Interobserver Variation of the Bolus-and-Burst Method for Pancreatic Perfusion with Dynamic – Contrast-Enhanced Ultrasound

Authors
Marcus Stangeland¹, Trond Engjom¹-², Martin Mezl³, Radovan Jirík⁴, Odd Helge Gilja¹-⁵, Georg Dimcevski¹-⁵, Kim Nylund³

Affiliations
1 Department of Clinical Medicine, University of Bergen, Bergen, Norway
2 Department of Medicine, Haukeland Universitetssjukehus, Bergen, Norway
3 Dept. of Biomedical Engineering, Brno Univ. of Technology, Brno, Czech Republic
4 Academy of Sciences of the Czech Republic, Institute of Scientific Instruments, Brno, Czech Republic
5 National Center of Ultrasound in Gastroenterology, Haukeland University Hospital, Bergen, Norway

Key words
interobserver, dynamic contrast-enhanced ultrasound, perfusion, pancreas

ABSTRACT

Purpose Dynamic contrast-enhanced ultrasound (DCE-US) can be used for calculating organ perfusion. By combining bolus injection with burst replenishment, the actual mean transit time (MTT) can be estimated. Blood volume (BV) can be obtained by scaling the data to a vessel on the imaging plane. The study aim was to test interobserver agreement for repeated recordings using the same ultrasound scanner and agreement between results on two different scanner systems.

Materials and Methods Ten patients under evaluation for exocrine pancreatic failure were included. Each patient was scanned two times on a GE Logiq E9 scanner, by two different observers, and once on a Philips IU22 scanner, after a bolus of 1.5 ml Sonovue. A 60-second recording of contrast enhancement was performed before the burst and the scan continued for another 30 s for reperfusion. We performed data analysis using MATLAB-based DCE-US software. An artery in the same depth as the region of interest (ROI) was used for scaling. The measurements were compared using the intraclass correlation coefficient (ICC) and Bland Altman plots.

Results The interobserver agreement on the Logiq E9 for MTT (ICC = 0.83, confidence interval (CI) 0.46–0.96) was excellent. There was poor agreement for MTT between the Logiq E9 and the IU22 (ICC = −0.084, CI −0.68–0.58). The interobserver agreement for blood volume measurements was excellent on the Logiq E9 (ICC = 0.9286, CI 0.7250–0.98) and between scanners (ICC = 0.86, CI = 0.50–0.97).

Conclusion Interobserver agreement was excellent using the same scanner for both parameters and between scanners for BV, but the comparison between two scanners did not yield acceptable agreement for MTT. This was probably due to incomplete bursting of bubbles in some of the recordings on the IU22.
Introduction

Dynamic contrast-enhanced ultrasonography (DCE-US) imaging is a real-time, low-cost application that can be used to measure perfusion in tissue. An ultrasound contrast agent (UCA) is given intravenously and a selected area is observed over time. Ultrasound intensity in a region of interest (ROI) plotted over time results in a time intensity curve (TIC). Since ultrasound data are relatively noisy, it is common to fit the data to mathematical models from which a number of amplitude or temporal features can be derived [1, 2].

DCE-US is not included in clinical guidelines [3, 4], which is most likely due to standardization difficulties [5, 6]. Briefly summarized, differences in settings of the ultrasound scanner, patient characteristics, injection technique, ultrasound contrast behavior, presence of large vessels in the ROI, region selection and tissue motion are all factors that may introduce variability [5–8]. Studies with strict standardization criteria show promising results [9], while there is high variability and low reproducibility when attempts are made to compare different equipment [10].

The two main approaches for contrast administration when performing DCE-US are disruption replenishment [11] and bolus tracking [9]. With bolus tracking the evolution of tissue concentration is observed over time after a bolus injection of UCA. The method is easy to apply and allows the estimation of absolute perfusion parameters. Unfortunately, patient characteristics and injection technique influence the temporal-based parameters. These differences can be described by the arterial input function (AIF). The AIF is mixed with tissue function in TIC, however. This makes it challenging to measure and model the actual tissue perfusion or tissue residue function (TRF) [12, 13]. In the burst replenishment technique, the UCA is given as a steady infusion, the UCA in the ROI is destroyed with a burst of ultrasound with a high mechanical index and tissue reperfusion is observed. This process can be done repeatedly and has good reproducibility. However, absolute perfusion parameters cannot be obtained [11].

Recently, a third approach was suggested by Jiřík et al., proposing a new technique for ultrasound perfusion analysis: the bolus-and-burst method in combination with a pharmacokinetic model based on the AIF and TRF [13]. This method combines bolus tracking with burst replenishment, i.e., a burst is applied when the tissue concentration has reached its plateau, and allows for the estimation of the AIF and the true mean transit time of the tissue [13, 14]. The method has been further extended to use a fully parametric pharmacokinetic model [15]. It allows absolute quantification of the mean capillary transit time (MTT), blood volume (BV), and hence also blood flow (BF) using the central volume theorem (BF = BV/MTT).

Absolute quantification in the bolus-and-burst method can be achieved by a scaling procedure using the TIC in a vessel (artery or vein). Absolute quantification can lead to a reduced variability of perfusion-parameter estimates caused by differences in ultrasound scanner type and settings [6, 13, 16].

The bolus-and-burst method has been used to measure absolute blood flow in the intestine and pancreas [14, 17, 18]. Perfusion measurements with DCE-US could prove useful as a noninvasive technique for diagnosing chronic pancreatitis and the loss of exocrine function [17, 19]. However, as we have previously demonstrated, perfusion imaging of the pancreas is challenging and combined with the complexity of the method this has implications for the interobserver agreement of measurements in the clinical application of the model (17). In the present study, the aim was to estimate interobserver agreement in a population with mixed pancreatic exocrine function using the same ultrasound scanner and agreement between two different scanner systems.

Materials and Methods

Subjects

During a one-year period (April 2014-April 2015), patients admitted to pancreatic ultrasound due to various pancreatic diseases or suspicion thereof were prospectively recruited to the study. As the study was designed to look at technical reproducibility between repeated injections by two investigators and on two ultrasound scanners, a preliminary B-mode scan of the pancreas was performed prior to DCE-US and subjects with insufficient visualization of the pancreas were not included. Also, patients were not included if both investigators were not available. No patients were excluded from the study after DCE-US had been performed. We included 10 patients with suspected chronic pancreatitis and cystic fibrosis with suspected exocrine failure. The subjects represent the whole range from severe pancreatic failure to a normal functioning pancreas.

The protocol was approved by the Regional Committee for Medical and Health Research Ethics in Western Norway (approval number: REK: 2010/2857-7) and the study was performed in accordance with the Helsinki Declaration. All subjects signed an informed consent form.

Patient characterization

Patient records were reviewed and all subjects were interviewed. The age and sex of the patient, body mass index, and final diagnosis were documented (Table 1).

Dynamic contrast-enhanced ultrasonography

The subjects fasted >4 h before the examination. Scanning of the pancreas was performed with the subjects in a supine position using a transverse or oblique epigastric probe position.

A GE Logic E9 scanner (GE Healthcare, Milwaukee, WI, USA) and an IU22 ultrasound scanner (Philips, Best, the Netherlands) and their corresponding curvilinear probes for abdominal use (C1-5 for Logiq E9 and CS-1 for IU22) were applied.

The general default contrast settings were used on each scanner and the method for contrast detection was amplitude modulation. Scanner settings such as gain, dynamic range of the video, color map and imaging depth remained constant throughout the study. On the Logiq E9 the send frequency of the transducer was 2.5 MHz, the mechanical index (MI) was 0.10 during DCE-US and 0.85 during a 10-frame burst while the dynamic range was 66 and the color map was 2/0. On the IU22 the send frequency of the transducer was 1.7 MHz, the MI was 0.05 during DCE-US and 0.75 during a 3-frame burst while the dynamic range was 50, the gray map was 2 and the chroma map was 3. Gain was adjusted to minimize background signal from the tissue and was thus adjusted for each study.
The first examination was performed on the Logiq E9 by observer 1 (TE). The second examination was performed on the same scanner by observer 2 (KN). Observer 2 was blinded to the study performed by observer 1 and did not know the probe placement nor the imaging plane used by observer 1. The third examination was performed on the IU22 by observer 1 who could not be blinded to his first examination on the Logiq E9. He aimed to reproduce the imaging planes between the two examinations. There was an interval of 10 min between the examinations, allowing the contrast to dissipate. For each examination 1.5 ml of the ultrasound contrast agent SonoVue (Bracco, Milan, Italy) was given as a bolus over 2 s and flushed with 10 ml of saline over 4 s through an IV line with a diameter of 1.1 mm placed in the left cubital fossa.

The examiner located a part of the pancreas with minimal movement and interference from air shadowing, also avoiding large vessels between the probe and the pancreas. A recording of the contrast enhancement was performed for 60 s before the burst and the scan continued for a further 30 s. Thus, 90-second video recordings were acquired for each bolus injection. The videos were exported in DICOM format, which could be uploaded to the DCE-US software.

Perfusion analysis
The DCE-US (http://www.isibrno.cz/perfusion/) software is MATLAB-based and structured according to the analysis process. To reduce the size of datasets and thus computational load, the sequences were down-sampled to about 3 frames per second. Subsequently, a manual movement correction and single frame exclusion procedure were performed as described in more detail in a previous paper [18]. After the manual correction, the video data was re-linearized using information from the color conversion map and the video dynamic range. Following the conversion from video to intensity-data linear to UCA concentration, an ROI was defined (Fig. 1, panels a-c). The ROI was chosen in a region of the pancreas where movement and other artifacts were minimized. Finally, an artery in close relation to the ROI was identified (Fig. 1, panels d-f) and the area under the TIC curve of the estimated AIF was equalized to the area under the curve of this artery for scaling [14].

The manual motion correction and perfusion analysis were performed by KN, who was blinded to the pancreatic function data and the genotype of the cystic fibrosis patients.

Statistics
Interobserver agreement for the parameters mean transit time and blood volume was examined using the intra-class correlation coefficient (ICC). Since blood flow is a product of these parameters, it was not included in the analysis of interobserver agreement. The ICC has values between 0 and 1 and is considered poor if 0–0.2, fair if 0.2–0.4, good if 0.4–0.75 and excellent if > 0.75. Interobserver variability was also examined with the Pearson correlation coefficient and the Bland Altman plot with limits of agreement. To test for fixed bias, a single sample T-test was performed, and to test for proportional bias, a linear regression analysis was performed. The data analysis was performed using IBM SPSS Statistics software version 23 (IBM Inc., Armonk, NY, USA).

Results
Patients with cystic fibrosis (n = 3), chronic pancreatitis (n = 4) and a normal pancreas (n = 3) were included. The patients represent the whole range from severe exocrine failure to a normal pancreas.

Perfusion parameters
Blood flow (BF) in mL/min/100 mL, blood volume (BV) in mL/100 mL and mean capillary transit time (MTT) in seconds were calculated in all subjects. The results are displayed in Table 2. One recording by the IU22 scanner was lost as it was not stored properly on the scanner. There was excellent interobserver agreement for MTT on the Logiq E9 with an ICC of 0.83 (95 % confidence interval (CI) 0.46–0.96). Between the Logiq E9 and the IU22, there was poor agreement for MTT (ICC = 0.084, CI = 0.68–0.58). The interobserver agreement for blood volume measurements was excellent between observers on the Logiq E9 (ICC = 0.92, CI 0.72–0.98) and between scanners (ICC = 0.86, CI 0.50–0.97).

<table>
<thead>
<tr>
<th>Patient</th>
<th>Diagnosis *</th>
<th>Gender (F = female, M = male)</th>
<th>Age (years)</th>
<th>F-elastase (µg/g)</th>
<th>BMI (kg/m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SCP</td>
<td>M</td>
<td>25</td>
<td>125</td>
<td>21.8</td>
</tr>
<tr>
<td>2</td>
<td>SCP</td>
<td>F</td>
<td>79</td>
<td>148</td>
<td>21.1</td>
</tr>
<tr>
<td>3</td>
<td>CP</td>
<td>M</td>
<td>52</td>
<td>122</td>
<td>22.6</td>
</tr>
<tr>
<td>4</td>
<td>CF</td>
<td>F</td>
<td>16</td>
<td>N/A</td>
<td>25.3</td>
</tr>
<tr>
<td>5</td>
<td>SCP</td>
<td>M</td>
<td>76</td>
<td>146</td>
<td>N/A</td>
</tr>
<tr>
<td>6</td>
<td>CP</td>
<td>M</td>
<td>59</td>
<td>95</td>
<td>24.2</td>
</tr>
<tr>
<td>7</td>
<td>CF</td>
<td>F</td>
<td>40</td>
<td>42</td>
<td>29.0</td>
</tr>
<tr>
<td>8</td>
<td>CP</td>
<td>F</td>
<td>78</td>
<td>83</td>
<td>22.9</td>
</tr>
<tr>
<td>9</td>
<td>CP</td>
<td>F</td>
<td>39</td>
<td>66</td>
<td>22.7</td>
</tr>
<tr>
<td>10</td>
<td>CF</td>
<td>F</td>
<td>56</td>
<td>419</td>
<td>17.0</td>
</tr>
</tbody>
</table>

* SCP = suspected chronic pancreatitis, CP = chronic pancreatitis, CF = cystic fibrosis

Table 1 Patient demographics in a study using ultrasound and contrast agents to improve pancreatic imaging.
There was a significant correlation between observer 1 and observer 2 on the Logiq E9 for MTT ($p = 0.001$) and BV ($p < 0.001$) with correlation coefficients of 0.87 and 0.93, respectively (Fig. 2, panels a and c). For the comparison between the Logiq E9 and IU22 using the same examiner (TE), there was no correlation for the MTT, but significant correlation for blood volume ($r = 0.88, p = 0.002$, Fig. 2, panels b and d).

Bland Altman plots of the MTT and BV were drawn comparing recordings between operators (Fig. 3, panel a and b,) and different scanners (Fig. 3, panels c and d). There was no fixed or proportional bias between observer 1 and 2 on the Logiq E9. However, between the Logiq E9 and the IU22, there was a fixed bias for the blood volume calculation with the IU22 giving consistently lower values. There was also a fixed bias for the MTT.

Discussion

In this study, we found excellent interobserver agreement for repeated measurements of all perfusion parameters using the same ultrasound scanner with the bolus-and-burst method. Between the scanners used in this study, the interobserver agreement was excellent for blood volume, but poor for mean transit time.

The interobserver agreement of the bolus-and-burst method in our study is comparable to that of the burst replenishment technique as applied by Atri et al. [20]. They investigated tumors with variable vascularity, testing different approaches and measured ICC for blood volume estimates between 0.78 and 0.97.

Another study by Gauthier et al. of the bolus method applied on a skin tumor model in nude mice demonstrated an intraobserver variability with a coefficient of variation between 9.06–15.79 for AUC from a time intensity curve [21]. However, this was an experimental study on mice with a strict setup, while our examinations were performed on patients in a clinical setting.

As mentioned in the introduction, only one study has previously addressed the variability of measurements between scanners [10]. They used the Vuebox software which analyzes time intensity curves. This software provides both amplitude-related variables, such as area under the curve, and time-related variables, such as mean transit time, rise time and fall time. This method does not correct for the arterial input function and has no internal scaling to
Table 2  Mean transit time (MTT, s), blood volume (BV, ml/100 ml tissue) and blood flow (BF, ml/min/100 ml tissue) measurements for every patient in a study on the pancreas for the Logiq E9 (LE9-1 = observer 1, LE9-2 = observer 2) and the IU22 ultrasound scanner (IU22). The data from patient 4 on the IU22 was lost as it was not saved correctly on the ultrasound scanner.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Diagnosis</th>
<th>MTT LE9-1 (s)</th>
<th>MTT LE9-2 (s)</th>
<th>MTT IU22 (s)</th>
<th>BV LE9-1 (ml/100 ml)</th>
<th>BV LE9-2 (ml/100 ml)</th>
<th>BV IU22 (ml/100 ml)</th>
<th>BF LE9-1 (ml/min/100 ml)</th>
<th>BF LE9-2 (ml/min/100 ml)</th>
<th>BF IU22 (ml/min/100 ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SCP</td>
<td>4.0</td>
<td>3.5</td>
<td>3.4</td>
<td>2.6</td>
<td>3.0</td>
<td>1.2</td>
<td>38.9</td>
<td>51.0</td>
<td>20.62</td>
</tr>
<tr>
<td>2</td>
<td>SCP</td>
<td>3.6</td>
<td>4.9</td>
<td>1.6</td>
<td>2.0</td>
<td>2.0</td>
<td>0.3</td>
<td>32.8</td>
<td>23.7</td>
<td>10.9</td>
</tr>
<tr>
<td>3</td>
<td>CP *</td>
<td>11.7</td>
<td>12.7</td>
<td>2.0</td>
<td>2.7</td>
<td>2.6</td>
<td>1.3</td>
<td>12.2</td>
<td>12.3</td>
<td>40.2</td>
</tr>
<tr>
<td>4</td>
<td>CF</td>
<td>10.4</td>
<td>16.2</td>
<td>NA</td>
<td>2.5</td>
<td>2.1</td>
<td>2.6</td>
<td>20.7</td>
<td>16.6</td>
<td>14.9</td>
</tr>
<tr>
<td>5</td>
<td>SCP *</td>
<td>7.3</td>
<td>7.5</td>
<td>1.2</td>
<td>1.8</td>
<td>2.2</td>
<td>1.2</td>
<td>20.2</td>
<td>24.1</td>
<td>12.6</td>
</tr>
<tr>
<td>6</td>
<td>CP</td>
<td>5.3</td>
<td>5.5</td>
<td>5.8</td>
<td>2.7</td>
<td>2.6</td>
<td>2.4</td>
<td>41.0</td>
<td>41.8</td>
<td>43.2</td>
</tr>
<tr>
<td>7</td>
<td>CF</td>
<td>3.9</td>
<td>3.7</td>
<td>3.3</td>
<td>2.7</td>
<td>2.6</td>
<td>2.4</td>
<td>28.0</td>
<td>30.6</td>
<td>166.3</td>
</tr>
<tr>
<td>8</td>
<td>CP *</td>
<td>10.6</td>
<td>8.4</td>
<td>1.5</td>
<td>4.9</td>
<td>4.3</td>
<td>4.2</td>
<td>28.0</td>
<td>30.6</td>
<td>166.3</td>
</tr>
<tr>
<td>9</td>
<td>CP</td>
<td>8.8</td>
<td>8.2</td>
<td>8.5</td>
<td>1.7</td>
<td>1.3</td>
<td>1.0</td>
<td>11.5</td>
<td>9.4</td>
<td>6.7</td>
</tr>
<tr>
<td>10</td>
<td>CF</td>
<td>5.0</td>
<td>4.4</td>
<td>4.9</td>
<td>3.2</td>
<td>3.0</td>
<td>3.1</td>
<td>38.0</td>
<td>41.6</td>
<td>38.0</td>
</tr>
</tbody>
</table>

* In these patients there was residual tissue enhancement in the pancreas after the burst on the IU22 ultrasound scanner (SCP = suspected chronic pancreatitis, CP = chronic pancreatitis, CF = cystic fibrosis)

Fig. 2  Panels a and b show correlations for blood volume (BV) between observer 1 and 2 on the Logiq E9 and between Logic E9 and IU22, respectively, in a study using ultrasound and contrast agents to improve pancreatic imaging. Panels c and d show the corresponding correlations for mean transit time (MTT).
absolute values. Although the software in this study was calibrated to the settings of ultrasound probes and scanners, they did not achieve good interobserver agreement. In our study, the interobserver agreement for blood volume measurements between scanners was excellent although there was a fixed bias between the scanners. One explanation for such a bias could be differences in dynamic range between the scanners.

For the whole group of patients, there was no relationship between the MTTs measured on the IU22 and the Logiq E9. While going through recordings from the IU22, it became clear through visual inspection that for some patients (n = 3) the burst was not sufficient for clearing all the bubbles in the ROI. In these patients, the mean transit time was very short, reflecting this incomplete bursting. In the patients where the burst did remove all microbubbles (patients no. 1, 2, 6, 7, 9 and 10), there was good concordance between the IU22 and Logiq E9 (r = 0.94, p = 0.016).

The bolus-and-burst method with internal scaling could potentially solve many of the problems with variability in perfusion analysis as the MTT estimate is corrected using the AIF. Furthermore, the issues of scanner variability are reduced by internal scaling. Nevertheless, some standardization of scanner settings and contrast use are needed. The mechanical index should be low (0.05–0.19) so that bubble destruction is minimized and the dynamic range of the exported videos should be at least 45 dB to reduce errors in converting the videos to linear data [7, 22]. The issue with sampling a 2D region in 3D tissue and out-of-plane motion was not addressed specifically in this paper. However, the interobserver agreement was excellent even though the second observer was blinded to the first observer’s plane selection.

Limitations of study

The pancreas is located deep in the upper part of the abdomen and is frequently hidden behind air-containing organs such as the colon, the stomach or the duodenum. Consequently, there are challenges regarding CEUS recordings since the pancreas is situated relatively close to the diaphragm and abdominal aorta scanning will be disturbed by respiratory movements and pulsation [18, 23]. To deal with these challenges, we chose to include only patients with optimal scanning conditions judged by preliminary ultrasound scanning of the pancreas. However, this approach introduces a selection bias. We still argue that these measures were necessary to reduce the disturbances introduced by poor scanning quality in the final analysis. Putting quality requirements on scanning conditions is not rare in diagnostic decision making from sonographic imaging.

In scanning protocols where a burst is applied, this should be standardized to ensure that the bubbles on the imaging plane are sufficiently destroyed. For simplicity, the default settings for contrast detection were used in each scanner as were the settings for the burst. Upon analysis, it became apparent that the burst on the IU22 was not sufficient in some patients. In retrospect, we should have increased the burst length and MI on the IU22.
The manual motion correction that was performed was cumbersome and time-consuming. A number of out-of-plane frames were excluded. This was done by the same investigator to reduce variability for this part of the analysis. Scaling was also challenging as it required the identification of an artery at the same depth as the ROI. Together, these two steps may account for some of the individual variations between the observers. Furthermore, the manual exclusion process may introduce selection bias in the analysis. In the future these steps should be automated to reduce user effects and improve the applicability of the technique.

Repeated contrast boluses may affect results since there could be leftover UCA bubbles in the circulation from the previous bolus. We tried to avoid this by waiting 5–10 min between the boluses and by removing the microbubbles from microcirculation by applying repeated high MI bursts between the bolus injections. Upon inspection, there was no visible UCA in the ROI before the next bolus was given. We believe that this factor has minimal influence due to the fact that the perfusion calculations are adjusted to baseline noise before the arrival of the UCA in the ROI.

Conclusion

We find very good correlation and good repeatability with the same scanner system. Improvement of the software aiding adequate standardization and enhanced and simplified motion correction and scaling process may eliminate some of the observed variability. We have demonstrated that ultrasound imaging of the pancreas in general and contrast-enhanced ultrasound in particular pose challenges regarding image quality, reproducibility and standardization [17]. The process of movement correction and exclusion of image disturbances and the analysis itself in the bolus-and-burst software are manual and time-consuming and involve several steps that are likely to reduce interobserver agreement. Forthcoming software improvements to reduce the number of manual steps and to automatize tissue recognition are highly warranted to increase the clinical applicability of the method.

Acknowledgements

We are indebted to the personnel at the gastrolab at Haukeland University Hospital without whom this study could not have been performed.

conflicts of interest

Odd Helge Gilja: Lecture fees GE Healthcare, Takeda AS, MEDA AS
Kim Nylund: Lecture fees MEDA AS, Ferring Pharmaceuticals

References


