-dimensional filters in abdominal CT

VGR regression coefficients from the abdominal CT experiment are presented in Table 2. As the rating scale here was not inverted, positive values now indicate improved image quality. The absolute values of the

coefficients for dependence on the mAs setting are now higher than in Table 1, and vary more between the criteria. The strongest mAs dependence was found for image noise, and the weakest for delineation of the common bile duct. The effect of applying filtering was again strongest for image noise, and weakest for delineation of the common bile duct. Furthermore, it tended to be larger for the 3D filter than for the 2D filter.

With 2D filtering, an mAs reduction of 41% was estimated for image noise and 32–36% for the other criteria. For 3D filtering, the estimated reduction for image noise exceeded 50%, and for the other criteria was 42–43%.

Discussion

As pointed out above, visual grading experiments may be a good alternative to ROC studies, for example, in particular in the phase when imaging or post-processing procedures are being developed and preliminarily evaluated. The effort needed to organise and perform such a study is moderate, and for projection radiography visual grading data have been shown to correlate with physical measures of image quality as well as with diagnostic performance evaluated with methods based on ROC [21–23].

The present study shows that the parametric VGR models can be used not only to ascertain whether significant differences are present between different schemes for acquisition or post-processing, but also to obtain direct numerical estimates of how much one can expect the mAs setting, and thus the effective dose, to be reduced or increased if one is exchanged for another. This should be useful, particularly in the phase when protocols are optimised. The results suggest that in

Table 1. Estimated potential relative reduction in mAs setting when using 2D adaptive filter in coronary CT angiography

Image quality criterion	Regression coefficients according to Equation (3) (95% confidence limits)		Estimated mAs reduction according to Equation (5) (95% confidence limits)
	log(mAs) (a)	2D adaptive filter (b)	
1: Visually sharp reproduction of the thoracic aorta	−2.52 (−2.88; −2.16)	−0.45 (−0.78; −0.11)	16% (6%; 27%)
2: Visually sharp reproduction of the aortic wall	−2.53 (−2.82; −2.24)	−0.75 (−1.07; −0.44)	26% (17%; 34%)
3: Visually sharp reproduction of the heart	−2.54 (−2.91; −2.18)	−0.74 (−1.12; −0.36)	25% (15%; 36%)
4: Visually sharp reproduction of the LMCA	−2.52 (−2.81; −2.24)	−0.61 (−0.91; −0.30)	21% (13%; 30%)
5: Noise sufficiently low for diagnosis	−2.74 (−3.04; −2.44)	−0.77 (−1.07; −0.46)	24% (17%; 32%)

2D, two-dimensional; LMCA, left main coronary artery.

Figure 3. Predicted probability of obtaining a score of 1 or 2 for Criterion 4, “Visually sharp reproduction of the left main coronary artery”, at varying mAs settings for filtered and unfiltered images. (a) With linear axes, a curved relationship is found. (b) With logarithmical transformation of the horizontal axis and logit transformation of the vertical axis, the curves are transformed into two parallel lines. In both cases, mAs values required for reaching a probability of 80% for unfiltered images (209 mAs) and filtered images (165 mAs) are indicated by dashed vertical lines.

coronary CTA, the application of 2D adaptive filters to the acquired slices can be expected to permit investigators to reduce the effective dose by 20–25%. Some caution should be applied, though, when drawing conclusions from this study, since the low-dose and normal-dose images were acquired in different cardiac phases. Similarly, in abdominal CT, application of 2D or 3D adaptive filters to an image volume may be expected to result in dose reductions of about 30% and 40%, respectively.

The regression coefficients can also be given more intuitive interpretations. In the first example (Table 1), where the coefficients are negative since high image quality corresponds to low scoring values, a value of around -2.5 of the coefficient for $\log(\text{mAs})$ implies, according to Equation (3), that if the mAs setting is doubled, then the odds $[p/(1-p)]$ for a score below a certain threshold (indicating higher image quality) are multiplied by $2^{2.5} \approx 5.7$. A coefficient of -0.7 for the 2D filter, on the other hand, means that the addition of such a filter is expected to multiply the odds for a lower value by $\exp(0.7) \approx 2$.

The coefficient values in the second example (Table 2) have larger absolute values than in Table 1, which could be interpreted as a stronger dependence on mAs setting, as well as on filtering. A $\log(\text{mAs})$ coefficient of around 6 implies that a doubling of the mAs for one of the image stacks in the comparison would lead to a dramatic

increase by a factor of $2^6 = 64$ in the odds for a score favouring that stack. By the same token, coefficient values around 2.5 and 3 for 2D and 3D filters could be translated to the odds being multiplied by $\exp(2.5) \approx 12$ and $\exp(3) \approx 20$, respectively. Apart from all other differences in image acquisition and diagnostic task, we would like to point out the differing situations for the reviewers in single-image and image-pair experiments. The higher absolute values of the coefficient in the latter case might be an expression of a greater ability to detect differences in image quality with an image-pair experiment.

Our approach, which includes the identities of patients and reviewers in the same regression equation as acquisition and post-processing parameters, makes the statistical analysis somewhat challenging. The situation where both the patients and the reviewers are random samples from two larger populations, and thus should be treated as random effects [13], necessitates a more sophisticated statistical algorithm for the ordinal logistic regression than those available in standard software [14]. Our solution was to use the GLLMM method [15, 16]. An additional statistical challenge concerns the calculation of the confidence limits of the dose reduction estimates. However, this problem could easily be solved with the “delta method” [17], which is integrated in the Stata software.

The basic assumptions of our parametric statistical model may certainly be questioned. The most central

Table 2. Estimated potential relative reduction in mAs setting when using 2D and 3D adaptive filters in abdominal CT

Image quality criterion	Regression coefficients according to Equation (7) (95% confidence limits)			Estimated mAs reduction according to Equation (8) (95% confidence limits)	
	$\log(\text{mAs})$ (a)	2D adaptive filter (b_1)	3D adaptive filter (b_2)	2D adaptive filter	3D adaptive filter
1. Delineation of pancreas	5.79 (4.98; 6.59)	2.62 (2.18; 3.06)	3.17 (2.70; 3.06)	36% (33%; 40%)	42% (39%; 45%)
2. Delineation of veins in liver	5.64 (4.86; 6.43)	2.22 (1.82; 2.62)	3.19 (2.72; 3.66)	32% (29%; 36%)	43% (40%; 46%)
3. Delineation of the common bile duct	4.76 (4.06; 5.47)	2.07 (1.69; 2.46)	2.61 (2.18; 3.04)	35% (31%; 39%)	42% (38%; 46%)
4. Image noise	6.53 (5.72; 7.34)	3.40 (2.91; 3.89)	4.64 (4.06; 5.21)	41% (38%; 43%)	51% (48%; 54%)
5. Overall diagnostic acceptability	6.18 (5.36; 7.00)	2.59 (2.18; 3.00)	3.48 (2.99; 3.97)	34% (31%; 37%)	43% (40%; 46%)

2D, two-dimensional; 3D, three-dimensional.

concept in ordinal logistic regression is the “proportional odds” or “parallel regression” assumption, which in our case states that the different cumulative scores for situations with different post-processing correspond to parallel lines in a graph with the axes transformed as in Figure 3b. Although not carried out in this study, one could, in principle, test the validity of this assumption by designing a more complex study including more mAs levels, and then applying a test designed specifically for testing the “proportional odds assumption”, such as Brant’s test [24]. With the limited empirical data available in our two examples, however, this type of test is hardly meaningful.

We would also like to point out that the linear relationship between log(mAs) and the log odds postulated by the model can be expected to hold within only a limited interval. Interpolation between actually studied mAs values therefore seems safer than extrapolation outside the studied interval.

Regardless of the formal validity of the model, the estimates of potential dose reduction produced by our method should be confirmed by experiments where the predicted dose reduction is actually tested and evaluated with either visual grading or ROC. The method we suggest may therefore find its greatest use in pilot studies when the degree of dose reduction attainable by a certain measure needs to be estimated before the main study testing the diagnostic value is designed.

The approach used here could possibly be generalised to handle other situations where imaging procedures are optimised. In particular, in MRI, a multitude of acquisition parameters need to be selected simultaneously. The balance between image quality and acquisition time is often analogous to that between image quality and effective dose in CT, and similar reasoning to that in conjunction with Figure 3 might be attempted. However, the dependence of image quality on a certain parameter (e.g. the flip angle) need not be monotonic, and one could envisage a situation in which a VGR model including a quadratic term might help in the optimal selection of such a parameter.

In conclusion, we have in this article proposed a new way of analysing visual grading experiments with VGR that results in direct estimates of the potential for dose reduction, and which seems to be useful particularly for guiding researchers in designing subsequent studies evaluating diagnostic value.

Acknowledgments

This paper was presented at the MIPS XIV conference, Dublin, Ireland, in August 2011.

The authors are grateful for assistance with the post-processing provided by Contextvision AB, Linköping, Sweden.

References

- Månsson LG. Methods for the evaluation of image quality: a review. *Radiat Prot Dosimetry* 2000;90:89–99.
- European Communities. European guidelines on quality criteria for computed tomography. Report EUR 16262 EN. Luxembourg, Luxembourg: Office for Official Publications of the European Communities; 2000.
- Gordon R. A tutorial on ART (algebraic reconstruction techniques). *IEEE Trans Nucl Sci* 1974;NS-21:78–93.
- Herman GT, Lent A. Iterative reconstruction algorithms. *Comput Biol Med* 1976;6:273–94.
- Beekman FJ, Kamphuis C. Ordered subset reconstruction for X-ray CT. *Phys Med Biol* 2001;46:1835–44.
- Silva AC, Lawder HJ, Hara A, Kujak J, Pavlicek W. Innovations in CT dose reduction strategy: application of the adaptive statistical iterative reconstruction algorithm. *AJR Am J Roentgenol* 2010;194:191–9.
- Pontana F, Pagniez J, Flohr T, Faivre JB, Duhamel A, Remy J, et al. Chest computed tomography using iterative reconstruction vs filtered back projection (Part 1): Evaluation of image noise reduction in 32 patients. *Eur Radiol* 2011;21:627–35.
- Kachelriess M, Watzke O, Kalender WA. Generalized multi-dimensional adaptive filtering for conventional and spiral single-slice, multi-slice, and cone-beam CT. *Med Phys* 2001;28:475–90.
- Båth M, Månsson LG. Visual grading characteristics (VGC) analysis: a non-parametric rank-invariant statistical method for image quality evaluation. *Br J Radiol* 2007;80:169–76.
- Svensson E. Ordinal invariant measures for individual and group changes in ordered categorical data. *Stat Med* 1998;17:2923–36.
- Ledenius K, Svensson E, Stålhammar F, Wiklund LM, Thilander-Klang A. A method to analyse observer disagreement in visual grading studies: example of assessed image quality in paediatric cerebral multidetector CT images. *Br J Radiol* 2010;83:604–11.
- Smedby Ö, Fredrikson M. Visual grading regression: analysing data from visual grading experiments with regression models. *Br J Radiol* 2010;83:767–75.
- Snijders TAB. Fixed and random effects. In: Everitt BS, Howell DC, eds. *Encyclopedia of statistics in behavioral science*. Chichester, UK: Wiley; 2005. pp. 664–5.
- Smedby Ö, Fredrikson M, De Geer J, Sandborg M. Visual grading regression with random effects. *Proc SPIE Med Imaging* 2012;0001:831805.
- Rabe-Hesketh S, Pickles A, Skrondal A. GLLAMM: A general class of multilevel models and a Stata program. *Multilevel Modelling Newsletter* 2001;13:17–23.
- Skrondal A, Rabe-Hesketh S. Multilevel logistic regression for polytomous data and rankings. *Psychometrika* 2003;68:267–87.
- Oehlert GW. A note on the delta method. *Am Statistician* 1992;46:27–9.
- De Geer J, Smedby Ö, Sandborg M, Persson A. The efficacy of 2D, non-linear noise reduction filtering in cardiac imaging: a pilot study. *Acta Radiologica* 2011;52:716–22.
- Granlund GH, Knutsson H. *Signal processing for computer vision*. Dordrecht, The Netherlands: Kluwer Academic Publishers; 1995.
- Borgen L, Kalra MK, Lærum F, Hachette IW, Fredriksson CH, Sandborg M, et al. Application of adaptive nonlinear 2D and 3D post-processing filters on reduced dose abdominal CT. *Acta Radiologica* 2012;53:335–42.
- Sandborg M, Tingberg A, Dance DR, Lanhede B, Almen A, McVey G, et al. Demonstration of correlations between clinical and physical image quality measures in chest and lumbar spine screen-film radiography. *Br J Radiol* 2001;74:520–8.
- Herrmann C, Sund P, Tingberg A, Keddache S, Månsson L, Almén A, et al. Comparison of two methods for evaluating image quality of chest radiographs. *Proc SPIE Med Imaging* 2000;3981:251–7.
- Tingberg A, Båth M, Håkansson M, Medin J, Besjakov J, Sandborg M, et al. Evaluation of image quality of lumbar spine images: a comparison between FFE and VGA. *Radiat Prot Dosimetry* 2005;114:53–61.
- Brant R. Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics* 1990;46:1171–8.