Development and Validation of an Automatic Segmentation Algorithm for Quantification of Intracerebral Hemorrhage

Moritz Scherer, MD; Jonas Cordes, MSc; Alexander Younisi, MD; Yasemin-Aylin Sahin, MS; Michael Götz, MSc; Markus Möhlenbruch, MD; Christian Stock, MSc, PhD; Julian Bösel, MD; Andreas Unterberg, MD, PhD; Klaus Maier-Hein, MSc, PhD; Berk Orakcioglu, MD

Background and Purpose—ABC/2 is still widely accepted for volume estimations in spontaneous intracerebral hemorrhage (ICH) despite known limitations, which potentially accounts for controversial outcome-study results. The aim of this study was to establish and validate an automatic segmentation algorithm, allowing for quick and accurate quantification of ICH.

Methods—A segmentation algorithm implementing first- and second-order statistics, texture, and threshold features was trained on manual segmentations with a random-forest methodology. Quantitative data of the algorithm, manual segmentations, and ABC/2 were evaluated for agreement in a study sample (n=28) and validated in an independent sample not used for algorithm training (n=30).

Results—ABC/2 volumes were significantly larger compared with either manual or algorithm values, whereas no significant differences were found between the latter (P<0.0001; Friedman+Dunn’s multiple comparison). Algorithm agreement with the manual reference was strong (concordance correlation coefficient 0.95 [lower 95% confidence interval 0.91]) and superior to ABC/2 (concordance correlation coefficient 0.77 [95% confidence interval 0.64]). Validation confirmed agreement in an independent sample (algorithm concordance correlation coefficient 0.99 [95% confidence interval 0.98], ABC/2 concordance correlation coefficient 0.82 [95% confidence interval 0.72]). The algorithm was closer to respective manual segmentations than ABC/2 in 52/58 cases (89.7%).

Conclusions—An automatic segmentation algorithm for volumetric analysis of spontaneous ICH was developed and validated in this study. Algorithm measurements showed strong agreement with manual segmentations, whereas ABC/2 exhibited its limitations, yielding inaccurate overestimations of ICH volume. The refined, yet time-efficient, quantification of ICH by the algorithm may facilitate evaluation of clot volume as an outcome predictor and trigger for surgical interventions in the clinical setting. (Stroke. 2016;47:2776-2782. DOI: 10.1161/STROKEAHA.116.013779.)

Key Words: computed tomography • computer-assisted image analysis • intracerebral hemorrhage • machine learning • volumetric analysis

Intracerebral hemorrhage (ICH) is a special form of stroke and is characterized by a high initial mortality and severe neurological morbidity of surviving patients.1,2 Various methods have been described to estimate outcome, mortality, or hematoma progression based on clinical parameters or imaging surrogates. The role of surgery in the treatment of ICH still remains inconclusive, despite recent large randomized trials, however.3–9

Part of the controversy about the value of surgical interventions might be explained by the fact that hematoma volume, a key prognostic factor of interest, has rarely been measured in precise volumetric fashion in surgical outcome studies.10–14 Instead, clot volume is commonly estimated by the formula ABC/2 until today, owing to the unparalleled applicability and time efficiency of this method.4 Nevertheless, a significant quantification error has been described in larger, irregular-shaped hematomas or in case of intraventricular hemorrhage.13,15,16 Because surgical evacuation is frequently chosen in the latter cases of large and irregular hematomas, it can be hypothesized that ABC/2 measurements have imposed bias especially on the putative surgical cohort in past studies. For careful reevaluation of ICH volume as an outcome surrogate, advancements in ICH quantification methods hold a high potential utility for clinical practice and research.

Received April 13, 2016; final revision received August 17, 2016; accepted August 29, 2016.

From the Department of Neurosurgery (M.S., A.Y., Y.-A.S., A.U., B.O.), Institute of Medical Biometry and Informatics (IMBI) (C.S.), and Department of Neurology (J.B.), University Hospital Heidelberg, Germany; Junior Group Medical Image Computing, German Cancer Research Center (DKFZ), Heidelberg, Germany (J.C., M.G., K.M.-H.); and Division of Neuroradiology, Heidelberg University Hospital, Germany (M.M.).

The online-only Data Supplement is available with this article at http://stroke.ahajournals.org/lookup/suppl/doi:10.1161/STROKEAHA.116.013779/-/DC1.

Correspondence to Moritz Scherer, MD, Im Neuenheimer Feld 400, 69120 Heidelberg, Germany. E-mail moritz.scherer@med.uni-heidelberg.de

© 2016 American Heart Association, Inc.

Stroke is available at http://stroke.ahajournals.org DOI: 10.1161/STROKEAHA.116.013779
The aim of this study was to develop an automatic segmentation algorithm for brain computerized tomography (CT) that can keep up with ABC/2 in terms of applicability and time efficiency, while producing more accurate volumetric data in a contemporary sample of spontaneous, supratentorial ICH. We performed agreement analysis for results retrieved from the segmentation algorithm, manual segmentations, and ABC/2 to compare all 3 different quantification techniques and sought to validate our segmentation algorithm for use in subsequent clinical series.

Patients and Methods

Patients

From a consecutive database of 374 spontaneous ICH cases treated at neurological and neurosurgical departments of Heidelberg University Hospital from 2008 to 2015, 60 cases were randomly selected to enter this study. Random case selection was performed according to initial CT scans after the onset of neurological symptoms, blinded to clinical data regarding therapy, performed hematoma evacuation, or outcome. ICH associated with vascular malformations, tumor, ischemic stroke, or trauma was excluded. The first 30 cases formed the study sample to derive the segmentation algorithm, and the second 30 cases entered the validation sample.

From 30 spontaneous ICH cases assigned to the study sample, 2 showed preexistent parenchymal damage because of a history of either traumatic brain injury or extensive ischemic stroke, which impeded unequivocal manual volume segmentations. Both cases were excluded from the study, and n=58 cases were used for further analysis.

Scans were performed on different scanners, were unenhanced, had different gantry tilt, and slices varied from 2 to 10 mm in thickness. Image analysis and volumetric analysis was performed using the Medical Imaging and Interaction Toolkit (MITK), a free, open-source research tool for medical image analysis (http://www.mitk.org).17

Image Analysis Workflow

This segmentation project consisted of 2 phases. Manual image segmentations of the study sample were used to derive a learning-based approach for an automatic segmentation algorithm. Read-out data from the study sample were compared with respective manual segmentations and ABC/2 in statistical agreement analysis. In a second phase, the algorithm was tested in an independent validation sample not used for algorithm training.

Manual Segmentations

For manual segmentations, across-slice segmentation of intracranial volumes was performed on axial planes for the supratentorial region. Segmentation included ICH volume, subarachnoid space, and brain parenchyma. Segments were traced manually or using a semiautomatic threshold-based region-growing tool. Labeled voxels from every applicable slice were added with the MITK image statistics application to yield volume measures in milliliter.

ICH volume was additionally estimated by ABC/2.4,18 For this purpose, the slice illustrating the largest overall clot diameter was identified on axial planes and digitally measured to the next millimeter in MITK (A). The applicable clot diameter perpendicular to A was acquired accordingly (B). The number of slices containing the clot were than multiplied by the slice thickness to yield (C). Using the formula (A×B×C)/2, the clot volume was approximated and transformed to milliliter. In cases of intraventricular hemorrhage, intraventricular hemorrhage fractions were not considered for diameter measurement.

Manual image segmentation and ABC/2 were performed by 2 independent raters (M. Scherer and A. Younis, specialized in ICH care) blinded to patient identity and clinical data. Measurements of both independent raters were averaged to yield mean values of manual segmentations for analysis of agreement.

Automatic Segmentation Algorithm

Based on a machine learning methodology, the automatic segmentation approach was realized as a voxel-wise random forest classification task. Random forest classification combines several distinct trained decision trees in an ensemble, and majority voting was applied to obtain the target class labels. Random data selection (bagging) and random node optimization were applied during tree growth to induce variations of individual trees. Data used during training and classification phase were based on local descriptive gray value image features, including first- and second-order statistics and texture features. Each feature describes the neighborhood properties of a voxel and was applied for each case CT. The determined features in combination with manual segmentations with volume labeling for ICH, subarachnoid space, and brain parenchyma were used for training of a random forest classifier for automated imaging analysis. Post-processing of prediction maps included identification of the predominant clot, removal of isolated voxels not assigned to a designated volume or different to its neighbors (remove/smooth island effect), and smoothing of volume boundaries. Smoothing was realized by Gaussian smoothing of the probability maps. To remove islands, morphological operations were applied. The segmentation algorithm provided with a volume readout in milliliter by summing up the voxel volumes for each of the target-class labels. The reported results for the study sample were obtained by k-fold cross validation (k=5). For each of the k-folds in the validation process, the algorithm produced probability maps of the target classes and derived prediction maps for respective target volumes. Means of three repetitive runs were used as final volumes.

After training of the algorithm and analysis of the study sample were completed, a graphic user interface was created to implement the algorithm into MITK as a ready-to-use application. The validation sample was subsequently analyzed using the algorithm application independent from further machine learning steps. In the same manner, new samples can be analyzed with the algorithm for research purposes.

Statistical Analysis

ICH volume segmentations were first analyzed using standard descriptive statistics. Limits of agreement were calculated and Bland–Altman plots were drawn to assess interrater agreement and agreement between the different segmentation approaches.19 Correlation and agreement between raters and the segmentation approaches were further assessed by Pearson correlation coefficients, concordance correlation coefficients (CCC), and coverage probabilities. For the estimates of the CCC and coverage probabilities, lower one-sided 95% confidence limits were calculated.20 A nonparametric repeated measures analysis of variance (Friedman test followed by Dunn’s multiple comparison test) and Wilcoxon signed-rank tests were used to test for differences in measured volumes. P values <0.05 were considered statistically significant. The analyses were conducted in the R language and environment for statistical computing (version 3.3.0) and PRISM software (version 5.0c, GraphPad Inc, CA). The R-package Agreement was used to compute agreement statistics (http://www.R-project.org).

Results

Demographic Data

Analysis was performed in n=58 cases consisting of 22 female and 36 male patients. Basic demographic information for both samples in this study is given in Table 1. Both samples did not show significant differences regarding age, hematoma location, ventricular involvement, use of anticoagulants, or surgical evacuation. Moreover, volumetric characteristics were comparable in both samples.
ICH Quantification

CCCs between independent raters (manual segmentations CCC 0.99, 95% confidence interval [CI] 0.98 and ABC/2 CCC 0.97, 95% CI 0.95) indicated excellent interrater agreement for manually acquired data (see Figure I in the online-only Data Supplement). Manual segmentations were regarded as the volumetric reference for further analysis.

Results from ICH quantification for the study sample and the validation sample are given in Table 1 and illustrated in Figure 1. The segmentation algorithm yielded the smallest mean clot volumes in this study, which did not differ significantly from respective manual segmentations, however. In contrast, ABC/2 produced significantly larger mean clot sizes ($P<0.0001$, Friedman test followed by Dunn’s multiple comparison test), overestimating ICH volume by 23% to 29% (validation sample and study sample, respectively).

Time to generate algorithm segmentations from CT scans took ≈20 to 30 seconds per case (depending on the number of slices per scan and computer system used).

Correlation and Agreement Analysis in the Study Sample

Algorithm Versus Manual Segmentation

Strong correlation and agreement between results obtained from the segmentation algorithm and manual segmentations was illustrated by a CCC of 0.95 (95% CI 0.91), Pearson $r=0.96$, $P<0.0001$. Mean deviation from true ICH volume was –3.1 mL (95% limits of agreement –16.0 to 20.0 mL; Table 2). This difference did not reach statistical significance ($P=0.06$, Wilcoxon signed-rank test) but indicated a trend toward underestimation. Best agreement applied below 40 mL, with expanding 95% limits of agreement for larger hematomas. See Figure 2A and Figure II in the online-only Data Supplement.
Supplement for illustration. Coverage probability in Table 2 suggests that hematoma volumes can correctly be approximated to at least 10 mL in 71% of cases and to at least 20 mL in 97% using the automatic segmentation algorithm.

**ABC/2 Versus Manual Segmentation**

Mean deviation between ABC/2 and manual segmentations was 13.0 mL (95% limits of agreement −29.3 to 55.6 mL; Table 2). This was statistically significant (P=0.0008, Wilcoxon signed-rank test), indicating systematic overestimation of ICH volume by ≈30% compared with the manual reference. ABC/2 had weaker agreement with manual segmentations compared with the algorithm indicated by a CCC of 0.77 (95% CI 0.64), Pearson r=0.84, P<0.0001. Agreement decreased particularly in hematomas exceeding 40 to 50 mL (Figure 2B). The probability that hematoma volume was accurately approximated by ABC/2 to at least 10 mL was 29%; to at least 20 mL, it was 54%.

**Validation of the Segmentation Algorithm: Validation Sample**

The independent validation sample confirmed results from the study sample (see Table 1): volumes retrieved from the segmentation algorithm did not show a significant difference to manual segmentations (P=0.29; Wilcoxon signed-rank test). In contrast, ABC/2 volumes were significantly larger than manual and algorithm measurements (P=0.004 and P=0.003, respectively; Wilcoxon signed-rank test). Agreement with manual correspondents was corroborated for algorithm (CCC 0.99, 95% CI 0.98, Pearson r=0.99, P<0.0001) and ABC/2 (CCC 0.82, 95% CI 0.72, Pearson r=0.91, P<0.0001). See Figure 3 and Figure II in the online-only Data Supplement for illustration.

**Table 2. Correlation and Agreement Statistics**

<table>
<thead>
<tr>
<th>Agreement Statistics</th>
<th>Study Sample</th>
<th>Validation Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Differnces, mL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range (min, max [abs])</td>
<td>−16.0, 20.0 [36.0]</td>
<td>−29.5, 76.1 [105.6]</td>
</tr>
<tr>
<td>Mean</td>
<td>−3.1</td>
<td>13.5</td>
</tr>
<tr>
<td>Median</td>
<td>−3.8</td>
<td>9.7</td>
</tr>
<tr>
<td>IQR</td>
<td>−8.0, −0.7</td>
<td>2.4, 21.7</td>
</tr>
<tr>
<td>95% LOA, mL (low, high [abs])</td>
<td>−19.8, 13.7 [33.5]</td>
<td>−29.3, 56.4 [85.7]</td>
</tr>
<tr>
<td>CCC (95% CI)</td>
<td>0.95 (0.91)</td>
<td>0.77 (0.64)</td>
</tr>
<tr>
<td>CP &lt;10 mL (95% CI)</td>
<td>0.71 (0.59)</td>
<td>0.29 (0.22)</td>
</tr>
<tr>
<td>CP &lt;15 mL (95% CI)</td>
<td>0.89 (0.77)</td>
<td>0.42 (0.33)</td>
</tr>
<tr>
<td>CP &lt;20 mL (95% CI)</td>
<td>0.97 (0.89)</td>
<td>0.54 (0.43)</td>
</tr>
</tbody>
</table>

Agreement statistics for comparisons of different measurement techniques for ICH volume. Coverage probabilities (CP) give the probability for the test value to approximate the reference value by a given amount. CCC indicates concordance correlation coefficient; 95% CI, lower one-sided 95% confidence limit; CP, coverage probability; ICH, intracerebral hemorrhage; IQR, interquartile range; and LOA, limits of agreement.
Stroke November 2016

ABC/2 Versus Algorithm Segmentation

CCCs show concordant agreement for ABC/2 with either manual or algorithm segmentations (Table 2), which is further illustrated by respective coverage probabilities to detect ICH volume differences in a clinical context. Based on ABC/2 being the clinical standard of ICH volumetry, this implicates that algorithm could replace manual measurements without significantly compromising on accuracy.

Merging all patients in this study, algorithm segmentations were found to be closer to their respective manual segmentations than ABC/2 in 52/58 cases (89.7%). In the remaining n=6 cases, ICH volumes retrieved by the algorithm were indifferent from ABC/2 (P=0.44; Wilcoxon signed-rank test). Use of antithrombotic drugs (19/58 cases) was not associated with reduced accuracy of the algorithm compared with cases without such medication (P=0.60; Mann–Whitney test).

Discussion

In this study, we report the development and validation of an automatic segmentation algorithm that provides time-efficient ICH volumetry in brain CT. Agreement analysis for 3 different volumetric techniques and validation of results in an independent data set confirm the accuracy of algorithm segmentations comparable to time-consuming manual segmentations and reveals limited validity of the commonly used ABC/2 method for estimating ICH volume.

ABC/2 exhibited known limitations in ICH quantification. Especially in larger hematomas, we observed disagreement between raters as well as systematic volume overestimation by 30% on average (Figure 2B). In 2 trials evaluating the benefit of surgery versus initial conservative treatment in ICH, larger and superficial hematomas were significantly more likely to cross over and receive surgery, despite initial randomization to conservative treatment. Especially, those hematomas were particularly exposed to inaccurate ABC/2 measurements, however, which implies significant volume bias within the surgical cohort. Regarding our hypothesis, this finding raises the possibility that inaccuracy of ABC/2 is a contributor to inconclusive results regarding surgical outcome after ICH. Moreover, it illustrates why it is of clinical interest to improve methods of ICH quantification to allow for a careful reevaluation of ICH volume as an outcome surrogate.

Previous approaches to automatic ICH segmentation have been driven primarily by Hounsfield unit thresholds. However, those segmentations are susceptible to misclassification caused by threshold overlap, hardening artifacts.
adjacent to bone, time dependence of clot density, and its heterogeneity under antithrombotic treatment. Moreover, segmentation algorithms have struggled with variation of slice thickness. The segmentation algorithm presented here aimed to combine multiple existent segmentation methodologies in a random forest–based learning approach, which sought to mitigate the methodological limitations mentioned above and improve segmentation results compared with ABC/2.

Although other automatic ICH segmentation techniques have been described in literature, few were validated in comparison to other applicable segmentation methodologies. Nowinski and colleagues have compared agreement of different segmentation algorithms with respective manual segmentations. CCCs for algorithms based on clustering, graph theory, and modified thresholding were 0.87, 0.91, and 0.97, respectively, and they are comparable to the CCCs calculated for our segmentation algorithm. This supports our approach to combine multiple segmentation strategies within a random forest model to tackle challenging segmentation tasks and is a promising distinction of our algorithm.

After development of the algorithm in the study sample, we were able to reproduce results in the independent validation sample. This sought to validate our method based on n=28 reference cases also for other data sets and indicates that our segmentation algorithm is robust to overfitting of manual data on which it is based. The apparently superior agreement for the algorithm in the validation sample (Figure 3A versus Figure 2A and Table 2) should be carefully balanced against potential effects of selection bias in our small sample, which could render this a coincidental observation. Nevertheless, comparability of both samples for the purpose of validation was indicated by comparable baseline characteristics (Table 1) and concordant shortcomings in agreement observed for ABC/2 and manual segmentations (Figure 3B versus Figure 2B).

Our study has limitations. First, it represents only a small sample of spontaneous ICH cases. The main reason to reduce case numbers in this study was the time-consuming preparation of high-resolution manual segmentations required for algorithm training. Random selection resulted in composition of a contemporary ICH sample regarding age, antithrombotic drugs, treatment, and clot size. Selection did not account for balance of different hematoma locations or occurrence of intraventricular hemorrhage, however, which could have negatively influenced the accuracy of ABC/2 in agreement analysis. Nevertheless, strong agreement of algorithm and manual segmentations simultaneously supports the algorithm’s validity, despite a potential selection bias.

Accuracy of segmentations is known to correlate with clot size and decreased particularly in clots >30 to 40 mL in this study (Figures 2 and 3). Direct comparison of algorithm and ABC/2 measurements (Figure 2C) suggests independent errors affecting each respective method because any mutual error-causing disagreement from manual segmentations would be cancelled out in a Bland–Altman comparison. Although error is primarily driven by size, shape, and location in ABC/2, we identified increased misclassification in hematomas exhibiting multiple clot stages and inhomogeneous density in the segmentation algorithm. Because hematoma size and irregularity in shape or density are often closely correlated, the potential for increased error in larger clots is reasonable for both methods, yet based on different mechanisms.

As for the segmentation algorithm, we did not observe a reduction in accuracy associated with antithrombotic drug use. However, our random sample was not balanced regarding this variable to draw final conclusions about the accuracy of the algorithm in patients on antithrombotic drugs. Because antithrombotic drugs, in general, and oral anticoagulants, in particular, are typical causes of ICH presenting with different stages of clot formation and density inhomogeneity, those drugs should, therefore, be reminded as a possible source of compromised accuracy for the algorithm along with traumatic or vascular causes of ICH.

Automated image analysis has the potential to facilitate image interpretation and could supersede manual measurements in the future. As a preliminary step in this regard, we developed a software tool that has the potential to replace ABC/2 and make time-consuming manual segmentations unnecessary. In a clinical context, this tool offers the opportunity for reevaluation of clot size as a trigger for surgical interventions and as a variable for patient outcome under avoidance of a volume bias previously induced by ABC/2. This might depict an approach to resolve current controversy in ICH treatment. However, the application of our segmentation algorithm is currently limited to research purposes, and its clinical validity and its prognostic value have to be evaluated in future clinical studies.

Conclusions

For volumetric analysis of spontaneous ICH, an automatic segmentation algorithm is developed and validated in this study. Algorithm measurements showed strong agreement with manual segmentations, while limitations of the commonly used ABC/2 method were illustrated yielding inaccurate overestimations of ICH volume. The refined, yet time-efficient, quantification of ICH by the segmentation algorithm provides an image analysis tool for evaluation of clot volume as an outcome predictor and trigger for surgical interventions in the clinical setting in the future.

Acknowledgments

We thank Oliver Roth for retrieving ICH data from patient records.

Disclosures

None.

References


Development and Validation of an Automatic Segmentation Algorithm for Quantification of Intracerebral Hemorrhage

Moritz Scherer, Jonas Cordes, Alexander Younsi, Yasemin-Aylin Sahin, Michael Götz, Markus Möhlenbruch, Christian Stock, Julian Bösel, Andreas Unterberg, Klaus Maier-Hein and Berk Orakcioglu

*Stroke*. 2016;47:2776-2782; originally published online October 4, 2016;

doi: 10.1161/STROKEAHA.116.013779

*Stroke* is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Copyright © 2016 American Heart Association, Inc. All rights reserved.

Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the World Wide Web at:

http://stroke.ahajournals.org/content/47/11/2776

Data Supplement (unedited) at:

http://stroke.ahajournals.org/content/suppl/2016/10/20/STROKEAHA.116.013779.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in *Stroke* can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:

http://www.lww.com/reprints

Subscriptions: Information about subscribing to *Stroke* is online at:

http://stroke.ahajournals.org/subscriptions/
Development and Validation of an Automatic Segmentation Algorithm for Quantification of Intracerebral Hemorrhage

Moritz Scherer¹ M.D, Jonas Cordes² M.Sc., Alexander Younsi¹ M.D., Yasemin-Aylin Sahin¹ M.S., Michael Götz² M.Sc., Markus Möhlenbruch³ M.D., Christian Stock⁴ M.Sc., Ph.D., Julian Bösel⁵ M.D., Andreas Unterberg¹ M.D. Ph.D., Klaus Maier-Hein² M.Sc. Ph.D., Berk Orakcioglu¹ M.D.

¹Dpt. of Neurosurgery, University Hospital Heidelberg, Germany
²Junior Group Medical Image Computing, German Cancer Research Center (DKFZ), Heidelberg, Germany
³Division of Neuroradiology, Heidelberg University Hospital, Heidelberg, Germany
⁴Institute of Medical Biometry and Informatics (IMBI), University Hospital Heidelberg, Germany
⁵Dpt. of Neurology, University Hospital Heidelberg, Germany
Online Supplement Figure 1. Interrater agreement in manual measurements

(A) Interrater agreement between independent raters M.S. and A.Y for manual segmentations illustrated in Bland-Altman-Plots. (B) Agreement for ABC/2 measurements of ICH volume.
Online Supplement Figure II. Pearson correlation for algorithm and manual segmentations

(A) correlation in the study sample and (B) correlation in the validation sample