


# Quantitative MRCP Imaging: Accuracy, Repeatability, Reproducibility, and Cohort-Derived Normative Ranges

Marc H. Goldfinger, PhD,<sup>1\*</sup>  Gerard R. Ridgway, PhD,<sup>1</sup> Carlos Ferreira,<sup>1</sup> Caitlin R. Langford, PhD,<sup>1</sup> Lin Cheng, PhD,<sup>1</sup> Arina Kazimianec, MSc,<sup>1</sup> Andrea Borghetto, BSc,<sup>1</sup> Thomas G. Wright, PhD,<sup>1</sup> Gary Woodward, PhD,<sup>1</sup> Neelam Hassanali, PhD,<sup>1</sup> Rowan C. Nicholls, MSc. (Res),<sup>1</sup> Hayley Simpson, PGCE,<sup>1</sup> Tom Waddell, MSc,<sup>1</sup> Siddarth Vikal, MSc,<sup>1</sup> Marija Mavar, MSc,<sup>1</sup> Soubera Rymell, PGCert,<sup>1</sup> Ioan Wigley, BSc,<sup>1</sup> Jaco Jacobs, PhD,<sup>1</sup> Matt Kelly, PhD,<sup>1</sup> Rajarshi Banerjee, PhD; MD,<sup>1</sup> and J. Michael Brady, FRS, FEng, FMedSci<sup>1,2</sup>

**Background:** Magnetic resonance cholangiopancreatography (MRCP) is an important tool for noninvasive imaging of biliary disease, however, its assessment is currently subjective, resulting in the need for objective biomarkers.

**Purpose:** To investigate the accuracy, scan/rescan repeatability, and cross-scanner reproducibility of a novel quantitative MRCP tool on phantoms and in vivo. Additionally, to report normative ranges derived from the healthy cohort for duct measurements and tree-level summary metrics.

**Study Type:** Prospective.

**Phantoms/Subjects:** Phantoms: two bespoke designs, one with varying tube-width, curvature, and orientation, and one exhibiting a complex structure based on a real biliary tree.

Subjects

Twenty healthy volunteers, 10 patients with biliary disease, and 10 with nonbiliary liver disease.

**Sequence/Field Strength:** MRCP data were acquired using heavily T<sub>2</sub>-weighted 3D multishot fast/turbo spin echo acquisitions at 1.5T and 3T.

**Assessment:** Digital instances of the phantoms were synthesized with varying resolution and signal-to-noise ratio. Physical 3D-printed phantoms were scanned across six scanners (two field strengths for each of three manufacturers). Human subjects were imaged on four scanners (two field strengths for each of two manufacturers).

**Statistical Tests:** Bland–Altman analysis and repeatability coefficient (RC).

**Results:** Accuracy of the diameter measurement approximated the scanning resolution, with 95% limits of agreement (LoA) from –1.1 to 1.0 mm. Excellent phantom repeatability was observed, with LoA from –0.4 to 0.4 mm. Good reproducibility was observed across the six scanners for both phantoms, with a range of LoA from –1.1 to 0.5 mm. Inter- and intraobserver agreement was high. Quantitative MRCP detected strictures and dilatations in the phantom with 76.6% and 85.9% sensitivity and 100% specificity in both. Patients and healthy volunteers exhibited significant differences in metrics including common bile duct (CBD) maximum diameter (7.6 mm vs. 5.2 mm  $P = 0.002$ ), and overall biliary tree volume (12.36 mL vs. 4.61 mL,  $P = 0.0026$ ).

**Data Conclusion:** The results indicate that quantitative MRCP provides accurate, repeatable, and reproducible measurements capable of objectively assessing cholangiopathic change.

Evidence Level: 1

Technical Efficacy: Stage 2

J. MAGN. RESON. IMAGING 2020;52:807–820.

View this article online at [wileyonlinelibrary.com](http://wileyonlinelibrary.com). DOI: 10.1002/jmri.27113

Received Oct 24, 2019, Accepted for publication Feb 18, 2020.

\*Address reprint requests to: M.H.G. Gemini One, 5520 John Smith Drive, Oxford, OX4 2LL, United Kingdom. E-mail: [marc.goldfinger@perspectum.com](mailto:marc.goldfinger@perspectum.com)

From the <sup>1</sup>Perspectum Ltd, Oxford, UK; and <sup>2</sup>Department of Oncology, Medical Sciences Division, University of Oxford, Oxford, UK

Additional supporting information may be found in the online version of this article

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

**H**EPATOBILIARY DISEASES (HBDs) are a major cause of morbidity and mortality around the world, and their accurate diagnosis and monitoring is key to delivering appropriate and cost-effective care for patients.<sup>1</sup> In HBD, chronic inflammation, injury, bacteria, and microbiota have all been implicated in altering the biliary ducts, which may eventually lead to a cycle of hardening, obstruction, and destruction.<sup>2</sup> While pharmaceutical treatments exist for some HBDs, these are often not effective in the long term and none exist for primary sclerosing cholangitis (PSC)—a chronic condition with unpredictable rates of progression.<sup>1,3</sup>

Endoscopic retrograde cholangiopancreatography (ERCP), introduced in the 1970s, became the standard clinical method for the assessment of the biliary ducts.<sup>4</sup> However, ERCP is invasive and poses a significant risk of procedural complications including pancreatitis, with approximately 0.2–1% procedural mortality and a complication rate of 9.8–15.9%.<sup>4–6</sup>

Subsequently, magnetic resonance cholangiopancreatography (MRCP), which uses heavily T<sub>2</sub>-weighted MR sequences to highlight fluid-filled biliary structures, was shown to allow assessment of the biliary ducts to the same standard as ERCP.<sup>7,8</sup> As it is noninvasive and poses no risk to patients, MRCP currently plays an essential role in the diagnosis and monitoring of HBD.

Despite this advantage, MRCP is limited by its qualitative nature, which necessitates operator interpretation and results in high rates of inter- and intrarater variation. A recent study by Zenouzi et al demonstrated that even with high-quality MRCP images, the interpretation of follow-up MRI/3D-MRCP in PSC differed significantly among highly experienced radiologists and clinicians.<sup>9</sup> Furthermore, no standard MRCP imaging sequence exists, resulting in variability of image quality.<sup>10</sup> This lack of standardization further compounds the interpretive inconsistency across practitioners and institutions using MRI for the diagnosis and monitoring of many HBDs.

A recent position statement from the International PSC Study Group identified the lack of standardization as an area of unmet need for imaging techniques in PSC. In addition, the need to detect disease and cancer complications early, and to predict disease progression, was emphasized, both of which are difficult to do with current (qualitative) methods.<sup>11</sup> This represents a huge challenge for the standardization of good clinical practice and evaluation of multicenter clinical trials.

MRCP can also be used in the diagnosis and monitoring of pancreatic diseases. Sugiyama et al demonstrated the ability of MRCP to detect dilatations, strictures and irregularities of the main pancreatic duct (PD) in 2007.<sup>12</sup> In 2014, Manikkavasakar et al also showed the advantages of using MRCP in chronic pancreatitis, from detecting dilated and irregular side duct branches in early stages of chronic pancreatitis, while in late stages of chronic pancreatitis, MRCP can detect ductal dilatation, strictures, and irregularities.<sup>13</sup>

Quantitative MRI techniques for the assessment of the liver parenchyma have been developed and show good correlation with histologically assessed liver fibrosis and inflammation. These techniques use standardized liver MRI protocols and have been shown to predict clinical endpoints.<sup>14,15</sup> At present, quantitative techniques have not been extensively applied to HBD, as manual measurements of ducts are time-consuming and only practical for a small number of measurements on the main ducts. In order to address this, we developed both a standardized imaging protocol and image processing software that enhances conventional MRCP. The result is a 3D model of the biliary tree which not only improves visualization, but more important, provides novel quantitative measures for the direct assessment of ductal anatomy. This enables high-definition evaluation of duct diameter throughout the biliary tree and PDs, as well as tree volume, gallbladder volume, and the identification of candidate strictures and dilatations. Standardized computational analysis has the potential to improve conventional MRCP imaging in assessing HBD pathology and tracking the progression of disease to improve clinical care.

In this article, we report the performance profile of the application of quantitative MRCP imaging across MRI manufacturers and demonstrate its clinical potential in a cohort of healthy volunteers and patients.

## Materials and Methods

### Participants

The study was reviewed and given favorable opinion by the South Central – Oxford C Research Ethics Committee (REC reference: 17/SC/0459) in the United Kingdom. Potential participants were identified from a pool of individuals who had taken part in previous studies or had expressed an interest in participating in studies through our website. The participant cohort consisted of 40 participants aged over 18, and included 20 healthy volunteers, 10 with independently diagnosed parenchymal liver disease, and 10 with independently diagnosed biliary disease. Participant demographics are shown in Table 1, including body mass index (BMI), sex, age, and diagnostic details, and further, in Tables S2 and S3 in the supplementary file. All participants were willing and able to give informed consent, which was obtained in accordance with the standards of Good Clinical Practice (ICH-GCP). Participants were asked to personally sign and date the latest approved version of the Informed Consent Form before any study specific procedures were performed.

### MRCP Acquisition

MRI was performed in accordance with the imaging centers (Cambridge, Leiden, Oxford, UK) procedures regarding safety requirements and contraindications between July and September 2018. Prior to image acquisition, participants fasted for a minimum of 4 hours to reduce fluid secretions within the stomach and duodenum, reduce bowel peristalsis, and promote gallbladder distension. Approximately 15–20 minutes before scanning, participants

**TABLE 1. Participant Demographics and Details of Pathology**

Age	Sex	Weight	Height	BMI	Diagnosis
39	M	86.4	1.76	27.89	Healthy
45	F	69	1.7	23.88	Healthy
35	F	60	1.63	22.58	Healthy
38	F	67	1.72	22.65	Healthy
30	F	65.8	1.75	21.49	Healthy
37	F	55.5	1.6	21.68	Healthy
31	F	67	1.63	25.22	Healthy
25	F	75.5	1.74	24.94	Healthy
24	M	86	1.81	26.25	Healthy
31	F	72.4	1.62	27.59	Healthy
37	F	98	1.67	35.14	Healthy
33	M	63	1.76	20.34	Healthy
24	F	87	1.61	33.56	Healthy
32	M	74	1.75	24.16	Healthy
26	F	68	1.6	26.56	Healthy
24	M	74	1.78	23.36	Healthy
35	M	105	1.95	27.61	Healthy
36	M	70	1.8	21.60	Healthy
26	M	71	1.74	23.45	Healthy
31	M	73	1.78	23.04	Healthy
30	F	57	1.46	26.74	PSC
47	M	85	1.8	26.23	PSC
63	M	75	1.76	24.21	PSC
65	M	81	1.79	25.28	PSC
47	F	89	1.72	30.08	PSC
34	M	80	1.8	24.69	PSC
53	F	103	1.62	39.25	PBC
61	F	71	1.6	27.73	PBC
50	F	52.2	1.59	20.65	PBC
67	F	71.3	1.6	27.85	PBC
36	F	89	1.59	35.20	NAFLD and PSC
29	F	78.5	1.74	25.93	NAFLD
34	M	110	1.75	35.92	NAFLD
57	M	86.7	1.76	27.99	NAFLD
39	M	83.8	1.71	28.66	NAFLD
52	F	77	1.77	24.58	NAFLD and HC

TABLE 1. Continued

Age	Sex	Weight	Height	BMI	Diagnosis
58	M	83	1.8	25.62	HC
25	M	77	1.87	22.02	HC
39	F	64	1.74	21.14	HCV
21	F	58.6	1.65	21.52	Veno-occlusive disease

HC = hemochromatosis, HCV = hepatitis C virus, PSC = primary sclerosing cholangitis, PBC = primary biliary cholangitis, NAFLD = nonalcoholic fatty liver disease.

consumed 200–400 mL of pineapple juice to reduce signal intensity of overlapping fluid within the stomach and duodenum.<sup>16</sup>

All subjects were scanned twice each on the Siemens Prisma and Siemens AvantoFit scanners (Erlangen, Germany) ( $4 \times 40$  scans), respectively, and 10 subjects were scanned twice each on the GE Discovery and GE Optima scanners (Milwaukee, WI) ( $4 \times 10$  scans), respectively, yielding a potential 200 MRCP scans per biliary and gallbladder scan sequence. MRCP data were acquired using 3D multishot fast/turbo spin echo acquisitions with very long echo train lengths and short echo spacing, to generate heavily T<sub>2</sub>-weighted 3D volumetric images. Parameters employed include an echo train length of 199 for Siemens 1.5T, 200 for Siemens 3T, and 120 for GE scanners. The echo time (TE) was 477 msec for Siemens 1.5T, 604 msec for Siemens 3T, and, ~570 msec and 534 msec for GE 3T and for GE 1.5T, respectively. Sixty contiguous slices were acquired with a voxel resolution of  $1.1 \times 1.1 \times 1.1$  mm and an acquisition matrix of  $256 \times 256$  for all scanners. Two images were acquired with different coronal oblique orientations, one to cover the largest extent of the biliary tree and PD, the other to cover the full extent of the gallbladder. Data were acquired with respiratory gating, which was achieved using navigator tracking. Data were acquired during the expiration phase of the breathing cycle, so that the repetition time (TR) varied with breathing rate. Fat suppression techniques were used to suppress signal from fat and parallel imaging techniques were used to reduce scanning time.

### Quantitative Image Analysis

The MRCP quantitative software (MRCP+) processes 3D MRCP acquisitions deriving a quantitative parametric model of the biliary tree and PDs. The first stage of the procedure is enhancement of tubular structures, using anisotropic diffusion,<sup>17</sup> followed by Frangi's multiscale Hessian-based analysis.<sup>18</sup> A proprietary modification to Otsu's thresholding algorithm<sup>19</sup> was used for iso-surfacing connected components. A trained operator distinguished pancreatobiliary components of interest from neighboring components, such as gastrointestinal structures and blood vessels. Next, an intelligent path search algorithm<sup>20</sup> was applied using features from the Frangi analysis together with features from gradient vector flow<sup>21</sup> and other information, which determines an initial path through part of the tree, then recursively follows branches arising from the initial path, until the entire tree is traversed. The diameter of each duct was calculated from perpendicular cross-sections at all points of the tree, enabling subvoxel accuracy for the duct centerlines and

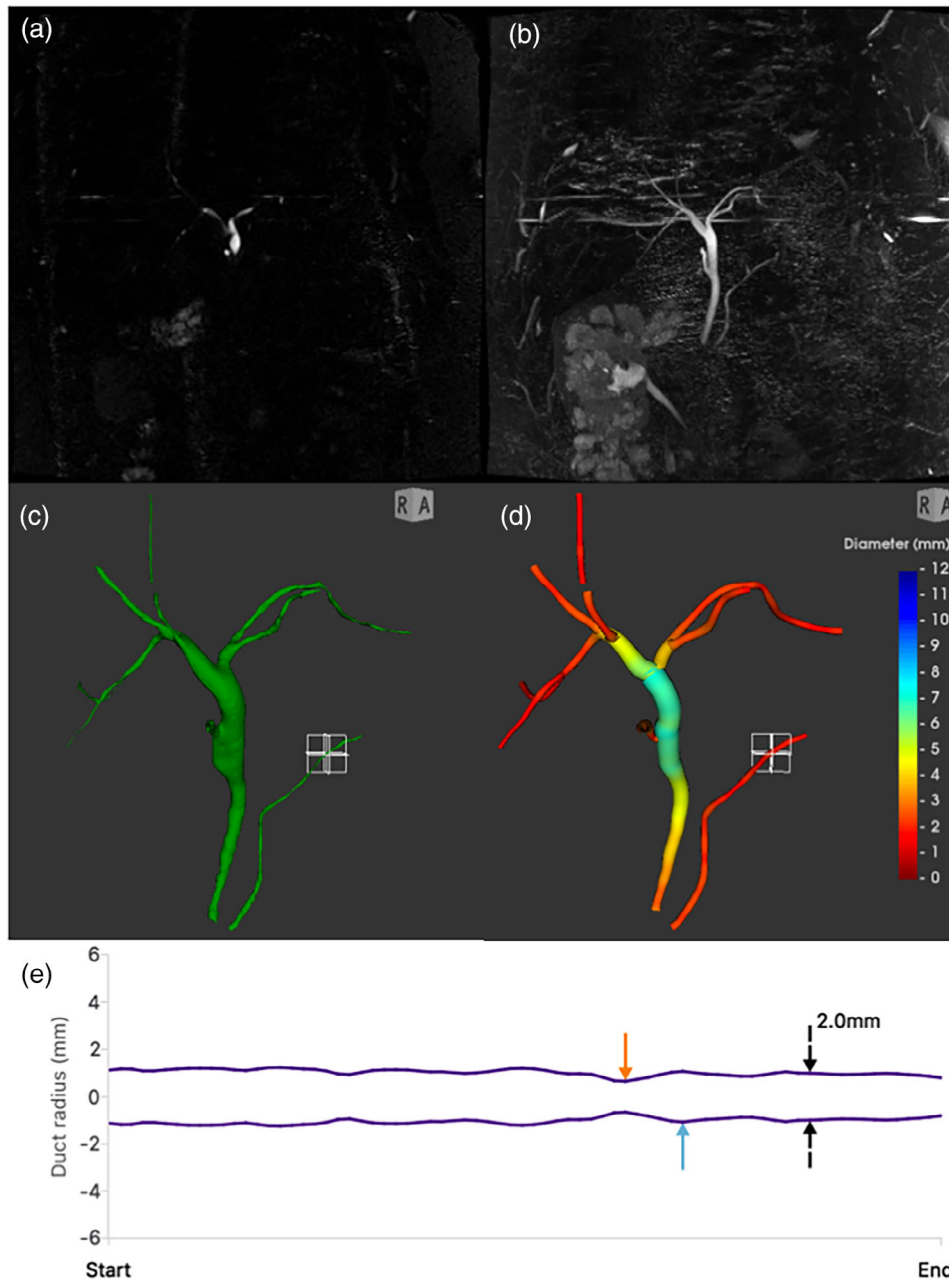
diameter measurements. Multiple quantitative metrics can then be derived from the constructed model, such as diameter percentiles. Additionally, the total volume of the selected pancreatobiliary structures and of the gallbladder were quantified. The resultant model is a color-coded 3D rendering of the biliary tree with interactive plots showing the variation in diameter along each duct (Fig. 1).

### Duct Labeling and Anatomical Considerations

For "normal" pancreatobiliary anatomy, visible in the 3D quantitative model, the operator duct labeling conventions were as follows: common bile duct (CBD); common bile and common hepatic duct (CHD); right hepatic bile duct (RHBD), labeled from the main left/right bifurcation to the first right hemiliver bifurcation (eg, into right anterior and right posterior sectional ducts), or if the RHBD was not visible (cases with anatomical variants) or was too short to model/quantify, then the right anterior bile duct (RABD) and right posterior bile duct (RPBD) were labeled; left hepatic bile duct (LHBD): the LHBD was modeled up to the first left hemiliver branch (ie, in the anatomical variants where a right hemiliver branch comes from the left hepatic duct, the modeled LHBD extends beyond this, eg, to the bifurcation into left medial and left lateral sectional ducts); cystic duct (CD); PD. If ducts were present, but in multiple disconnected pieces, the most representative portion was labeled. If ducts had artifacts or unrepresentative portions (eg, gastrointestinal contamination, artifactual widening near junctions), gaps were introduced in the biliary tree to avoid contamination.

### Measurement Accuracy

Accuracy was assessed using two custom-designed phantoms (Fig. 2), each with known underlying geometry: one with 27 simple tubes exhibiting varying tube-width and curvature ("tubewidth" phantom), and one containing a more anatomically realistic tree structure based on modifying the model from a previously scanned and processed clinical case ("clinical" phantom). Both phantoms had underlying mathematical specifications of the centerline coordinates and diameters, as well as triangulated surfaces that interpolated between the specified points to model duct-like tubes. Simulated MR data were then used to investigate accuracy in a controlled and flexible setting. To do this, the triangulated surface model was voxelized at an upsampled resolution before downsampling to simulate the partial volume effect, then Rician noise was approximated by adding Gaussian noise in quadrature. The tubewidth phantom (which repeated each of nine tube types along the three cardinal



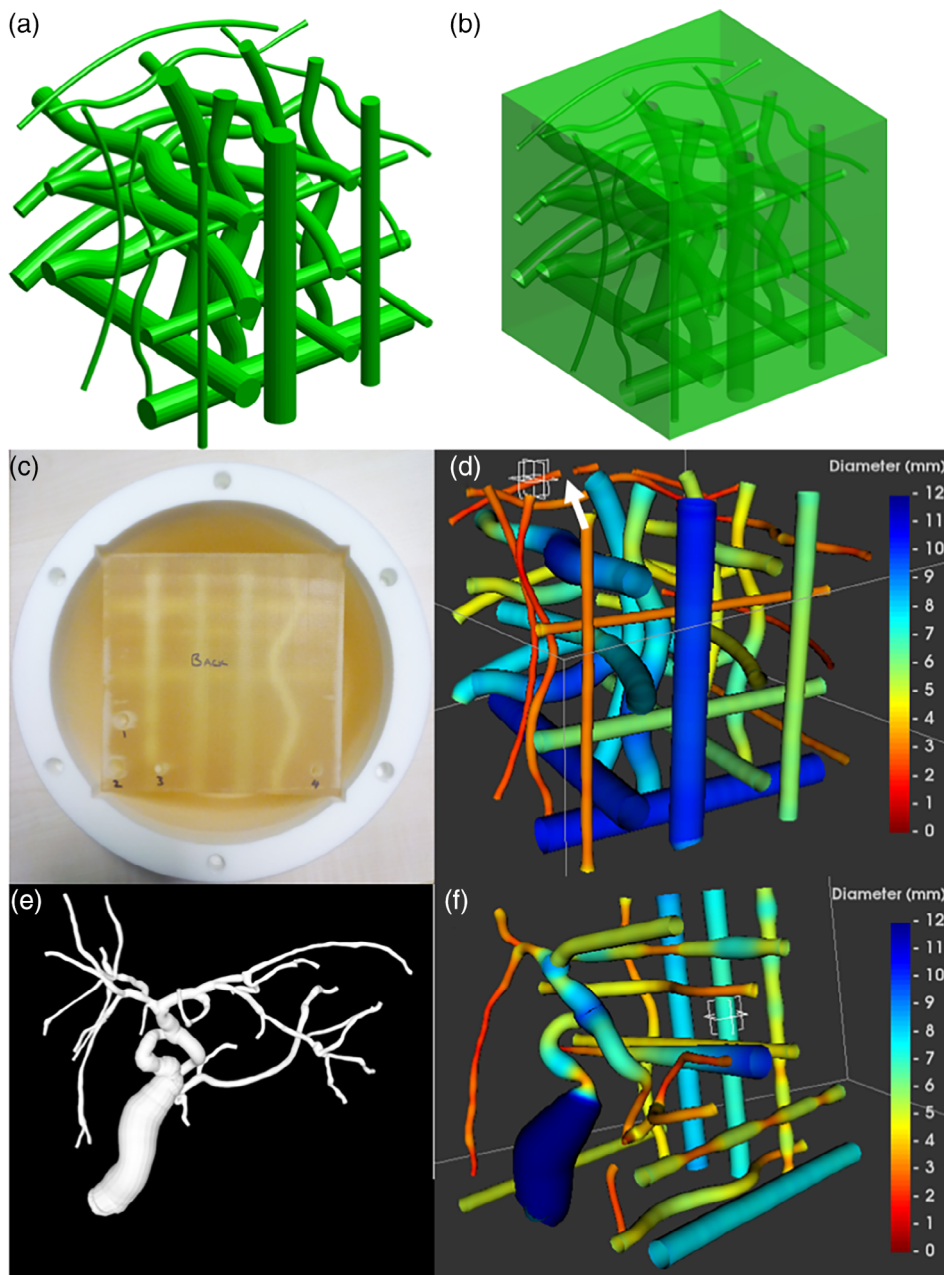
**FIGURE 1: Overview of conventional MRCP and quantitative MRCP for an example case. (a) Coronal slice through a 3D MRCP acquisition, near the bifurcation. (b) Maximum intensity projection (MIP) over all slices. (c) Surface rendering of enhanced ducts. (d) Quantitative MRCP model, with biliary tree and pancreatic duct (PD) colored by diameter. (e) A plot of the modeled diameter profile along the length of the selected PD; the orange and blue arrows denote automatically identified points where the diameter is more than 30% narrower or wider than adjacent regions.**

axes) was used to investigate varying resolution and anisotropy, while the clinical phantom (which included realistically challenging junctions and branches) was used to explore the impact of the varying signal-to-noise ratio (SNR).

The triangulated surfaces from the two phantom designs were converted to hollow regions within a model of a cube, and STL format versions of these were 3D-printed to give physical versions of the phantoms; the resultant cubes can then be placed in separate 3D-printed housings. The hollow regions were fluid-doped with a 1-mM solution of  $\text{NiCl}_2$  such that the expected  $T_1$  was 1050 msec and  $T_2$  was 800 msec, which are both within the ranges of values

for bile reported by Håkansson et al.<sup>22</sup> Accuracy of the 3D printing process was verified using Vernier caliper measurements of duct diameter at the point of entry/exit on the cube surface (Fig. 2).

Accuracy of the quantitative MRCP analysis was defined as proximity of the diameter measurements to the underlying “ground truth,” which were the specifications for the two phantoms. Statistical analysis focused on so-called “stably” matched points; for a given point in the specification, the closest point in the results was found, and then from this point the closest point in the specification was checked, and if this returned to the originally considered specification point, then the match was considered “stable.” This restriction



**FIGURE 2: Quantitative MRCP phantoms.** (a) Mathematical surface model generated from specified centerline coordinates and tube diameters for the “tubewidth” phantom; synthetic 3D voxel arrays can be generated from this at varying resolutions. (b) A version of the model with the same tubes as hollow voids within a cube. (c) The cube has been 3D-printed and is bathed in fluid within a 3D-printed housing that can then be inserted into the scanner. (d) Example results of modeling a scan of the cube; note that the break visible to the right of the crosshair (white arrow) is due to an air bubble trapped in the fluid; the corresponding point will not be counted as a stable match for the evaluation. (e) A modeled surface from a clinical case was used as a starting point for the more anatomically realistic portion of the “clinical” phantom. (f) Example modeled results from a scan of the clinical phantom, showing the additional artificial tubes.

aimed to exclude points where the fluid-embedded 3D-printed phantom contained trapped air bubbles that corrupted the signal at that point.

**Measurement Precision**

Precision is defined in terms of repeatability and reproducibility. Repeatability, performed on the Siemens 3T Prisma scanner, equates to the difference between two acquisitions of the same patient

(or 3D-printed phantom) under the same MRI scanner, roughly 10 minutes apart. The participant or phantom was removed from the scanner, then returned and rescanned, in order to induce realistic positional variation. Reproducibility equates to the difference between the reference scanner (Siemens 3T Prisma scanner [Oxford, UK]) and nonreference scanners (Siemens 1.5T Avanto-fit [Oxford, UK], GE 3T 750 Discovery [Cambridge, UK] and GE 1.5T 450 W Optima [Cambridge, UK]; additionally for the phantoms Philips



1.5T and 3T Ingenia scanners). For the phantom, repeatability and reproducibility were assessed for pointwise diameter measurements, similar to the assessment of accuracy, with stable matches now referring to pairs of stably matched centerline points between two scans, rather than from the specification to one scan. For in vivo data, repeatability and reproducibility were assessed for duct and tree level metrics.

Operator variability was assessed using the same acquired in vivo images. Intraoperator variability was assessed using repeated independent modeling of the same images by one operator, and interoperator variability was assessed comparing this operator's first results to a second operator's independently produced result. In line with the FDA 510(k) clearance and CE marking for the quantitative MRCP service, the operators are radiographers, familiar with hepatopancreatobiliary anatomy and pathologies, and trained to use the quantitative MRCP software.

### Statistical Analysis

#### Accuracy and Precision

Bland–Altman analysis (bias and limits of agreement, LoA) was performed on quantitative MRCP pilot data acquired during development on the reference scanner and single internal operator ( $n = 16$ , on–off–on scanner repeats) using a prototype of the quantitative MRCP software. Based on this pilot, 95% confidence intervals (CIs) for the bias and LoA values were estimated for an assumed  $n$  of 40 patients, allowing for possible 25% dropout and  $n$  of 10 for the GE manufacturer subset analyses. These suggested an adequate sample size for the calculations of performance metrics within 95% CIs for duct median widths within  $\pm 1$  mm. The difference between two replicates were investigated using Bland–Altman analysis by estimating bias, LoA, and the corresponding 95% CIs. The repeatability coefficient (RC), the maximum difference that only 5% of measurement pairs will exceed, was calculated according to standard formulas.<sup>23,24</sup> Inter-/intraobserver differences were assessed by Bland–Altman analysis. The anticipated predicted worst-case scenario was assessed as two separate acquisitions on a single patient on different scanner manufacturers, different field strengths, different scan date, and location processed by different operators. In vivo biliary tree metrics were compared using a Wilcoxon test.

#### Quantitative MRCP Reference Intervals

MRCP images of the participants on the reference scanner were used for the derivation of quantitative reference intervals. The MRCP images were processed using the quantitative MRCP software by a single operator. For normally distributed variables, the 95% prediction interval (95% PI) was calculated as:

$$95\% \text{ PI} = \text{mean} \pm t_{0.975, n-1} \sqrt{\frac{n+1}{n}} \text{sd}$$

where  $t_{0.975, n-1}$  is the 97.5% quantile of a Student's  $t$ -distribution with  $n-1$  degrees of freedom. For nonnormally distributed variables, using the nonparametric method as per the IFCC and NCCLS recommendations, the observations were ranked according to size, and the 2.5 and 97.5 percentiles were obtained as the 0.025 ( $n+1$ ) and 0.975 ( $n+1$ ) ordered observations. These RI thus reflect the 2.5

and 97.5 percentiles,<sup>25–27</sup> which given the sample size, reduce to the minimum and maximum of the sample.

## Results

### Digital Phantom Accuracy

The clinical synthetic phantoms all revealed a stable match above 85% and all had slope of bias at  $-0.03$ , where the bias was 0.0 save for one synthetic case. The LoA ranged from  $-0.4$  to 0.7 (Table S4 in the supplementary file). The tubewidth phantom revealed a stable match of 100% for all phantoms, with a bias in the range of 0.0–0.4 across different slice thicknesses and in-plane resolutions and the LoA in the range of  $-0.6$  (the most extreme value for lower LoA) to 1.3 (the most extreme value for higher LoA). The slope of bias for all phantom analysis was within  $\pm 0.12$ . All results can be viewed in Fig. 3 and Table S5 in the supplementary file.

### 3D Printed Phantom Accuracy

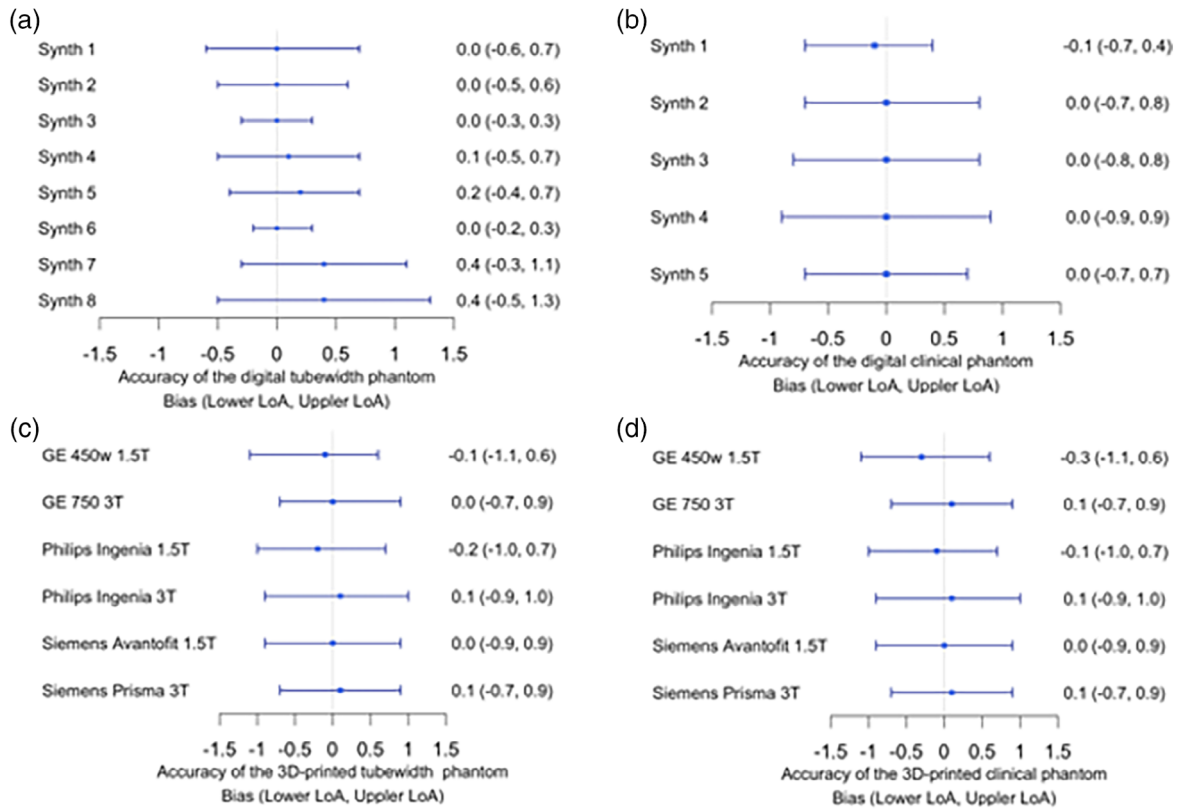
The 3D-printed clinical phantom exhibited 80–85% stable matches, LoA ranges from  $-1.1$  to 1.0, bias range of  $-0.3$  to 0.1, and a slope of bias of  $-0.03$  to  $-0.06$  across scanner models. The 3D-printed tubewidth phantom revealed excellent accuracy with 98–100% stable matches, LoA ranging from  $-1.1$  to 1.0, bias ranging from  $-0.2$  to 0.0, and a slope of bias ranging from  $-0.03$  and  $-0.07$ . All results are listed in Tables S6 and S7 in the supplementary files and can be visualized in Fig. 3. It can be concluded that in general there is an overestimation at lower tubewidths; however, the trend (reflected by the slope of bias) was considered acceptable.

### Repeatability and Reproducibility of Phantoms

Repeatability on the Siemens Prisma 3T scanner had a bias of 0 mm and stable match of 98% for both the tubewidth and clinical phantom. The tubewidth phantom LoA ranged between  $-0.3$  mm to 0.3 mm, while the clinical phantom LoA was found to be in the range of  $-0.4$  mm to 0.4 mm. The slope of bias was  $-0.01$  for the tubewidth phantom, while the slope for the clinical phantom was 0.01. Reproducibility (scanner vs. reference) of the clinical phantom across the five scanners revealed a range of slopes (with the largest slope at  $-0.5$  mm and the smallest at 0 mm) and a range of upper LoA (0.3–0.5 mm) and lower LoA ( $-0.6$  mm to  $-1.1$  mm). Moreover, reproducibility of the tubewidth phantom across the five scanners also presented a range of slopes ( $-0.2$  mm to 0.0 mm) and LoA (with the highest value of 0.7 mm and the lowest value of 0.4 mm for the upper LoA and the lower LoA ranging from  $-0.8$  mm to  $-0.5$  mm). All the above data are illustrated in Fig. 4 and Tables S8–S11 in the supplementary document.

### Detection Sensitivity of Duct Caliber Changes in Phantoms

Three ducts in the clinical phantom contained controlled “stricture-like” and “dilatation-like” changes in diameter with



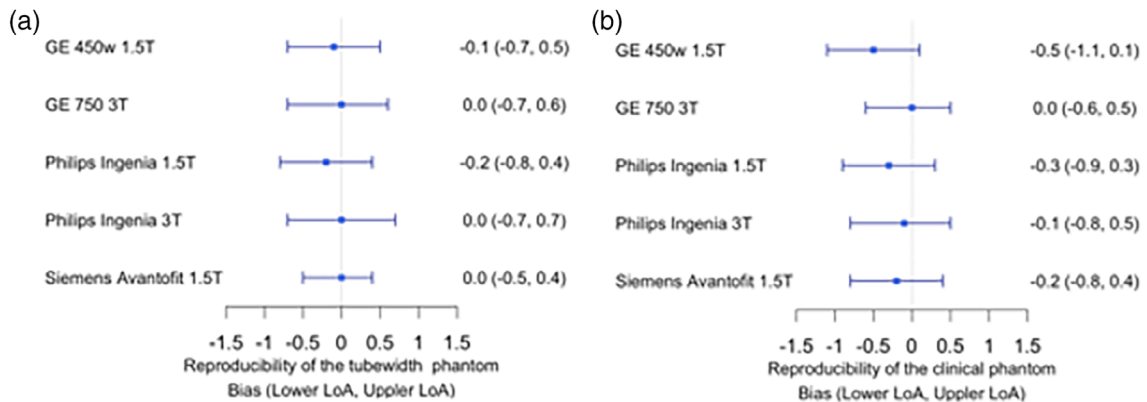
**FIGURE 3:** Forest plots summarizing the Bland–Altman analyses for the accuracy of diameter measurements on the (a) digital tubewidth and (b) digital clinical phantoms and (c) 3D-printed tube-width and (d) 3D-printed clinical phantoms

a range of magnitudes (at or above our chosen minimum relative change of 30%). The detection of these was assessed in synthetic digital phantom instances and in the scanned 3D-printed phantom. In the synthetic cases, six of the seven strictures were detected, resulting in a sensitivity of 85.7%. Between 7–9 dilatations were detected, resulting in a sensitivity from 77.8–100%. The scanned printed phantom revealed a range of stricture detection from 4–6, with a sensitivity ranging from 57.1–85.7%. Dilatation detection ranged from 6–9, with a sensitivity ranging from 66.7–100%. Overall, across all models of the synthetic or printed phantoms,

stricture sensitivity was 76.6% and dilatation sensitivity was 85.9%. No false positives were detected for strictures or dilatations in any of the relevant ducts. All above data are illustrated in Table S12 in the supplementary document.

***In Vivo-Derived Metrics***

Of the 200 biliary sequence scans taken, 180 scan reports were available for statistical analysis, a scan success rate of 90%. Of the 200 gallbladder sequence scans taken, 195 scans reports were available for analysis, a scan success rate of 97.5%.



**FIGURE 4:** Forest plots summarizing the Bland–Altman analyses of reproducibility across scanners for the (a) tubewidth 3D-printed phantom and (b) clinical 3D-printed phantom.

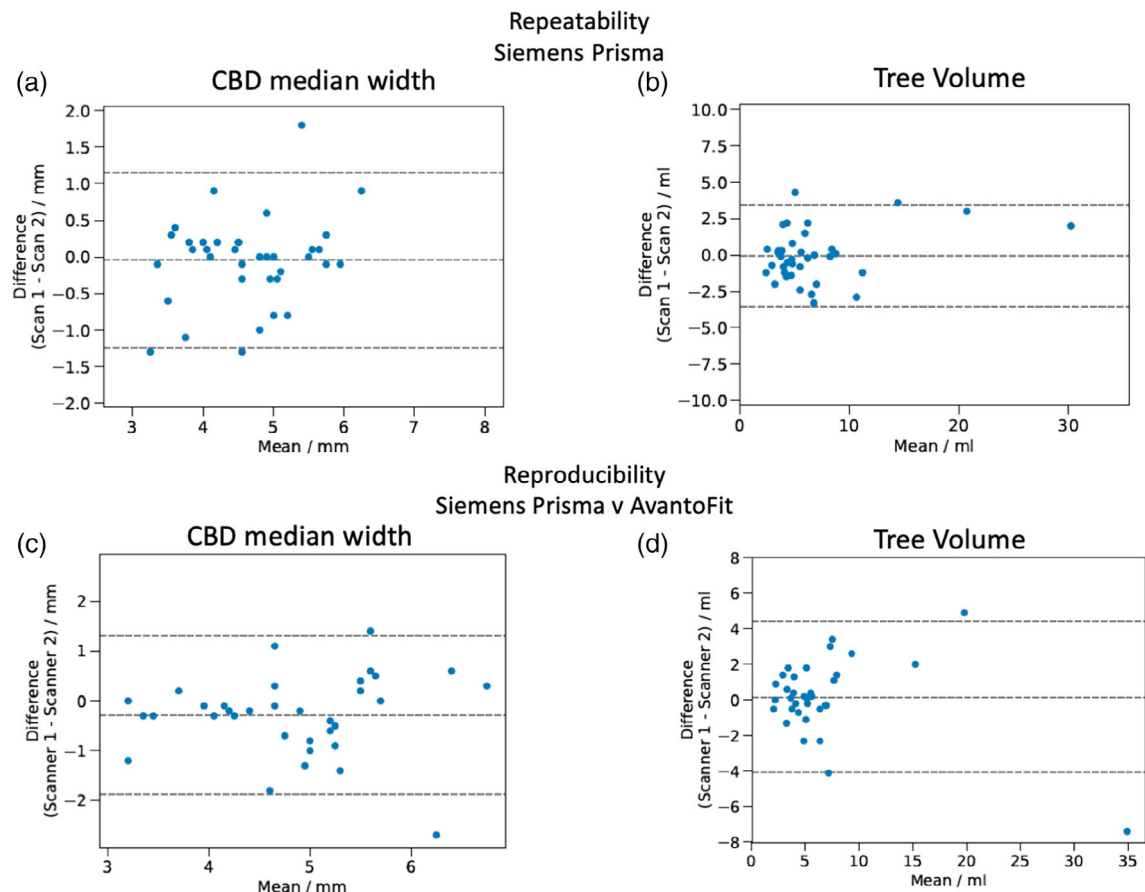


### Repeatability and Reproducibility of Biliary Metrics

Quantitative MRCP revealed moderate repeatability in the reference scanner, in gallbladder volume (bias  $-3.0$  mL, 95% LoA of  $-13.2$  to  $7.3$  mL, RC =  $\pm 14.5$  mL) and biliary tree volume (bias  $-0.1$  mL, 95% LoA of  $-3.5$  to  $3.4$  mL, RC =  $\pm 4.9$  mL). Moderate reproducibility was also observed across scanners in most measurements, including biliary tree volume, an example being Siemens Prisma vs. GE discovery 750 bias  $0.7$  mL, 95% LoA of  $-1.5$  to  $2.8$  mL, RC =  $\pm 3.0$  mL. The Bland-Altman scatterplots of repeatability and reproducibility of the biliary tree volume measurements can be seen in Fig. 5b,d, the remainder of the data are listed in Tables S13 and S14 in the supplementary files. Interoperator agreement was high for gallbladder volume (bias  $0$  mL, 95% LoA of  $-0.7$  to  $0.6$  mL), tree volume (bias  $-1.6$  mL, 95% LoA of  $-6.8$  to  $3.6$  mL), and other metrics, as can be seen in Table S15 of the supplementary tables. Quantitative MRCP revealed a high intraoperator agreement in most measurements such as the gallbladder (bias  $0.0$  mL, 95% LoA of  $0.0$  mL to  $0.0$  mL) and biliary tree volume (bias  $0.2$  mL, 95% LoA of  $-0.6$  to  $0.9$  mL), with other metrics listed in Table S16 in the supplementary files.

### Repeatability and Reproducibility of Individual Duct Metrics

Here we report the metrics that revealed the highest and lowest variation, which we refer to as the best and worst repeatability metrics for the Siemens Prisma 3T. For the assessment of the duct median, the best metrics were identified in the RPBD, with a bias of  $-0.1$  mm, 95% LoA ranging from  $-0.7$  to  $0.5$ , RC =  $\pm 0.8$  mm, and the worst metrics from the CD with a bias of  $-0.2$  mm and 95% LoA in the range of  $-2.0$  to  $1.7$ , RC =  $\pm 2.6$  mm. For the assessment of the duct minimum, the best metrics were identified in the RPBD with a bias of  $-0.1$  mm, 95% LoA in the range of  $-1$  to  $0.7$ , RC =  $\pm 1.2$  mm, and the worst metrics from the CD with a bias of  $-0.1$  mm and 95% LoA in the range of  $-2.2$  to  $1.9$ , RC =  $\pm 2.9$  mm. For the assessment of the duct IQR, the best metrics were identified in the PD, with a bias of  $0.1$  mm, 95% LoA in the range of  $-1.7$  to  $1.8$ , RC =  $\pm 0.8$  mm, and the worst metrics from the CD with a bias of  $-0.1$  mm and 95% LoA in the range of  $-2.2$  to  $1.9$ , RC =  $\pm 2.4$  mm. For the assessment of the duct maximum, the best metrics were identified in the RABD, with a bias of  $0.0$  mm, 95% LoA ranging from  $-1.1$  to  $1.2$ , RC =  $\pm 1.6$  mm,



**FIGURE 5:** Bland-Altman scatterplots reveal moderate repeatability of biliary tree metrics with examples shown for (a) the CBD median ( $0.0$  mm bias,  $-1.2 - 1.1$  mm LoA, RC =  $\pm 1.7$  mm) and (b) total biliary tree volume ( $-0.1$  mL bias,  $-3.5 - 3.4$  mL LoA, RC =  $\pm 2.2$  mL). Reproducibility of quantitative MRCP comparing the results between Siemens Prisma and Siemens AvantoFit in both (c) the CBD median ( $-0.3$  mm bias,  $-1.9 - 1.3$  LoA, RC =  $\pm 2.3$  mm) and (d) total tree volume ( $0.2$  mL bias,  $-4.1 - 4.4$  mL LoA, RC =  $\pm 4.9$  mL).

and the worst metrics from the CBD, with a bias of 0.1 mm and 95% LoA in the range of -2.9 to 3.2, RC = ±4.2 mm.

We also identified the metrics with the smallest and greatest variation as regards reproducibility across scanners. For the assessment of the duct median, the best metrics were identified in the RPBD (Siemens Prisma vs. GE Discovery 750) with a bias of 0.0 mm (LoA -0.5 to 0.5, RC = ± 0.7 mm) and the worst metrics from the CD (Siemens Prisma vs. Siemens AvantoFit) with a bias of -0.1 mm (LoA -2.8 to 2.6, RC = ±3.4 mm). For the assessment of the duct minimum, the best metrics were identified in the LHBD (Siemens Prisma vs. GE Optima 450 W) with a bias of 0.0 mm (LoA -0.4 to 0.5, RC = ±0.7 mm), and the worst metrics from the CD (Siemens Prisma vs. Siemens AvantoFit) with a bias of -0.1 mm (LoA -2.8 to 2.6, RC = ±3.8 mm). For the assessment of the duct IQR, the best metrics were identified in the PD (Siemens Prisma vs. GE Optima 450 W) with a bias of -0.1 mm, (LoA -0.5 to 0.4, RC = ±0.6 mm), and the worst metrics from the CD (Siemens Prisma vs. GE Optima 450 W) with a bias of -0.2 mm (95% LoA -2.3 to 1.9, RC = ±3.0 mm). For the assessment of the duct maximum, the best metrics were identified in the CBD (Siemens Prisma vs. GE Optima 450 W) with a bias of 0.0 mm, (LoA -1.0 to 1.0, RC = ±1.5 mm), and the worst metrics from the CD (Siemens Prisma vs. GE Discovery 750) with a bias of 0.3 mm (LoA -3.5 to 4.0, RC = ± 3.8 mm).

**Normative Ranges Derived From a Cohort of Healthy Controls**

Ranges for biliary tree metrics, derived from a cohort of healthy controls can be seen in Table 2. Ranges for individual duct metrics from healthy controls can be seen in Table 3.

**In Vivo Results of Healthy Controls and Hepatobiliary Disease**

The results show large variability of ductal measurements among the human healthy and biliary disease cohort. Initial analyses revealed that the 10 biliary disease patients had a

significantly larger total average biliary tree volume compared to healthy volunteers (12.36 mL vs. 4.61 mL, P = 0.0026). The maximum value of the CBD of individuals with biliary disease was found to be significantly greater than the healthy volunteers (7.6 mm vs. 5.2 mm P = 0.002) in addition to the median CBD diameter (P = 0.005). The total percentage of ducts in the biliary tree with a diameter of less than 3 mm was found to be significantly greater in healthy volunteers (P = 0.029), while the converse was found for the percentage of ducts between 5–7 mm, whereby the individuals with biliary disease showed a greater percentage (P = 0.0018) (Figs. 6 and 7).

**Discussion**

The aim of this study was to report on the accuracy, reproducibility, and repeatability of our quantitative MRCP software in both phantoms and human volunteers, across varying scanner field strengths, and across models. These results suggest that quantitative MRCP, especially as compared with current qualitative MRCP, could reduce subjectivity, enable measurement of duct diameter throughout the biliary tree, and automatically detect candidate strictures and dilatations with a high degree of sensitivity and specificity. Quantitative MRCP therefore improves upon the current “gold standard” of noninvasive imaging of the biliary tree.

The results from the healthy controls revealed a range of duct volumes and allowed normative ranges for biliary tree features to be established. The potential utility of these measurements goes beyond basic reporting, and with external validation may provide important reference points for clinicians and radiologists when assessing the biliary tree within the clinic.

When compared to other direct imaging modalities within the same patients, MRCP consistently reported significantly different measures of mean duct diameter.<sup>28,29</sup> For example, measuring the bile ducts at the porta hepatis via ultrasound revealed an upper limit of normal of ~6 mm<sup>30</sup> and a mean of diameter 4.0 mm<sup>31</sup>; similar measurements using ERCP revealed average maximal and midportion diameters of the CHD of 6.1 mm and 5.3 mm and 6.4 mm and 5.5 mm for the CBD,<sup>30</sup> whereas employing MRCP revealed mean CBD diameters between 4.13 mm and 4.6 mm.<sup>32,33</sup> Such differences between MRCP, ultrasound, and ERCP imaging techniques can be attributed to compliant bile duct distension during contrast application, where the literature has reported a 2.3 mm discrepancy between ultrasound (transverse diameter) and ERCP diameter measurements.<sup>29</sup> Additionally, with traditional nonquantitative scans, there is no ability to accurately measure the diameter of ducts. In these cases, ducts are assumed to be uniform cylindrical tubes, yet in reality are variable in size and oval in 70% of patients.

**TABLE 2. Biliary Tree Reference Ranges Compiled From Healthy Volunteers**

Measurement	Reference range
Tree volume (mL)	1.2–8.8
Gallbladder volume (mL)	6.7–39.8
3–5 mm (%)	3.0–37.0
5–7 mm (%)	0.0–14.0
Greater than 7 mm (%)	0.0–3.0
Less than 3 mm (%)	60–95

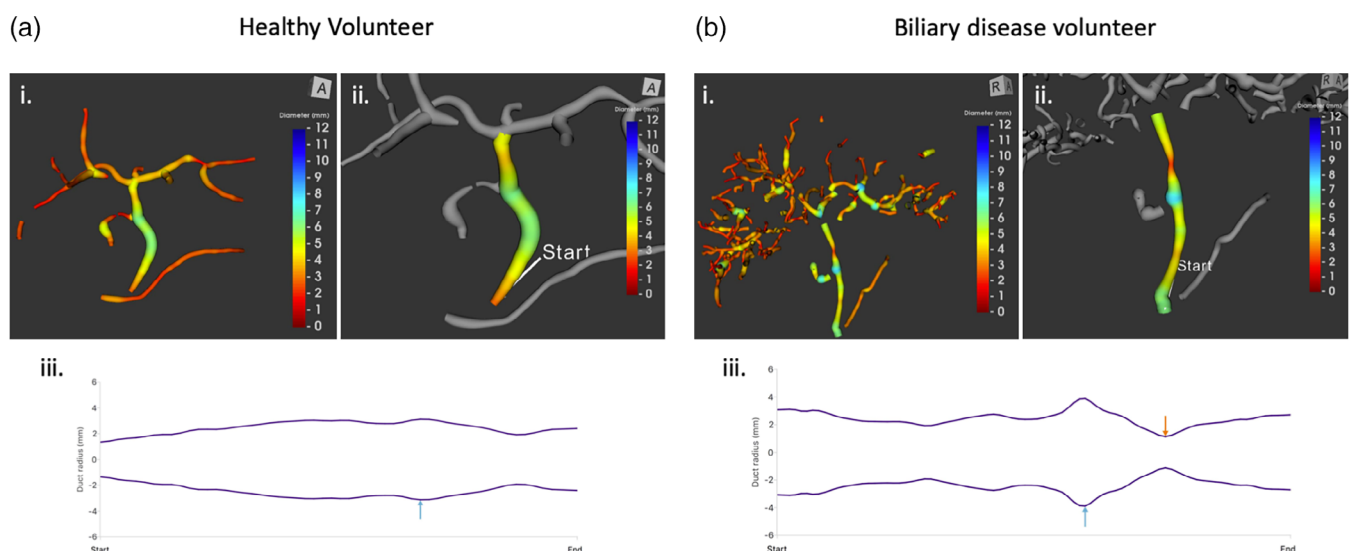
**TABLE 3. Reference Ranges for Individual Ducts, With Interquartile Range (IQR) as a Scalar Range, Maximum (max), Median and Minimum (min) Values**

Duct	Reference range			
	IQR (mm)	Max (mm)	Median (mm)	Min (mm)
CBD	0.1–2.3	3.3–8.9	2.6–6.4	1.7–4.1
CD	0.0–3.5	1.9–7.5	1.3–4.7	0.7–3.6
LHBD	0.0–2.3	2.9–6.2	2.3–5.2	1.8–4.7
PD	0.2–1.5	1.7–5.8	1.4–4.2	0.6–3.2
RABD	0.0–1.5	2.5–5.1	1.9–4.2	1.3–3.9
RHBD	0.1–1.5	2.6–6.2	2.3–5.3	1.7–5.1
RPBD	0.2–1.5	2.6–4.6	1.9–3.3	1.0–2.9

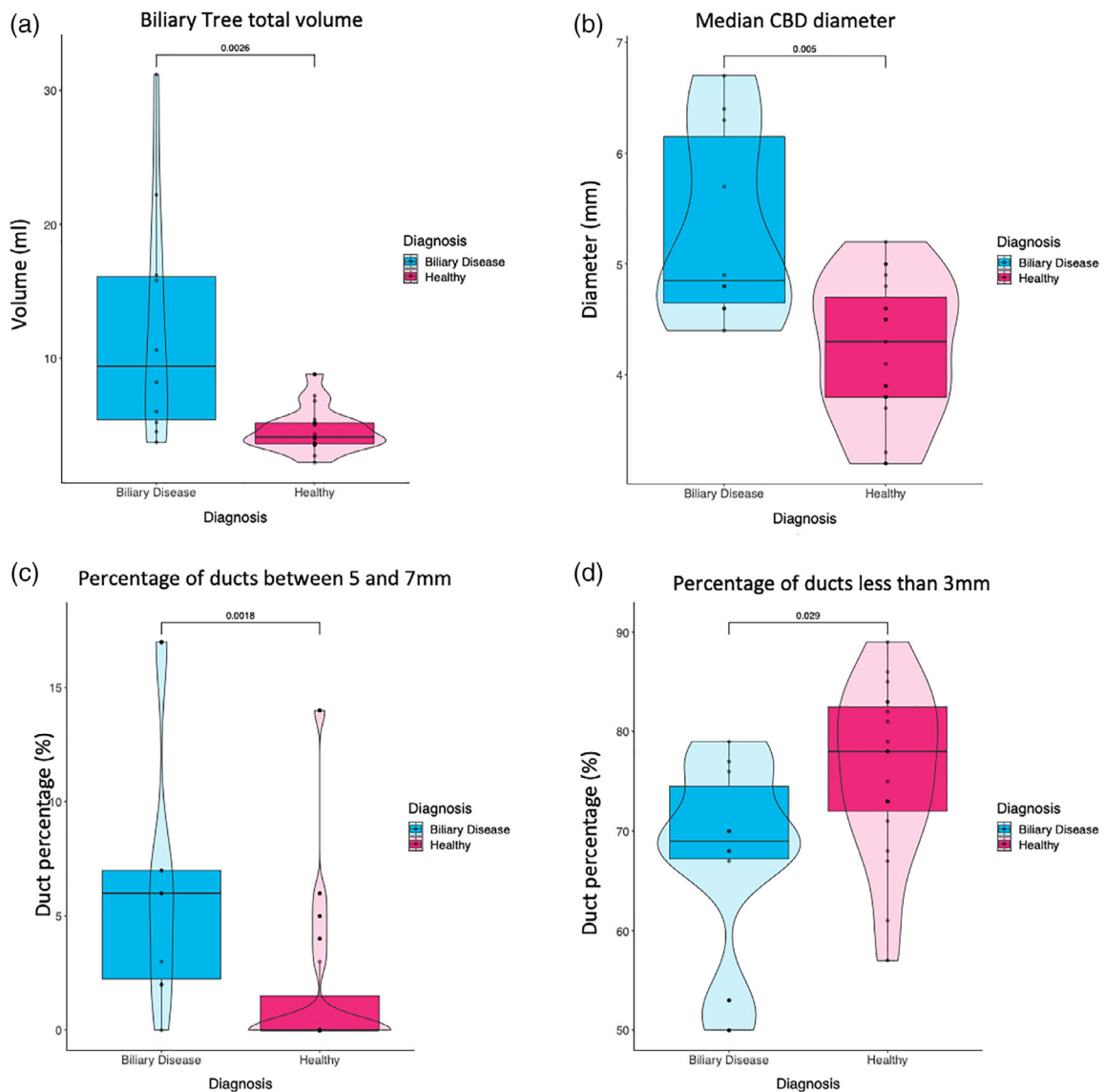
These results highlight the need to establish consistency across the modalities used to establish reference intervals and indicate that direct comparisons across imaging modalities must be interpreted with caution. Quantitative MRCP measures the actual 3D duct diameter, without assumptions, generating more accurate measurements and does so in a standardized, automated method minimizing human error.

It is important to recognize the limitations of each imaging modality. In MRCP, strictures are defined as an area of narrowing (<1.5 mm for CBD or <1 mm for CHD) or nonvisualization of a duct, with proximal dilatation.<sup>3</sup> The spatial resolution of MRCP (1 mm) is less than that of ERCP (0.1–0.5 mm),<sup>34</sup> and therefore may miss

severe strictures. That said, published studies have shown no significant difference between MRCP and ERCP for the detection of bile duct abnormalities such as the presence of strictures, dilatations, and biliary stones.<sup>35</sup> However, there is still greater confidence in diagnoses made by ERCP because of its superior resolution compared to traditional MRCP, despite a patient preference for MRCP over ERCP, as MRCP is not associated with mortality.<sup>8</sup> Furthermore, there is a dearth of measurements reported in the literature beyond CBD and CHD. Our quantitative analysis tool measures duct diameter throughout MRCP-visible intra- and extrahepatic ducts and has high sensitivity to detecting subtle cholangiopathic changes in the diameter of ducts. Future work will focus on further



**FIGURE 6: Comparison of healthy and biliary diseased patients (a) healthy participant (i) total tree volume of 4.1 mL, (ii) isolation and analysis of the CBD, with quantitative results: Median: 5.2 Min 2.7 mm Max 6.2 IQR 4.1–5.8, (iii) CBD diameter profile revealing slight increase (>30%) at blue arrow. (b) Participant with biliary disease, (i) total tree volume 22.2 mL, (ii) isolation and analysis of the CBD, with quantitative results: Median 4.8, Min 2.3 Max 7.8, IQR 4.4–5.8, (iii) CBD diameter profile revealing a stricture (blue) and apparent dilatation (orange).**



**FIGURE 7:** (a) Total biliary tree volume analysis revealed biliary diseased patients to have a larger overall biliary tree compared to healthy volunteers ( $P = 0.0026$ ). (b) The CBD of individuals with biliary disease were found to be significantly larger than the healthy volunteers ( $P = 0.005$ ). (c) The total percentage of ducts in the biliary tree whose diameter was between 5–7 mm were significantly increased in individuals with biliary disease ( $P = 0.0018$ ), (d) whereas the total percentage of ducts in the biliary tree with a diameter of less than 3 mm were found to be significantly greater in healthy volunteers ( $P = 0.029$ ).

characterizing strictures and dilatations, which would be difficult without the aid of computational procedures.

The results from the human cohort revealed substantial variability between those with and without hepatobiliary disease. Quantitative MRCP was also able to identify regions of variation in duct diameter that may reflect candidate strictures and dilatations. In diseases such as PSC, guidelines define a dominant stricture as “a stenosis with a diameter of <1.5 mm in the CBD or of <1 mm in the hepatic ducts.”<sup>3</sup> As yet, the basis for this cutoff is unclear, and is likely based on the resolution of imaging modalities. Quantitative MRCP has the potential to advance this field by defining precisely the parameters for classification of a stricture or dilatation. For example, by setting a percentile

change in tube diameter relative to a normal reference interval, or changes within individual ducts relative to duct diameter, metrics can be standardized and classified more meaningfully in clinical settings. Such standardized measurements could more accurately identify and diagnose strictures and dilatations, in addition to quantifying their severity, and provide clinicians with information related to symptoms reported by patients in the clinic.

Our findings show that individuals with biliary disease present with significantly different biliary metrics compared to healthy volunteers. Measurement of total biliary tree volume revealed that volunteers with biliary disease contained significantly more “MR-apparent” ducts, in addition to a greater percentage of ducts identified with a diameter between

5–7 mm. This may be due to inflammation within the biliary tree, especially within PBD, which results in increasing diameter of bile ducts. This is further supported by the decrease in the total percentage of ducts that are <3 mm seen in PBD patients. Further, the larger diameter of the CBD in the biliary disease cohort may indicate the presence of dilatations within the duct, increasing its size relative to healthy volunteers. These quantitative metrics provide an objective assessment of the biliary system across different disease states, which may be used to stratify individuals and aid in diagnosis. Investigating a larger cohort of patients with various biliary diseases will enable us to identify disease-specific metrics that can aid in diagnosis and monitoring of disease progression and validate our findings.

This study was designed to characterize the technical performance of our quantitative MRCP software. A larger cohort of healthy volunteers would be required to validate the normative ranges reported from the healthy cohort in this study. Due to the complex and variable nature of pancreatobiliary anatomy seen in MRCP, criteria were imposed on acquired images for the metrics to be considered in quantitative MRCP processing. If the ducts were not clearly visible or reliable metrics could not be obtained, then these individual ducts were not used for analysis. This primarily concerned the cystic and PDs and the complex structures sometimes seen at the bifurcation point, which are often difficult to image on MRCP. This was most common within the CD which, anatomically, is complex and can vary between individuals. For example, it has been observed that the CD can either join the intrahepatic right or left hepatic ducts rather than the CBD and even when joining the CBD, the location the duct joins can also vary, with some joining near the ampulla. The length of the CD can also be very short, making it difficult to distinguish and model, which may explain why the CD was commonly found to be the “worst metric” when analyzed for repeatability and reproducibility.

Anatomical variations in duct morphology and anatomy have been observed in some pathological cohorts and are associated with clinical events such as stone formation and biliary tract injuries.<sup>36</sup> While inherent anomalies of ducts are uncommon, in patients who are already suffering from disorders such as HBD the possibility of variation increases, especially within the PD. This may have an unquantified impact on quantitative evaluation of this duct. Variations in the configurations of the PD may present operators with difficulty when modeling, and agreement may differ on labeling. Additionally, intrahepatic duct anatomy is complex with many common and uncommon intra- and extrahepatic variants. The “typical” branching pattern is only seen in 50–60% of the population, leaving a significant proportion of the population with variation in the branching pattern of the intrahepatic ducts.<sup>37</sup> These variances introduce potential difficulties when measuring precision between operators and scans when labeling the 3D tree.

## Conclusion

Quantitative MRCP provides objective and accurate noninvasive measurements that are employable across a range of MRI scanners without the need for additional hardware. High repeatability and reproducibility of this technique allows for comprehensive assessment of the biliary tree over time, providing investigators with an objective tool for accurate monitoring of biliary morphology. Illustrated by the differences observed in these measurements between patients with biliary disease and those without, this novel method has the potential to improve both clinical management and the execution of interventional trials.

## Conflict of Interest

All authors affiliated with Perspectum Diagnostics are employees, and several have stock options within the company. J.M.B. and S.V. are listed on relevant patent applications.

## REFERENCES

1. Chapman R, Fevery J, Kalloo A, et al. AASLD practice guidelines. Diagnosis and management of primary sclerosing cholangitis. *Hepatology* 2010;51(2):660-678.
2. Giordano DM, Pinto C, Maroni L, Benedetti A, Marzioni M. Inflammation and the gut-liver axis in the pathophysiology of cholangiopathies. *Int J Mol Sci* 2018;19:3003.
3. Lindor KD, Kowdley KV, Harrison ME. ACG clinical guideline: Primary sclerosing cholangitis. *Am J Gastroenterol* 2015;110:646-659.
4. Sackmann M, Beuers U, Helmberger T. Biliary imaging: Magnetic resonance cholangiography versus endoscopic retrograde cholangiography. *J Hepatol* 1999;30:334-338.
5. Alberto E, García O, Brizuela RA, et al. Factores de riesgo relacionados con pancreatitis e hiperamilasemia posterior a colangiopancreatografía endoscópica retrograda. *Rev Colomb Gastroenterol* 2015;27:580.
6. Kochar B, Akshintala VS, Afghani E, et al. Incidence, severity, and mortality of post-ERCP pancreatitis: A systematic review by using randomized, controlled trials. *Gastrointest Endosc* 2015;81:143-149.
7. Lee MG, Lee HJ, Kim MH, et al. Extrahepatic biliary diseases: 3D MR cholangiopancreatography compared with endoscopic retrograde cholangiopancreatography. *Radiology* 1997;202:663-669.
8. Kaltenthaler EC, Walters SJ, Chilcott JB, Blakeborough A, Vergel YB, Thomas SM. MRCP compared to diagnostic ERCP for diagnosis when biliary obstruction is suspected: A systematic review. *BMC Med Imaging* 2006;6:9.
9. Zenouzi R, Welle CL, Venkatesh SK, Schramm C, Eaton JE. Magnetic resonance imaging in primary sclerosing cholangitis—current state and future directions. *Semin Liver Dis* 2019;39(3), 369-380.
10. Mandarano G, Sim J. The diagnostic MRCP examination: Overcoming technical challenges to ensure clinical success. *Biomed Imaging Interv J* 2008;4:16-17.
11. Schramm C, Eaton J, Ringe KI, Venkatesh S, Yamamura J. IPSCSG for the MRI working group of the: Recommendations on the use of magnetic resonance imaging in PSC—A position statement from the International PSC Study Group. *Hepatology* 2017;66:1675-1688.
12. Sugiyama M, Haradome H, Atomi Y. Magnetic resonance imaging for diagnosing chronic pancreatitis. *J Gastroenterol* 2007;42:108-112.
13. Manikkavasakar S, AlObaidy M, Busireddy KK, et al. Magnetic resonance imaging of pancreatitis: An update. *World J Gastroenterol* 2014; 20:14760-14777.

14. Banerjee R, Pavlides M, Tunnicliffe EM, et al. Multiparametric magnetic resonance for the noninvasive diagnosis of liver disease. *J Hepatol* 2014;60:69-77.
15. Pavlides M, Banerjee R, Sellwood J, et al. Multiparametric magnetic resonance imaging predicts clinical outcomes in patients with chronic liver disease. *J Hepatol* 2016;64:308-315.
16. Bittman ME, Callahan MJ. The effective use of acai juice, blueberry juice and pineapple juice as negative contrast agents for magnetic resonance cholangiopancreatography in children. *Pediatr Radiol* 2014;44:883-887.
17. Weickert J. *Anisotropic diffusion in image processing*. Vol 1. Stuttgart: Teubner; 1996. p 59-60.
18. Frangi AF, Niessen WJ, Vincken KL, Viergever M. Multiscale vessel enhancement filtering. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Berlin, Heidelberg: Springer; 1998, p 130-137.
19. Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern* 1996;9:62-66.
20. Hart PE, Nilsson NJ, Raphael B. A formal basis for the heuristic determination of minimum cost paths. *IEEE Trans Syst Sci Cybern* 1968;4:100-107.
21. Xu C, Prince JL. Snakes, shapes, and gradient vector flow. *IEEE Trans Image Process* 1998;7:359-369.
22. Håkansson K, Christoffersson JO, Leander P, Ekberg O, Håkansson HO. On the appearance of bile in clinical MR cholangiopancreatography. *Acta Radiol* 2002;43(4):401-410.
23. Bland JM, Altman DG. Applying the right statistics: Analyses of measurement studies. *Ultrasound Obstet Gynecol* 2003;22:85-93.
24. Raunig DL, McShane LM, Pennello G, et al. Quantitative imaging biomarkers: A review of statistical methods for technical performance assessment. *Stat Methods Med Res* 2014;24:27-67.
25. Solberg HE. Approved recommendation (1986) on the theory of reference values. Part 1. The concept of reference values. *J Clin Chem Clin Biochem* 1987;25:337-342. (*Clin Chim Acta* 1987;165:111-118; *Labmedica* 1987;4:27-31; *Ann Biol Clin* 1987;45:237-241).
26. Sasse EA. Objective evaluation of data in screening for disease. *Clin Chim Acta* 2002;315:17-30.
27. Sasse EA, Dumas BT, Miller WG, D'Orazio P, Eckfeldt JH, Evans SA, Graham GA, Myers GL, Parsons PJ, Stanton NV. How to define and determine reference intervals in the clinical laboratory; approved guideline — second edition serving the World's Medical Science Community through voluntary consensus. Vol 20(13); 2000.
28. Rösch T, Meining A, Frühmorgen S, et al. A prospective comparison of the diagnostic accuracy of ERCP, MRCP, CT, and EUS in biliary strictures. *Gastrointest Endosc* 2002;55:870-876.
29. Kitazono MT, Qayyum A, Yeh BM, Chard PS, Ostroff JW, Coakley FV. Magnetic resonance cholangiography of biliary strictures after liver transplantation: A prospective double-blind study. *J Magn Reson Imaging* 2007;25:1168-1173.
30. Horrow MM. Ultrasound of the extrahepatic bile duct: Issues of size. *Ultrasound Q* 2010;26:67-74.
31. Lal N, Mehra S, Lal V. Ultrasonographic measurement of normal common bile duct diameter and its correlation with age, sex and anthropometry. *J Clin Diagn Res* 2014;8:AC01-AC04.
32. Peng R, Zhang L, Zhang X-M, et al. Common bile duct diameter in an asymptomatic population: A magnetic resonance imaging study. *World J Radiol* 2015;7:501-508.
33. Hung CR, Huang AC, Chen YC, Lii JM, Chen RC. Common bile duct diameter measurement by magnetic resonance cholangiopancreatography. *J Radiol Sci* 2011;36:16-21.
34. Tang Y, Yamashita Y, Arakawa A, et al. Pancreaticobiliary ductal system: Value of half-Fourier rapid acquisition with relaxation enhancement MR cholangiopancreatography for postoperative evaluation. *Radiology* 2000;215:81-88.
35. Polistina FA. Accuracy of magnetic resonance cholangiography compared to operative endoscopy in detecting biliary stones, a single center experience and review of literature. *World J Radiol* 2015;7:70-78.
36. Khayat MF, Al-Amoodi MS, Aldaqal SM, Sibiany A. Abnormal anatomical variations of extrahepatic biliary tract, and their relation to biliary tract injuries and stones formation. *Gastroenterol Res* 2014;7:12-16.
37. Choi JW, Kim TK, Kim KW, et al. Anatomic variation in intrahepatic bile ducts: An analysis of intraoperative cholangiograms in 300 consecutive donors for living donor liver transplantation. *Korean J Radiol* 2003;4:85-90.