

NOTICE: this is the author's version of a work that was accepted for publication in European Journal of Radiology. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in European Journal of Radiology, Vol. 83 (2), 2014. DOI: 10.1016/j.ejrad.2013.11.003

Evaluating data capture methods for the establishment of diagnostic reference levels in CT scanning

Running Head: Data capture methods for diagnostic reference levels in CT

Rachael E Moorin (PhD)^{1,2} r.moorin@curtin.edu.au

David AJ Gibson* (BscHlthSci(Hons))² david.gibson@uwa.edu.au

Rene K Forsyth (BSciMedIm(Hons))³ rene.forsyth@postgrad.curtin.edu.au

Max K Bulsara (PhD)⁴ max.bulsara@nd.edu.au

C D'Arcy J Holman(PhD)² darcy.holman@uwa.edu.au

¹ Centre for Population Health Research, Faculty of Health Science, Curtin University, GPO Box U1987, Perth Western Australia, 6845, Australia. Ph: (+618) 9266 1854

²Centre for Health Services Research, School of Population Health, University of Western Australia, 35 Stirling Highway, Crawley, Perth Western Australia, 6009, Australia. Ph: (+61 8) 6488 1261

³Department of Medical Imaging Science, Curtin University, GPO Box U1987, Perth, Western Australia, 6845, Australia. Ph: (+618) 9266 1854

⁴Institute of Health Research, University of Notre Dame, 19 Mouat Street, Fremantle, Western Australia, 6959, Australia. Ph: (+618) 9433 0297

*Corresponding author.

Evaluating data capture methods for the establishment of diagnostic reference levels in CT scanning

Abstract

Objective: Concerns about the radiation dose associated with CT scanning have led to a call for establishment of diagnostic reference levels. Self-complete surveys have been used extensively to gather this information however, departmental Radiological Information System's/Picture Archive Communication System's (RIS/PACS) also hold this information. We compared dosimetry derived from survey with that using RIS/PACS's.

Methods: Technical data were collected from a large metropolitan tertiary hospital in WA using both data collection methods for a range of adult CT scanning examinations. Radiation dose was calculated from both datasets and the results evaluated for several indexes of inter-rater agreement.

Results: Radiation dose calculated using self-report survey data differed both systematically and proportionally from that calculated using RIS/PACS data. Differences were not consistent across CT examination type and thus not amenable to simple correction. The disparity was greater and more variable for organ dose than effective dose due to reliance of survey data on "generic" anatomical start and stop limits compared with actual data available on RIS/PACS.

Conclusions: The bias observed in our study indicates care should be taken when interpreting the results of studies measuring radiation dose using self-complete surveys. The availability of electronic databases that include information required for the evaluation and monitoring of CT radiation dose provides the opportunity to capture better quality data in a cost-effective manner. We recommend that national and local databases are established that routinely capture these data so as to facilitate the development and monitoring of radiation dose associated with CT scanning.

Keywords: CT scan, Radiation Dosage, Radiation Dosimetry, Survey Methods, Radiology Information Systems

Introduction

CT scanning has always been recognised as a relatively high radiation dose procedure, but in the early days there was no viable alternative. No other modality could compete with CT on the diagnostic accuracy of brain scans; and when CT of the rest of the body first began, its use was largely limited to cancer patients where radiation dose was less of a concern given the balance of risks [1]. Today, CT is used extensively in benign disease and young patients, for whom radiation protection considerations hold greater weight [1, 2].

Awareness and concern regarding the high burden of radiation dose to the population stemming from CT began in the late 1980s following a survey conducted by the UK National Radiological Protection Board [3]. It found, despite comprising only 2% of all medical imaging examinations, CT contributed a disproportionately large proportion (20%) of the collective population dose from diagnostic imaging [3, 4]. The survey also found radiation doses from CT were trending in a direction at odds with the decreasing collective patient doses received from conventional x-ray examinations [3]. Rather, CT examinations had grown steadily to become a significant source of population exposure from medical radiology [4]. A follow-up study showed CT in the UK had doubled both its contribution to the population effective dose (40%) and the proportion of all radiological examinations (4%) by 1995 [5]. Similar findings were reported in other developed countries such as Switzerland, [6] Germany, [7] the Netherlands [8] and the US [9]. Mettler *et al* [9] found CT in the USA accounted for 11% of diagnostic radiological procedures in 1999, compared with only 6% in 1990. This study also showed CT accounted for 67% of the total effective radiation dose from diagnostic imaging in 1999 [9]. Of particular concern, Mettler *et al* [9] found 11% of CT examinations were performed on children. In 1998, the Royal College of Radiologists (UK) estimated CT contributed approximately one half of the collective dose from all examinations, with the average in UK CT workload to be a 24% increase per year against only 2% for all other medical and dental radiological procedures [10].

Concerns have been expressed about the radiation dose associated with CT scanning, which have led to guidelines, advising on clinical indications for utilisation or diagnostic reference levels (DRL's) for the radiation dose received from each type of examination [1, 5, 11]. Due to paucity of data, estimates of radiation dose and cancer risk from CT scanning to underpin the creation of DRL's have been undertaken

using self-report surveys combined with 'typical' CT protocol / machine settings [12-14]. However, this data capture technique has inherent limitations typical of all self-complete surveys such as response bias, low response fraction, poor generalizability (due to the data captured not being a representative sample either purposefully or otherwise) and, if the instrument has not been appropriately validated, poor validity and reliability.

Databases of CT scanner radiation dose facilitated by Radiological Information Systems (RIS) and Picture Archiving Communication Systems (PACS) could enable comprehensive routine institutional benchmarking, optimisation of CT protocols and quality control [15]. These databases would also enable more accurate patient specific dose estimation than possible from data sources available previously, which would ultimately lead to more accurate patient specific risk assessment to better inform imaging decisions [15, 16]. However, while the data required are captured in individual department RIS/PACS, extracting these data efficiently remains problematic. Thus self-report survey methods are currently used extensively for capture of national and regional CT dose information.

The aim of this study was to compare the concordance (reliability and validity) of the radiation dose derived from a recognised international self-complete survey data capture instrument with dose obtained from data extracted from RIS/PACS (i.e. the gold standard) for a range of CT scanning examination types undertaken in the same public hospital.

Materials and Methods

Technical data on local CT practice were collected from a large metropolitan tertiary (teaching) hospital in Western Australia using (i) a self-complete survey and (ii) information contained in the dose report obtained from the PACS database for diagnostic CT scanning examinations performed on adults using the same make and model 64 slice scanner in the following six clinical scenarios.

1. Routine Head for trauma or stroke
2. Chest for lung cancer (known or suspected metastases)
3. Chest for pulmonary embolism
4. High Resolution Chest for chronic obstructive pulmonary disease
5. Abdomen / Pelvis (Abscess)
6. Chest/Abdomen /Pelvis for lymphoma staging or follow up

This study was approved by the Western Australia Department of Health Human Research Ethics Committee, which exempted the study from requiring individual patient consent.

Data collection: Survey data

Technical data were collected by means of a questionnaire completed by the chief CT radiographer in early 2011. The questionnaire was identical to the survey used by the national survey of doses from CT in the UK in 2003 [12], amended only in terms of the clinical scenarios examined (addition of ‘chest’ for pulmonary embolus and removal of ‘abdomen’ for liver metastases) and limited to examinations performed on adults. The questionnaire sought technical data in relation to examinations performed on ‘average-sized’ patients to represent ‘usual practice’. Technical information was collected for each clinical scenario (excluding the scout view) consisting of separate scanning sequences (phases) where appropriate, each representing a single helical exposure or a series of similar axial exposures using identical scan conditions. The data included various technical parameters such as kilovoltage (kV), milliamperage (mA), tube rotation times, collimation widths, pitch, scanning method and typical anatomical reference start-stop positions of the scan. Respondents were asked to report the average volume weighted CT dose index (CTDI_{vol}) and dose–length

product (DLP) based on up to ten standard ('average-sized') adult patients undergoing each examination type, as performed by the 2003 UK dose survey [12].

Data collection: Picture Archiving Communication System

For each scanning scenario, data pertaining to a random sample of 20 adult cases (or all the cases available if less than 20) identified from the departmental PACS data base, performed between 1st January and 30th April 2011 inclusive, were collected by the researchers. Protocol information (excluding the scout view) consisted of separate scanning sequences (phases) where appropriate for each protocol. The data collected included various technical parameters such as kilovoltage (kV), milliamperage (mA), tube rotation times, collimation widths, pitch, scanning method, volume weighted CT dose index (CTDI_{vol}) and dose-length product (DLP) for each case. CT radiographers were interviewed to acquire typical anatomical start-stop locations for each scanning scenario.

Radiation dosimetry

For both sets of data (survey and PACS), dosimetry involved reporting of CTDI_{vol}, DLP values and calculation of organ and effective dose for each sequence using scan settings provided by the data. Scanner-specific dosimetrics published by ImPACT are included within the CT patient dosimetry calculator [17]. For sequences performed with automatic tube current modulation, doses were calculated using reported values, where available, of (average) tube current or current-time product (mAs) including the effects of modulation. For survey data the reported anatomical start-stop locations were used to determine the placement of the scanning field on the ImPACT dosimetry calculator. However, for the PACS data the average scan length of the cases collected, obtained by dividing DLP by CTDI_{vol}, was used to modify the reported anatomical start stop locations for each clinical scenario. For both data collections, in circumstances when ImPACT calculated values of CTDI_{vol} differed to those reported, the organ and effective doses were corrected by the ratio of reported and the ImPACT calculated CTDI_{vol}. For example, if the reported CTDI_{vol} was 20mGy but the ImPACT software calculated 18mGy (from the imputation of other scanning parameters) then the subsequent effective dose values were multiplied by 1.11 (20/18). Protocol (cumulative) values of DLP and effective dose were also calculated on the basis of summation over

all routine sequences reported for each clinical scenario (Sequence 1 DLP + Sequence 2 DLP + Sequence 3 DLP = Protocol DLP). For the PACS data the average and standard deviation of the CTDIvol, DLP organ and effective doses of the cases collected were reported.

Analysis of inter-rater agreement

Evaluations of the agreement of organ dose and effective dose derived using PACS versus self-reported survey data were conducted under three scenarios: (i) the radiation doses calculated from PACS and survey data both adjusted to one decimal place; (ii) where the two values differed by ≤ 1 the PACS value was reset to be equal to survey values; and (iii) where the two values differed by ≤ 5 the PACS value was reset to be equal to survey values.

The reliability (inter-rater agreement) of the two data collection instruments as input to radiation dosimetry calculations was assessed using the intraclass correlation coefficient (ICC) separately for each clinical scenario. We used the ICC in two distinct ways. The first ICC evaluated the absolute agreement of the organ and effective doses produced for single measures, this is the reliability of the ratings for one, typical, single dose [18, 19]. The second ICC used was with respect to average measures, the reliability of all the organ and effective dose produced by the different data collection instruments averaged together [18, 19]. In addition, the concordance correlation coefficient (ρ_c) was used to determine the degree to which pairs of organ / effective doses resulting from the two data collection instruments related to a line of best fit (set to 45° through the origin) [20]. The data were evaluated for both precision (how far each observation deviated from the best-fit line) and accuracy (how far the best-fit line deviated from the 45° line through the origin). The values of the concordance correlation coefficient were interpreted using a descriptive scale suggested by McBride [21]: <0.9 : poor; $0.90-0.95$: moderate; $0.95-0.99$: substantial and > 0.99 almost perfect.

The equality of measurements (ie radiation doses) from two different analytical methods (PACS versus survey) was tested using Passing and Bablok regression. This is a linear regression procedure with no special assumptions regarding the distribution of the samples and the measurement errors [22]. This regression model was used to evaluate if the radiation doses calculated by the two methods differed by a constant

amount (systematic differences) and for the presence of proportional differences in the values produced by the two methods.

Finally, Bland-Altman plots were used to graphically compare the radiation doses produced using data from the two data collection methods [23]. For this study the differences between the radiation doses produced by the two methods were plotted against the radiation doses produced by PACS data since this method is the reference or "gold standard" [24].

Results

As shown in Table 1 the CT technical information for the six clinical scenarios investigated varied in terms of number of sequences, CTDIvol and DLP across the data collection methods. Data on actual examinations undertaken, sourced from the PACS, showed that all scenarios except 'head' and 'abdomen/pelvis' had a median number of sequences greater than one. Only a single sequence was reported on the self-complete survey for all scenarios. Comparison of CTDIvol showed the average across the cases collected using PACS data was substantially lower than self-reported in three scenarios. The greatest difference occurred in 'chest' for chronic obstructive pulmonary disease with a CTDIvol reported nearly three times larger than the PACS mean CTDIvol. The DLP was consistently lower using the average of the PACS data compared with self-report.

Table 2 presents the organ and effective radiation dose computed using the two methods of data collection. The PACS data produced substantially lower ($> 2\text{mSv}$) effective doses in three scenarios ('chest' for pulmonary embolism, chronic obstructive pulmonary disorder and 'abdomen/pelvis'). Conversely, the PACS data produced substantially higher ($>2\text{mSv}$) effective doses in two scenarios ('chest' for lung cancer and 'chest/abdo/pelvis').

The individual organ doses did not follow the pattern described for the whole body effective dose. PACS data produced higher doses in organs near the limits of the scanning field, because the scan lengths observed in the PACS data were substantially longer than those calculated using the self-reported anatomical start-stop

reference points. For example, in the 'chest' for lung cancer and pulmonary embolism scenarios the radiation dose to the thyroid was higher using the PACS data by 38.9 and 29.7mGy respectively.

Table 3 presents the analysis of agreement between the two data capture methods. There was moderate agreement, according to the ICC and concordance (pc) for some scenarios, such as 'head', 'abdo/pelvis' and 'chest/abdo/pelvis.' In all three of these scenarios the precision (p) was high (≥ 0.98), but the accuracy (C_b) was only moderate. Both the ICC and concordance were poor in the remaining three scenarios. Particularly poor agreement was observed on the 'chest' for chronic obstructive pulmonary disorder scenario, where the ICC and pc were < 0.6 . While the precision in this scenario was perfect (1.00), the accuracy was only 0.56, indicating that the pairs of organ doses produced by the two methods were strongly linearly related, but departed substantially from the 45degree line of concordance.

Differences in the typology of the agreement was reflected in the regression analysis (Table 4), which evaluated the degree of systematic and proportional difference observed in the dosimetry produced using the two data capture methods. In the 'chest' for chronic obstructive pulmonary disorder scenario, there were only non-significant minor systematic differences between the two methods ($A = 0.00$ [-0.10 to 0.056]). There were significant proportional differences ($B = 2.64$ and the 95% confidence interval did not include 1), indicating a non-constant significant difference in the organ dose produced. In contrast, there were no systematic differences and only non-significant proportional differences in the 'head' scenario, indicating the dosimetry produced by the two data capture methods were in good agreement.

When all the scenarios were evaluated together (either using effective dose or organ dose) the agreement was poorer than when evaluated individually. This is shown in both tables 3 and 4, but is most apparent in Figure 1. Panel A of Figure 1 shows the Bland-Altman plot for effective dose comparing the differences between the two data capture methods with the dosimetry produced using PACS (ie, the reference or gold-standard method). In Figure 1 and 2, horizontal lines are drawn at the mean difference the limits of agreement ($+ / - 1.96$ standard deviation) and represent the 95% confidence intervals around the agreement. The Bland-Altman plot can assist in identifying relationships between the differences and the magnitude of

measurements (between the survey and PACS data). The technique aids in identifying the types of systematic bias to aid in the identification of potential corrections. It can be seen that the mean (-1.7) of the comparison did not lie at zero indicating non equal variances; however, all the points lay within the 95% confidence intervals so there was no evidence of a systematic error. Panel B of Figure 1 shows the Bland-Altman plot for organ doses with those from each scenario identified by a different symbol. While the mean lay closer to the zero line (-1.1) than in panel A, the data points followed different gradients across scenarios. This indicated that the relationship between the dosimetry produced by two data capture methods had systematic bias. For example, the 'chest' (chronic obstructive pulmonary disease) values in panel C of Figure 2 had a steep negative gradient indicating bias. The 'chest' (pulmonary embolus) values in panel B of Figure 2 fanned out, indicating systematic bias where the variation in measures were for larger PACS generated values.

Discussion

The results of our study show radiation dosimetry calculations differed, both systematically and proportionally, when using self-report survey data and PACS data sources. Importantly, the study found the differences were not consistent across CT examination types, indicating the disparity cannot be adjusted using a simple correction factor. The disparity was greater and more variable for organ dose than effective dose, most likely resulting from the reliance of the survey data on 'standardised' anatomical start and stop limits compared with actual start and stop limits in the PACS data. The PACS data also provided information regarding the variance of the radiation dose unavailable using the survey instrument, because the survey instrument required the respondent to provide a single CTDIvol and DLP measure averaged from 10 'average' adult patients, whereas the PACS data provided individual level data for each randomly selected case.

The observed discordance between the organ dose calculated from each data source is an important finding since individual or population risk estimates are derived using organ doses rather than effective dose [25]. Controversies persist about the effect of low-level radiation exposure and how these risks should be reported, especially to the public ; however it is generally accepted there is no safe level of exposure (ie, the linear-no-

threshold model) [25]. In light of concerns regarding the increasing use of CT scanning and the expansion of its use in high risk populations such as children, several studies have attempted to quantify the risk to the population of CT scanning based upon radiation doses generated from surveys [13, 26]. In light of our results demonstrating the discordance of the data collected using self-complete surveys with a more rigorous data source, care should be taken when interpreting the results of these studies.

In contrast, guidance regarding the establishment of DRLs recommend they be based on the 75th percentile of the radiation dose measured across providers' derived dose data from a minimum of 10 standard sized patients undergoing standard examinations [27]. Historically DRLs have been articulated in terms of both the effective dose and the DLP. DLPs have to date been generated exclusively from self-complete surveys [12, 28, 29] and, while our study has shown the data are not concordant with data obtained from PACS, the bias with respect to effective dose does appear to follow a more consistent pattern allowing the potential for correction to be applied.

Strengths and Limitations

This study used a survey instrument applied extensively in the United Kingdom for collection of the self-report data [12], without modification (except in terms of the clinical; scenarios examined) and is also recommended by the European Commission on Radiation Protection [30]. It represents a widely accepted example of this data collection method. Data from the PACS were extracted manually from the same institution and over the same time period as the survey. In addition, all CT examinations included in the study were undertaken on the same make and model of CT scanner and by the same technical staff. Data extracted from PACS were a random sample of 20 examinations on adult patients (ie double the number recommended for generation of DRLS [27]) and were limited to examinations most closely matched the clinical scenarios on the survey after consultation with CT technical and clinical staff. The survey was completed by a senior radiographer who specialised in CT prior to collection of the PACS data, and the same individual was consulted regarding matching of PACS examination codes to the clinical scenarios used in the survey.

Since the study was conducted at a single hospital site, the results may not represent concordance between the two data collection methodologies in different sites; thus it is unknown if the disparities observed would have been larger or smaller if multiple providers were included in the study. The study was also restricted in the number of CT examinations evaluated; however, a range of examination methodology and anatomical locations were included. This study was limited to an evaluation of reliability and assumed PACS data as the gold standard. Reliability is only one of the potential limitations to self-complete survey data collection instruments and issues such as response bias and response fraction were not evaluated. In our study, due to the manual nature of the PACS data extraction, we required the provider to agree to participate in the study and hence there would have been no benefit in terms of response fraction. However, if the recommendation of the European Commission on Radiation Protection [30] is realised (ie that national authorities responsible for population dose surveys gather electronic information on patient doses from RIS/PACS into national databases established for the development and monitoring of DRLs and population dose estimates) then this data collection method would be far superior in terms of both response fraction and response bias. Our study did not evaluate the cost differences between the two methodologies; however, since collection of PACS data for our study was done manually (because data were not able to be extracted automatically), the PACS data collection was significantly more expensive to undertake compared with the survey. This additional cost needs to be considered when evaluating the relative advantages and disadvantages of the two data collection methods. Should national or even local databases become established, then the cost of such data would be significantly reduced. Implementing dose databases for DRL and audit purposes can maintain the anonymity of patients by only collating or aggregating the data required for dose calculation. Such a system would be complicit with Australian health data and biospecimen law [31], ensuring no breach of legal obligations is made and the ethical calculus strongly favours the public interest in radiation protection and avoidance of unnecessary harm to patients. Data from both the survey and PACS data were analysed in an identical manner to produce organ and effective doses and this method has been used previously and is well established. Analysis of the dosimetry data for reliability was undertaken using the most appropriate advanced statistical methods and provided a range of indicators of concordance.

Conclusion

Due to increasing concerns about the radiation dose from CT scanning examinations, robust methods for collection of CT dose data are required. Currently self-report surveys are used extensively for this purpose but suffer from limitations. The increasing availability of electronic databases to capture information required for the evaluation and monitoring of CT radiation dose provides an opportunity for better quality data in a cost-effective manner, eliminating many of the potential biases. This study has confirmed data extracted from RIS/PACS is superior to self-reported survey data and has shown survey data contains both proportional and systematic bias not consistent across CT examinations. We recommend national and local databases are established to routinely capture aggregated and anonymous CT dose data for the development and monitoring of DRLs and surveillance of population radiation dose.

References

- [1] Golding SJ, Shrimpton PC. Radiation Dose in CT: are we meeting the challenge? *The British Journal of Radiology*. 2002;75:1-4.
- [2] Naik KS, Ness LM, Bowker AMB, Robinson PJA. Is computed tomography of the body overused? An audit of 2068 attendances in a large acute hospital. *British Journal of Radiology*. 1996;69:126-31.
- [3] Hart D, Wall BF. UK population dose from medical X-ray examinations. *European Journal of Radiology*. 2004;50:285-91.
- [4] Shrimpton PC, Wall BF, Hart D. Diagnostic medical exposures in the UK. *Applied Radiation and Isotopes*. 1999;50:261-9.
- [5] Shrimpton PC, Wall BF. The increasing importance of x-ray computed tomography as a source of medical exposure. *Radiation Protection Dosimetry*. 1995;57:413-5.
- [6] Aroua A, Burnand B, Vader JP, Valley JF. Nation-wide survey on radiation doses in diagnostic and interventional radiology in Switzerland in 1998. *Health Physics*. 2002;83:46-55.
- [7] Kaul A, Bauer B, Bernhardt J, Nosske D, Veit R. Effective doses to members of the public from the diagnostic application of ionizing radiation in Germany. *European Radiology*. 1997;7:1127-32.
- [8] Brugmans MJP, Buijs WCAM, Geleijns J, Lembrechts J. Population exposure to diagnostic use of ionizing radiation in the Netherlands. *Health Physics*. 2002;82:500-9.
- [9] Mettler FAJ, Wiest PW, Locken JA, Kelsey CA. CT scanning: patterns of use and dose. *Journal of Radiological Protection*. 2000;20:353-9.
- [10] Rehani MM. CT: Caution on radiation dose. *Indian Journal of Radiology and Imaging*. 2000;10:19-20.
- [11] Wall BF. Implementation of DRLs in the UK. *Radiation Protection Dosimetry*. 2005;114:183-7.
- [12] Shrimpton PC, Hillier MC, Lewis MA, Dunn M. National survey of doses from CT in the UK:2003. *The British Journal of Radiology*. 2006;79:968-80.
- [13] Pearce MS, Little MP, McHugh K, Lee C, Kim KP, Howe NL, et al. Radiation exposure from CT scans in childhood and subsequent risk of leukaemia and brain tumours: a retrospective cohort study. *The Lancet*. 2012;Published online June 7 DOI: 10.1016/S0140-6736(12)60815-0.
- [14] Moorin R, Forsyth R, Gibson D, Fox R. Radiation dosimetry assessment of routine CT scanning protocols used in Western Australia. *Journal of Radiological Protection*. 2013;33:295-312.
- [15] Sodickson A, Warden GI, Farkas CE, Ikuta I, Prevedello LM, Andriole KP, et al. Automated anatomy-specific CT radiation exposure extraction for quality assurance and radiation monitoring. *Radiology*. 2012;264:397-405.
- [16] Sodickson A. CT radiation risks coming into clearer focus. *British Medical Journal*. 2013;346:f3102.
- [17] Imaging and Performance and Assessment of CT scanners (ImPACT). CT patient dosimetry Excel spreadsheet. Imaging and Performance and Assessment of CT scanners evaluation centre of the DH Medicines and Healthcare products Regulatory Agency (MHRA), <http://www.impactscan.org>; 2001.
- [18] Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin* 1979;86:420-8.
- [19] McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychological Methods*. 1996;1:30-46.
- [20] Lin LI-K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989;45:255-68.
- [21] McBride. A proposal for strength-of-agreement criteria for Lin's Concordance Correlation Coefficient. NIWA Client Report : HAM2005-062. Hamilton, New Zealand: National Institute of Water & Atmospheric Research Ltd; 2005.
- [22] Passing H, Bablok W. A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in Clinical Chemistry, Part I. *Journal of Clinical Chemistry, Clinical Biochemistry*. 1983;21:709-20.
- [23] Bland JM, Altman DG. Measuring agreement in method comparison studies. *Statistical Methods in Medical Research* 1999;8:135-60.
- [24] Krouwer JS. Why Bland-Altman plots should use X, not (Y+X)/2 when X is a reference method. *Statistics in Medicine* 2008;27:778-80.

- [25] Committee to Assess Health Risks from Exposure to Low Levels of Ionizing Radiation. Health risks from exposure to low levels of ionizing radiation: BEIR VII-Phase 2. Board on Radiation Effects Research, Division on Earth and Life Studies, National Research Council.; 2006.
- [26] Mathews JD, Forsythe AV, Brady Z, Butler MW, Goergen SK, Byrnes GB, et al. Cancer risk in 680 000 people exposed to computed tomography scans in childhood or adolescence: data linkage study of 11 million Australians. *British Medical Journal*. 2013;346:f2360.
- [27] Institute of Physics and Engineering in Medicine (IPEM). Guidance on the establishments and use of diagnostic reference levels for medical x-ray examinations. IPEM Report No 88. York: Institute of Physics and Engineering in Medicine; 2004.
- [28] Friberg EG, Widmark A, Ryste Hauge IH. National collection of local diagnostic reference levels in Norway and their role in optimization of X-ray examinations. Osteras, Norway: Norwegian Radiation Protection Authority; 2008.
- [29] Kharuzhyk SA, Matskevich SA, Filjustin AE, Bogushevich EV, Ugolkova SA. Survey of computed tomography doses and establishment of national diagnostic reference levels in the Republic of Belarus. *Radiation Protection Dosimetry*. 2010;139:367-70.
- [30] European Commission: Dose Datamed Project. European Guidance on Estimating Population Doses from Medical X-Ray Procedures, Radiation Protection No.154. Chilton, Didcot, Oxfordshire: Health Protection Agency; 2008.
- [31] Holman CDJ. Anonymity & Research. Health Data and Biospecimen Law in Australia. Perth: Uniprint, The University of Western Australia; 2012.

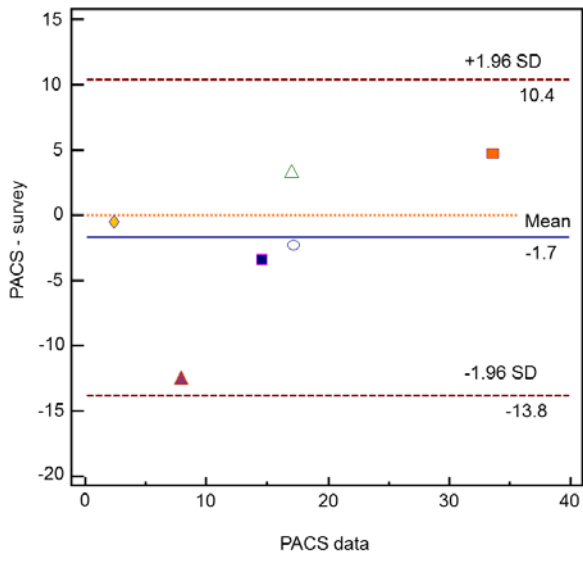
Figure Labels

Figure 1 Bland-Altman analysis of the agreement in effective dose (a) and organ radiation dose (b) using survey data compared with data extracted from departmental picture archival communication system for all CT examination types evaluated

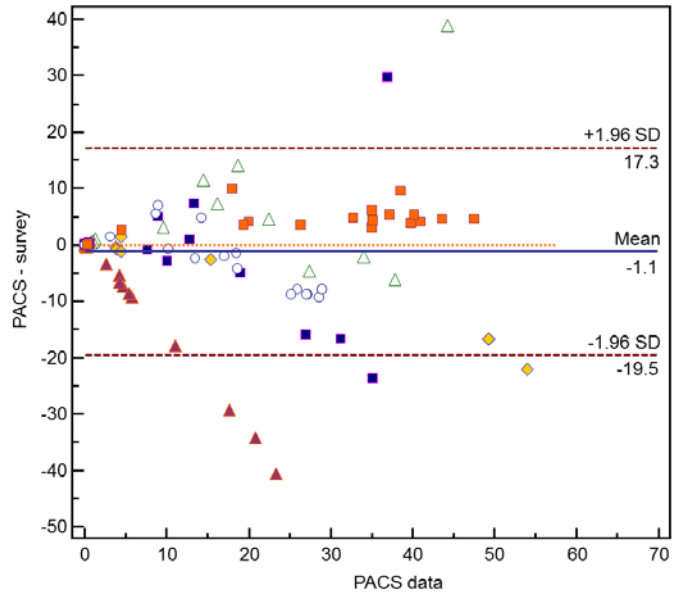
Figure 2 Bland-Altman analysis of the agreement in organ radiation dose produced using survey data compared with data extracted from departmental picture archival communication system according to type of CT examination

Figure 1.

A. Effective doses

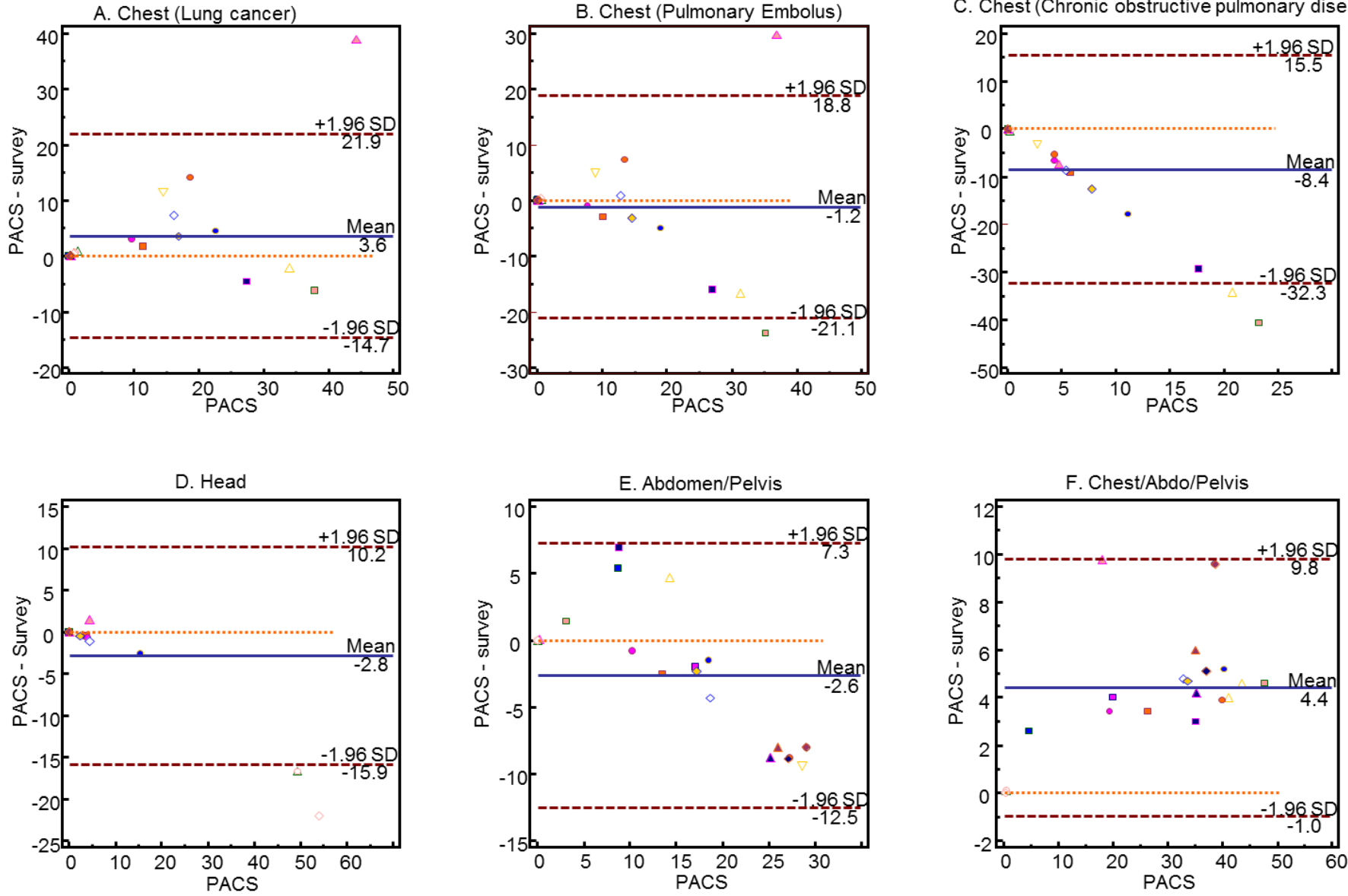


B. Organ radiation doses



CT examination type ◆ Head △ Chest (Lung cancer) ■ Chest (Pulmonary embolus) ▲ Chest (Chronic obstructive pulmonary disease)
 ○ Abdomen/Pelvis ■ Chest/Abdo/Pelvis

Figure 2 Bland-Altman analysis of the agreement in organ radiation dose produced using survey data compared with data extracted from departmental picture archival communication system according to type of CT examination



- Bladder
- Bone Marrow
- Bone Surface
- △ Brain
- Breast
- ▲ Colon
- ◆ Effective dose (mSv)
- ◇ Eye lenses
- Gonads
- Liver
- △ Lung
- ▲ Ovaries
- ◆ Prostate
- ◇ Remainder organs
- Salivary Glands (Brain)
- Skin
- ▽ Stomach
- Testes
- ▲ Thyroid
- ◆ Uterus
- Oesophagus

Table 1 Comparison of CT scanning parameters and radiation dosimetry reported using self-complete survey with data extracted from departmental picture archival communication system

Examination	Self-report survey data							Data captured from the departmental picture archive communication system						
	Anatomical Positions Top	Anatomical Positions Bottom	Seq [*]	kV [†]	Pitch [‡]	CTDIvol [§]	DLP	Number of cases	Seq [¶]	kV ^{**}	Pitch ^{††}	CTDIvol ^{‡‡} Mean (StDev)	DLP ^{§§} Mean (StDev)	
Head for stroke or trauma	Top of head	Base of skull	1	120	0.423	86.70	1256	20	1	120	0.423	64.6 (28.6)	1207.8 (647.3)	
Chest for lung cancer	Above lung apices	Base of lung/adrenals	1	120	1.078	26.60	1131	20	2	120	1.078	22.5 (11.1)	404.9 (458.6)	
Chest for pulmonary emboli	Above lung apices	Base of lung/adrenals	1	120	0.673	35.20	968	20	3	120	0.673	47.8 (23.2)	738.2 (329.8)	
Chest for chronic obstructive pulmonary disease	Above lung apices	Base of lung/adrenals	1	140	0.516	38.10	1048	20	2	140	0.516	13.7 (9.4)	489.0 (340.9)	
Abdomen/Pelvis for liver metastases	Diaphragm	Symphysis pubis	1	120	0.891	26.70	1226	20	1	120	0.891	19.6 (8.7)	903.4 (382.7)	
Chest, Abdomen and Pelvis for lymphoma	Above lung apices	Symphysis pubis	1	120	1.078	24.80	1720	10	2	120	1.078	27.6 (6.1)	431.6 (832.4)	

*Seq = Self-reported number of sequences (phases) usually undertaken for the examination.

†kV= Self-reported kilovoltage usually used or the examination.

‡Self-reported pitch (mm) usually used for the examination.

§CTDIvol = Self-reported average Volumetric CT Dose Index, measured in mGy for the examination.

||DLP = Self-reported average Dose Length Product, measured in mGy.cm for the examination.

¶Seq = Median number of sequences (phases) observed for the cases collected.

**Median kilovoltage observed across the cases collected.

††Median pitch (mm) observed across the cases collected.

‡‡Mean (standard deviation) of the Volumetric CT Dose Index (mGy) observed for the cases collected. Note where multiple sequences (phases) were observed the CTDIvol for the examination was the sum of the sequence specific CTDIvol.

§§Mean (standard deviation) of the Dose Length Product (mGy.cm) observed for the cases collected. Note where multiple sequences (phases) were observed the DLP for the examination was the sum of the sequence specific DLP.

|||Only 10 cases were identified over the data capture time frame.

Both data collections were based on examinations performed on the same model and make of CT scanner.

Table 2 Organ radiation dose (mGy) and whole body effective radiation dose (mSv) of CT scanning protocols obtained using self-complete survey and data extracted from departmental picture archival communication system as input to radiation dosimetry calculations

	Head (Stroke)			Chest (Lung Cancer)			Chest (Pulmonary Embolism)			High Resolution Chest (COPD)			Abdomen/Pelvis (Metastases)			Chest/ Abdo/Pelvis (Lymphoma staging)		
	Organ dose (mGy) [§]			Organ dose (mGy) [§]			Organ dose (mGy) [§]			Organ dose (mGy) [§]			Organ dose (mGy) [§]			Organ dose (mGy) [§]		
	Self-report*	PACS [†]	Diff [‡]	Self-report*	PACS [†]	Diff [‡]	Self-report*	PACS [†]	Diff [‡]	Self-report*	PACS [†]	Diff [‡]	Self-report*	PACS [†]	Diff [‡]	Self-report*	PACS [†]	Diff [‡]
Gonads	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.1	0.0	0.0	0.0	0.0	19.0	17.0	-2.0	16.0	20.0	4.0
Bone Marrow	4.3	3.8	-0.5	9.6	11.4	1.8	13.0	10.1	-2.9	15.0	5.8	-9.2	16.0	13.5	-2.5	23.0	26.4	3.4
Colon	0.0	0.0	0.0	0.2	0.8	0.7	0.2	0.4	0.2	0.3	0.1	-0.2	34.0	26.0	-8.0	29.0	35.0	6.0
Lung	0.2	0.2	0.0	36.0	34.0	-2.0	48.0	31.3	-16.7	55.0	20.8	-34.2	9.6	14.3	4.7	39.0	43.6	4.6
Stomach	0.0	0.0	0.0	3.0	14.5	11.5	4.0	9.0	5.0	6.0	2.7	-3.3	38.0	28.6	-9.4	37.0	41.0	4.0
Bladder	0.0	0.0	0.0	0.0	0.3	0.3	0.0	0.0	0.0	0.0	0.0	0.0	37.0	29.0	-8.0	29.0	38.6	9.6
Breast	0.0	0.0	0.0	32.0	27.4	-4.6	43.0	27.0	-16.0	47.0	17.7	-29.3	2.0	8.9	6.9	32.0	35.0	3.0
Liver	0.0	0.0	0.0	4.6	18.7	14.1	6.1	13.4	7.3	9.7	4.3	-5.4	36.0	27.2	-8.8	36.0	39.9	3.9
Oesophagus	0.1	0.1	0.0	44.0	37.9	-6.1	59.0	35.2	-23.8	64.0	23.2	-40.8	1.7	3.1	1.4	43.0	47.6	4.6
Thyroid	2.9	4.4	1.5	5.4	44.3	38.9	7.2	36.9	29.7	12.0	4.7	-7.3	0.1	0.2	0.1	8.2	18.0	9.8
Skin	4.6	3.9	-0.7	6.5	9.6	3.1	8.7	7.7	-1.0	11.0	4.3	-6.7	11.0	10.2	-0.8	16.0	19.4	3.4
Bone Surface	18.0	15.4	-2.6	18.0	22.5	4.5	24.0	19.0	-5.0	29.0	11.1	-17.9	20.0	18.5	-1.5	35.0	40.2	5.2
Brain	66.0	49.3	-16.7	0.3	1.3	1.0	0.4	0.6	0.3	0.6	0.2	-0.4	0.0	0.0	0.0	0.4	0.5	0.1
Salivary Glands (Brain)	66.0	49.3	-16.7	0.3	1.0	0.7	0.4	0.6	0.3	0.6	0.2	-0.4	0.0	0.0	0.0	0.4	0.5	0.1
Remainder organs	5.5	4.4	-1.1	8.8	16.1	7.3	12.0	12.9	0.9	14.0	5.4	-8.6	23.0	18.7	-4.3	28.0	32.8	4.8
<u>Other organs of interest</u>																		
Eye lenses	76.0	54.0	-22.0	0.3	0.7	0.4	0.0	0.5	0.5	0.6	0.3	-0.2	0.0	0.0	0.0	0.4	0.4	0.1
Testes	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.3	8.7	5.4	2.0	4.6	2.6
Ovaries	0.0	0.0	0.0	0.0	0.2	0.2	0.1	0.1	0.0	0.1	0.0	0.0	34.0	25.2	-8.8	31.0	35.2	4.2
Uterus	0.0	0.0	0.0	0.0	0.2	0.1	0.0	0.1	0.1	0.1	0.0	0.0	36.0	27.1	-8.9	32.0	37.1	5.1
Prostate	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	37.0	29.0	-8.0	29.0	38.6	9.6
Whole body Effective dose (mSv) 	2.9	2.4	-0.5	13.3	16.9	3.6	17.7	14.6	-3.1	20.4	7.8	-12.6	19.5	17.2	-2.3	28.9	33.6	4.7

*Organ and effective radiation dose calculated using the self-reported (survey) dose information.

†Mean of the organ and effective radiation dose effective doses calculated across the cases using case specific scanning and dose information.

‡Difference in radiation dose calculated using picture archival communication system (PACS) data compared with that calculated using self-report data. A negative number indicates radiation dose is lower using PACS data.

§Organ dose reported in mGy – the amount of radiation received by the organ specified not modified by the tissue weighting factor for the organ.

||Effective dose reported in mSv – The whole body effective dose is the sum of the modified (using ICRP 103 tissue weighting factor for each organ) organ dose. This is a measure of the cancer risk to the individual due to ionizing radiation taking into account both the type of radiation and the radio-sensitivity of each organ being irradiated.

Note anatomical start and stop positions used during calculation of the radiation dose for survey data were those reported on the survey. For the PACS data CT staff were interviewed and the reported anatomical start stop locations modified to match the average scan lengths for cases collected for each protocol.

Table 3 Evaluation of the agreement of organ dose and effective dose derived using picture archival communication systems (PACS) versus self-reported survey data

Intra class correlation coefficient (ICC)	All examinations						Head CT			Chest (Lung cancer)			Chest (PE)			Chest (COPD)			Abdo/Pelvis			Chest/Abdomen/Pelvis		
	Effective dose (mSv)		Organ doses (mGy)				Organ doses (mGy)			Organ doses (mGy)			Organ doses (mGy)			Organ doses (mGy)			Organ doses (mGy)					
	ICC*	95% CI	ICC*	95% CI	ICC*	95% CI	ICC*	95% CI	ICC*	95% CI	ICC*	95% CI	ICC*	95% CI	ICC*	95% CI	ICC*	95% CI	ICC*	95% CI	ICC*	95% CI		
PACS vs Survey																								
Single Measures [†]	0.81	0.18	0.97	0.84	0.78	0.89	0.95	0.86	0.98	0.75	0.47	0.89	0.78	0.53	0.91	0.58	0.14	0.81	0.91	0.76	0.97	0.94	0.19	0.99
Average Measures [‡]	0.90	0.30	0.99	0.92	0.88	0.94	0.97	0.92	0.99	0.86	0.64	0.94	0.88	0.69	0.95	0.73	0.25	0.90	0.95	0.86	0.98	0.97	0.32	0.99
PACS mod 1 vs Survey																								
Single Measures [†]	0.81	0.18	0.97	0.84	0.78	0.89	0.95	0.86	0.98	0.75	0.47	0.89	0.78	0.53	0.91	0.57	0.15	0.81	0.91	0.76	0.97	0.94	0.12	0.99
Average Measures [‡]	0.90	0.30	0.99	0.92	0.88	0.94	0.97	0.92	0.99	0.86	0.64	0.94	0.88	0.70	0.95	0.73	0.26	0.90	0.95	0.86	0.98	0.97	0.21	0.99
PACS mod 5 vs Survey																								
Single Measures [†]	0.83	0.30	0.97	0.85	0.79	0.89	0.94	0.85	0.98	0.76	0.50	0.90	0.79	0.53	0.91	0.57	0.15	0.81	0.93	0.83	0.97	0.96	0.86	0.98
Average Measures [‡]	0.91	0.46	0.99	0.92	0.88	0.94	0.97	0.92	0.99	0.87	0.67	0.95	0.88	0.70	0.95	0.73	0.26	0.90	0.96	0.90	0.99	0.98	0.92	0.99
Concordance correlation coefficient (ρ _c)	ρ _c	95% CI	ρ _c	95% CI	ρ _c	95% CI	ρ _c	95% CI	ρ _c	95% CI	ρ _c	95% CI	ρ _c	95% CI	ρ _c	95% CI	ρ _c	95% CI	ρ _c	95% CI	ρ _c	95% CI		
PACS vs Survey (ρ_c)	0.78	0.14	0.96	0.84	0.79	0.89	0.94	0.92	0.96	0.74	0.47	0.88	0.78	0.57	0.89	0.56	0.43	0.66	0.91	0.84	0.94	0.94	0.88	0.97
Precision (ρ)	0.81		0.86				1.00			0.77			0.82			1.00			0.98			0.99		
Accuracy (C _b)	0.96		0.98				0.94			0.96			0.94			0.56			0.93			0.95		
PACS mod 1 vs Survey (ρ_c)	0.78	0.14	0.96	0.84	0.79	0.89	0.94	0.92	0.96	0.74	0.48	0.88	0.78	0.57	0.89	0.55	0.43	0.66	0.91	0.84	0.94	0.94	0.87	0.97
Precision (ρ)	0.81		0.86				1.00			0.77			0.82			1.00			0.98			0.99		
Accuracy (C _b)	0.96		0.98				0.94			0.96			0.94			0.55			0.93			0.95		
PACS mod 5 vs Survey (ρ_c)	0.81	0.17	0.97	0.85	0.79	0.89	0.94	0.91	0.96	0.76	0.50	0.89	0.78	0.57	0.89	0.55	0.43	0.66	0.93	0.86	0.96	0.95	0.89	0.98
Precision (ρ)	0.84		0.87				1.00			0.78			0.82			0.99			0.97			0.97		
Accuracy (C _b)	0.97		0.98				0.94			0.97			0.95			0.56			0.96			0.99		

PACS: Radiation dose values calculated using PACS data (to one decimal place).

PACS mod 1: PACS dose values <= 1 reset to be equal to survey values.

PACS mod 5: PACS dose values <= 5 reset to be equal to survey values.

*The intra class correlation coefficient measures the degree of absolute agreement among the radiation dose estimated using the two measurement instruments providing input data to the radiation dose calculation (PACS vs Survey).

†Estimates the reliability of single measurements.

‡Estimates the reliability of the averages of all measurements.

ρ_c is the concordance correlation coefficient. It evaluates the degree to which pairs of observations (ie radiation dose calculated using PACS and survey data) fall on the 45° line (ie line of absolute agreement) through the origin.

ρ is the Pearson correlation coefficient, which measures how far each observation deviates from the best-fit line, and is a measure of precision.

C_b is a bias correction factor that measures how far the best-fit line deviates from the 45° line through the origin, and is a measure of accuracy.

Table 4 Passing and Bablok regression testing the equality of the organ and effective doses calculated based on picture archival communication system (PACS) versus self-reported survey CT examination data

All examinations				
	Effective dose (mSv)		Organ doses (mGy)	
Regression equation	y = 5.98 + 0.72 x		y = 0.00 + 1.06 x	
	Point estimate	95% CI	Point estimate	95% CI
Systematic differences^{A*}	5.98	-310.00 to 23.50	0.00	-0.031 to 0.00
Proportional differences^{B†}	0.72	-0.40 to 20.67	1.06	0.91 to 1.17
Random differences^{RSD‡}	4.60	-9.02 to 9.02	6.45	-12.65 to 12.65
Cusum test for linearity	P=0.99		P=0.02 [§]	
Head CT				
	Organ doses (mGy)		Chest (Lung cancer) Organ doses (mGy)	
Regression equation	y = 0.00 + 1.22 x		y = -0.12 + 0.60 x	
	Point estimate	95% CI	Point estimate	95% CI
Systematic differences^{A*}	0	0.00 to 0.00	-0.12	-0.42 to -0.01
Proportional differences^{B†}	1.22	1.13 to 1.34	0.60	0.27 to 0.89
Random differences^{RSD‡}	2.00	-3.92 to 3.92	7.80	-15.29 to 15.29
Cusum test for linearity	P=0.05		P=0.52	
Chest (PE)				
	Organ doses (mGy)		Chest (COPD) Organ doses (mGy)	
Regression equation	y = -0.12 + 1.16 x		y = 0.00 + 2.64 x	
	Point estimate	95% CI	Point estimate	95% CI
Systematic differences^{A*}	-0.12	-0.47 to 0.00	0.00	-0.10 to 0.056
Proportional differences^{B†}	1.16	0.68 to 1.54	2.64	2.57 to 2.69
Random differences^{RSD‡}	6.89	-13.51 to 13.51	0.28	-0.54 to 0.54
Cusum test for linearity	P=0.25		P=0.07	
Abdomen/Pelvis				
	Organ doses (mGy)		Chest/Abdomen/Pelvis Organ doses (mGy)	
Regression equation	y = -2.13 + 1.35 x		y = -1.35 + 0.91 x	
	Point estimate	95% CI	Point estimate	95% CI
Systematic differences^{A*}	-2.13	-5.04 to -0.07	-1.35	-3.50 to -0.04
Proportional differences^{B†}	1.35	1.24 to 1.51	0.91	0.87 to 0.98
Random differences^{RSD‡}	1.99	-3.90 to 3.90	1.82	-3.56 to 3.56
Cusum test for linearity	P=0.10		P=0.36	

Regression equation: the regression equation with the calculated values for A and B according to Passing & Bablok (1983) [23].

^{A*}Systematic differences. The intercept A is a measure of the systematic differences between the two methods. The 95% confidence interval for the intercept A tests the hypothesis that A=0. This hypothesis is accepted if the confidence interval for A contains the value 0. If the hypothesis is rejected, then it is concluded that A is significantly different from 0 and the methods differ by a constant amount.

^{B†}Proportional differences. The slope B is a measure of the proportional differences between the two methods. The 95% confidence interval for the slope B tests the hypothesis that B=1. This hypothesis is accepted if the confidence interval for B contains the value 1. If the hypothesis is rejected, then it is concluded that B is significantly different from 1 and there is a proportional difference between the two methods.

^{‡RSD}Random differences. The residual standard deviation (RSD) is a measure of the random differences between the two methods. 95% of random differences are expected to lie in the interval -1.96 RSD to +1.96 RSD. If this interval is large, the two methods may not be comparable.

Linear model validity: the Cusum test for linearity is used to evaluate how well a linear model fits the data. [§]Indicates significant deviation from linearity (ie p value <0.05).