<https://helda.helsinki.fi>

Software output from semi-automated planimetry can underestimate intracerebral haemorrhage and peri-haematoma oedema volumes by up to 41 %

Wu, Teddy Y.

2016-09

Wu , T Y , Sobowale , O , Hurford , R , Sharma , G , Christensen , S , Yassi , N , Tatlisumak , T , Desmond , P M , Campbell , B C V , Davis , S M , Parry-Jones , A R & Meretoja , A 2016 , ' Software output from semi-automated planimetry can underestimate intracerebral haemorrhage and peri-haematoma oedema volumes by up to 41 % ' , *Neuroradiology* , vol. 58 , no. 9 , pp. 867-876 . <https://doi.org/10.1007/s00234-016-1720-z>

<http://hdl.handle.net/10138/228321>

<https://doi.org/10.1007/s00234-016-1720-z>

publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Software output from semi-automated planimetry can underestimate intracerebral haemorrhage and peri-haematoma oedema volumes by up to 41 %

Teddy Y. Wu¹ · Oluwaseun Sobowale² · Robert Hurford² · Gagan Sharma³ · Søren Christensen⁴ · Nawaf Yassi¹ · Turgut Tatlisumak^{5,6} · Patricia M Desmond³ · Bruce C. V. Campbell¹ · Stephen M Davis¹ · Adrian R Parry-Jones² · Atte Meretoja^{1,7}

Received: 6 May 2016 / Accepted: 20 June 2016 / Published online: 5 July 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract

Introduction Haematoma and oedema size determines outcome after intracerebral haemorrhage (ICH), with each added 10 % volume increasing mortality by 5 %. We assessed the reliability of semi-automated computed tomography planimetry using Analyze and Osirix softwares.

Methods We randomly selected 100 scans from 1329 ICH patients from two centres. We used Hounsfield Unit thresholds of 5–33 for oedema and 44–100 for ICH. Three raters segmented all scans using both softwares and 20 scans repeated for intra-rater reliability and segmentation timing. Volumes reported by Analyze and Osirix were compared to volume estimates calculated using the best practice method, taking

effective individual slice thickness, i.e. voxel depth, into account.

Results There was excellent overall inter-rater, intra-rater and inter-software reliability, all intraclass correlation coefficients >0.918. Analyze and Osirix produced similar haematoma (mean difference: Analyze – Osirix = 1.5 ± 5.2 mL, 6 %, $p \leq 0.001$) and oedema volumes (-0.6 ± 12.6 mL, –3 %, $p = 0.377$). Compared to a best practice approach to volume calculation, the automated haematoma volume output was 2.6 mL (–11 %) too small with Analyze and 4.0 mL (–18 %) too small with Osirix, whilst the oedema volumes were 2.5 mL (–12 %) and 5.5 mL (–25 %) too small, correspondingly. In scans with variable slice thickness, the volume underestimations were larger, –29 %/–36 % for ICH and –29 %/–41 % for oedema. Mean segmentation times were $6:53 \pm 4:02$ min with Analyze and $9:06 \pm 5:24$ min with Osirix ($p < 0.001$).

Conclusion Our results demonstrate that the method used to determine voxel depth can influence the final volume output markedly. Results of clinical and collaborative studies need to be considered in the context of these methodological differences.

Electronic supplementary material The online version of this article (doi:10.1007/s00234-016-1720-z) contains supplementary material, which is available to authorized users.

✉ Atte Meretoja
atte.meretoja@unimelb.edu.au

¹ Department of Medicine at the Royal Melbourne Hospital, University of Melbourne, Parkville, Victoria, Australia

² Centre for Vascular and Stroke Research, Manchester Academic Health Sciences Centre, Salford Royal Hospital Trust, University of Manchester, Manchester, UK

³ Department of Radiology, the Royal Melbourne Hospital, University of Melbourne, Parkville, Victoria, Australia

⁴ Stanford Stroke Center, Stanford University, Palo Alto, CA, USA

⁵ Institute of Neuroscience and Physiology, Sahlgrenska Academy at the University of Gothenburg, Gothenburg, Sweden

⁶ Department of Neurology, Sahlgrenska University Hospital, Gothenburg, Sweden

⁷ Department of Neurology, Helsinki University Hospital, PK Admin, Haartmaninkatu 4E, P.O. Box 220, 00029 HUS Helsinki, Finland

Keywords Intracerebral haemorrhage · Oedema · Planimetry · Validation · Reliability

Introduction

Intracerebral haemorrhage (ICH) is a highly fatal stroke subtype, and the only proven treatments are blood pressure control and stroke unit care [1, 2]. Primary injury results from direct mechanical pressure of the haematoma in a volume-dependent manner [3]. Secondary injury ensues from peri-haematoma oedema [4], with emerging evidence of oedema's

association with poor clinical outcome [5]. Effective oedema management may improve outcome and is the target of several early phase clinical trials [6–9]. Accurate volumetric assessment of both ICH and oedema is therefore vital to the analysis of clinical and collaborative studies.

The volume of tissue with oedema and ICH can be quantified on computed tomography (CT) using Hounsfield unit (HU) threshold semi-automated segmentation [10–14]. From a technical perspective, the most accurate measure of volume of a segmented region is the cumulative total volume of the voxels contained within this region. In turn, the individual voxel volume is determined by the product of voxel area and voxel depth. The voxel width and height, which determine the voxel area, are easily accessible in the Digital Imaging and Communications in Medicine (DICOM) header. The voxel depth that should be used for volume calculation is the distance between slice centres. This information is generally not directly available in the DICOM header, but must be calculated based on other spatial information in the header. The method with which voxel depth is calculated by imaging analysis software has not been explored in detail in published studies. Several softwares are available for segmentation, but no reports have directly assessed the reliability of planimetric measurements between softwares.

The aims of this study were two fold. First, we assessed the reliability and time taken for segmentation using semi-automated planimetry between Analyze 12.0 (Biomedical Imaging Resource, Mayo Clinic) and Osirix 6.5 (Pixmeo, Geneva, Switzerland). Secondly, we analysed the impact of different methods for voxel depth estimation on the final estimated volumes. We hypothesised that the segmented volumes and times taken for segmentation are different between Analyze and Osirix and that the volumes are influenced by the method of voxel depth determination.

Methods

Image selection

A convenient sample of 100 cases was chosen for this study. The baseline scans of randomly selected patients from the combined database of 1329 consecutive ICH patients from Helsinki University Hospital, Finland, and Salford Royal Hospital, UK, were used after institutional approval. Fifty scans were randomly chosen from each centre. The Helsinki ICH study is a retrospective analysis of consecutive ICH patients admitted to Helsinki University Central Hospital between January 2005 and March 2010 [15]. Patients from Salford Royal Hospital were ICH patients treated at the centre between January 2013 and May 2015.

Image processing and segmentation

The de-identified images were transferred in the DICOM format to a central workstation. The DICOMs were then converted into Neuroimaging Informatics Technology Initiative (NIfTI) format using a DCM-to-NIfTI conversion tool before loading on Analyze. The NIfTI allows the individual DICOM images to be saved and loaded from one single file. For Osirix, the DICOMs were loaded directly.

Segmentation steps

The semi-automated method reported by Volbers et al. was used [10]. A limit boundary was placed around the ICH and oedema complex, following which a seed point was placed within the region of interest (ROI). For oedema segmentation, the lower HU limit is fixed at 5, with the upper boundary manually adjusted to a maximum of 33-HU range using visual inspection and comparison to the contralateral hemisphere for background leucoaraiosis. For ICH segmentation, the HU range was kept within 44–100 and manual editing of ROI was allowed at the rater's discretion.

Video demonstrations of the segmentation process on each of the softwares are available in the Electronic supplementary material.

Rater workflow

All 100 scans were segmented using Analyze and Osirix by each rater (T.Y.W., O.S., and R.H.) independently and blinded to patient's clinical details. T.Y.W. is a neurologist with 6 years of experience in stroke, O.S. is a neurosurgery registrar with 6 years of experience in stroke, and R.H. is a clinical academic fellow with 3 years of experience in stroke imaging. Twenty scans (ten from Helsinki and ten from Salford) were chosen randomly and repeat segmentation was performed in these for intra-rater assessment. Intra-rater assessment was performed after a minimal interval of 7 days. The time taken to perform segmentation for each software was measured for the intra-rater scans.

ROI volume calculation

The volume estimates were determined by automated software output ('Sampling Options' in Analyze and 'ROI manager' in Osirix) and then by subjecting the software outputs to the "best practice" volume estimation. The best practice volume was calculated using an in-house script developed using Matlab (The MathWorks, Inc., MA, USA; [Supplementary Table](#)). To this end, the ROI object files from Analyze were used whilst the ROI segmented on Osirix was exported in DICOM format using an in-house plug-in ([Supplementary File](#)). The Matlab script

determined the number of voxels within the ROI, which was then multiplied by the voxel volume. The voxel volume was derived from the product of voxel width and voxel height (DICOM header 0028, 0030) and voxel depth. The voxel depth may or may not be equivalent to the slice thickness presented in the DICOM header (0018, 0050), which is not well defined in the DICOM standard and used differently between scanner manufacturers. Therefore, the in-house script uses the best practice method for determining inter-slice distance, which is to determine the normal vector to the image slice orientation (identical for all slices) using ‘Image Orientation Patient’ (DICOM header 0020, 0037).

The location of each slice ‘Image Position Patient’ (IPP; DICOM header 0020, 0032) is then projected onto the slice normal vector, producing one value per slice expressing the location of the slice on the slice normal as a scalar value (Zpos). This calculation is not influenced by gantry tilt which can overestimate the voxel depth by approximately 5 % (Supplementary Fig. 1).

The Zpos was then used to calculate the voxel depth using the equation,

$$\text{Voxel depth} = [\text{Zpos}(n+1) - \text{Zpos}(n)]/2 + [\text{Zpos}(n) - \text{Zpos}(n-1)]/2$$

where n is the image slice, $(n-1)$ the image slice below and $(n+1)$ the slice above. For the first and last slices of the scan, we considered the thickness to be the same as the second slice and the penultimate slice, respectively. This calculation approximates the slice distance not just between slices with identical distance between them but also in the interface where slice thickness changes from, for example, 7.5 to 5 mm, where an overlap (Fig. 1a) can be present. Our in-house method adjusts for any slice overlap and is direction-insensitive (Fig. 1b). We will demonstrate using a case illustration the impact of varying voxel depth on volume output. We developed this approach because the Z position within IPP is direction- and gantry tilt-sensitive and can overestimate the volume when slice overlap is present.

Statistical analysis

Intraclass correlation coefficients (ICCs) were calculated using two-way random effects model and used to measure inter-rater and intra-rater reliability for ICH and oedema. An ICC was considered moderate agreement if 0.41–0.60, substantial agreement if 0.61–0.80 and excellent if 0.81–1.00 [16]. Bland–Altman plots were used to assess for systematic bias. Paired t test or one-way analysis of variance (ANOVA) was used to compare means, where appropriate. A p value <0.05 was considered significant. All statistics were performed using SPSS 22 (IBM, Armonk, NY).

Reporting standards

This study was reported in accordance with the Guidelines for Reporting Reliability and Agreement Studies [17].

Results

Patient characteristics

Of the 100 scans included in this study, 87 were performed within 24 h of ictus. The median age was 69 years, 48 % were men, baseline median National Institutes of Health Stroke Scale score was 10, median time to scan was 2.8 h, and ICH location was lobar in 36 %, deep in 43 % and infratentorial in 10 %.

Volumetric output and agreement statistics

The mean $\pm$ standard deviation haematoma volumes segmented on Analyze were 23.4 ± 29.6 and 20.9 ± 25.6 mL for oedema. The overall inter-rater ICCs were 0.994 for ICH and 0.952 for oedema; the intra-rater ICCs were 0.998 for ICH and 0.983 for oedema. The mean volumes for haematoma and oedema performed on Osirix were 21.9 ± 27.0 and 21.5 ± 24.4 mL, respectively. The inter-rater ICCs were 0.997 for ICH and 0.944 for oedema, and the intra-rater ICCs were 0.996 for ICH and 0.918 for oedema. The inter-software ICCs were 0.991 for ICH and 0.932 for oedema (Table 1). ICH volume was 1.5 mL (± 5.2 mL, 6 %, $p \leq 0.001$) larger in Analyze and oedema was 0.6 mL (± 12.5 mL, 3 %, $p = 0.377$) larger in Osirix.

The difference in the ICH or oedema volumes obtained between raters was not statistically significant (Table 2). The Bland–Altman plots for inter-rater agreement (Fig. 2 and Supplementary Fig. 2), presented as volume difference between raters 1 and 2, raters 2 and 3, and raters 1 and 3, demonstrated bias values of 0.6, -1.1 and -0.5 mL, respectively, for ICH and 0.5, -0.4 and 0.1 mL, respectively, for oedema segmented on Analyze. The inter-rater bias values for Osirix were 1.4, -1.4 and 0 mL for ICH and 6.7, -1.8 and 4.9 mL for oedema.

The mean segmentation times were $6:53 \pm 4:02$ min with Analyze and $9:06 \pm 5:24$ min with Osirix, with a difference of $2:13 \pm 3:10$ min ($p < 0.001$) in favour of Analyze.

Comparing in-house method with Analyze automated output

The ICCs between our in-house method and Analyze’s automated output were 0.986 and 0.980 respectively for ICH and oedema. The automated output underestimated ICH by 2.6 mL (-11 %, $p < 0.001$) and oedema by 2.5 mL (-12 %, $p < 0.001$; Table 3 and Figs. 3 and 4). When analysis was

Fig. 1 Different methods of voxel depth calculation. Four representative consecutive brain slices with a gantry tilt of 16.5°. Slices 9 and 8 represent junction between thicker supratentorial (7.5 mm) and thinner infratentorial (5.0 mm) slices. Voxel depth derived from the mid-position of consecutive slices results in overestimation of the overlap

region (a). In-house method using half distance of inter-slice midpoints (b) results in no overestimation by inclusion of the overlap region. Software-automated output utilises the infratentorial voxel depth for the entire series (c) for volume calculation, resulting in underestimation

restricted to scans with variable slice thickness ($n = 50$), the underestimation increased to 6.4 mL (−29 %, $p < 0.001$) for

ICH and 6.0 mL (−29 %, $p < 0.001$) for oedema. The automated output from Analyze overestimated ICH by 1.3 mL

Table 1 Inter-rater, intra-rater and inter-software intraclass correlation coefficients

	Analyze (<i>n</i> = 100)		Osirix (<i>n</i> = 100)		<i>n</i> = 100
	Inter-rater ICC (95 % CI)	Intra-rater ICC (95 % CI)	Inter-rater ICC (95 % CI)	Intra-rater ICC (95 % CI)	Inter-software ICC (95 % CI)
All ICH	0.994 (0.992–0.996)	0.998 (0.996–0.999)	0.997 (0.996–0.998)	0.996 (0.993–0.998)	0.991 (0.989–0.993)
All oedema	0.952 (0.933–0.966)	0.983 (0.971–0.990)	0.944 (0.922–0.961)	0.918 (0.863–0.951)	0.932 (0.915–0.946)

(+5 %, $p < 0.001$) and oedema by 1.0 mL (+5 %, $p < 0.001$) for scans with uniform thickness ($n = 50$).

Comparing in-house method with Osirix automated output

The analysis was performed on 98 patients as automated output failed in two patients with ROIs limited to a single CT image slice. The ICCs between the in-house method and Osirix automated output were 0.983 for ICH and 0.921 for oedema (Table 3). The automated output underestimated ICH by 4.0 mL (−18 %, $p < 0.001$) and oedema by 5.5 mL (−25 %, $p < 0.001$). The underestimations in scans with variable slice thickness were 7.5 mL (−36 %, $p < 0.001$) for ICH and 9.4 mL (−41 %, $p < 0.001$) for oedema. The differences for scans with uniform slice thickness were −0.3 mL (−1 %, $p = 0.045$) for ICH and −1.7 mL (−8 %, $p < 0.001$) for oedema. The Bland–Altman plots for these comparisons are in Figs. 3 and 4.

Influence of calculated voxel depth/slice thickness on volume output

The detailed workflow for ROI volume calculation for a case with variable slice thickness and a case with uniform thickness is represented in the supplementary online spreadsheet. The non-gantry-adjusted voxel depth from the infratentorial region was used as the effective voxel depth in both softwares to calculate volume in the automated output when scans were loaded as a single series (Fig. 1c). When compared to our in-house method, there were underestimations of 12 and 24 % using automated outputs from Analyze and Osirix,

respectively, in the case with variable thickness. For the case with uniform thickness, Analyze overestimated by 1 % whilst Osirix underestimated volume by 7 %.

Discussion

The main findings of our study are twofold. Firstly, there is excellent inter-rater, intra-rater and inter-software reliability in semi-automated segmentation of ICH and oedema. Secondly, our results demonstrate that the automated outputs from the two softwares consistently underestimate volumes in scans with variable slice thickness by up to 41 %.

Reliability of semi-automated planimetry

We demonstrated excellent inter-rater and intra-rater ICCs in both Analyze and Osirix using a validated semi-automated planimetry approach [10]. The reliability of the ICH measurement in our study is consistent with that reported in the literature (inter-rater and intra-rater ICCs >0.900) [12, 14, 18]. Divani et al. demonstrated in a simulated cadaveric ICH model that Analyze software and Medical Imaging Processing, Analysis and Visualization (Center for Information Technology, National Institutes of Health, Bethesda, MD) software produced less error than the ABC/2 method when compared to the actual volume of simulated blood injected, but the authors did not directly compare the volumes produced by the different softwares [19]. To our knowledge, our report is the first to directly compare ICH and oedema volume outputs from different softwares.

Table 2 Mean volume output of intracerebral haemorrhage and oedema

	Analyze (<i>n</i> = 100)		Osirix (<i>n</i> = 100)	
	ICH, mean ± SD (mL)	Oedema, mean ± SD (mL)	ICH, mean ± SD (mL)	Oedema, mean ± SD (mL)
Rater 1	23.4 ± 28.7	21.1 ± 27.3	22.4 ± 27.5	25.4 ± 28.2
Rater 2	22.8 ± 28.5	20.6 ± 21.4	21.0 ± 26.1	18.7 ± 18.8
Rater 3	23.9 ± 31.8	21.0 ± 27.9	22.4 ± 27.5	20.5 ± 25.0
<i>p</i> value ^a	0.97	0.99	0.912	0.132

ICH intracerebral haemorrhage, SD standard deviation

^a One-way ANOVA

Fig. 2 Bland–Altman plots for haematoma segmentation. Bland–Altman plots with intraclass correlation coefficients (*ICC*) for semi-automated planimetry of hematoma segmentation on Analyze and Osirix. *Solid line* represents the mean difference and *dotted lines* the limits of 95 % agreement. Analyze outliers: the two outliers with approximately 120 mL ICH

both had large right frontal hematoma with heterogeneous density. The other outlier had moderate movement artefact together with moderate ventricular hemorrhage. Osirix outliers: The outliers with <50 mL ICH had large volume ventricular hemorrhage and the other outliers had heterogeneous ICH density

The oedema inter-rater ICCs are also in agreement with the largest oedema study to date—the pooled analysis of the

Intensive Blood Pressure Reduction in Acute Cerebral Haemorrhage Trials (INTERACT) using the same semi-

Table 3 Reliability statistics and volume difference between in-house technique and automated outputs from Analyze and Osirix

		Automated output, mean $\pm$ SD (mL)	In-house output, mean $\pm$ SD (mL)	Difference (mL)	<i>p</i> value ^a	ICC (95 % CI)
Analyze						
Full sample (<i>n</i> = 100)	ICH	20.8 $\pm$ 28.5	23.4 $\pm$ 29.6	−2.6 (−11 %)	<0.001	0.986 (0.983–0.989)
	Oedema	18.4 $\pm$ 23.7	20.9 $\pm$ 25.6	−2.5 (−12 %)	<0.001	0.980 (0.975–0.984)
Variable thickness (<i>n</i> = 50)	ICH	15.4 $\pm$ 17.7	21.8 $\pm$ 25.1	−6.4 (−29 %)	<0.001	0.968 (0.956–0.977)
	Oedema	14.6 $\pm$ 17.3	20.6 $\pm$ 24.5	−6.0 (−29 %)	<0.001	0.961 (0.947–0.972)
Uniform thickness (<i>n</i> = 50)	ICH	26.3 $\pm$ 35.5	25.0 $\pm$ 33.5	1.3 (+5 %)	<0.001	0.999 (0.999–0.999)
	Oedema	22.2 $\pm$ 27.3	21.2 $\pm$ 26.9	1.0 (+5 %)	<0.001	0.999 (0.998–0.999)
Osirix						
Full sample (<i>n</i> = 98) ^b	ICH	18.4 $\pm$ 22.3	22.4 $\pm$ 27.1	−4.0 (−18 %)	<0.001	0.983 (0.979–0.987)
	Oedema	16.5 $\pm$ 20.8	22.0 $\pm$ 24.5	−5.5 (−25 %)	<0.001	0.921 (0.901–0.937)
Variable thickness (<i>n</i> = 49) ^b	ICH	13.6 $\pm$ 16.2	21.1 $\pm$ 23.6	−7.5 (−36 %)	<0.001	0.961 (0.946–0.972)
	Oedema	13.5 $\pm$ 17.0	22.9 $\pm$ 25.1	−9.4 (−41 %)	<0.001	0.957 (0.940–0.969)
Uniform thickness (<i>n</i> = 49) ^b	ICH	23.3 $\pm$ 31.2	23.6 $\pm$ 30.2	−0.3 (−1 %)	0.045	0.999 (0.999–0.999)
	Oedema	19.4 $\pm$ 23.8	21.1 $\pm$ 23.9	−1.7 (−8 %)	<0.001	0.949 (0.930–0.963)

ICC intraclass correlation coefficient, *ICH* intracerebral haemorrhage, *SD* standard deviation

^a Paired *t* test

^b Automated output in Osirix failed in two patients (one with variable slice thickness and one with uniform slice thickness), with the region of interest limited to one image slice; therefore, 98 patients were used in this analysis

Fig. 3 Bland–Altman plots for haematoma volume difference between software output and in-house method. Bland–Altman plots comparing haematoma volume produced by software output from Analyze (*top panels*) by uniform ($n = 50$) or variable ($n = 50$) image slice thickness. Haematoma volume comparisons with Osirix (*bottom panels*) were

available for 49 patients with uniform slice thickness and 49 patients with variable slice thickness as automated output failed in two patients with the region of interest limited to one single image slice. *Solid line* represents the mean difference and *dotted lines* the limits of 95 % agreement

automated approach ($n = 1138$, MISTar 3.2 software; oedema inter-rater ICCs of 0.91 for INTERACT 1 and 0.93 for INTERACT 2; ICH ICCs not reported) [5]. Other threshold-based planimetry segmentation, including the methods proposed by McCourt et al. (inter-rater ICC for oedema at 24 h of 0.99) [13] using an upper HU limit for oedema of 23 and the edge detection technique by Urdy et al. (inter-rater and intra-rater ICCs both 0.99) [11], have also used Analyze and demonstrated excellent reliability. Osirix was also used in the Minimally Invasive Surgery and rt-PTA in ICH Evacuation (MISTIE) phase II study for the assessment of ICH and oedema volumes, but rater reliability assessment was not reported [20].

Although excellent ICCs and similar volumes were demonstrated between Analyze and Osirix, there is more inter-rater

oedema variability with Osirix. The variability in oedema volume in Osirix could have resulted from the need to ‘black out’ brain regions ([Online Supplementary Video](#)) outside of the peri-haematoma region, which may interfere the rater’s ability to distinguish oedema from leucoaraiosis without visibility of the contralateral hemisphere.

Implications for clinical studies

We observed 3–4 mL or 11–18 % underestimation of ICH volume and 3–6 mL or 12–25 % underestimation of oedema volume using the automated software outputs. These are substantial errors in measurement, as a 1 mL increase in ICH volume has been associated with a 5 %

Fig. 4 Bland–Altman plots for oedema volume difference between software output and in-house method. Bland–Altman plots comparing oedema volume produced by software output from Analyze (*top panels*) by uniform ($n = 50$) or variable ($n = 50$) image slice thickness. Oedema volume comparisons with Osirix (*bottom panels*) were available for 49

patients with uniform slice thickness and 49 patients with variable slice thickness as automated output failed in two patients with the region of interest limited to one single image slice. *Solid line* represents the mean difference and *dotted lines* the limits of 95 % agreement

increase in death and dependency [21] and a 10 % increase in ICH volume with a 5 % increase in overall mortality [22]. The >6 mL difference in ICH between our in-house output and the automated output in scans with variable slice thickness may influence the outcome of clinical studies. The 6 mL difference is defined in some ongoing ICH clinical trials as the marker of significant haematoma expansion [23]. Furthermore, there is emerging evidence of the volume-dependent effect of perihematoma oedema on outcome [5, 24, 25]. The largest analysis of oedema on outcome comes from the pooled analysis of the INTERACT studies, demonstrating that absolute oedema growth at 24 h was independently associated with increased odds of death or dependency at 90 days (OR = 1.17 (1.02–1.33), $p = 0.025$) [5]. In this

analysis, per millilitre of oedema growth is associated with a 3 % increase of poor outcome, and a 10 mL growth increased the risk of poor outcome by 40 %. We have identified significant volume underestimation in patients with variable slice thickness, and given the volume-dependent effect of ICH and oedema on outcome, the results of future studies will need to be considered in the context of these findings.

We have also demonstrated a 2-min (24 %) difference in segmentation time per scan in favour of Analyze over Osirix. Analyze is available commercially for an annual licence renewal fee or outright purchase, whilst the basic version of Osirix is a downloadable freeware. The cost of software and time taken for segmentation may influence the choice of software by researchers.

Failure of automated output in variable slice thickness

Our results indicate that software-generated volume output is unreliable in scans with variable slice thickness. Analyze underestimates volume on average by nearly 30 % whilst Osirix by approximately 40 %. Our analysis indicates that the software uses the voxel depth of the infratentorial slices for volume calculation when the scans are loaded as one single series (Fig. 1c and [Supplementary Excel File](#)). A solution to this issue is loading the supratentorial and infratentorial regions separately, which can be performed using the ‘DICOM tool’ function in Analyze and manually in Osirix. This approach still requires the software to adjust for effect of gantry tilt on voxel depth, which can additionally overestimate by including the overlap region at the interchange between supratentorial and infratentorial regions. This approach would also increase segmentation and processing time and impact on research output productivity. The automated output from Osirix uses the average area between adjacent image slices multiplied by voxel depth to calculate volume ([Supplementary Excel File](#)), and consequently, ROIs present on a single image slice could not be derived from the automated output and is an additional source of underestimation. Previous studies have also included patients with variable slice thickness [10, 14, 25]. In the Efficacy of Nitric Oxide in Stroke Study, the investigators were unable to obtain automated output from Osirix in scans with variable slice thickness and had to multiply the surface area by scan thickness to derive volume [14], whilst other studies did not specifically address how voxel depth was calculated [10, 25]. Our in-house method remains an estimate of the factual volume, but minimises potential for error.

Limitations

This study is limited by the inclusion of mainly early scans, with the majority of the scans performed within 24 h. Oedema increases in volume and evolves over a period of 2 weeks, whilst haematoma resorption takes place over the same time. Our study could not address whether semi-automated segmentation is also reliable for late oedema. We also did not have magnetic resonance imaging (MRI) scans to correlate with CT-based segmentation as MRI is not routinely performed for stroke diagnosis at the study centres. The reliability of the threshold-based approach against MRI has already been demonstrated previously [10, 11]. Thirdly, our in-house calculation requires additional technical expertise and manipulation beyond the planimetric segmentation, which may not be available to all research centres. A new version of Osirix (7.5) was made available after completion of the study, but the approach to segmentation is unchanged. Since completion of this work, we have modified the in-house Osirix plug-in (developed by S.C.) to enable automated volumetric output using the best practice method, which is available for download with

this manuscript. Our work will help others in the future by providing codes and plug-in to achieve the corrections needed to take true slice thickness into account.

Conclusion

Our results demonstrate that semi-automated volumetric segmentation of haematoma and oedema provides consistent inter-rater, intra-rater and inter-software results and can be performed in a timely manner. The method used to determine voxel depth can substantially influence volumetric measurement, and this is of critical importance to the accuracy of multicentre studies.

Acknowledgments TYW is supported by grants from the Neurological Foundation of New Zealand and the Royal Melbourne Hospital Neuroscience Foundation. AP-J is supported by a National Institute for Health Research Clinician Scientist Award. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health. AM is supported by grants from National Health and Medical Research Council (Australia), the Academy of Finland and the Finnish Medical Foundation.

Compliance with ethical standards We declare that all human and animal studies have been approved by the Helsinki University Hospital and have therefore been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments. We declare that given no identifiable patient data is presented, Helsinki University Hospital waived informed consent for this observational registry study.

Conflict of interest We declare that we have no conflict of interest.

References

1. Hemphill JC 3rd, Greenberg SM, Anderson CS et al (2015) Guidelines for the management of spontaneous intracerebral hemorrhage: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* 46(7):2032–2060
2. Steiner T, Al-Shahi Salman R, Beer R et al (2014) European Stroke Organisation (ESO) guidelines for the management of spontaneous intracerebral hemorrhage. *Int J Stroke* 9(7):840–855
3. Broderick JP, Brott TG, Duldner JE, Tomsick T, Huster G (1993) Volume of intracerebral hemorrhage. A powerful and easy-to-use predictor of 30-day mortality. *Stroke* 24(7):987–993
4. Urday S, Kimberly WT, Beslow LA et al (2015) Targeting secondary injury in intracerebral haemorrhage–perihematomal oedema. *Nat Rev Neurol* 11(2):111–122
5. Yang J, Arima H, Wu G et al (2015) Prognostic significance of perihematomal edema in acute intracerebral hemorrhage: pooled analysis from the intensive blood pressure reduction in acute cerebral hemorrhage trial studies. *Stroke* 46(4):1009–1013
6. Yeatts SD, Palesch YY, Moy CS, Selim M (2013) High Dose Deferoxamine in Intracerebral Hemorrhage (HI-DEF) trial: rationale, design, and methods. *Neurocrit Care* 19(2):257–266

7. Fu Y, Hao J, Zhang N et al (2014) Fingolimod for the treatment of intracerebral hemorrhage: a 2-arm proof-of-concept study. *JAMA Neurol* 71(9):1092–1101
8. Kollmar R, Juettler E, Huttner HB et al (2012) Cooling in Intracerebral Hemorrhage (CINCH) trial: protocol of a randomized German–Austrian clinical trial. *Int J Stroke* 7(2):168–172
9. Gonzales NR, Shah J, Sangha N et al (2013) Design of a prospective, dose-escalation study evaluating the Safety of Pioglitazone for Hematoma Resolution in Intracerebral Hemorrhage (SHRINC). *Int J Stroke* 8(5):388–396
10. Volbers B, Staykov D, Wagner I et al (2011) Semi-automatic volumetric assessment of perihemorrhagic edema with computed tomography. *Eur J Neurol* 18(11):1323–1328
11. Urdy S, Beslow LA, Goldstein DW et al (2015) Measurement of perihematomal edema in intracerebral hemorrhage. *Stroke* 46(4):1116–1119
12. Kosior JC, Idris S, Dowlatshahi D et al (2011) Quantomo: validation of a computer-assisted methodology for the volumetric analysis of intracerebral haemorrhage. *Int J Stroke* 6(4):302–305
13. McCourt R, Gould B, Gioia L et al (2014) Cerebral perfusion and blood pressure do not affect perihematoma edema growth in acute intracerebral hemorrhage. *Stroke* 45(5):1292–1298
14. Krishnan K, Mukhtar SF, Lingard J et al (2015) Performance characteristics of methods for quantifying spontaneous intracerebral haemorrhage: data from the Efficacy of Nitric Oxide in Stroke (ENOS) trial. *J Neurol Neurosurg Psychiatry* 86(11):1258–1266
15. Meretoja A, Strbian D, Putaala J et al (2012) SMASH-U: a proposal for etiologic classification of intracerebral hemorrhage. *Stroke* 43(10):2592–2597
16. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33(1):159–174
17. Kottner J, Audige L, Brorson S et al (2011) Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *Int J Nurs Stud* 48(6):661–671
18. Rodriguez-Luna D, Boyko M, Subramaniam S et al (2016) Magnitude of hematoma volume measurement error in intracerebral hemorrhage. *Stroke* 47(4):1124–1126
19. Divani AA, Majidi S, Luo X et al (2011) The ABCs of accurate volumetric measurement of cerebral hematoma. *Stroke* 42(6):1569–1574
20. Mould WA, Carhuapoma JR, Muschelli J et al (2013) Minimally invasive surgery plus recombinant tissue-type plasminogen activator for intracerebral hemorrhage evacuation decreases perihematomal edema. *Stroke* 44(3):627–634. doi:10.1161/STROKEAHA.111.000411
21. Delcourt C, Huang Y, Arima H et al (2012) Hematoma growth and outcomes in intracerebral hemorrhage: the INTERACT1 Study. *Neurology* 79(4):314–319
22. Davis SM, Broderick J, Hennerici M et al (2006) Hematoma growth is a determinant of mortality and poor outcome after intracerebral hemorrhage. *Neurology* 66(8):1175–1181
23. Meretoja A, Churilov L, Campbell BC et al (2014) The spot sign and tranexamic acid on preventing ICH growth—AUStralia Trial (STOP-AUST): protocol of a phase II randomized, placebo-controlled, double-blind, multicenter trial. *Int J Stroke* 9(4):519–524
24. Appelboom G, Bruce SS, Hickman ZL et al (2013) Volume-dependent effect of perihematomal oedema on outcome for spontaneous intracerebral haemorrhages. *J Neurol Neurosurg Psychiatry* 84(5):488–493
25. Volbers B, Willfarth W, Kuramatsu JB et al (2016) Impact of perihemorrhagic edema on short-term outcome after intracerebral hemorrhage. *Neurocrit Care* 24(3):404–412