European Radiology (2019) 29:5307–5313
https://doi.org/10.1007/s00330-019-06055-5

PHYSICS

X-ray examination dose surveys: how accurate are my results?

Stephen Taylor[1] · Alain Van Muylem[2] · Nigel Howarth[3] · Pierre Alain Gevenois[4] · Denis Tack[5]

Received: 29 October 2018 / Revised: 13 January 2019 / Accepted: 31 January 2019 / Published online: 15 March 2019

Abstract

Objectives To determine the variabilities of dose-area-products (DAP) of frequent X-ray examinations collected for comparison with diagnostic reference levels (DRLs).

Methods DAP values of chest, abdomen, and lumbar spine examinations obtained on devices from two manufacturers were collected in three centers over 1 to 2 years. The variability of the average DAP results defined as the 95% confidence interval in percentage of their median value was calculated for increasing sample sizes, each examination and center. We computed the sample sizes yielding variabilities lower or equal to 25% and 10%. The effect of narrowing patient selection based on body weight was also investigated (ranges of 67–73 Kg, or 60–80 Kg).

Results DAP variabilities ranged from 75 to 170% of the median value when collecting small samples (10 to 20 DAP). To reduce this variability, larger samples are needed, collected over up to 2 years, regardless of the examination and center. A variability ≤ 10% could only be reached for chest X-rays, requiring up to 800 data. For the abdomen and lumbar spine, the lowest achievable variability was 25%, regardless of the body weight selection, requiring up to 400 data.

Conclusion Variabilities in DAP collected through small samples of ten data as recommended by authorities are very high, but can be reduced down to 25% (abdomen and lumbar spine) or even 10% (chest) through a substantial increase in sample sizes. Our findings could assist radiologists and regulatory authorities in estimating the reliability of the data obtained when performing X-ray dose surveys.

Key Points

- *Low but reasonable variabilities cannot be reached with samples sized as recommended by regulatory authorities. Higher numbers of DAP values are required to reduce the variability.*
- *Variabilities of 10% for the chest and 25% for abdomen and lumbar spine examinations are achievable, provided large samples of data are collected over 1 year.*
- *Our results could help radiologists and authorities interpret X-rays dose surveys.*

Keywords Radiation protection · Radiography · Surveys and questionnaires

✉ Denis Tack
denis.tack@skynet.be

[1] Department of Radiology, Hôpital Ambroise Paré, Boulevard Président Kennedy 2, 7000 Mons, Belgium

[2] Department of Pneumology, Hôpital Erasme, Route de Lennik 808, 1070 Brussels, Belgium

[3] Department of Radiology, Clinique des Grangettes, 7 Chemin des Grangettes, 1224 Chêne-Bougeries, Geneva, Switzerland

[4] Department of Radiology, Hôpital Erasme, Route de Lennik 808, 1070 Brussels, Belgium

[5] Department of Radiology, EPICURA, Clinique Louis Caty, Rue Louis Caty 136, 7331 Baudour, Belgium

Abbreviations

AEC	Automatic exposure control
CTDIvol	Volume computed tomographic dose index
DAP	Dose area product
DRL	Diagnostic reference level
ESD	Entrance surface dose
EU	European Union
PA	Posterior-anterior

Introduction

In the European Union (EU), two directives implemented through national regulations have imposed dose surveys [1,

2]. These surveys produced diagnostic reference levels (DRLs) available since the early 2000s [3, 4]. The EU published practical methodological recommendations based on those applied in the UK. However, their application in EU states remains at the discretion of national authorities. These recommendations were to collect data in ten consecutive so-called standard weight patients (70 ± 3 Kg) for a given X-ray examination. As the recruitment of such patients was difficult to implement, national authorities adapted their recommendations by increasing the sample size up to 50 consecutive patients, while increasing the accepted weight range to 60–80 Kg or considering all patients regardless of their weight [3, 4]. In early surveys, population- and examination-dependent exposures have been computed by pooling data but the numbers of measurements per examination and per device remained unknown [4–8].

On CT equipped with automatic exposure control (AEC), a wide variability in dose descriptors has been reported recently, when samples were sized as prescribed by national and international authorities [3, 9, 10]. In order to obtain an acceptable degree of variability in collected dose descriptors, sample sizes should be increased by up to ten times compared with current recommendations [10]. As with CT equipped with AEC, X-ray examinations are performed with photoelectric cells, which also adjust the X-ray exposure to the patient's attenuation. We hypothesized that the variability in dose descriptors of X-ray examinations collected from samples of small size could be higher than anticipated. The aim of this study was therefore to determine this variability in frequently performed X-ray examinations as a function of sample sizes, allowing the reader to know the accuracy of his own results for further comparison with DRLs.

Materials and methods

Patients and acquisitions

This retrospective multicenter study received approval from local ethical committees and informed consent was waived.

The dose area product (DAP) of chest (posterior-anterior (PA) alone or combined with lateral view), lumbar spine (with a variable number of views), and abdomen (with a variable number of views) X-ray examinations delivered to patients aged 18 years or more, by devices from two manufacturers installed in three centers (centers A, B, and C), were collected. Repeated examinations in the same patient were considered as new data. The DAP measurement devices were calibrated by the manufacturers every 3 months and annually audited by the national agency. The legal tolerance for measurements was 25% but that measured by our local radiophysicists was < 5%.

In center A, DAP values were collected from 15,438 examinations performed between July 2014 and June 2016, on four identical digital direct detector devices (Luminos, Siemens Healthineers), set with identical parameters. In center B, DAP values were collected from 39,936 examinations performed between May 2014 and April 2016 on four identical devices (Luminos, Siemens Healthineers), set with identical parameters. In center C, DAP values were collected from 9466 examinations performed between August 2015 and July 2016 on one device (DigitalDiagnost, Philips Healthcare) and from 13,691 examinations performed on two identical devices (Juno DRF, Philips Healthcare), set with identical parameters. In addition, the patient's weight was collected in order to create the so-called standard weight patients subgroups (i.e., 60–80 Kg and 67–73 Kg) [3, 9].

Statistical analysis

Our data was initially tested for normality and log-normality for each series of results by a Lilliefors variant of the Kolmogorov-Smirnov test. The distributions all significantly differed from a normal as well as from a log-normal distribution (with a p value < .001 indicating statistical significance).

For a given sample of size n, the sampling distribution of the mean DAP cannot be reliably anticipated. In each center for each of the three examinations, this distribution was thus estimated empirically, by randomly drawing with replacement 2000 samples of size n from the total population of DAPs. This process is illustrated in Fig. 1. For each sample size, the variability of means was reflected by the 95% confidence interval (CI95) of the mean sampling distribution. Given the great differences in DAP values between centers and types of examination, CI95 was expressed as a percentage of the median (med) of the mean sampling distribution (CI95/med). This process was repeated for increasing sample sizes (n = 10, 20, 30, 40, 50, 60,100, 200, 300, 400, 500, 1000, etc.). The sample sizes required to achieve variabilities—corresponding to a CI95/med lower than 10% (n_{10}) and 25% (n_{25})—were computed by log-log (log(IC95/med) vs. log(n)) regression fittings. The time periods required to collect n_{10} and n_{25} data were estimated by a rule of three from the number of data collected during the time frame of the study in each center, for each type of examination.

The open source PASCAL compiler FreePascal IDE was used to carry out this procedure.

Results

The numbers of collected examinations and views per examination from our three centers are shown in Table 1 with corresponding mean, standard deviation, median, and quartiles of DAP values of chest (posterior-anterior (PA) alone or combined with lateral view), abdomen (with a variable number of views), and lumbar spine (with a variable number of views) X-rays. The numbers of collected DAP values of each examined

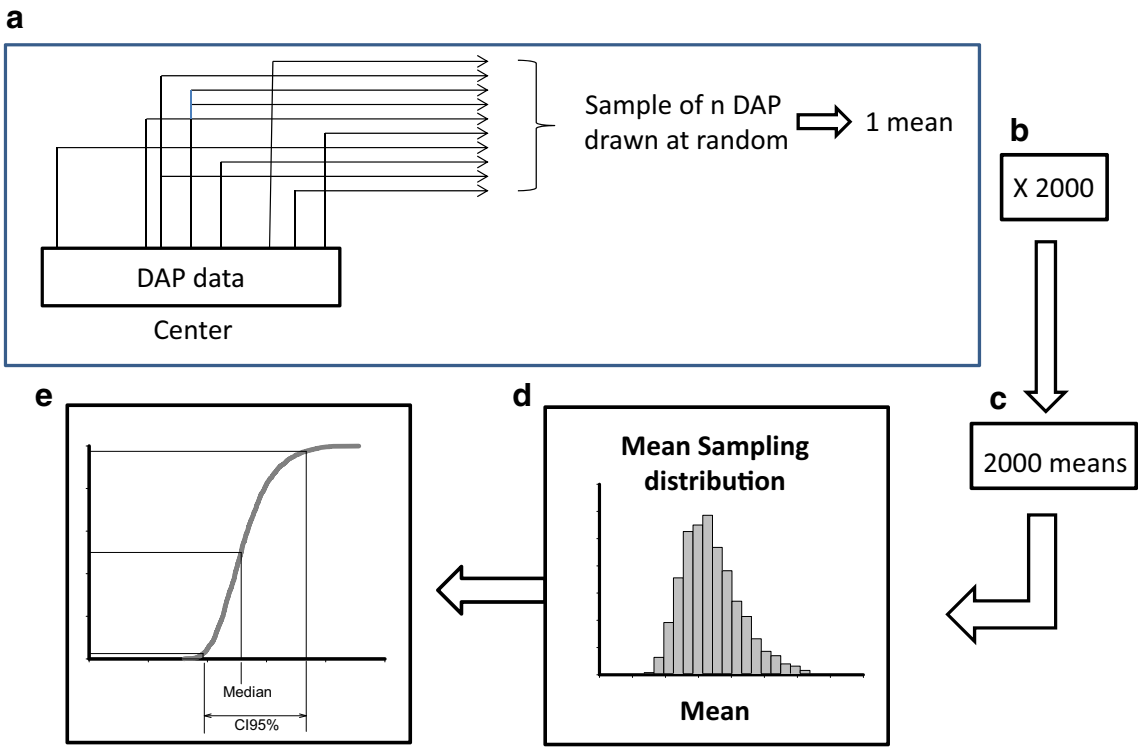

Fig. 1 For a given center and a given examination, a sample of *n* DAP values was randomly drawn (with replacement) from the set of DAP values, given one sampled mean (**a**). By repeating this process 2000 times (**b**), 2000 sampled means were obtained (**c**), deriving the mean sampling distribution (**d**) from which median and 95% confidence interval were computed (**e**). A to E was performed for $n = 10$, 20, 30, 40, 50, 60, 100, 200, 300, 400, 500, and 1000

region, the variability (CI95/med) for all obtained examinations, as well as the number of examinations required to achieve variabilities $\leq 10\%$ or $\leq 25\%$ for each examination and in each center are displayed as example values in Table 2.

For all examined regions, the variability in DAP values decreased as sample size increased. As an example, Fig. 2 illustrates in center A the DAP variability for an abdomen view with no body weight selection. For small sample sizes

Table 1 Number of examinations, views per examination, and DAP values per examination in each center

		Chest		Abdomen	Lumbar spine
		Postero-anterior view	Postero-anterior and lateral views		
Center A	Number of patients	2403	2112	1452	1275
	Number of views, mean (range)	1 (1–1)	2 (2–2)	1 (1–5)	6 (1–10)
	DAP (mean ± SD; mGy.cm^2)	63 ± 51	271 ± 196	1447 ± 1812	7533 ± 7753
	DAP median (mGy.cm^2)	52	220	889	5028
	DAP first–third quartiles (mGy.cm^2)	34–64	144–326	461–1744	2636–9501
Center B	Number of patients	15,682	15,318	2987	478
	Number of views, mean (range)	1 (1–1)	2 (2–2)	2 (1–12)	4 (1–9)
	DAP (mean ± SD; mGy.cm^2)	50 ± 32	232 ± 205	1944 ± 2927	4147 ± 3946
	DAP median (mGy.cm^2)	41	175	1164	2920
	DAP first–third quartiles (mGy.cm^2)	30–57	112–278	576–2332	1668–5092
Center C	Number of patients	3738	3106	1479	431
	Number of views, mean (range)	1 (1–1)	2 (2–2)	2 (1–6)	8 (1–14)
	DAP (mean ± SD; mGy.cm^2)	295 ± 302	1228 ± 1226	3762 ± 4298	22,888 ± 18,203
	DAP median (mGy.cm^2)	209	871	2514	17,909
	DAP first–third quartiles (mGy.cm^2)	142–356	536–1458	1409–4590	10,069–29,930

Table 2 Number of collected examinations, variability of DAP values in each explored body region according to body weight category in each center, and required number of examinations to reach variability of DAP values ≤ 10% and ≤ 25% with time period (in months) for collecting required number of examinations

Body weight selection			Center A				Center B				Center C			
			Number of collected examinations	Variability	*N* for variability ≤ 10% (months)	*N* for variability ≤ 25% (months)	Number of collected examinations	Variability	*N* for variability ≤ 10% (months)	*N* for variability ≤ 25% (months)	Number of collected examinations	Variability	*N* for variability ≤ 10% (months)	*N* for variability ≤ 25% (months)
No weight selection	Chest	PAr view	2403	5%	628 (7 m)	91 (1 m)	15,682	2%	561 (1 m)	89 (0.25 m)	3738	5%	877 (3 m)	149 (0.5 m)
		PA and lateral	2112	7%	987 (12 m)	168 (2 m)	15,318	3%	1035 (2 m)	154 (0.25 m)	3106	7%	1424 (6 m)	228 (1 m)
	Abdomen		1452	11%	ID	296 (5 m)	2987	8%	2126 (18 m)	370 (3 m)	1479	11%	ID	287 (3 m)
	Lumbar spine		1275	11%	ID	282 (6 m)	478	19%	ID	289 (15 m)	431	16%	ID	246 (7 m)
60–80 Kg	Chest	PAr view	1131	6%	343 (8 m)	49 (2 m)	3999	3%	363 (3 m)	58 (1 m)	1074	8%	650 (8 m)	107 (2 m)
		PA and lateral	1026	9%	795 (19 m)	134 (4 m)	3963	4%	718 (5 m)	107 (8 m)	886	11%	ID	147 (2 m)
	Abdomen		742	14%	ID	229 (8 m)	1021	15%	ID	343 (9 m)	470	16%	ID	181 (5 m)
	Lumbar spine		632	11%	ID	139 (6 m)	145	27%	ID	ID	144	37%	ID	ID
67–73 Kg	Chest	PAr view	347	8%	236 (17 m)	45 (4 m)	1265	5%	278 (6 m)	47 (1 m)	320	13%	ID	80 (4 m)
		PA and lateral	303	15%	ID	106 (9 m)	1249	7%	534 (11 m)	82 (2 m)	263	17%	ID	123 (6 m)
	Abdomen		220	23%	ID	185 (21 m)	297	25%	ID	287 (24 m)	127	30%	ID	ID
	Lumbar spine		212	21%	ID	190 (22 m)	49	40%	ID	ID	42	54%	ID	ID

ID insufficient data

(i.e., from 10 to 20 examinations) without any weight selection, and all types of examinations, the boundaries of the CI95 of the mean DAP distribution ranged from 0.6 to 3.3 times the median value of that distribution. Figure 3 shows variabilities (CI95/med) as a function of the sample size in the three centers for a PA chest view, combined PA and lateral chest views, abdomen, and lumbar spine. Symbols are CI95/med's computed from mean sampling distributions and continuous and dotted lines are log-log regression fittings on which n_{10} and n_{25} were computed. The coefficient of determination of the fitting was always greater than 0.985. The numbers of collected DAP values of each examined region, the variability (CI95/med) for all obtained examinations, as well as n_{10} and n_{25} for each examination and in each center are displayed as example values in Table 2. The estimated number of months required to collect n_{10} and n_{25} DAPs is indicated between brackets.

Discussion

This study shows the following: (1) the number of examinations required to lower the variability can be very high, regardless of the region examined, the body weight category, the absolute DAP values, and the center; (2) should a very low variability (such as 10%) be required, except for chest X-rays, the large number of examinations needed is likely to require a time period longer than reasonably achievable; a variability of 25% can be achieved more rapidly; (3) the body weight selection can help reduce the number of examinations required but not the time necessary for data collection; (4) collecting examinations as recommended by national regulatory authorities (i.e., samples of 10–50 DAP values) yields variabilities ranging from 75 to 170%; and (5) collecting examinations over the time period recommended by authorities (i.e., 3 months) yields variabilities up to 45%. These results deserve further discussion at the levels of radiology departments, national regulatory authorities, and dose descriptors.

At the level of radiology departments, comparisons, in accordance with legal requirements, between data collected on small-sized samples and DRLs are questionable as small-sized samples could induce huge variability. Indeed, the range of DAP values from a given center could include the DRL itself, regardless of its value, even with low and appropriately optimized radiation doses. Furthermore, ranges of DAP values collected in our three centers overlap, although those from center C are higher than DRLs (i.e., national 75th percentiles), while those in centers A and B are even lower than national 25th percentiles. This observation could have two consequences. On the one hand, if in a given center, DAPs collected from a small sample of examinations are higher than DRLs, while true DAP values (i.e., those collected from all examined patients) are actually lower than DRLs, a need for further dose

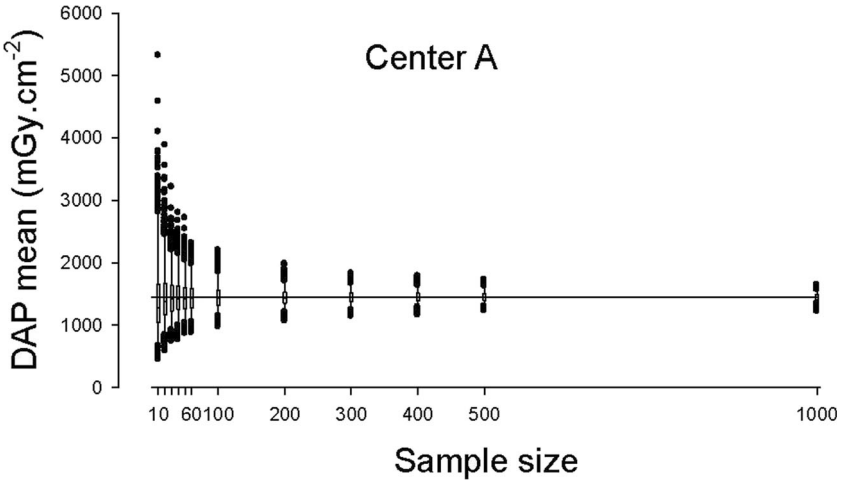

Fig. 2 The sampled distribution of mean DAP values, for posterior-anterior chest X-rays (2000 samples) as a function of sample size. For each sample size, the box represents the inter-quartile range and the whiskers represent the 95% confidence interval; the closed circles are the values above percentile 97.5 and under percentile 2.5. Inside each box, the horizontal line is the median. The continuous horizontal line is the mean DAP value of the whole population

reduction could be recommended inappropriately, with a risk of deterioration of image quality. On the other hand, if local DAPs are lower than DRLs, while their true values are higher than DRLs, dose optimization could be deemed inappropriately as unnecessary, with a subsequent risk of high radiation dose delivery. At the level of radiology departments, dedicated dose-tracking software should be used to collect large numbers of examinations easily [11], providing daily monitoring of the doses delivered and alerts for extreme dose values.

At the level of national regulatory authorities, attention should be paid to the interpretation of radiation dose delivered by individual devices and comparisons with DRLs. This study shows that surveys based on legally sized samples almost certainly do not reflect the truly delivered doses. As a consequence, further work is needed to ensure that data considered by national regulatory authorities are sufficient to provide unbiased DRLs, regardless of the device-dependent time period necessary for their collection.

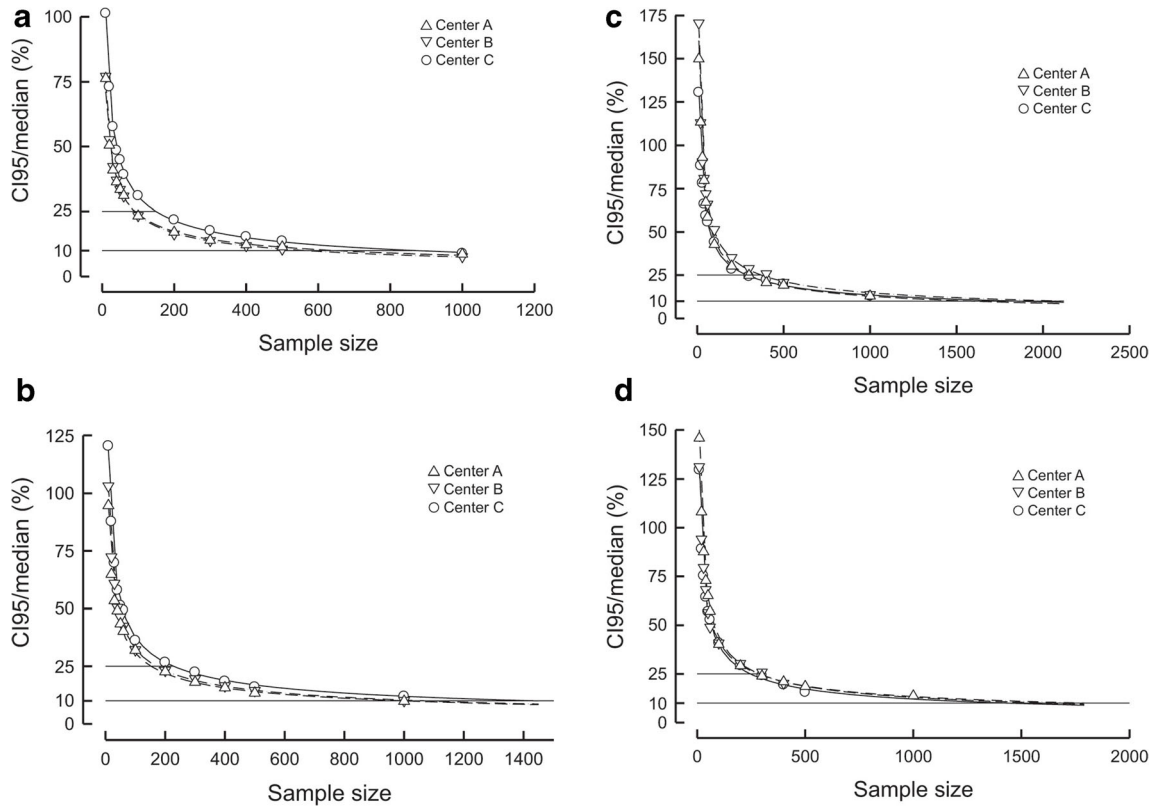

Fig. 3 **a** Poster-anterior chest X-rays, centers A, B, and C, CI95/med as a function of the sample size; **b** Combined postero-anterior and lateral chest X-rays, centers A, B, and C, CI95/med in function of the sample size; **c** Plain Film abdominal examinations, centers A, B, and C, CI95/med as a function of the sample size; **d** Lumbar spine examinations, centers A, B, and C, CI95/med as a function of the sample size

At the level of dose descriptors, the size-specific dose estimate (SSDE) was introduced to quantify CT-delivered radiation dose more appropriately than volume computed tomographic dose index (CTDIvol) and has been reported to be associated with a lower patient-dependent variability than CTDIvol [12–14]. For radiographic examinations however, as no size-specific dose descriptor has been proposed, there is currently no other available method than DAP and entrance surface dose (ESD) measurements [8]. As variabilities in CT dose have been recently reported as much higher than anticipated when based on small sample sizes, the present study shows that variabilities are even higher with radiography than with CT [10]. Body weight selection could reduce the variability [3, 9] but not shorten the time periods needed to reach the expected variability. In other words, reducing the time frame of dose surveys inevitably induces higher data variabilities, and those considered as low and acceptable in the seminal report on radiation dose surveys were almost certainly underestimated [8].

This study has limitations. First, data from several radiology devices were pooled in order to increase the number of available examinations. However, any eventual bias would have been minimal, as only data from devices with identical settings were pooled. Second, devices from no more than two manufacturers were considered. Although larger numbers of patients are required to achieve a lower variability with these two manufacturers, further verifications with other equipment are needed. Interestingly, although the absolute DAP values were very different between manufacturers (centers A and B were lower than the 25th percentiles of national dose distributions whereas center C was above DRLs), the variabilities of data were very similar in the three centers and for both manufacturers. Third, we did not investigate factors that could contribute to variabilities other than the sample size and weight. Other factors such as differences in body habitus, clinical indications, and local practices could indeed impact them and should be further addressed. To note, including consecutive patients is the actual legal requirement in our country. Fourth, we considered some individual examinations for which the number of views was obviously above that usually obtained but rules of dose surveys preclude us from excluding them from collected samples. Interestingly, such outliers reflect local practices but as the variability was defined as the 95th confidence interval divided by the median, these outliers (i.e., within the 2.5% highest or the 2.5% lowest DAP values) were excluded from our analyses. Fifth, all clinical indications were included for a given examination in accordance with the current survey rules. Narrowing selection through clinical indications as suggested for CT examinations [11] could indeed reduce the variabilities but not the time period necessary to collect them, as suggested in Table 2 for weight-based stratification. Sixth, we introduced weight-based stratification of our data—as historically suggested—by considering standard weight patients. However, as overweight patients are at risk of low back pain, such stratification would not reflect the disease distribution in our population [10].

In conclusion, variabilities in radiation doses collected as recommended by national regulatory authorities through the EU directives are high. With the current rules, comparisons between collected doses and DRLs are at risk of both over- and underestimation. Sample sizes should therefore be increased. The variabilities reported here as a function of sample sizes and regions examined could assist radiologists and regulatory authorities in estimating the reliability of X-ray dose surveys.

Funding The authors state that this work has not received any funding.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Denis Tack.

Conflict of interest The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article.

Statistics and biometry One of the authors has significant statistical expertise.

Informed consent Written informed consent was waived by the Institutional Review Board.

Ethical approval Institutional Review Board approval was obtained.

Methodology
- retrospective
- observational
- multicenter study

References

1. The council of the European Union (1997) Council Directive 97/43 of 30 June 1997 on Health protection of individuals against the dangers of ionizing radiation in relation to medical exposure, and repealing directive 84/466/Euratom. Available via http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:31997L0043. Accessed 19 Apr 2018
2. The Council of the European Union (2013) Council Directive 2013/59/EURATOM of 5 December 2013 laying down basic safety standards for protection against the dangers arising from exposure to ionising radiation, and repealing Directives 89/618/Euratom, 90/641/Euratom, 96/29/Euratom, 97/43/Euratom and 2003/122/Euratom. Available via http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32013L0059&from=EN. Accessed 19 Apr 2018
3. European commission (1999) Radiation Protection 109: guidance on diagnostic reference levels for medical exposures. Available via

https://ec.europa.eu/energy/sites/ener/files/documents/109_en.pdf. Accessed 19 Apr 2018
4. National Radiological Protection Board (NRPB) (2002) Doses to Patients from Medical X-ray Examinations in the UK – 2000 Review. Availabe via https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/405115/2002_NrpbW14.pdf. Accessed 19 Apr 2018
5. Mettler FA Jr, Huda W, Yoshizumi TT, Mahesh M (2008) Effective doses in radiology and diagnostic nuclear medicine: a catalog. Radiology 248:254–263
6. Børretzen I, Lysdahl KB, Olerud HM (2007) Diagnostic radiology in Norway – trends in examination frequency and collective effective dose. Radiat Prot Dosimetry 124:339–347
7. Advisory Committee on Radiological Protection (1997) Radiation Doses from Medical Diagnostic Procedures in Canada. Available via http://www.nuclearsafety.gc.ca/eng/about/past/timeline-dev/resources/documents/infohistorical/info-0670-E.pdf. Accessed 19 Apr 2018
8. Health Protection Agency – Centre for Radiation, Chemical and Environmental Hazards Radiation Protection Division (2007) Doses to Patients from Radiographic and Fluoroscopic X-ray Imaging Procedures in the UK – 2005 Review. Available via https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/431134/HPA-RPD-029.pdf. Accessed 19 Apr 2018
9. European commission (2008) Radiation Protection N°154: European Guidance on Estimating Population Doses from Medical X-Ray Procedures. Available via http://www.ddmed.eu/_media/background_of_ddm1:rp154.pdf. Accessed 19 Apr 2018
10. Taylor S, Van Muylem A, Howarth N, Gevenois PA, Tack D (2017) CT dose survey in adults: what sample size for what precision? Eur Radiol 27:365–373
11. Vañó E, Miller DL, Martin CJ et al (2017) ICRP publication 135: diagnostic reference levels in medical imaging. Ann ICRP 46:1–144
12. American Association of Physicists in Medicine (2011) Size-specific dose estimates (SSDE) in pediatric and adult body CT examinations. Task Group 204. College Park, Md: American Association of Physicists in Medicine
13. Bankier AA, Kressel HY (2012) Through the looking glass revisited: the need for more meaning and less drama in the reporting of dose and dose reduction in CT. Radiology 265:4–8
14. Christner JA, Braun NN, Jacobsen MC, Carter RE, Kofler JM, McCollough CH (2012) Size-specific dose estimates for adult patients at CT of the torso. Radiology 265:841–847