Published in final edited form as:

Acad Radiol. 2014 June ; 21(6): 733–742. doi:10.1016/j.acra.2014.02.006.

Thymic measurements in pathologically proven normal thymus and thymic hyperplasia: Intra- and interobserver variability

Tetsuro Araki, M.D.¹, Lynette M. Sholl, M.D.³, Victor H. Gerbaudo, Ph.D., M.S.H.C.A.¹, Hiroto Hatabu, M.D., Ph.D.¹, and Mizuki Nishino, M.D.^{1,2}

¹ Department of Radiology, Brigham and Women's Hospital 75 Francis St. Boston MA, 02215, USA 450 Brookline Ave. Boston MA, 02215, USA

² Department of Imaging, Dana-Farber Cancer Institute, Harvard Medical School 75 Francis St. Boston MA, 02215, USA 450 Brookline Ave. Boston MA, 02215, USA

³ Department of Pathology Brigham and Women's Hospital 75 Francis St. Boston MA, 02215, USA

Abstract

Objective: Determine the intra- and interobserver variability of thymic measurements on computed tomography (CT) in patients with pathological diagnosis of thymic hyperplasia or normal thymus.

Materials and Methods: Thirty-three patients with pathological diagnosis of thymic hyperplasia (n=25) or normal thymus (n=8) who had identifiable thymus gland on CT were retrospectively studied. Two radiologists independently measured thymic size and CT attenuation. Concordance correlation coefficients (CCCs) and Bland-Altman plots were used to assess intra- and interobserver agreements.

Results: The intra- and interobserver agreements of thymic diameters and the lobe length were moderate, with CCCs ranging 0.73-0.89 and 0.72-0.81, respectively. Higher agreement was noted among patients whose measurements were performed on the same CT image in two independent measurements, with intraobserver CCC 0.95 for diameters and length. After providing readers with an instruction for consistent selection of CT image for measurements, the intra- and interobserver agreement improved, resulting in CCCs ranging 0.81-0.92 and 0.77-0.85 for diameters and length, respectively. Thymic lobe thickness had the least agreement. CT attenuation measurements were highly reproducible, with CCCs ranging 0.88-0.97. In patients with thymic CT attenuation >30HU, the attenuation measurements were more reproducible with narrower 95% limits of agreement.

© 2014 The Association of University Radiologists. Published by Elsevier Inc. All rights reserved

Corresponding author: Mizuki Nishino, M.D. Department of Radiology, Brigham and Women's Hospital and Dana-Farber Cancer Institute 450 Brookline Ave. Boston MA, 02215, USA Mizuki_Nishino@DFCI.HARVARD.EDU Phone: 617-582-7163 Fax: 617-582-8574.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conclusion: Thymic size measurements had moderate to high intra- and interobserver agreement, when the instruction for consistent selection of images were provided to the readers. CT attenuation was highly reproducible, with higher reproducibility for thymic glands with >30HU. Awareness of thymic measurement variability is necessary when interpreting measured values of normal thymus and thymic pathology on CT.

Keywords

Thymus; computed tomography; measurement; intraobserver variability; interobserver variability

INTRODUCTION

Thymus demonstrates unique morphological transition according to age, reaching its maximum weight at puberty and subsequently undergoing “involution”, which is defined as a decrease in size and weight with advancing age [1-3]. The assessment of thymic size on imaging in relation to patient’s age is an important component of radiologic interpretation of thymus glands and thymic pathology. Measurements of normal and abnormal thymus glands using computed tomography (CT) were extensively studied in early 1980s [4-8]. Baron et al studied 154 patients without clinical basis to suspect thymic abnormality who underwent chest CT, and described width, thickness and attenuation of the normal thymus in different age groups ranging from 6 to over 50 years old [4]. Francis et al revisited the age-related changes of thymic size in 309 patients without clinical evidence of thymic abnormality and 23 patients with thymic abnormalities, and described transverse and anterior-posterior (AP) diameters, craniocaudal extent, and thickness of thymus gland in each age group [6]. As demonstrated by these earlier studies, given the unique feature of involution over time, thymic measurements should be interpreted relative to the “normal” measurement values in the corresponding age group of patients.

Ackman et al recently studied the AP diameter, lobe thickness, CT attenuation, and morphological features of thymus in subjects who were 20-30 years old and have presumably normal thymus, and reported that the thymus gland had significantly higher attenuation and more commonly had quadrilateral shape in females compared to males [9]. In our previous study of 31 patients with pathologically proven thymic hyperplasia, the size of the thymic gland with hyperplasia was significantly larger compared to the age-matched mean values of normal thymus [10]. The study also demonstrated that the thymic CT attenuation higher than 41.2 HU on contrast-enhanced CT is suggestive of lymphoid hyperplasia rather than true hyperplasia. The size and the CT attenuation of thymic gland is an important factor to define “normal” thymus and diagnose thymic pathology [11, 12].

The thymus has a unique morphology with a bilobed configuration, which can manifest on imaging as pyramidal, triangular, arrowhead or quadrilateral shape [4, 6, 9, 10]. Such morphology requires dedicated measurement methods specific for thymus, as described by Baron et al and Francis et al [4, 6], that are more complicated than the ordinal measurement methods such as the longest diameter measurements or bidimensional measurements that are commonly used. It is well known that measurements on CT are subject to intra- and interobserver measurement variability, and this has been very well studied in tumor

measurements for the purpose of response assessment to anti-cancer therapy [13, 14]. While intra- and interobserver reproducibility of ultrasonographic measurements of AP and transverse diameters in infantile thymus have been examined [15], no prior reports have systematically investigated the reproducibility of thymic measurements on CT in adults. Given the importance of size and CT attenuation measurements in imaging diagnosis of thymic pathology and the specific method of thymic size measurements due to its unique morphology, studying the variability of size and CT attenuation measurements of the thymus is an important next step to further improve the application of the results of investigations of thymic imaging.

The purpose of the study is to determine the intra- and interobserver variability of thymic measurements on chest CT in patients with pathologically proven normal thymus and thymic hyperplasia.

MATERIALS AND METHODS

Subjects

The original cohort was identified from the retrospective review of the pathology database for patients who underwent thymectomy between 2001 and 2012, and included 11 patients with pathologically proven normal thymus and 29 patients with pathologically proven thymic hyperplasia, totaling 40, who had preoperative CT imaging available for review. CT images of these patients were evaluated for visibility of thymus by consensus review of two board-certified chest radiologists (M.N. and H.H.). Seven patients (3 patients with normal thymus and 4 patients with thymic hyperplasia) had no identifiable thymus gland on CT (i.e., their thymus glands were indistinguishable from anterior mediastinal fat tissue and therefore not measurable) and were excluded from this measurement study. Therefore, a total of 33 patients, consisting of 8 patients with pathologically confirmed normal thymus (2 males and 6 females, median age 25.5 years, age range: 22-47) and 25 patients with pathologically confirmed thymic hyperplasia (7 males and 18 females, median age: 35 years, age range: 19-68) were included in this study and retrospectively studied with the approval of the institutional review board.

Initial sets of thymic measurements

Thymic size and CT attenuation measurements were independently performed by two board-certified radiologists (T.A., Radiologist 1; M.N., Radiologist 2). The size measurements included the transverse and AP diameters of the gland and the length and thickness of each lobe, and were performed using a caliper-type measurement tool on the Picture Archiving Communication System (PACS) workstation (Centricity, GE Healthcare, Waukesha, WI) (Figure 1). The CT attenuation of the thymus gland was measured using an oval region of interest covering the maximum area of thymus gland excluding the surrounding fat. All images were reviewed on PACS, using a mediastinal window setting (level = 50; width = 350). Radiologist 1 performed the measurements twice with a 2-week interval, and Radiologist 2 performed the measurements once. For each set of measurements, a total of 33 patients' chest CT scans were presented in a different random order, and the readers were blinded for pathological diagnosis as to normal thymus or hyperplasia.

For the initial sets of measurements, the radiologists were instructed to perform measurements as shown in Figure 1, and were not given any additional specific instruction for the selection of the most appropriate axial CT image for thymic measurements. The series and image numbers of the axial CT image used for measurements were recorded for each patient during the two measurements by Radiologist 1.

Second sets of measurements after providing the instruction for CT image selection

After both radiologists completed the initial sets of measurements (two for Radiologist 1 and one for Radiologist 2), the series and image numbers of CT images used for the two measurements by Radiologist 1 were reviewed by the two radiologists (M.N. and T.A.). The cases measured on the same image during two independent measurements and the cases measured on the different images were compared, to develop an instruction for the selection of the appropriate axial CT image to measure thymus gland. The instruction was as follows: “select the axial CT image demonstrating the maximum AP diameter of the thymus gland to help consistent selection of the representative image for thymic measurements”. The reason behind this instruction was, given the triangular and bilobed configuration of the thymus, the maximum AP diameter is demonstrated on the image showing the maximum length of one of the lobes (usually left lobe since it extends further than the right lobe [4]), which is measured as the maximum distance between the apex and the most posterior part of the lobe (Fig 1). After being provided with the instruction, the two radiologists performed the second sets of thymic measurements, following the same procedures described in the initial sets. Radiologist 1 performed the measurements twice, 2 weeks apart, and Radiologist 2 performed the measurements once. The second sets of measurements were performed 2 months after the completion of the initial set.

Statistical Analysis

Measurement variability was assessed using the concordance correlation coefficient (CCC), mean relative difference, and 95% limits of agreement. CCC was used to assess reproducibility of two measurements, as described previously [16]. Assuming two measurements have mean u_1 and u_2 , with variance σ_1^2 , σ_2^2 , and covariance σ_{12} , $CCC = (2\sigma_{12})/[\sigma_1^2 + \sigma_2^2 + (u_1 - u_2)^2]$. CCC is composed of a measure of precision (how far each pair of measurements deviates from the best-fit line through the data) and a measure of accuracy (the distance between the best-fit line and the 45 degree line through the origin). A CCC value of 1 indicates perfect agreement and -1 indicates perfect reversed agreement [16, 17]. Bland-Altman plots were used to visually demonstrate agreement in the two measurements, plotting the absolute difference between the two measurements ($M_1 - M_2$; M_1 =Measurements 1, M_2 =Measurement 2) over the average of the two measurements [18, 19]. In each set of measurements, two measurements by Radiologist 1 were used to assess intraobserver variability, and the first measurement by Radiologist 1 and the measurement by Radiologist 2 were used to assess interobserver variability.

RESULTS

Intraobserver and interobserver variability of the initial sets of measurements

Table 1 summarizes of the intraobserver variability of the thymic size measurements, including transverse and AP diameters of the thymus gland, the length and thickness of each lobe, and CT attenuation of the gland. CCCs for transverse and AP diameters, and the length of each lobe ranged from 0.73 to 0.89, indicating moderate agreements [16]. The measurements of lobe thickness had the least agreement, with CCC of 0.61 for the right lobe and 0.65 for the left lobe. CCC for the CT attenuation measurement was 0.91, indicating higher agreement compared to size measurements.

Table 2 summarizes of the interobserver variability between measurements by two radiologists. CCCs of the transverse and AP diameters, and length of each lobe ranged from 0.72 to 0.81, indicating moderate agreements. The measurements of lobe thickness had the least agreement, with CCC of 0.56 for the right lobe and 0.59 for the left lobe. The CT attenuation measurements had CCC of 0.90, which was higher than CCCs of size measurements.

It was indicated that thymic gland with higher CT attenuation may have less variability in attenuation measurements. When divided into two groups at 30HU based on the attenuation greater than 30HU had the 95% limits of interobserver agreement were (−11.9, 14.1), while 23 patients with the average attenuation of less than 30HU had the 95% limits of agreement of (−37.7, 36.9).

Variability and the selection of representative CT image for measurement

To further assess the impact of CT image selection for thymic measurements on reproducibility, the series and image numbers of the axial CT images used in the two measurements by Radiologist 1 were compared. The two independent measurements were performed on the same CT image in 16 patients (49%), and were on one image above or below in 5 patients (15%), and two or more images above or below in 12 (36%). CCCs were higher in 16 patients whose two measurements were performed on the same image (range: 0.76 - 0.99), compared to those in 17 cases where the two measurements were done on different images (range: 0.36 - 0.86) (Table 3). Notably, the size measurements except for thickness had CCCs of 0.95 or above, demonstrating high reproducibility in the 16 patients whose two measurements were performed on the same image (Table 3). The 95% limits of agreement for the 16 patients were narrower, mostly within the range of ± 10 mm.

Intraobserver and interobserver variability of the second sets of measurements

The second sets of measurements performed after the instruction of selecting the axial CT image demonstrating the maximum AP diameter of the thymus gland were compared. On two independent measurements by Radiologist 1, the measurements were performed on the same image in 18 patients (55%), and on the same image or on one image above or below in 28 patients (85%). On two independent measurements by Radiologists 1 and 2, the measurements were done on the same image in 12 patients (36%), and were on the same image or one image above or below in 24 patients (73%).

The intra- and interobserver variability of the second sets of measurements are shown in Tables 4 and 5, respectively. The intraobserver agreement has improved on the second set compared to the first set, with CCCs ranging from 0.81-0.92 for size measurements (Table 4). The interobserver agreement for size measurements has also slightly improved in most of the measures, demonstrating moderate agreement with the CCC ranging from 0.77-0.85, except for the left lobe thickness with CCC of 0.57 (Figs. 2 and 3).

DISCUSSION

The present study investigated the intra- and interobserver variability of the thymic size and attenuation measurements on CT in patients with normal thymus or thymic hyperplasia on pathology. The intra- and interobserver agreements of size measurements were moderate, except for the lobe thickness which has the least agreement, in the initial sets of measurements performed without instruction for image selection. After providing the instruction on how to select appropriate image for measurements, the intraobserver agreements improved and became moderate to high for all size measures. Interobserver agreements also improved to moderate for all measures except for the left lobe thickness. CT attenuation measurements showed high intra- and interobserver agreements. Awareness of the measurement variability of thymus gland and the importance of consistent selection of the appropriate image for measurements are essential for radiologists during radiologic interpretation of thymus gland and its pathology.

Both intra- and interobserver agreements for the initial sets of thymic size measurements, before providing the instruction for CT image selection, were moderate in diameters of the gland and lobe length, and were poor for lobe thickness. However, in the 16 patients whose measurements were performed on the same CT image during the two independent measurements, the agreement for diameters and length were high with CCCs of 0.95 or greater, and the agreement for thickness was moderate. The results indicated that the measurement technique itself could be highly reproducible for the most part, especially for diameters and length, as long as the same CT image is consistently selected to perform the measurements. However, in the initial sets of measurements, the same CT image was selected for the two measurements by a radiologist in only half of the patients. In other half of the patients, different images were used for the two measurements, resulting in increased variability. It was clear that the consistent selection of the same image is essential for the reproducible thymic measurements, and the methodology of selecting the appropriate CT image for thymic measurements needs to be established.

After reviewing the cases with two measurements performed on the same image versus the cases with measurements performed on different images, the selection of the axial CT image demonstrating the maximum AP diameter of the thymus gland was proposed to ascertain consistency of the selection of the representative image for thymic measurements. The instruction was developed by careful review of the cases and the images used for two independent measurements by Radiologist 1 in the first set, because no such instruction or guideline was provided in the published literature. In the second sets of measurements performed after providing the instruction, the number of cases measured on the same image increased from 49% to 55%, and the number of cases measured on the same image or on one

image above or below, increased from 64% to 85%, in two independent measurements by Radiologist 1. Between the measurements by two radiologists after the instruction, 73% of cases were measured either on the same image or on the image one above or below. The results indicate the instruction helped to improve the consistent selection of CT images for thymic measurements.

In the second sets of measurements, moderate to high intraobserver agreements were observed. Interobserver agreement was also moderate except for left lobe thickness. These results highlight the importance of consistent image selection for thymic size measurements in order to improve the reproducibility of the measurements. McErlean et al mentioned the importance of image selection in their recent study of measurement variability in oncologic population [20]. However, the impact of image selection was not well investigated in their study, because readers were allowed to select a different image as an option, and mostly preselected images were used for measurements. Further investigation to establish a robust and reliable methodology for consistent selection of the appropriate CT image for thymic measurements is needed. In addition, when assessing the change of thymic size on consecutive scans, for example in cases of suspected rebound hyperplasia, it is essential to perform measurements size by size on two (or more) scans to minimize change due to measurement variability and to accurately detect true biological change of the gland.

The measurement of the lobe thickness had the least agreement, with intraobserver CCC of 0.61 and 0.65, and interobserver CCC of 0.56 and 0.59 for each lobe in the initial sets of measurements. The agreements improved to the moderate level in the 2nd sets, except for the interobserver agreement for left lobe thickness which retained CCC of 0.57. The low intra- and interobserver agreements for thymic lobe thickness are likely due to the complex morphology of the gland, and the difficulty of finding the standardized and accurate way to identify the maximum lobe thickness on axial images, and performing precise measurements using a caliper type measurement tool on PACS. Our results indicate that the thymic lobe thickness, although often described in the previous literature studying normal and pathological thymus [4-6, 9, 10], cannot be a reliable marker to differentiate normal and abnormal thymus on imaging, given the wide variability of the measurements.

In order to provide practical results for clinical application, we further assessed the intra- and interobserver variability of the thymic size measurements using the Bland-Altman plots, which are commonly used in the investigations of measurement variability [18, 19]. The plots visually demonstrate the differences of two measurements, plotted over the average of the two, with the 95% limits of agreement. The awareness of the range of measurement variability is important for thymic size measurements, which often contribute to radiological decision-making. The knowledge is particularly useful when radiologists face cases of “enlarged thymus” based on the imaging studies performed at other institutions, or are asked to evaluate thymic enlargement over time.

It is also worth pointing out that the difference of size measurements in millimeters was used in the Bland-Altman plots in the present study, instead of the relative difference in percentage (%), which is often used in the studies of variability of tumor measurements [19, 21]. This was because the interpretation of thymic size is based on the actual values, rather

than evaluating the percent change over time. On the other hand, in oncologic imaging, the relative difference (%) is used, because the purpose of tumor measurements is to assess changes of tumor burden over time during therapy to document the results of cancer treatment.

The CT attenuation measurements of the thymus had high intra- and interobserver agreements, with higher CCCs (0.91 for intraobserver and 0.90 for interobserver agreements) than those for size measurements. However, relatively wide 95% limits of agreements were noted both for intraobserver agreement (−24.5, 33.4 HU) and interobserver agreement (−31.7, 31.8 HU). Thymic glands with CT attenuation higher than 30 HU had less variability in attenuation measurements compared to those with CT attenuation less than 30HU. The result is suggestive that a thymic CT attenuation higher than 30HU is more reliable with less variability compared to that of lower values, which may be explained by heterogeneous fat contents in thymic glands, leading to higher variability.

While there are many previous reports on thymic size measurements [4, 6, 9, 10], few studies investigated the measurement variability of adult thymus glands on CT. Therefore, it is somewhat difficult to compare the present results with the prior reports. The agreements of thymic measurements, especially in the initial sets of measurements, were lower than those reported in tumor measurements, performed in lung cancer patients, where CCCs were mostly above 0.9 and often close to 1 [14]. This is expected because, in tumor measurements, the lesions were chosen based on the suitability for accurate and consistent measurements on follow-up scans, and often included lesions such as discrete lung nodules, rather than apyramidal shaped organ like the thymus. Unidimensional (the longest diameter) or bidimensional (the longest diameter and the longest perpendicular diameter) measurements are performed in tumor measurements, as opposed to the diameters of the gland as well as the thickness and length of the lobes as performed in thymic measurements. Measurement of thymus is far more complicated than just measuring the longest diameters of a discrete nodule or mass.

Furthermore, in tumor measurements, the target lesions are indicated by providing the series and image numbers to readers, so that the measurements can be performed on the same lesions on baseline and follow-up scans [13, 14, 17]. Indicating a target lesion with series and image numbers is an appropriate and necessary approach for the assessment of therapeutic tumor response; however, in thymic measurements in clinical practice, radiologists most likely select an axial plane that seems to best represent the gland based on their own visual assessment, and perform measurements if necessary. For this reason, we chose not to provide the series and image numbers in advance in this thymic measurement study; instead, we proposed the methodology as to how to select the optimal axial CT image for thymic measurements, which resulted in improved measurement reproducibility.

The limitations of the present study include its retrospective design and a small number of patients. The number of patients with pathologically confirmed normal thymus is particularly limited. While it is possible to expand the cohort without requiring pathological confirmation, such approach will provide us the risk of including subjects with apparently normal thymus harboring pathology and excluding subjects with apparently abnormal

thymus but with normal histology. Since this is the first study which aimed to define intra- and interobserver variability of thymic measurements on CT in adult subjects, we chose to investigate a small but pathologically confirmed cohort of patients. The results of the present study should be interpreted in the context of the study design.

In conclusion, the present study provided the intra- and interobserver variability of thymic size and attenuation measurements on CT. Providing readers with an instruction for selection of CT image for thymic measurements improved the intra- and interobserver agreement, resulting in moderate to high reproducibility for thymic diameters and length. The lobe thickness consistently demonstrated lowest agreement. CT attenuation measurements of the thymus gland were moderately to highly reproducible, with higher reproducibility for the thymic gland measuring higher than 30HU. Awareness of the thymic measurement variability and the importance of image selection are essential when interpreting measured values of normal thymus and thymic pathology on CT imaging.

Acknowledgments

The investigator, M.N. was supported by 1K23CA157631 (NCI).

References

1. Nishino M, Ashiku SK, Kocher ON, et al. The Thymus: A Comprehensive Review. *Radiographics*. 2006; 26:335–348. [PubMed: 16549602]
2. Nasser F, Eftekhari F. Clinical and Radiologic Review of the Normal and Abnormal Thymus: Pearls and Pitfalls. *Radiographics*. 2010; 30:413–428. [PubMed: 20228326]
3. Steinmann, G. The Human Thymus. Springer; New York City, New York: 1986. Changes in the human thymus during aging; p. 43-88.
4. Baron R, Lee J, Sagel S, Peterson R. Computed tomography of the normal thymus. *Radiology*. 1982; 142:121–125. [PubMed: 7053521]
5. Baron R, Lee J, Sagel S, Levitt R. Computed tomography of the abnormal thymus. *Radiology*. 1982; 142:127–134. [PubMed: 7053522]
6. Francis I, Glazer G, Bookstein F, Gross B. The thymus: reexamination of age-related changes in size and shape. *American journal of roentgenology*. 1985; 145:249–254. [PubMed: 3875220]
7. Moore AV, Korobkin M, Olanow W, et al. Age-related changes in the thymus gland: CT-pathologic correlation. *American Journal of Roentgenology*. 1983; 141:241–246. [PubMed: 6603111]
8. Keats, TE.; Siström, C. 7th. Mosby Inc; Maryland Heights, Missouri: 2001. Atlas of radiologic measurement.
9. Ackman JB, Kovacina B, Carter BW, et al. Sex difference in normal thymic appearance in adults 20-30 years of age. *Radiology*. 2013; 268:245–253. [PubMed: 23440318]
10. Araki T, Sholl LM, Gerbaudo VH, Hatabu H, Nishino M. Imaging characteristics of pathologically proven thymic hyperplasia: identifying features that can differentiate true from lymphoid hyperplasia. *AJR American Journal of Roentgenology*. 2014; 202:471–478. [PubMed: 24555583]
11. Shimosato, Y.; Mukai, K. 3rd. Vol. 21. Armed Forces Institute of Pathology; Washington, DC: 1997. Tumors of the mediastinum: atlas of tumor pathology; p. 158-168.
12. Lavini, C. Springer; New York City, New York: 2008. Thymus gland pathology: clinical, diagnostic, and therapeutic features.
13. Zhao B, James LP, Moskowitz CS, et al. Evaluating Variability in Tumor Measurements from Same-day Repeat CT Scans of Patients with Non-Small Cell Lung Cancer. *Radiology*. 2009; 252:263–272. [PubMed: 19561260]

14. Nishino M, Guo M, Jackman DM, et al. CT tumor volume measurement in advanced non-small-cell lung cancer: performance characteristics of an emerging clinical tool. *Academic radiology*. 2011; 18:54–62. [PubMed: 21036632]
15. Hasselbalch H, Nielsen MB, Jeppesen D, Pedersen JF, Karkov J. Sonographic measurement of the thymus in infants. *European radiology*. 1996; 6:700–703. [PubMed: 8934137]
16. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989; 45:255–268. [PubMed: 2720055]
17. Nishino M, Jackman DM, Hatabu H, et al. New Response Evaluation Criteria in Solid Tumors (RECIST) guidelines for advanced non-small cell lung cancer: comparison with original RECIST and impact on assessment of tumor response to targeted therapy. *AJR American journal of roentgenology*. 2010; 195:W221–228. [PubMed: 20729419]
18. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *Statistician*. 1983; 32:307–317.
19. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986; 1:307–310. [PubMed: 2868172]
20. McErlan A, Panicek DM, Zabor EC, et al. Intra- and Interobserver variability in CT measurements in oncology. *Radiology*. 2013; 269:451–459. [PubMed: 23824993]
21. Dewitte K, Fierens C, Stockl D, Thienpont LM. Application of the Bland-Altman plot for interpretation of method-comparison studies: a critical investigation of its practice. *Clinical chemistry*. 2002; 48:799–801. author reply 801-792. [PubMed: 11978620]

Fig. 1.

Thymic size measurements including length, thickness, and transverse and anterior-posterior (AP) diameters. With modification from Reference 10 (Araki T, Sholl LM, Gerbaudo VH, Hatabu H, Nishino M. Imaging characteristics of pathologically proven thymic hyperplasia: identifying features that can differentiate true from lymphoid hyperplasia. *AJR American Journal of Roentgenology* 2014; 202:471-478)

Fig. 2.

Bland-Altman plots demonstrating intraobserver variability of the second set of thymic measurements after providing the instruction for CT image selection. (Measurements of right lobe length and thickness (c, e) were performed in 32 cases, excluding one patient with round thymus gland located in the left side of the anterior mediastinum, which was considered to represent the left lobe.)

Fig. 3.

Bland-Altman plots demonstrating interobserver variability of the second set of thymic measurements after providing the instruction for CT image selection. (Measurements of right lobe length and thickness (c, e) were performed in 32 cases, excluding one patient with round thymus gland located in the left side of the anterior mediastinum, which was considered to represent the left lobe.)

Table 1Intraobserver variability of thymic measurements: the 1st set of measurements

	Intraobserver CCC [95% CI]	Mean relative difference (mm)	95% limits of agreement (mm)
Transverse diameter	0.73 [0.53-0.85]	-2.6	-19.4, 14.1
AP diameter	0.80 [0.63-0.89]	-0.3	-20.5, 19.9
Right lobe length	0.89 [0.79 -0.94]	-1.8	-12.7, 9.1
Left lobe length	0.76 [0.57 - 0.87]	-1.9	-23.1, 19.2
Right lobe thickness	0.61 [0.35 - 0.78]	0.7	-11.2, 12.6
Left lobe thickness	0.65 [0.41 - 0.81]	0.2	-9.0, 9.3
CT attenuation (HU)	0.91 [0.82- 0.95]	4.5 (HU)	-24.5, 33.4 (HU)

AP: Anterior-posterior

CI: confidence interval

HU: Hounsfield unit

Table 2

Interobserver variability of thymic measurements: the second sets of measurements

	Interobserver CCC [95% CI]	Mean relative difference (mm)	95% limits of agreement (mm)
Transverse diameter	0.81 [0.66- 0.90]	-1.2	-14.6, 12.2
AP diameter	0.72 [0.52 - 0.85]	-1.3	-24.8, 22.2
Right lobe length	0.80 [0.66 - 0.89]	-5.9	-17.0, 5.1
Left lobe length	0.73 [0.53 - 0.85]	0.2	-21.9, 22.4
Right lobe thickness	0.56 [0.29 - 0.75]	-2.8	-16.6, 11.0
Left lobe thickness	0.59 [0.34 - 0.76]	-2.1	-12.6, 8.4
CT attenuation	0.90 [0.80 - 0.95]	0.1 (HU)	-31.7, 31.8 (HU)

AP: Anterior-posterior

CI: confidence interval

HU: Hounsfield unit

Table 3

Intraobserver variability of thymic measurements in matched and unmatched CT image slice for measurements

	16 patients measured on the same image			17 patients measured on the different images		
	Interobserver CCC [95% CI]	Mean relative difference (mm)	95% limits of agreement (mm)	Interobserver CCC [95% CI]	Mean relative difference (mm)	95% limits of agreement (mm)
Transverse diameter	0.95 [0.87-0.98]	-1.5	-8.8, 5.8	0.52 [0.14-0.76]	-3.7	-26.1, 18.7
AP diameter	0.99 [0.96- 1.00]	-0.9	-6.3,4.4	0.47 [0.07- 0.74]	0.3	-27.8, 28.4
Right lobe length	0.96 [0.90- 0.99]	-1.4	-7.7, 4.8	0.80 [0.54- 0.92]	-2.1	-16.4, 12.2
Left lobe length	0.97 [0.93- 0.99]	-2.8	-8.1,2.5	0.47 [0.04 - 0.76]	-1.1	-30.5,28.3
Right lobe thickness	0.76 [0.48- 0.90]	-0.2	-10.7, 10.3	0.36 [-0.11-0.70]	1.6	-11.7, 14.9
Left lobe thickness	0.80 [0.53-0.93]	-0.8	-8.1,6.6	0.48 [0.04-0.76]	1.1	-9.4, 11.6
CT attenuation (HU)	0.99 [0.98-1.00]	0.6 (HU)	-6.9,8.1 (HU)	0.86 [0.66-0.94]	8.1 (HU)	-30.7, 46.9 (HU)

AP: Anterior-posterior

CI: confidence interval

HU: Hounsfield unit

Table 4

Intraobserver variability of thymic measurements: the 2nd set of measurements

	Intraobserver CCC [95% CI]	Mean relative difference (mm)	95% limits of agreement (mm)
Transverse diameter	0.81 [0.66-0.90]	0.2	-14.5, 15.0
Anterior-posterior diameter	0.85 [0.72-0.92]	2.6	-12.6, 17.8
Right lobe length	0.92 [0.84-0.96]	-0.2	-10.7, 10.2
Left lobe length	0.92 [0.85-0.96]	1.6	-9.7, 12.9
Right lobe thickness	0.91 [0.82-0.95]	-0.2	-5.3, 4.9
Left lobe thickness	0.83 [0.70-0.91]	0.2	-6.3, 6.7
CT attenuation (HU)	0.97 [0.94-0.99]	0.9 (HU)	-16.1, 18.0 (HU)

CI: confidence interval

HU: Hounsfield unit

Table 5

Interobserver variability of thymic measurements: the 2nd set of measurements

	Interobserver CCC [95% CI]	Mean relative difference (mm)	95% limits of agreement (mm)
Transverse diameter	0.78 [0.61-0.89]	0.2	-17.0, 17.4
AP diameter	0.77 [0.58-0.88]	2.8	-17.7, 23.2
Right lobe length	0.85 [0.72-0.92]	-0.2	-14.5, 14.2
Left lobe length	0.82 [0.68-0.91]	0.7	-18.7, 20.0
Right lobe thickness	0.80 [0.64-0.89]	-1.2	-9.2, 6.9
Left lobe thickness	0.57 [0.29-0.76]	0.1	-11.0, 11.2
CT attenuation (HU)	0.88 [0.78-0.94]	-2.9 (HU)	-38.4, 32.5 (HU)

CI: confidence interval

HU: Hounsfield unit