Comparison Between Manual and Semiautomated Volumetric Measurements of Pituitary Adenomas

Qasim Al Hinai, B.Sc., M.D., M.R.C.S.I.,¹ Kelvin Mok, M.Eng., P.Eng.,¹ Anthony Zeitouni, M.D., F.R.C.S.C.,² Bruno Gagnon, M.D., M.Sc.,³ Abdul Razag Ailan, M.D.,¹ Juan Rivera, M.D., F.A.C.E., C.S.P.Q.,⁴ Marc Tewfik, M.D.C.M., M.Sc., F.R.C.S.C.,² and Denis Sirhan, M.D., F.R.C.S.C.¹

ABSTRACT

Linear measurements have many limitations. The aim of this study is to compare manual and semiautomated volumetric measurements of pituitary adenomas. Magnetic resonance imaging (MRI) scans of 38 patients with pituitary adenomas were analyzed. Preoperative MRI was acquired on a 1.5 T. MRI volumes of the pituitary adenomas were obtained by two methods: manual (MA) and semiautomated (SA). The concurrent validity for SA and MA methods on 38 patients in the form of correlation coefficient was 0.97 ($p < 0.0001$). The intraobserver and the interobserver correlation coefficients for SA volumes were both 0.98, as for the intraobserver MA volumes were 0.98. Although the results of both methods are comparable, analysis of volumetric measurements by SA method is more time-efficient than MA segmentation. Precision in volumetric measurement techniques is likely to increase reliability of posttherapeutic monitoring of pituitary adenomas.

KEYWORDS: Pituitary adenoma, volumetric measurement, correlation coefficient, tumor size

Accurate determination of tumor size is important in surgical resection, treatment planning, and post-treatment follow-up. Measurement of the pituitary adenoma can be done by two methods: linear and volumetric. Volumetric measurement can be performed by manual (MA), semiautomated (SA), and automated methods. The aim of this study is to compare MA and SA volumetric measurements of pituitary adenomas and to demonstrate which method is superior in terms of time efficiency.

PATIENTS AND METHODS

Magnetic resonance imaging (MRI) scans of 62 adult patients with pituitary adenomas were retrospectively reviewed between May 2005 and August 2010 at the Montreal Neurological Institute and Hospital, McGill University Health Centre, Montreal, Canada. Global MRI gadolinium sequences are performed as a preoperative preparation. The patients were selected consecutively from the McGill University Health Centre pathology records. Out of 62 patients, 24 patients were

¹Divisions of Neurosurgery; ²Otolaryngology and Head and Neck; ³Medicine and Oncology; and ⁴Endocrinology, McGill University, Montreal, Canada.

Address for correspondence and reprint requests: Denis Sirhan M.D., Department of Neurosurgery, Montreal Neurological Hospital/Institute, McGill University, 109-3801 University Avenue, Montreal, QC H3A 2B4, Canada (e-mail: denis.sirhan@mcgill.ca).

Skull Base 2011;21:365–372. Copyright © 2011 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA. Tel: +1(212) 584-4662.

Received: March 4, 2011. Accepted after revision: July 5, 2011. Published online: September 16, 2011. DOI: <http://dx.doi.org/10.1055/s-0031-1287677>. ISSN 1531-5010.

excluded as there were no global MRI gadolinium sequences. The 38 patients who had the required sequences were included, and were found to harbor macroadenomas: microadenomas were not deliberately excluded. Our operating room (OR) waiting list is long and loss or deterioration of the vision prioritizes patients to the top of the OR list. Patients who are going to the OR obtain global MRI gadolinium sequences. Our study does not require that the entire spectrum of tumors be included because it looks only to show the utility of using SA measurements of volume for estimation of tumor size as compared with MA measurements. These are time-consuming sequences, which is a limitation of both techniques because both require these sequences. Other inclusion criteria consisted of adult patients who had undergone transnasal transsphenoidal resection of pituitary adenomas. Patients with a diagnosis other than pituitary adenomas, such as meningioma and chordoma, and patients who underwent craniotomy for pituitary adenomas were excluded from the study. Approval from the Institutional Review Board at our institution was obtained.

Image Acquisition

Preoperative MRI data were obtained for 38 patients on a 1.5-T GE Signa Excite MR scanner (General Electric, Milwaukee, WI) and to obtain high-resolution imaging

for neuronavigation. The following sequences were obtained: T1-weighted gradient-echo (GRE) gadolinium-enhanced image (T1-Gd) using a 256×256 matrix, field of view (FOV) of 25.6 cm, 1.0 mm slice thickness, echo time (TE) of 8 ms, repetition time (TR) of 23 ms, axial fast relaxation fast spin echo T2-weighted image (5.0 mm slice thickness, TE 86.58 ms, TR = 5.66 s, flip angle of 90 degrees, an axial PSE proton density image (5.0 mm slice thickness, TE 29.57 ms, TR = 3.22 s, flip angle of 90 degrees, and a sagittal T1-fluid-attenuated inversion recovery image (5.0 mm slice thickness, TE 20.9 ms, TR = 2.11 s, flip angle of 90 degrees, inversion time of 0.75 s). An 8-channel phased-array head coil was used for radiofrequency reception.

Volume Analysis

MRI voxel-based volumes of the pituitary adenomas were obtained by two methods: MA and SA segmentation of MR T1-Gd (Fig. 1). In the first method, each T1-Gd MRI volume was manually segmented by two investigators blinded to the scan identity. In each of the three planes (axial, coronal, and sagittal), voxels were labeled as T1-Gd hyperintense voxels using image-intensity thresholding and manual labeling. A region of interest (ROI) was delineated using a bounding box and thresholding for the image intensity of lesion tissue. Voxels within the ROI were manually corrected at a voxel-by-voxel basis. The

Figure 1 The figure shows four views: Three orthogonal views of a volumetric image and fourth view is used to show the segmented structures in three dimensions. The structures adjacent to the pituitary adenomas are not included in the segmentation.

total volume of labeled voxels was measured and reported as the lesion volume. In the second method, lesion tissue was identified using a SA segmentation procedure (ITK-SNAP University of Pennsylvania).¹ The approximate center of the ROI was delineated in the sagittal plane. Identification of voxel clusters in the lesion tissue was then completed by the algorithm, automatically labeling neighboring voxels with manually defined tissue intensity. The labeled region extends toward to outer extent of the lesions. Finally, the labeled volume was corrected for obvious errors by the rater. The segmented lesion volumes were measured in cubic centimeters (cc) for both methods. The two raters followed the same procedure and classification protocol in both MA and SA measurements, with agreement on exclusion criteria of nonlesion structures.

Measurements

Two investigators (Q.A.H. and K.M.) obtained the MA and SA volumes. The measurements were performed by the investigators to determine intra- and interobserver variability. Each patient had one measurable lesion. For each patient, the volume was obtained five times.

Statistical Methods

The SA and the MA intraobserver comparisons were calculated using the mean intrascan lesion volumes as estimated by the principal investigator in the reading of the adenomas in two different sessions blinded to previous reading and presented in different sequences. The SA interobserver correlation was calculated using the estimated volume as measured by two different investigators (Al Hinai and Mok) in a random sample of 22 different adenomas. Intra- and interobserver correlations were calculated using Pearson correlation coefficients. Concurrent validity of the SA and MA measures was assessed using Pearson correlation coefficient. Intra- and interobserver reliability was determined by the intraclass correlation coefficient (ICC). The ICC was calculated using the Statistical Analysis Software (SAS, Cary, NC) intricacy macro. A paired *t*-test was also conducted to compare the mean performance times between the SA and MA measures. All statistical analyses were performed using SAS 9.2.

RESULTS

Fig. 2 presents the correlation between the SA and MA methods based on 38 observations. The Pearson correlation coefficient is 0.97 with a 95% confident interval that includes almost all the values with the exception of two data points. The ICC of the intraobserver reliability for MA method was 0.98 as was the ICC of the intra- and interobserver reliability for SA method. As shown in Table 1, the mean time to perform the measurements of

Figure 2 Correlation of the semiautomatic and the manual measurement methods. Caption: The semiautomated measures are plotted against the manual ones. The linear shape of the scatter plot indicates that these measures have a linear relationship and may be correlated. The 95% prediction ellipse encompasses all values which are considered to be correlated in the ellipse with a confidence level of 95% (similar to the definition of confidence intervals). From this graph, most of the data, based on 38 observations, is correlated at 0.97 with the exception of 2 data points.

the adenomas by the SA was 5 minutes less than by MA method ($p < 0.0001$). The mean time to perform the measurements of the adenomas by the SA method was 5 minutes less than by the MA one ($p < 0.0001$). The mean (SD) of the MA volumes was 10.4 cc (8.1) and the mean (SD) of the time taken to do the MA volumes was 10.6 (0.8) minutes. However, the mean of the SA volumes was smaller at 9.1 (7.3) and the mean (SD) of the time taken to do the SA volumes was 5.6 (0.8) minutes. For pituitary adenomas < 4 cc, the Pearson correlation coefficient of MA and SA was 0.81 ($p = 0.0085$) whereas for adenomas > 4 cc, it was 0.95 ($p < 0.0001$).

DISCUSSION

The results of this study show that SA measurements of volume compare well with MA measurements in MRI of pituitary adenomas and suggest that our method of SA volume calculations is a reliable technique when compared with MA segmentation, the current gold standard.

It has long been recognized that accurate determination of tumor size is important for surgical

Table 1 Statistical Analysis

	MA (SD)	SA (SD)	Time Difference Between MA & SA
Volume	10.4 (8.09)	9.07 (7.29)	
Time	10.6	5.57	$p < 0.0001$

MA, manual; SA, semiautomated.

resection, treatment planning, and posttreatment follow-up. It is also clear that linear measurements do not consistently convey a good sense of tumor growth. As reliable volume calculations include the third power of the radius, what may seem a small linear change would in fact represent a large volume change. However, clinicians have resisted using volume measurements because they are time-consuming. The current study shows that SA measurements take half the time of MA volumetric measurements.

The pituitary gland receives both positive and negative hormonal feedback from the hypothalamus and the target organs. Hence, evaluation of normal pituitary gland volume is a topic of endocrinologic and clinical interest. Takano et al assessed the volume of the anterior, posterior, and whole pituitary gland in normal children and adolescents aged from 0 to 19 years (mean, 7.4 years) using thin-section three-dimensional (3D)-MRI. They found that for the phantom study the measurement fell within 25%, while the interobserver correlation coefficient was 0.771. Their clinical study showed that the pituitary volume in girls was significantly larger than that in boys ($p=0.01$).² Marziali et al assessed the normal anterior pituitary gland volumes in both girls and boys aged from 2 months to 10 years using 3D-MRI techniques.³ They found that the measurement error was 0.2 to 0.4% and that the interobserver Pearson correlation coefficient was 0.8 for the pilot study and 0.75 for the clinical study. It was concluded that these data may be useful for pediatricians in the evaluation of patients with neuroendocrine diseases, in particular growth hormone deficiency. On the other hand, Fink et al in their series of 254 normal children up to age 10 years using 3D-MRI found that the intraobserver reliability was high with a concordance correlation coefficient (CCC) of 0.986.⁴ These authors also observed a very strong agreement between the volumes measured from the coronal and sagittal reconstructions with a CCC of 0.986, and a weak correlation between pituitary volume and height or body mass index (Spearman correlation coefficient, 0.20 and 0.17, respectively). Furthermore, there was a weak correlation between pituitary volume and pituitary height, with a Spearman correlation coefficient of 0.35. Ertekin et al in their retrospective study of 28 subjects with normal pituitary gland used three methods of volumetric measurements: elliptic formula and stereological approaches (point-counting method and planimetry method).⁵ They determined that the coefficient of error for the volume derived from the point counting technique was 8.07% and found no significant difference between point-counting and planimetric methods ($p>0.05$). There was a statistically significant difference between the pituitary volume measurements obtained using stereological techniques and the elliptic formula ($p<0.05$). Also, there was a correlation between the point-counting and planimetric

techniques ($r=0.968$), point-counting and elliptic formula ($r=0.754$), planimetric and elliptic formula ($r=0.752$). There was a 26.14 and 29.71% underestimation of pituitary volume as measured by the elliptic formula compared with the point-counting and planimetric techniques, respectively.

In patients with pituitary adenomas, tumor size is the most widely used parameter to monitor treatment. Computer-aided detection and SA segmentation systems have been developed to overcome radiologists' errors in detecting lesions and to decrease variability in lesion size determination introduced by the subjectivity of measurements. Single or multiple modalities of volumetric measurements of pituitary adenomas can be used either preoperatively (pretreatment), as in our study, or postoperatively (posttreatment).

Assessment of the size of pituitary adenomas using volumetric measurement is of great importance in surgical planning and posttreatment monitoring. Pamir et al used MRI-based volumetric analysis of pituitary adenomas pre- and post-Gamma Knife radiosurgery (GKRS) with 3 years follow-up.⁶ Using univariate and multivariate analyses of volumetric data, they found that the parameters showing a significant effect on lesser volume reduction were cavernous sinus invasion ($p=0.0001$) and lower marginal doses ($p=0.015$). The p value of the parameter of cavernous sinus invasion was very small, showing the significance of this factor on outcome after GKRS. The p value of the marginal doses was also less than 0.05, showing that this factor has additional significance. Wowra and Stummer also used volumetric comparisons to demonstrate statistically significant volume reduction ($p=0.0003$) between pre- and post-GKRS in their series of 30 patients with non-functioning pituitary adenomas.⁷

Interestingly, Buhk et al in their series of 45 patients with growth hormone-secreting pituitary adenomas with 24 months follow-up did not find a significant change ($p=0.54$) with regard to the adenoma volumetric measurements before and after pegvisomant therapy.⁸ Several tumors actually increased in size, which the authors attributed to the rebound effect after withdrawal of somatostatin analogs. In addition, Benesch et al did a magnetic resonance volumetric study for monitoring intramuscular bromocriptine treatment in macroprolactinomas.⁹ They found that the accuracy of volume calculations of phantom study was within an error range of 4% using two-dimensional spin-echo images and 2% using 3D gradient echo (GE) images, and for the anatomical specimen, the error was $<3\%$ using 3D GE images. Intrarater variability was $<4\%$. The above data are summarized in Table 2.

There is no standard method to measure the volume of the pituitary gland. McLachlan et al performed a postmortem study in 50 adults who died from nonendocrine diseases and without evidence of raised

Table 2 Literature Review on Pituitary Volumetric Measurements

Serial No.	Author, Year	Study	No. of Pts	Population Studied	Timing	Technique	Aspect	Correlation
1	Benesch et al, ⁹ 1995	N/A	2	Macroprolactinoma	Pre- and postbromocriptine therapy	MA and automated	To assess volumetric changes of pituitary prolactinomas	Accuracy of volume calculations of phantom was within an error range of 4% using 2D SE images and 2% using 3D GE images. For anatomical specimen, the error was <3% using 3D GE images. Intrarater variability was <4%.
2	Takano et al, ² 1999	N/A	199	Normal subjects	N/A	MA and SA	To determine the normal development of pituitary volume	Phantom study: Measurement errors within 25%, no statistically significant difference in errors between the two volumetric sequences, interobserver correlation coefficient was 0.771. Clinical study: Pituitary volume in girls was significantly larger than that in boys (p , 0.01)
3	Wowra et al, ⁷ 2002	Prospective	30	Pituitary adenomas	Pre- and post-RSx	GammaPlan software (MA)	Reduction in tumor size	$p = 0.0003$ significant
4	Marzali et al, ³ 2004	N/A	95	Normal subjects	N/A	MA	N/A	Pilot study: Measurement error of 0.2–0.4%. Pearson interobserver correlation coefficient was 0.8. Clinical study: Pearson interobserver correlation coefficient was 0.75 CCC of 0.986, significant
5	Fink et al, ⁴ 2005	Prospective	254	Normal subjects (children)	N/A	MA	N/A	Univariate analysis: p value = 0.0001 and 0.015. Multivariate analysis: p value = 0.001 and 0.050 significant
6	Pamir et al, ⁶ 2007	Prospective	100	Pituitary adenomas	Pre- and post-RSx	GammaPlan software(MA)	Reduction in tumor size	$p = 0.54$ not significant (due to medication)
7	Buhk et al, ⁸ 2010	Prospective	45	Pituitary adenomas	Pre- and postpegvisomant	MA and computer aided	No change	

Table 2 (Continued)

Serial No.	Author, Year	Study	No. of Pts	Population Studied	Timing	Technique	Aspect	Correlation
8	Ertekin et al, ⁵ 2011	Retrospective	28	Normal subjects	N/A	Elliptic formula, point-counting method, and planimetry method (= MA)	To obtain accurate normal pituitary gland volume	Coefficient of error for the volume derived from the point counting technique was 8.07%. No significant difference was found between point counting and planimetric methods ($p > 0.05$). There was a 26.14 and 29.71% underestimation of pituitary volume as measured by the elliptic formula compared with the point counting and planimetric techniques
9	Al Hinaï, present study, 2011	Retrospective	38	Pituitary adenomas	Preop	MA & SA	To compare MA and SA volumetric measurements	Pearson correlation coefficient between SA and MA methods is 0.97. ICC of the intraobserver reliability for MA method was 0.98 as was the ICC of the intra- and interobserver reliability for SA method. Mean time to perform measurements by SA method was 5 minutes less than by MA one ($p < 0.0001$).

2D, two-dimensional; 3D, three-dimensional; CCC, concordance correlation coefficient; ICC, intraclass correlation coefficient; GE, gradient echo; MA, manual volumes; N/A, not applicable or not available; RSx, radiosurgery; SA, semiautomated volumes; SE, spin echo.

intracranial pressure.¹⁰ They found that the sellar volume, derived from the product of length, height, and floor width or from the product of lateral area and floor width, correlated well with gland volume. It is noteworthy to mention that the conclusion made by McLachlan et al was based on linear measurements. Future postmortem studies based on objective volumetric measurements are needed to develop a gold standard method for obtaining an accurate volume.

In our study, the Pearson correlation coefficient for SA and MA was significant at 0.969 ($p < 0.0001$). The ICC of the intraobserver and interobserver reliability of the MA volumes and ICC of the intraobserver and interobserver of the SA volumes demonstrated significant correlations. Our study concludes that while both techniques yield comparable results, the SA technique is faster and reduced time by half. The results of our study show significant correlations compared with the results of MA and SA volumetric measurements done by Takano et al.² However, our study had a small sample size and the population was subjects with pituitary adenomas, whereas the study done by Takano et al had a large sample size and the population was normal subjects. Most volumes were greater with MA compared with SA method. Computer software algorithms are still being developed. The measurements and value of the information obtained are of great significance to the clinician and are superior to the basic linear measurements. As the software becomes more developed, its ability to distinguish normal tissue from tumor will improve.

There are several limitations of this study. First, it is a retrospective analysis of a series of adenomas documented from pathological report. Second, it includes preoperative images only. We did not evaluate the value of the SA method in the residual tissue in postoperative images which could enable physicians to detect residual adenomas and measure its volume which will assist in the future follow-up and management. Third, because of the availability of the required sequences (global MRI gadolinium), only macroadenomas were included in the study. Future studies should include tumors across the spectrum of microadenomas. These would have to be prospective studies since MRI sequences are not routinely performed unless the tumor is considered for surgery.

CONCLUSION

As most pituitary lesions are not perfectly spherical, 3D size determinations should be preferred over

linear measurements. Performance of volumetric measurements by SA method is more time-efficient than MA measurement of volumes. Widespread use of volumetric measurement techniques may give the clinician a more accurate sense of tumor growth and improve clinical decision-making.

ACKNOWLEDGMENT

The authors thank Mrs. Wendy Blue for assistance in the preparation of this article.

REFERENCES

1. Yushkevich PA, Piven J, Hazlett HC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* 2006;31(3):1116–1128
2. Takano K, Utsunomiya H, Ono H, Ohfu M, Okazaki M. Normal development of the pituitary gland: assessment with three-dimensional MR volumetry. *AJNR Am J Neuroradiol* 1999;20(2):312–315
3. Marziali S, Gaudiello F, Bozzao A, et al. Evaluation of anterior pituitary gland volume in childhood using three-dimensional MRI. *Pediatr Radiol* 2004;34(7):547–551
4. Fink AM, Vidmar S, Kumbra S, et al. Age-related pituitary volumes in prepubertal children with normal endocrine function: volumetric magnetic resonance data. *J Clin Endocrinol Metab* 2005;90(6):3274–3278
5. Ertekin T, Acer N, Turgut AT, Aycan K, Özçelik O, Turgut M. Comparison of three methods for the estimation of the pituitary gland volume using magnetic resonance imaging: a stereological study. *Pituitary* 2011;14(1):31–38
6. Pamir MN, Kiliç T, Belirgen M, Abacıoğlu U, Karabekiroğlu N. Pituitary adenomas treated with gamma knife radiosurgery: volumetric analysis of 100 cases with minimum 3 year follow-up. *Neurosurgery* 2007;61(2):270–280; discussion 280
7. Wowra B, Stummer W. Efficacy of gamma knife radiosurgery for nonfunctioning pituitary adenomas: a quantitative follow up with magnetic resonance imaging-based volumetric analysis. *J Neurosurg* 2002;97(5, Suppl):429–432
8. Buhk JH, Jung S, Psychogios MN, et al. Tumor volume of growth hormone-secreting pituitary adenomas during treatment with pegvisomant: a prospective multicenter study. *J Clin Endocrinol Metab* 2010;95(2):552–558
9. Benesch H, Felber SR, Finkenstedt G, Kremser C, Stockhammer G, Aichner FT. MR volumetry for monitoring intramuscular bromocriptine treatment in macroprolactinomas. *J Comput Assist Tomogr* 1995;19(6):866–870
10. McLachlan MS, Williams ED, Fortt RW, Doyle FH. Estimation of pituitary gland dimensions from radiographs of the sella turcica. A post-mortem study. *Br J Radiol* 1968; 41(485):323–330