Diagnostic reference levels in interventional radiology

2000 Phys. Med. Biol. 45 3833

(http://iopscience.iop.org/0031-9155/45/12/323)

Phys. Med. Biol. **45** (2000) 3833–3846. Printed in the UK PII: S0031-9155(00)16335-3

Diagnostic reference levels in interventional radiology

N W Marshall, C-L Chapple and C J Kotre

Regional Medical Physics Department, Newcastle General Hospital, Westgate Road, Newcastle upon Tyne NE4 6BE, UK

Received 14 August 2000

Abstract. Following the release of European Directive EU 97/43, radiodiagnostic facilities within the European Union are required to implement a system of patient dose reviews based on comparisons with European, national and local diagnostic reference levels (DRLs). Establishing these levels for typical interventional radiology examinations presents a problem as definition of 'typical' examinations can be difficult, patient numbers are limited and these procedures are often performed at a few specialist centres.

This paper uses dose–area product (DAP) gathered over a period of 3 years from 40 fluoroscopy rooms to investigate potential difficulties when it comes to forming diagnostic reference levels for interventional radiology. Comparison of DAP distributions with standard complex (fluoroscopy based) examinations such as barium enema reveals considerably more variation for interventional procedures. Two methods of forming a DRL are compared: pooled patient DAP distributions versus a distribution of DAP per room. The bootstrap resampling method is then applied to DAP distributions to form a confidence interval for the chosen DRL statistic. Potential error on a DRL formed at a local level from a limited number of patient dose readings and x-ray rooms is significant. The results are reviewed in the wider context of DRLs in general radiology. For complex examinations, it is suggested that the function of the DRL is best served by setting DRLs based on pooled size-corrected patient DAP distributions rather than distributions of average DAP per room.

1. Introduction

Various studies have surveyed patient doses from diagnostic radiology since the 1950s (Adrian Committee 1960, Johnson and Goetz 1986, Shrimpton *et al* 1986). Following the introduction of more complicated radiological techniques, monitoring has been extended to include modalities such as computed tomography (CT) (NRPB 1993) and interventional procedures (Vañó *et al* 1995, McParland 1998a, b). Past results from these surveys often form guideline doses for common diagnostic radiology examinations (Shrimpton *et al* 1986). This tendency will no doubt be reinforced by the recent European Medical Exposure Directive 97/43/EURATOM (European Commission 1997), which requires the establishment and use of diagnostic reference levels (DRLs). With the recent enactment of this Directive by European Union (EU) member states, the formal establishment of diagnostic reference levels for a range of diagnostic radiology examinations becomes a priority. Some historical perspective on the development of reference doses can be found in the recent paper by Wall and Shrimpton (1998) while advice on the formulation, use and frequency of revision of DRLs is provided in a recent ICRP publication (ICRP 1996). For example, it is stated that the initial value of the DRL might be a percentile point on the patient dose distribution for a given examination (ICRP 1996).

The intended use of the DRL will obviously affect the definition employed. Article 2 of the Directive 97/43/EURATOM defines DRLs as 'dose levels in medical radiodiagnostic

0031-9155/00/123833+14$30.00

Table 1. Size-corrected and weight selected (60 kg to 80 kg) DAP for ERCP procedures at seven centres.

	Size-corrected DAP (Gy cm^2)	Weight-selected DAP (Gy cm^2)
Number of patients	1736	937
Mean	14.5	13.6
1st quartile	4.23	4.30
2nd quartile	8.62	8.47
3rd quartile	16.4	16.2
Standard deviation	22.2	16.0

practices . . . for typical examinations for groups of standard-sized patients or standard phantoms for broadly defined types of equipment. These levels are expected not to be exceeded for standard procedures when good and normal practice regarding diagnostic and technical performance is applied.' While assessment of patient dose is relatively straightforward for simple radiographic examinations, it is less clear whether the DRL concept is applicable to more complex examinations such as interventional radiology procedures. Applying the EU Directive to interventional radiology forces us to consider the following points:

(a) Whether adequate definition of typical interventional radiology examinations is possible.
(b) The relative merits of size correction techniques and weight banding and their effects on patient sample size in interventional radiology.
(c) The type of distribution from which the DRL is taken.
(d) The number of patients required to define a DRL and the associated error on the statistic.

This paper discusses these points, beginning with problems of size correction and classification of examinations. Dose–area product (DAP) data for seven groups of interventional radiology procedures are presented and used to illustrate two different techniques of setting the DRL. The bootstrap resampling method (Efron and Tibshirani 1986) is then used to provide confidence intervals for various statistics of the dose distribution as a function of local sample size.

2. DAP data for interventional radiology

Several papers have reported DAP and entrance surface dose (ESD) data for interventional radiology procedures (Vañó *et al* 1995, Williams 1997, McParland 1998a, b). Dose–area product is suitable for monitoring patient doses in interventional radiology as the x-ray field size and focus–skin distance are subject to frequent change. Examination categories are kept fairly broad in the current study; data are listed for the more common procedures where sample sizes are reasonably large. The DAP data presented here were taken from an ongoing patient dose monitoring programme in the north of England. Forty fluoroscopy rooms were monitored using calibrated DAP chamber/meters, although not all the rooms are used for interventional radiology procedures. The method of collation dose records is detailed elsewhere (Broadhead *et al* 1995). Data were collected over a period of 3 years and corrected for patient height and weight using an established technique (Chapple *et al* 1995).

2.1. Size-corrected DAP for interventional radiology procedures

The principal aim of size correction and weight banding is to reduce the dispersion in the data due simply to spread in patient weight. In order to check the validity of the method

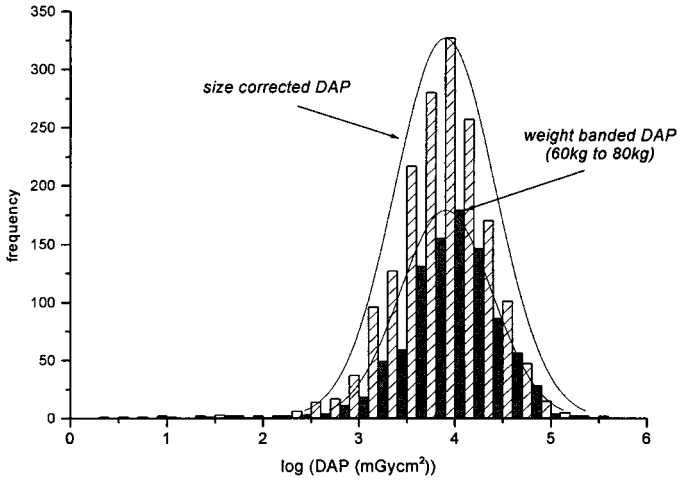

Figure 1. Size correction versus weight banding for ERCP procedures.

Table 2. (*a*) Size corrected dose–area product for seven groups of interventional radiology procedures (statistics are for full distributions).

	Angioplasty	Biliary intervention	Dilatation	Drainage	Embolization	ERCP	Stenting
n	766	153	139	234	128	1736	279
Mean	19.5	54.0	17.5	35.9	114	14.5	47.9
Median	11.6	38.0	5.9	15.9	82.1	8.6	30.3
3rd quartile	24.8	63.7	17.9	34.7	158	16.4	56.1
Standard deviation	24.1	70.1	30.5	58.8	104	22.2	59.9

Table 2. (*b*) Size corrected dose–area product for seven groups of interventional radiology procedures (statistics for mean DAP per room).

	Angioplasty	Biliary intervention	Dilatation	Drainage	Embolization	ERCP	Stenting
Number of rooms	11	6	6	8	5	7	7
Mean	21.1	51.7	19.5	26.4	85.9	13.9	42.6
Median	19.3	51.4	10.8	16.7	63.3	13.2	38.6
3rd quartile	23.9	70.5	17.4	19.8	123	15.5	47.1
Standard deviation	10.1	27.3	21.7	28.5	59.7	4.75	25.6

of size correction, both the size correction and weight banding (60 to 80 kg) methods were applied to the same set of DAP readings for ERCP procedures performed at seven centres. Table 1 shows the results of this analysis; the statistics are virtually identical although weight banding appears to reduce standard deviation of the data more successfully. An independent two-tailed t-test on the log-transformed DAP distributions indicated that the means of the two distributions were not significantly different at the 0.05 level ($p = 0.63$) and the two data sets are seen to overlap each other when plotted as histograms (figure 1). Size-corrected data are used throughout this work, as the weight banding method reduced the number of DAP results by factor of approximately 2.

Table 2(*a*) presents the basic size-corrected DAP data for seven types of interventional radiology procedures. These statistics were calculated from full patient DAP distributions,

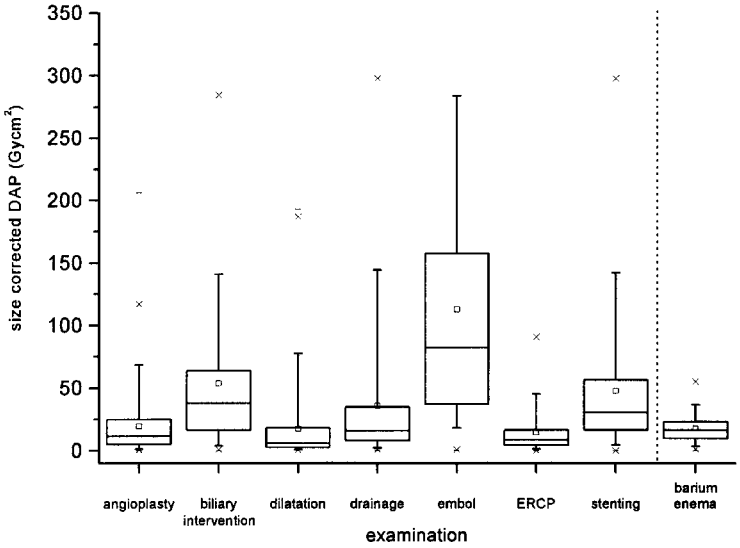

Figure 2. Box plots of size-corrected DAP for seven interventional radiology procedures along with results for 12 577 barium enema studies collected from eight x-ray rooms. (The central box marks the interquartile range, the single line shows the median, the single point (□) shows the mean and lines extending either side of the box indicate 5% and 95% points of distribution. The bottom two symbols mark the zeroth and first percentiles while the top two symbols mark 99th and 100th percentiles.)

for example angioplasty results for individual patients from 11 rooms were pooled to form one angioplasty data set. A graphical version of these data sets, in the form of box plots, is given in figure 2. Also shown in this figure is a plot for barium enema examinations in which 12 577 DAP measurements were pooled from eight x-ray rooms (mean DAP = 17.7 Gy cm^2). While the box plot for the barium enema examination reveals a reasonably symmetrical DAP distribution, the corresponding plots for the angioplasty procedures are far from symmetrical. Principal characteristics of these distributions are the long high-dose tails, indicating skewed distributions, together with a comparatively large interquartile range. As a consequence of this skew, the distribution means are greater than the medians. While DAP is not necessarily high for interventional radiology procedures, all the plots indicate considerable spread within the DAP data.

Figure 3 illustrates the greater spread often seen in interventional radiology DAP distributions. Box plots for angioplasty procedures from 11 rooms are shown in figure 3(*a*), while figure 3(*b*) shows box plots for barium enema examinations performed in eight rooms. Inter-room variability in the mean or median is similar for both these examinations, yet spread in DAP at a given centre is far greater for the angioplasty examination. Both DAP distributions are subject to the same sources of variation, such as patient weight (although size correction helps to reduce this), intercase complexity, differences in examination protocol between centres, x-ray system specification/performance and radiological technique (Warren-Forward *et al* 1998). The interventional radiology data in this study are subject to the additional uncertainty of whether the examination was appropriately classified before being recorded in the database. From these plots, it would appear that differences in intercase complexity are far greater for angioplasty examinations (and interventional radiology procedures in general).

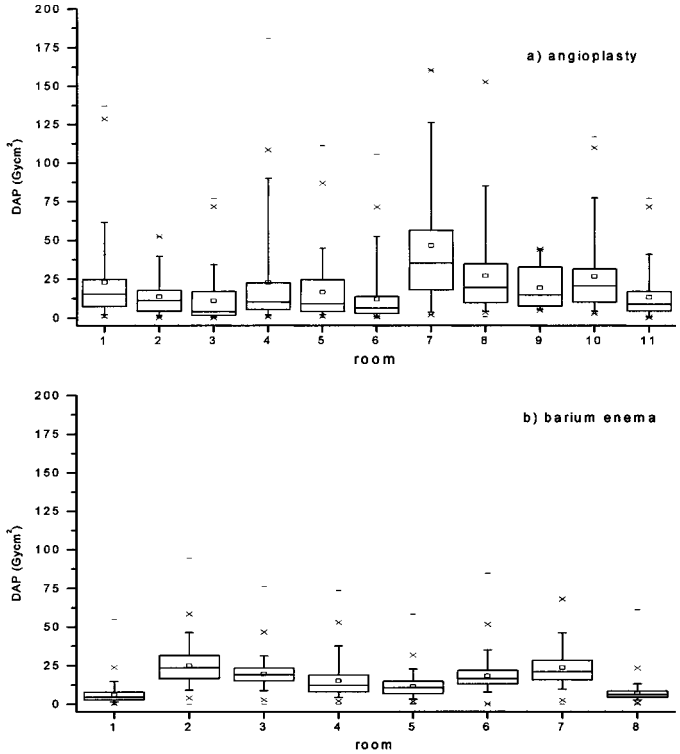

Figure 3. Box plots of size-corrected DAP for (*a*) angioplasty studies performed at 11 centres and (*b*) barium enema studies from eight centres.

2.2. Classification of procedures

A crucial phrase in the EU Directive definition is 'typical examination'. The DRL concept should work well for simple radiographic examinations such as thorax, abdomen and extremity radiographs, where the examinations can be well defined. Even the more complicated studies involving fluoroscopy such as barium contrast are amenable to definition (Hart *et al* 1996), yet applying the DRL concept to the area of interventional radiology raises several problems. Procedures are often non-standard, and complications can arise as a result of the involved and complex nature of the examinations. Because an aim of radiological intervention is to treat the patient (rather than just diagnose any problem which may be present) these cases are often clinically open ended, carrying on until the treatment is completed.

The problem of classification is illustrated using data from stenting procedures; figure 4(*a*) shows stent DAP results from six centres. Again, considerable variation is seen for stenting examinations, due in part to the following reasons. The condition of the vessel wall greatly affects the ease with which guide wires can be manipulated. Both vessel tortuosity and degree of plaque on the vessel wall add to the variability between cases. Position in the body, and associated problems of access, are further variables. Access to the vessel may be through a narrowing, the extent of which can vary considerably. Finally, case-specific problems, such as a tumour pressing on the vessel for example, may cause further variation. Protocols differ between centres with some units refusing to perform particularly difficult cases, preferring instead to send the patient for surgery or to a centre where the interventional radiology procedure is performed.

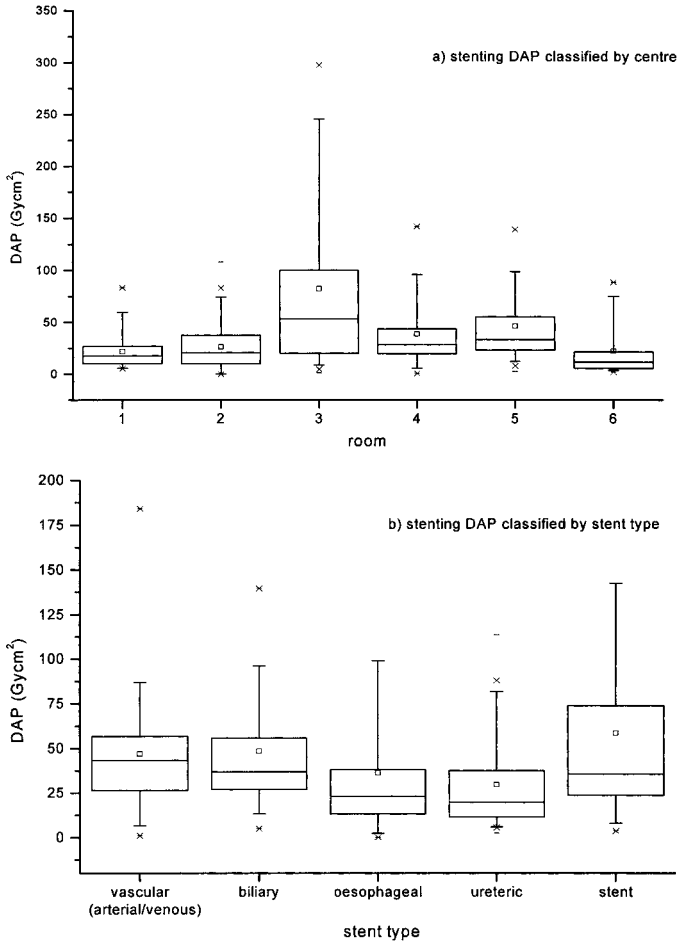

Figure 4. Box plots showing size-corrected DAP for stenting procedures (*a*) classified by x-ray room and (*b*) classified by stent type (unspecified stenting procedures are collected together in the group stent).

Returning to the stent results in figure 4(*a*), it can be seen that patient DAP is higher at centre 3; however, we cannot be sure if this is due to poorer x-ray equipment/radiological technique or whether more difficult stenting procedures are performed at this centre. To try and answer this, stent DAP was reclassified by stent type for five distinct categories plus a miscellaneous 'stent' group, for unspecified stenting procedures (figure 4(*b*)). Using a simple classification based on ease of access, tortuosity etc, we might expect DAP to be highest for vascular stents and lowest for ureteric and oesophageal stents. Although there is slight evidence for this in figure 4(*b*), significant differences are washed out due to the large variation within each stent group. Analysis of fluoroscopy time also showed little difference between the different stenting classes.

Two competing effects must be balanced when deciding upon examination grouping. Broad classes of procedure will produce large sample sizes and hence tend to reduce standard error on the mean. However, using broad categories (and/or inappropriate grouping) will tend to increase spread in the data and hence increase standard error on the mean (sample sizes being equal).

3. Establishing diagnostic reference levels

Current UK National Reference Levels for both simple radiographic projections and for more involved examinations such as barium contrast studies were taken from an extensive survey conducted by the NRPB in the 1980s (Shrimpton *et al* 1986). A frequency distribution of ESD per film for each radiograph type was plotted and the 75th percentile (third quartile) of this distribution was adopted as the DRL. While arbitrary, the use of the third quartile represents a pragmatic approach and is useful for isolating rooms which may have a problem either of radiographic technique or poor/faulty equipment. There are problems, however, with using a population-based parameter as a reference level and other parameters (for example perhaps calculated from a knowledge of x-ray techniques for simple examinations) should not be ruled out.

When assessing local performance, mean dose from a group of approximately 10 patients with a mean weight close to 70 kg can be compared with the DRL (ESD per radiograph) (IPSM 1992, Wall and Shrimpton 1998). The more recent survey by Hart *et al* (1996) used an alternative method of analysing dose data in which mean ESD for a typical patient (70 kg) was calculated from each set of patient dose measurements for a given x-ray room. Taking a third quartile from this type of distribution produces a DRL in the form of a mean dose per room for a given examination type. Hence mean room dose from a local survey can be compared with a similar DRL (i.e. average dose per room).

3.1. Individual patient dose versus mean room dose distributions

If forced to set a local DRL (in the absence of any national level) then two different approaches are available. Table 2(*a*) shows distribution statistics calculated from a dataset in which all the DAP results for individual patients from a number of x-ray rooms have been pooled and the statistics taken from this one dataset. Alternatively, DAP data can be averaged for a room and statistics taken from the resulting distribution of room means (table 2(*b*)). If the distribution and number of DAP values is similar for each room then we would expect these two methods to give similar results. For the data sets presented here, the mean and third quartile agree closely, although the median values differ significantly. Not surprisingly, agreement is poorest for embolization procedures where there are data from just five rooms.

The two methods generate distinctly different frequency distributions. If data from only a limited number of x-ray rooms are available, the distribution of mean room DAP will be poorly defined. Figure 5(*a*) illustrates this for drainage procedures from eight rooms. An advantage of this type of distribution is that poorly performing rooms are easily identified. Once the DAP measurements have been averaged and included in the distribution then all rooms carry the same statistical weight, i.e. they are just one point on the histogram. This technique is suitable for setting reference levels for common examinations performed on widely available x-ray equipment, i.e. we have data from many rooms. However, problems are encountered for the interventional radiology studies presented in this work where the number of x-ray rooms for special procedures is limited and there is significant case to case variation.

Plotting a distribution of drainage DAP data for individual patients pooled from seven rooms (figure 5(*b*)) produces a well defined (if highly skewed) histogram, although this method is not free from problems either. First, statistical parameters taken from this distribution will be biased towards results from rooms where the procedure of interest is most frequently performed. Second, the DRL may be self-referential if an insufficient number of x-ray rooms are used to form the distribution. However, this type of distribution makes efficient use of the available data and should be more stable, given the far greater number of data points in the histogram.

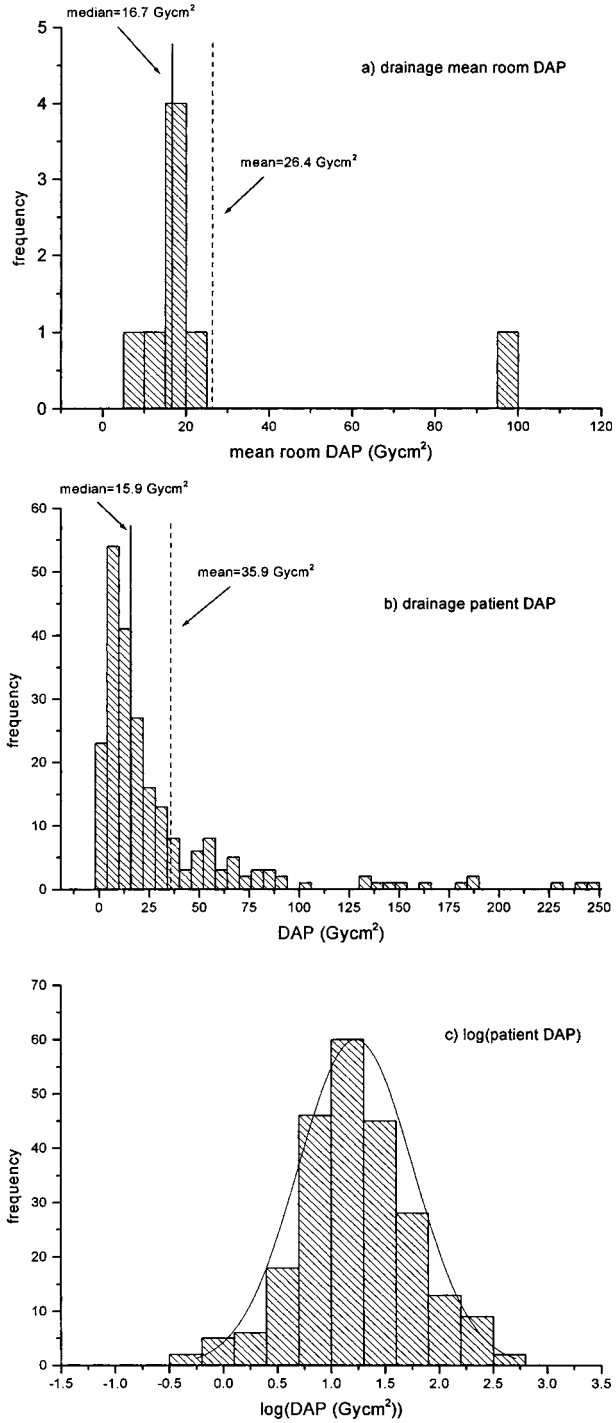

Figure 5. Histograms of size-corrected DAP for drainage procedures: (*a*) distribution as function of mean DAP for a room; (*b*) all DAP data pooled and plotted as patient DAP histogram; (*c*) logarithmic transformation of patient DAP distribution for ERCP studies.

Unless data from a reasonably large number of x-ray rooms are available, statistics taken from patient DAP distributions are probably the most reliable means of producing a DRL at a local level for complex examinations.

3.2. Uncertainty on the DRL statistic

The simplest option is to use national survey data when choosing a DRL (derived from a large number of x-ray units or patient examinations); however, data gathered locally may have to be used if national levels are unavailable. The parameter used as the DRL will suffer from considerable uncertainty if only a limited sample of the underlying population is available. When checking room performance, we must be sure that patient doses in the local sample are significantly greater (or not) than the DRL To do this consistently, whilst avoiding false alarms, requires an estimate of uncertainty on the DRL statistic itself.

3.2.1. Standard parametric technique. For normally distributed data, the standard error on the mean gives the uncertainty on the sample. A confidence interval may then be derived, marking the interval (on the measurement scale) in which the sample mean lies with probability p. Using the upper confidence limit on the DRL (instead of the just the DRL parameter itself) will guard against false alarms due to sampling limitations when setting up the DRL. Logarithmic transformation of the DAP distribution produces reasonably normal distributions for all the interventional radiology procedures listed in table 2(*a*). As an example, figure 5(*c*) presents the log-transformed ERCP distribution. Once normalized using the log-transform, a confidence interval can be constructed for the log-normal mean. This procedure is not free from difficulty, as the following analysis taken from Zhao and Gao (1997) demonstrates. If x is a log-normally distributed variable then the variable $y = \ln x$ is normally distributed with mean μ and variance σ^2. The mean of x, θ, will be given by:

$$\theta = \mathrm{e}^{(\mu+\sigma^2/2)}.$$

The mean of the log-transformed data is simply

$$\bar{y} = \sum_{i=1}^{n} \frac{y_i}{n}$$

with variance, s

$$s^2 = \sum_{i=1}^{n} \frac{(y_i - \bar{y})^2}{n-1}.$$

The obvious way of obtaining the (two-sided) confidence interval for the mean (θ) is to find the confidence interval for μ as

$$\bar{y} \pm Z_{(1-\alpha/2)} \frac{s}{\sqrt{n}}$$

where Z_α is the αth percentile of the standard normal distribution and n is the number of observations. Hence the antilogarithm of these limits gives the confidence interval for θ on the original scale:

$$\exp\left(\bar{y} \pm Z_{(1-\alpha/2)} \frac{s}{\sqrt{n}}\right).$$

This technique is often referred to as the naive method and is known to give incorrect results for distributions with large σ. Zhao and Gao (1997) describe an alternative method, due to Cox:

$$\exp\left[\bar{y} + \frac{s^2}{2} \pm Z_{1-\alpha/2}\sqrt{\left(\frac{s^2}{n} + \frac{s^4}{2(n-1)}\right)}\right].$$

Table 3. 90% two-sided confidence limits on the mean for stenting examination DAP.

	Parent distribution	Naive method	Cox's method	Bootstrap
Lower confidence limit	—	23.6	33.0	42.7
Mean	47.9	27.4	38.7	47.6
Upper confidence limit	—	31.2	44.3	52.5

Table 3 lists 90% two-sided confidence limits for the stenting DAP distribution using these two methods. While Cox's method is superior to the naive method, neither technique estimates the mean particularly well, implying that the untransformed stenting data are probably not log-normally distributed.

A more fundamental problem with the use of (parametric) confidence intervals is that the confidence limits are formed around the mean. Should we wish to define the DRL using an alternative statistic then we have to resort to the method of boostrap resampling (Efron and Tibshirani 1986) to provide confidence limits.

3.2.2. Bootstrap resampling. Figure 6(*a*) shows a patient DAP distribution for 766 angioplasty procedures performed at 11 centres. Although DAP results for a reasonable number of patients, x-ray systems, operators etc make up this distribution, it is ultimately a limited subset of the ideal angioplasty population. We take this to be the parent distribution D_{parent} for angioplasty examinations in the north of England and use a parameter (a_0, say) from this distribution to define a DRL. However, when there is only one such DAP distribution then estimating the accuracy of a_0 presents a problem. The bootstrap resampling method can be used to provide some idea of the error in the parameter when there is only one distribution available (Efron and Tibshirani 1986).

This technique is essentially a Monte Carlo method in which the parent distribution is randomly resampled (Press *et al* 1992). If the original data set has N data points, then n points are taken randomly (with replacement) from D_{parent} to produce data sets D_1, D_2, ... (subsamples). Every subsample will generate a different value for the parameter of interest i.e. a_1, a_2, ... and these will be distributed in some way around a_0. This technique assumes the data points in D_{parent} are independent and identically distributed (i.e. the sequential order of the points does not affect the process used to obtain the parameter a_0, for example addition for the mean or sorting for a percentile point).

The third quartile was taken to be the statistic of interest (Shrimpton *et al* 1986) for the DRL. Mean (x) and median (m) were also studied because of their use as estimators of the local sample when assessing local performance. Median for a sample of n values was defined as the $\frac{1}{2}(n+1)$th largest value, while third quartile was taken to be the $\frac{3}{4}(n+1)$th largest value.

The following procedure was then used to calculate 90% confidence limits on these statistics. One thousand subsamples were drawn from D_{parent} and the mean, median and third quartile calculated for each subsample. Each set of these parameters was then ordered and a two-sided 90% confidence interval calculated using the 5th and 95th percentiles, i.e. there is a 90% probably that the statistic of interest lies within this interval (Iwi *et al* 1999). Confidence intervals on the third quartile for the seven interventional radiology procedures are listed in table 4, along with subsample size used when drawing the samples from D_{parent}. Using the bold figures in this table (the upper confidence limit) instead of the third quartile gives a safer DRL, with less likelihood of any false alarms.

The intervals presented in this table reflect inherent spread in the parent distributions. ERCP, angioplasty, and dilatation distributions are all strongly skewed but are reasonably

Table 4. 90% two-sided bootstrap confidence limits for the third-quartile DAP for seven groups of interventional radiology examinations generated using 1000 resamples of the parent distribution.

Examination	Subsample size	Lower confidence limit (Gy cm^2)	Mean of third quartiles (Gy cm^2)	Upper confidence limit (Gy cm^2)
Angioplasty	500	21.9	24.8	**27.1**
Biliary intervention	100	53.7	66.2	**84.1**
Dilatation	100	12.2	18.1	**24.5**
Drainage	100	25.4	35.9	**50.7**
Embolization	100	127	160	**203**
ERCP	500	14.8	16.2	**17.9**
Stenting	250	50.1	55.7	**63.7**

Table 5. 90% two-sided confidence intervals (Gy cm^2) for the third-quartile DAP for seven interventional radiology examinations generated using 1000 resamples of the parent distribution (presented to two significant figures).

Subsample size	Angioplasty	Biliary intervention	Dilatation	Drainage	Embolization	ERCP	Stenting
5	9.8–40	29–170	5.1–73	13–140	70–290	7.5–47	27–150
10	7.6–30	24–82	3.3–25	11–49	61–200	6.2–20	22–66
20	13–37	41–94	7.3–29	19–64	89–220	9.4–25	34–82
50	16–31	48–82	9.1–25	21–51	110–200	11–21	38–68
100	19–31	54–84	12–25	25–51	130–200	13–20	46–68
250	21–28	—	—	—	—	14–19	50–64
500	22–27	—	—	—	—	15–18	—

smooth, with reasonable numbers of data taken from at least seven rooms. Consequently, large subsample sizes can be used and the confidence interval around the third quartile is fairly small. In contrast, the large confidence interval seen for embolization examinations is probably due to the parent distribution coming from just five rooms and hence being a rather limited and uneven sample of the full embolization population.

Subsample size (n) was varied in order to study the effect of local sample size on confidence interval. Figure 6(a) compares subsample third quartiles with the parent distribution third quartile for the angioplasty distribution. The extreme case of using just five samples to define the third quartile leads to significant error in the DRL, hence the broad confidence interval ($\sim$30 Gy cm^2) in table 5. All examinations have wide confidence intervals and these fall slowly as a function of subsample size, implying that a reasonable number of patient DAP data are required to establish a DRL for interventional radiology procedures. Even well defined distributions such as angioplasty and ERCP require at least 100 samples for a confidence interval of 10 Gy cm^2 or less. Other procedures, embolization or biliary intervention for example, have confidence intervals of approximately 70 and 30 Gy cm^2 respectively for a sample size of 100. These are extremely wide limits, considering that the size-corrected mean DAP for a barium enema examination is approximately 18 Gy cm^2.

Ultimately, the size of the confidence interval will be governed by the standard deviation of the parent distribution. This is demonstrated in figure 6(b), which shows that a large spread in the parent distribution leads to a correspondingly large confidence interval on a given estimator. Conversely, estimators describing less complex examinations (exhibiting less intercase variability) should have smaller confidence intervals.

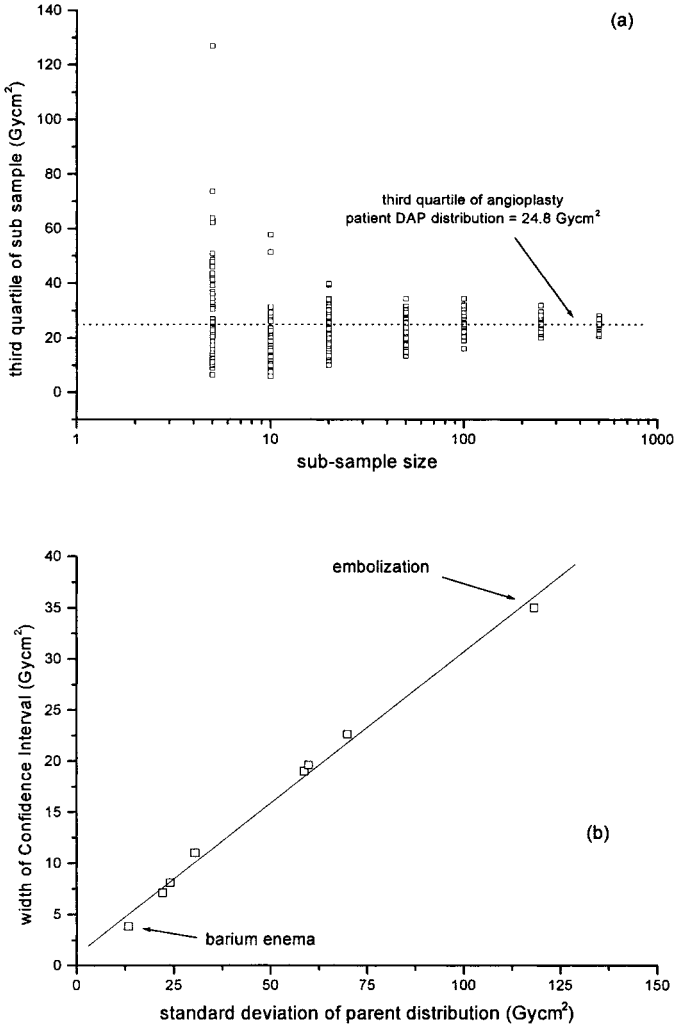

Figure 6. (*a*) Comparison of subsample third quartiles with parent distribution third quartile as a function of subsample size for angioplasty studies (100 bootstrap replications for each subsample size). (*b*) Two-sided 90% confidence interval as a function of standard deviation of parent distribution (subsample size = 100).

4. Conclusions

Definition of interventional radiological procedures is possible for the purposes of monitoring DAP, but variation between cases is extremely large for most of these examinations. Two different approaches can be used to produce DAP distributions for interventional examinations. Patient DAP data from various centres may be pooled to form one dataset or a histogram of mean patient DAP for a given room can be plotted. Taking the DRL from a histogram of patient DAP results, pooled from a number of rooms, makes efficient use of patient dose data, whereas the use of a mean room DAP distribution is suitable for simpler examinations where data from many x-ray rooms are available.

Taking a broader view of radiological procedures, from simple single radiographic views to complex interventional procedures such as those analysed here, it could be argued that the purpose of the DRL and hence its method of calculation should vary depending on the complexity of the examination. For simple film-screen examinations, variation in entrance surface dose on a given x-ray unit would be expected to be almost entirely due to differences in patient size. Variations arising from operator choices will be small, especially when examination protocols are in place. The purpose of the DRL in this case is to identify x-ray units and image receptors for which the standard patient dose, measured on a sampled basis for a small number of patients (IPSM 1992), lies at the upper end of the population of standard patient doses measured across a large number of x-ray units. For more complex examinations involving significant fluoroscopic contributions, however, the variation in dose (i.e. DAP) will tend to become dominated by operator choices and difficulty of individual cases. Interventional radiology represents the extreme case of this type of examination. For these procedures, DRLs based on pooled, size-corrected patient DAP can be used to identify routine high doses, including those due to operator, case profile and equipment effects.

Should a confidence interval be required for the DRL statistic, parametric techniques (in terms of logarithmic transformation of the data) can be difficult to apply due to the large skew and standard deviation of DAP data sets. Bootstrap resampling, on the other hand, allowed the generation of confidence intervals for these skewed data sets, for statistics other than the mean. The confidence interval fell slowly as function of sample size for interventional radiology examinations; approximately 100 patient DAP values (from several x-ray rooms) are needed for a reasonably well defined DRL. Instead of simply taking the 75th percentile as the reference level, we suggest that the upper confidence limit on the 75th percentile be used instead. This should, to some extent, allow for uncertainties inherent in the 75th percentile due to limitations of sampling.

While there is no intrinsic reason to reject the DRL concept for interventional radiology procedures, consideration of the errors arising from the skewed and disperse DAP distributions which are common in interventional radiology implies that a degree of caution should be exercised when examining or criticising performance of interventional radiology at a specific clinical centre.

Acknowledgments

We would like thank George Cowley of the Sunderland Royal Hospital X-ray Department and Jill Hetherington and Dr Roger Harrison of the Regional Medical Physics Department for their advice.

References

Adrian Committee 1960 *Radiological Hazards to Patients. Second Report of the Committee* (London: HMSO)

Broadhead D A, Chapple C-L and Faulkner K 1995 The impact of digital imaging on patient doses during barium studies *Br. J. Radiol.* **68** 992–6

Chapple C-L, Broadhead D A and Faulkner K 1995 A phantom based method for deriving typical patient doses from measurements of dose–area product on populations of patients *Br. J. Radiol.* **68** 1083–6

Efron B and Tibshirani R J 1986 Bootstrap methods for standard error, confidence intervals, and other methods of statistical accuracy *Statist. Sci.* **1** 54–75

European Commission 1997 Council Directive 97/43/EURATOM of June 30th 1997 on health protection of individuals against the dangers of ionising radiation in relation to medical exposure, and repealing Directive 84/466/Euratom *Official J. Eur. Commun.* no L 180/22

Hart D, Hillier M C, Wall B F, Shrimpton P C and Bungay D 1996 Doses to patients from medical x-ray examinations in the UK—1995 review *NRPB Report* R289 (London: HMSO)

ICRP 1996 Radiological protection and safety in medicine *ICRP Publication* 73 (Oxford: Elsevier)

IPSM 1992 *National Protocol for Patient Dose Measurements in Diagnostic Radiology* (Chilton: NRPB)

Iwi G, Millard R K, Palmer A M, Preece A W and Saunders M 1999 Bootstrap resampling: a powerful method of assessing confidence intervals for doses from experimental data *Phys. Med. Biol.* **44** N55–N62

Johnson D W and Goetz W A 1986 Patient exposure trends in medical and dental radiography *Health Phys.* **50** 107–16

McParland B J 1998a A study of patient radiation doses in interventional radiological procedures. *Br. J. Radiol.* **71** 175–85

——1998b Entrance skin dose estimates derived from dose–area product measurements in interventional radiological procedures *Br. J. Radiol.* **71** 1288–95

NRPB 1993 Protection of the patient in x-ray computed tomography *Doc. NRPB* **4** 43–74

Press W H, Teukolsky S A, Vetterling W T and Flannery B P 1992 *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge: Cambridge University Press)

Shrimpton P C, Wall B F, Jones D G, Fisher E S, Hillier M C, Kendall G M and Harrison R M 1986 A national survey of doses to patients undergoing a selection of routine x-ray examinations in English hospitals *NRPB Report* R200 (London: HMSO)

Vañó E, González L, Fernández J M and Guibelalde E 1995 Patient dose values in interventional radiology *Br. J. Radiol.* **68** 1215–20

Wall B F and Shrimpton P C 1998 The historical development of reference doses in diagnostic radiology *Radiat. Prot. Dosim.* **80** 15–20

Warren-Forward H M, Haddaway M J, Temperton D H and McCall I W 1998 Dose–area product readings for fluoroscopic and plain film examinations, including an analysis of the source of variation for barium enema examinations *Br. J. Radiol.* **71** 961–7

Williams J R 1997 The interdependence of staff and patient doses in interventional radiology *Br. J. Radiol.* **70** 498–503

Zhao X-H and Gao S 1997 Confidence intervals for the log-normal mean *Statist. Med.* **16** 783–90