

# UC San Diego

## UC San Diego Previously Published Works

### Title

Repeatability and accuracy of various region-of-interest sampling strategies for hepatic MRI proton density fat fraction quantification

### Permalink

<https://escholarship.org/uc/item/2jd8v03n>

### Journal

Abdominal Radiology, 46(7)

### ISSN

2366-004X

### Authors

Hong, Cheng William  
Cui, Jennifer Y  
Batakis, Danielle  
[et al.](#)

### Publication Date

2021-07-01

### DOI

10.1007/s00261-021-02965-5

Peer reviewed



Published in final edited form as:

*Abdom Radiol (NY)*. 2021 July ; 46(7): 3105–3116. doi:10.1007/s00261-021-02965-5.

## Repeatability and accuracy of various region-of-interest sampling strategies for hepatic MRI proton density fat fraction quantification

Cheng William Hong, MD MS<sup>1</sup>, Jennifer Y Cui, BS<sup>1,2</sup>, Danielle Batakis, BS<sup>1</sup>, Yang Xu, BS<sup>1</sup>, Tanya Wolfson, MA<sup>3</sup>, Anthony C Gamst, PhD<sup>3</sup>, Alexandra N Schlein, BS<sup>1</sup>, Lindsey M Negrete, MD<sup>1</sup>, Michael S Middleton, MD PhD<sup>1</sup>, Gavin Hamilton, PhD<sup>1</sup>, Rohit Loomba, MD MHSc<sup>4</sup>, Jeffrey B Schwimmer, MD<sup>5,6</sup>, Kathryn J Fowler, MD<sup>1</sup>, Claude B Sirlin, MD<sup>1</sup>

<sup>1</sup>Department of Radiology, University of California San Diego, San Diego, CA, USA

<sup>2</sup>School of Medicine, Wayne State University, Detroit, MI, USA

<sup>3</sup>Computational and Applied Statistics Laboratory, University of California San Diego, San Diego, CA, USA

<sup>4</sup>Division of Gastroenterology, Department of Medicine, University of California at San Diego

<sup>5</sup>Department of Pediatrics, School of Medicine, University of California San Diego, La Jolla, CA

<sup>6</sup>Department of Gastroenterology, Rady Children's Hospital San Diego, San Diego, CA, USA

### Abstract

**Purpose:** To evaluate repeatability of ROI sampling strategies for quantifying hepatic proton density fat fraction (PDFF) and to assess error relative to the 9-ROI PDFF.

**Methods:** This was a secondary analysis in subjects with known or suspected nonalcoholic fatty liver disease who underwent MRI for magnitude-based hepatic PDFF quantification. Each subject underwent three exams, each including three acquisitions (nine acquisitions total). An ROI was placed in each hepatic segment on the first acquisition of the first exam and propagated to other acquisitions. PDFF was calculated for each of 511 sampling strategies using every combination of 1, 2, ..., all 9 ROIs. Intra- and inter-exam intraclass correlation coefficients (ICCs) and repeatability coefficients (RCs) were estimated for each sampling strategy. Mean absolute error (MAE) was estimated relative to the 9-ROI PDFF. Strategies that sampled both lobes evenly ("balanced") were compared with those that did not ("unbalanced") using two-sample t-tests.

**Results:** The 29 enrolled subjects (23 male, mean age 24 years) had mean 9-ROI PDFF 11.8% (1.1–36.3%). With more ROIs, ICCs increased, RCs decreased, and MAE decreased. Of the 60 balanced strategies with 4 ROIs, all (100%) achieved inter- and intra-exam ICCs > 0.998, 55 (92%) achieved intra-exam RC < 1%, 50 (83%) achieved inter-exam RC < 1%, and all (100%) achieved MAE < 1%. Balanced sampling strategies had higher ICCs and lower RCs, and lower MAEs than

During the submission, please address correspondence to: Cheng William Hong, MD MS, 200 W. Arbor Drive #8756, San Diego, CA 92103-8756, USA, Phone: 858-246-2196, Fax: 619-471-0503, cwhong@ucsd.edu. If published, please address correspondence to: Claude B Sirlin, MD, 9500 Gilman Drive #0888, La Jolla, CA 92093-0888, USA, Phone: 858-246-2196, Fax 858-246-2221, csirlin@ucsd.edu.

unbalanced strategies in aggregate ( $p < 0.001$  for comparisons between balanced vs. unbalanced strategies).

**Conclusion:** Repeatability improves and error diminishes with more ROIs. Balanced 4-ROI strategies provide high repeatability and low error.

### Keywords

hepatic PDFF; repeatability; region-of-interest; sampling strategy; hepatic fat quantification; quantitative imaging biomarker; QIB

---

## Introduction

Proton density fat fraction (PDFF), based on chemical-shift-encoded magnetic resonance imaging (CSE-MRI), is a validated non-invasive quantitative imaging biomarker for hepatic fat content [1–8]. It has demonstrated excellent repeatability and reproducibility and is also accurate for quantifying fat content, using either histology or MR spectroscopy as reference standards [9–11]. To compute PDFF, the source CSE-MRI multi-echo data are reconstructed to generate parametric PDFF maps, which display the spatial distribution of PDFF throughout the imaged volume (e.g., the abdomen).

A composite hepatic PDFF value is derived from the PDFF map by averaging measurements made in the liver itself. As automated whole liver segmentation is not widely available at this time, this usually requires manual placement of regions of interest (ROIs) in the liver. Since hepatic fat content is spatially heterogeneous and since the right lobe usually has greater fat content than the left [12], multiple ROIs must be placed in a representative manner to sample the entire liver and derive a meaningful composite PDFF value [12–15]. The preferred strategy in prior publications and clinical trials is to place one ROI in each of the nine Couinaud segments (Figure 1) [16–21]. While this approach is rigorous, ensures representative sampling of the entire liver, and is appropriate for clinical trials, it is time-consuming, laborious, and impractical for routine clinical care [22].

Previous studies suggest that a less laborious sampling strategy of four ROIs with two ROIs in each hepatic lobe may provide adequately high accuracy relative to the conventional 9-ROI PDFF estimate [15,23]. However, showing high accuracy is not enough; it is also necessary to show high repeatability. Since the effect of placing fewer than nine ROIs on PDFF repeatability has not been rigorously studied, a strategy using fewer than 9 ROIs cannot be recommended for routine clinical care until this effect is understood.

Therefore, the primary purpose of this study was to evaluate intra- and inter-exam repeatability of hepatic PDFF quantification for different sampling strategies using various numbers of ROIs for hepatic PDFF quantification. A secondary purpose was to evaluate accuracy of those sampling strategies relative to the 9-ROI reference. We anticipated that sampling strategies with fewer than nine ROIs can achieve adequately high repeatability and accuracy and that sampling strategies that sample the right and left hepatic lobes evenly would have higher repeatability and accuracy than those that did not. In particular, we

focused on evaluating balanced 4-ROI strategies, based on their promising results in prior preliminary studies [15,23].

## Materials and Methods

### Study Design

Approved by the Institutional Review Board and compliant with the Health Insurance Portability and Accountability Act, this study was a secondary analysis of prospectively collected single-site data in research volunteers with obesity and known or suspected NAFLD who underwent confounder-corrected chemical-shift-encoded 3T MRI for magnitude-based MRI (MRI-M) quantification of PDFF. For the primary study, adult subjects provided written informed consent, and pediatric subjects provided written assent with parental consent. Demographic and anthropometric information was collected.

From August 2009 to October 2009, pediatric and adult subjects were recruited from hepatology and obesity clinics and through self-referral [24]. Eligibility criteria included known or suspected NAFLD, body mass index  $> 30 \text{ kg/m}^2$ , age  $> 8$  years, and willingness to undergo a research MRI. Exclusion criteria included contraindications to MRI, claustrophobia, and pregnancy.

Subjects underwent three same-day exams, where each exam comprised three MRI-M acquisitions (nine acquisitions total).

### MRI Acquisition

Subjects were scanned supine using a 3T Signa EXCITE HDxt MRI system (GE Healthcare, Waukesha, WI) with an eight-channel torso phased-array coil centered over the liver. A dielectric pad was placed between the coil and the abdomen.

Three exams were performed for each subject, each of which included a localizing sequence followed by three MRI-M acquisitions (nine MRI-M acquisitions total). Parameters for each MRI-M acquisition are summarized in Table 1. Briefly, an axial two-dimensional multi-echo spoiled gradient-recalled echo (SPGR) sequence was acquired with full liver coverage in one or two 18- to 30-s breath-holds. A low flip angle ( $10^\circ$ ) with 125-ms repetition time (TR) was used to minimize T1 bias [25–28]. Six echoes were obtained per TR at nominally out-of-phase and in-phase echo times to perform fat-water separation while accounting for T2\* signal decay and estimation of T2\* assuming monoexponential signal loss. Between exams, the subjects were removed from the scanner table for about five to ten minutes, the phase-array coil was re-attached, and the subjects were repositioned on the scanner table.

A previously described fitting algorithm was applied to the acquired six-echo source images pixel-by-pixel to create parametric PDFF maps [28], which corrected for T2\* signal decay and applied a multi-peak spectral model to account for the spectral complexity of fat [29]. Due to the fat-water ambiguity intrinsic to magnitude fitting, water was assumed to be the dominant signal in the liver. Corrections for phase errors were not performed as they are unnecessary for magnitude-based methods [30,31].

## MRI Post-processing and Analysis

Source images and the PDFFF maps were analyzed using OsiriX software (OsiriX Foundation, Geneva, Switzerland) by one of two trained research analysts (initials redacted during submission). For each subject, an analyst placed a 1-cm radius ROI in each of the nine hepatic segments while avoiding major vessels, bile ducts, liver edges, other organs, and artifacts on the first acquisition of the first exam. The fifth-echo source series was used to place ROIs as it consistently provided adequate anatomic delineation for this purpose. Co-localized ROIs were then propagated to the other eight acquisitions, manually adjusting ROI placement to ensure co-localization based on anatomical landmarks. The ROIs were then propagated onto the PDFFF maps, and the PDFFF values for each of the nine hepatic segments for each acquisition were recorded.

## ROI Sampling Strategies

For each acquisition of each subject, PDFFF was estimated 511 times by averaging the segmental PDFFFs using every combinatorial subset of up to 9 hepatic segments: all 9 combinations of 1 ROI, all 36 combinations of 2 ROIs, all 84 combinations of 3 ROIs, ..., and the single combination of 9 ROIs. The caudate lobe was considered to be part of the left hepatic lobe. Sampling strategies were classified as “balanced” or “unbalanced” as follows:

- Balanced = sampled both lobes and the number of ROIs sampled in the right and left hepatic lobes did not differ by more than one.
- Unbalanced = sampled both lobes and the number of ROIs sampled in the right and left hepatic lobes differed by two or more.

1-ROI sampling strategies are a special case and were not classified as either balanced or unbalanced.

## Statistical Analysis

Statistical analysis was performed using R version 3.5.1 statistical software (R: A language and environment for statistical computing. 2018. R Foundation for Statistical Computing, Vienna, Austria). Demographic, anthropometric, and PDFFF data were summarized descriptively. Repeatability was evaluated using intra-class correlation coefficients (ICCs) and repeatability coefficients (RCs). Pearson’s correlation was calculated between the absolute difference and the average of repeats, pairwise, for each sampling strategy, to inform our choice of RC, as several options for RC are available depending on the data distribution. Repeat measurements in all sampling strategies were examined for independence of mean and variance.

The intra-exam repeatability of each sampling strategy was computed by analyzing PDFFF values from the three acquisitions from the first exam. The inter-exam repeatability of each combinatorial strategy was computed by analyzing PDFFF values from the first acquisition of each of the three exams (Figure 2). For each of the above two analyses, subjects were excluded if any of the nine segmental ROIs could not be placed reliably on any of the relevant acquisitions (e.g., due to imaging artifacts or insufficient segmental volume for ROI

placement). Excluded vs. included subjects were compared using Wilcoxon-Mann-Whitney and Fisher's Exact tests, as appropriate.

We selected thresholds of ICC > 0.998 and RC < 1% *a priori* as desired repeatability benchmarks, because we aimed to identify sampling strategies that would be reliable in detecting even small (~1%) changes in PDFF if applied longitudinally. ICCs and RCs were summarized, both overall and between balanced and unbalanced strategies. ICCs and RCs were compared between balanced and unbalanced strategies using two-sample t-tests.

To assess accuracy, we used all acquisitions from all exams in which all nine segmental ROIs were measured. Acquisitions in which any segmental ROIs were missing were excluded. Figure 3 illustrates the data selection for this analysis. The average of PDFF in 9 segmental ROIs was considered a reference standard. The mean absolute error (MAE), which is the mean of absolute differences between each sampling strategy and the 9-ROI reference, as well as the Bland-Altman bias for each sampling strategy relative to the 9-ROI reference, were computed and summarized overall and between balanced and unbalanced strategies. MAEs were compared between balanced and unbalanced strategies using a two-sample t-test.

## Results

### Study Population

The cohort comprised 29 subjects (23 male) with mean age of 24 years (range: 12 – 59 years), mean PDFF of 11.8% (range: 1.1 – 36.3%), and mean weight of 104 kg (range: 66 – 128 kg).

### Intra-exam and Inter-exam Repeatability Analyses

Three subjects were excluded from the intra-exam analysis due to incomplete data. The excluded group did not differ significantly in age (mean of 22 years vs. 24 years,  $p = 0.81$ ), but weighed less (mean of 82.5 kg vs. 106.8 kg,  $p = 0.024$ ) and, at a trend level had lower PDFF (mean of 5.5% vs. 12.5%,  $p = 0.13$ ) and higher proportion of females (2/3 vs. 26/29,  $p = 0.10$ ).

Eight subjects were excluded from the inter-exam analysis due to incomplete data. The excluded group did not differ significantly in age (mean of 21.5 years vs. 24.9 years,  $p = 0.756$ ) or PDFF (9.3% vs. 12.7%,  $p = 0.35$ ), but had a higher proportion of females (4/8 vs. 2/21,  $p = 0.033$ ) and, at a trend level, weighed less (94.5 kg vs. 108.0 kg,  $p = 0.064$ ).

Since none of the Pearson's correlations between the absolute difference and the average of repeats were significant after piece-wise Bonferroni correction (with most not significant before the correction), i.e. no significant relationship between PDFF mean and variability was observed, we used RC as defined previously [32], which was extended to three repeats from the standard test-retest two-repeat setting using the results of Neyman and Scott [33].

Intra-exam and inter-exam ICCs increased and RCs decreased as the number of ROIs increased (Tables 2 and 3, Figure 4). In aggregate, balanced strategies were superior

compared to unbalanced strategies (Table 3) based on all repeatability metrics: intra-exam ICC (0.9986 vs. 0.9983,  $p < 0.001$ ), inter-exam ICC (0.9987 vs. 0.9985,  $p < 0.001$ ), intra-exam RC (0.887% vs. 0.974%,  $p < 0.001$ ), and inter-exam RC (0.904% vs. 0.956%,  $p < 0.001$ ).

For balanced 4-ROI strategies, the mean ICCs were 0.9987 (range, 0.9981 to 0.9991 for intra-exam and 0.9987 (range, 0.9981 to 0.9993) for inter-exam repeatability. Thus, all 60 balanced 4-ROI strategies achieved ICCs  $> 0.998$  for both intra- and inter-exam repeatability. The mean RCs of these strategies were 0.89% (range, 0.73% to 1.04%) for intra-exam and 0.91% (range, 0.69% to 1.12%) for inter-exam repeatability. Overall, 55/60 (92%) and 50/60 (83%) balanced 4-ROI strategies achieved intra-exam and inter-exam RCs of  $< 1\%$ , respectively.

By comparison, only 49/70 (70%) and 54/70 (77%) of balanced 3-ROI strategies achieved the *a priori* benchmark for intra- and inter-exam ICC and only 35/70 (50%) and 37/70 (53%) of balanced 3-ROI strategies achieved the *a priori* benchmark for intra- and inter-exam RC, respectively (Table 4).

### Accuracy Relative to the 9-ROI Reference Sampling Strategy

Two hundred and twenty-eight acquisitions were included in this analysis. Nineteen of the 29 study subjects had all 9 valid acquisitions, 4 subjects had 8 acquisitions, 1 subject had 7 acquisitions, 2 subjects had 6 acquisitions, and 2 subjects had 3 acquisitions. One subject had no valid acquisitions and was excluded. MAE decreased as the number of ROIs increased. (Table 5, Figure 5). In aggregate, balanced strategies had lower MAE than unbalanced strategies (0.418 vs. 0.618,  $p < 0.001$ ), and for the same number of ROIs, the balanced strategies informally had lower MAE compared to the unbalanced strategies (Table 5, Figure 5). Balanced 4-ROI strategies had MAE of  $0.41 \pm 0.29\%$  (range, 0.21% to 0.71%); thus, all 60 balanced 4-ROI strategies had MAE  $< 1.0\%$ .

No directional patterns of Bland-Altman bias were observed among the sampling strategies (Figure 6). Magnitude of bias decreased as the number of ROIs increased (Table 5, Figure 6). Balanced 4-ROI strategies had a mean bias of  $+0.08\%$  (range,  $-0.61\%$  to  $+0.68\%$ ); thus, all 60 balanced 4-ROI strategies had an absolute bias of  $< 1.0\%$ .

## Discussion

Our results are consistent with the expectation that both the repeatability and accuracy of ROI-sampling strategies improve as more ROIs are used. In addition, for any given number of ROIs, balanced approaches that sample a similar number of ROIs in each hepatic lobe tend to have higher repeatability and accuracy than those that do not. We focused on two types of repeatability, intra-exam (without subject repositioning) and inter-exam (with repositioning), and on two metrics of repeatability, ICC and RC. For the latter, we determined that the appropriate formula for RC assumes that variance and mean are independent and extended the formula to three repeated measures. For accuracy relative to the 9-ROI PDF, we focused on MAE and on Bland-Altman bias.

Prior studies have shown that in order to maximize repeatability and accuracy, as much liver area should be sampled as possible [15,34]. Hooker et al previously demonstrated that 9-ROI measurements have only slightly lower agreement than 27-ROI measurements [35]. Our study systematically evaluates the trade-off between the number of ROIs vs. repeatability and accuracy, and we found that fewer than 9 ROIs suffice. In particular, for any given number of total ROIs, balanced strategies sampling the same number of ROIs in each hepatic lobe consistently provide higher intra-exam and inter-exam repeatability and accuracy than unbalanced strategies, regardless of the precision and accuracy metric. All balanced 6-ROI sampling strategies met our criteria of ICC > 0.998 and RC < 1%, however, the improvement in precision and accuracy over 4 ROIs is incremental and may not justify the additional burden in routine applications. With caution, we suggest that the placement of 4 ROIs, with two in each lobe, suffices for clinical care in most contexts.

Hong et al previously demonstrated that 4-ROI sampling strategies, provided they sampled two ROIs in each hepatic lobe, could have limits of agreement < 1.5 absolute percentage points relative to 9-ROI PDFF [23]. Similarly, Vu et al found that relative to MR spectroscopy, PDFF estimation is more accurate when both hepatic lobes are sampled [15].

Based on our findings above, balanced 4-ROI strategies may provide a good tradeoff between laboriousness and maintaining repeatability and accuracy relative to an established reference standard. Nevertheless, the optimal choice of the number of ROIs depends on the context of use. We recognize that higher performance might be desired for certain applications, and depending on the exact performance benchmarks, users might desire a higher number of ROIs. The data summarized in Figures 4, 5, and 6 can help inform decisions about the minimum number of ROIs that might be needed to achieve various performance benchmarks.

Limitations of our study include the relatively small sample size, which in conjunction with the large number of ROI combinations, precludes formal statistical comparisons between individual sampling strategies (although comparisons between balanced vs. unbalanced strategies in aggregate were possible, and performed). This means that we are unable to recommend specific combinations of liver segments for ROI placement. Nevertheless, since agreement for all balanced 4-ROI strategies was high, the exact choice of ROIs can be made based on other considerations such as avoiding imaging artifacts, hepatic vasculature, and bile ducts. Some subjects were excluded due to incomplete data and there were some differences in characteristics between the excluded and included subjects. Co-localization of ROIs across acquisitions may overestimate PDFF repeatability relative to imaging protocols where ROIs are not similarly co-localized across longitudinal exams. We used 9-ROI PDFF as our reference standard, as that is the most commonly used method in clinical trials. Whole-liver segmentation for PDFF estimation, potentially using machine learning techniques, might be the most accurate approach but is still not widely available for research or clinical application at this time [36]. The single-center nature of our study in subjects with known or suspected NAFLD may limit generalization to populations in other geographical regions and subjects without NAFLD. Our study focused on repeatability, and further research is needed to assess reproducibility of PDFF estimated with fewer ROIs across vendor platforms and field strengths. We also focused on a magnitude-based PDFF



estimation technique, as that is still the most common MRI technique applied used for PDFF estimation in clinical trials; further research will be needed to validate our findings in complex-based CSE techniques, which are now commercially available on modern MRI systems.

In conclusion, while repeatability and agreement with the 9-ROI reference standard improve with increasing number of ROIs, for any given number of ROIs, balanced ROI sampling strategies which sample the right and left hepatic lobes more evenly have better repeatability and agreement with the 9-ROI reference standard than those which do not. Balanced 4-ROI sampling strategies may provide a reasonable compromise between laboriousness and performance for routine clinical care and potentially even in some clinical trials. Researchers and clinicians may select a different PDFF sampling strategy, using the results of our study, to achieve an appropriate tradeoff between laboriousness and performance required for a particular context of use.

### Grant support:

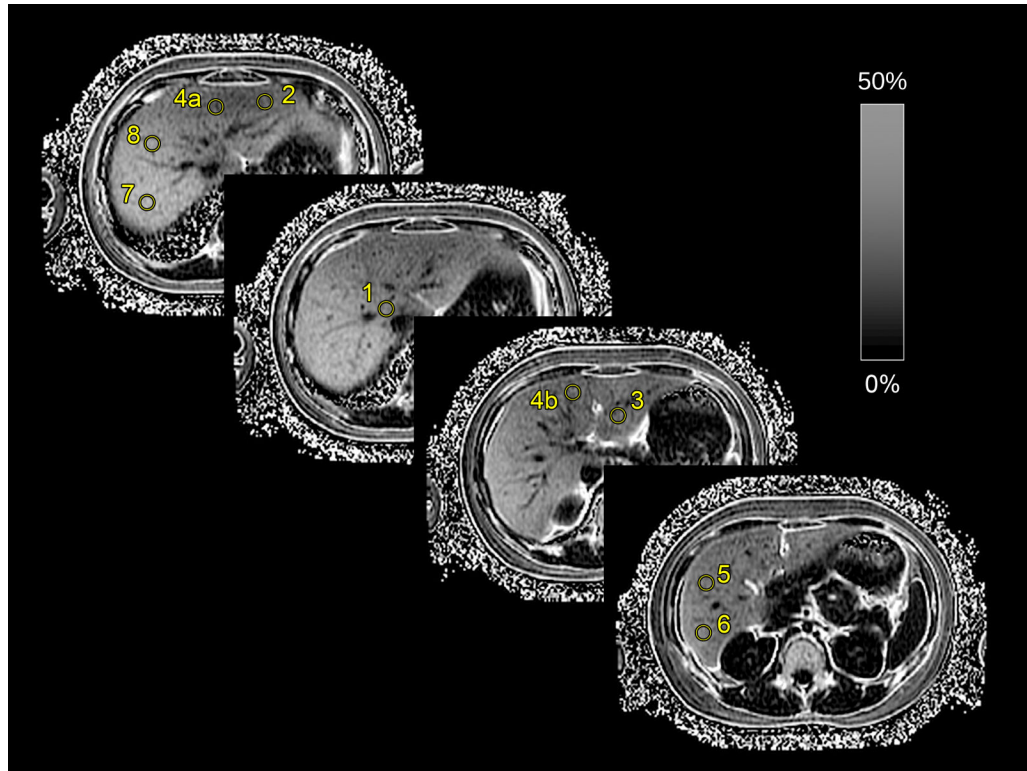
The authors would like to acknowledge grant support from the National Institutes of Health T32 EB005970-09 and R01 DK106419-02. We also acknowledge Gilead Sciences and GE Healthcare which provide research support to UCSD.

### References

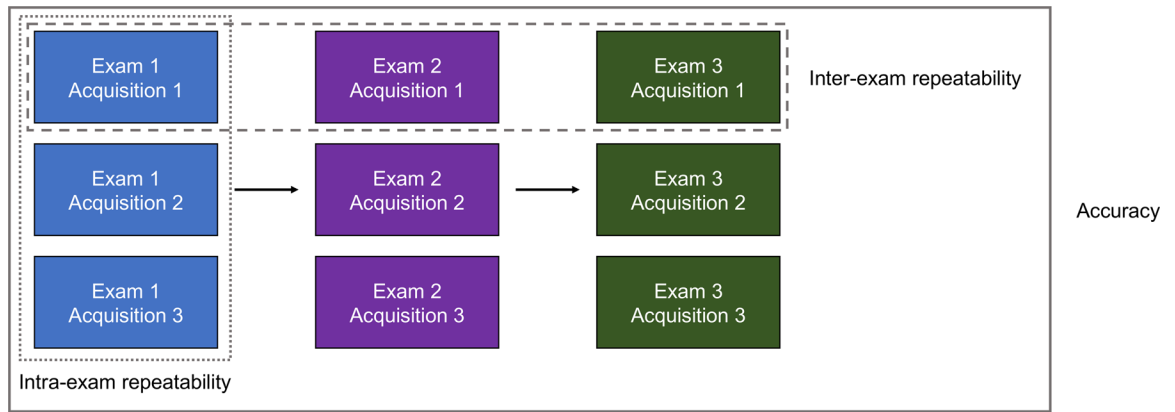
1. Reeder SB, McKenzie CA, Pineda AR et al. (2007) Water-fat separation with IDEAL gradient-echo imaging. *J Magn Reson Imaging* 25:644–52. 10.1002/jmri.20831 [PubMed: 17326087]
2. Reeder SB and Sirlin CB (2010) Quantification of liver fat with magnetic resonance imaging. *Magn Reson Imaging Clin N Am* 18:337–57, ix. 10.1016/j.mric.2010.08.013 [PubMed: 21094444]
3. Reeder SB, Cruite I, Hamilton G and Sirlin CB (2011) Quantitative assessment of liver fat with magnetic resonance imaging and spectroscopy. *J Magn Reson Imaging* 34:729–49. 10.1002/jmri.22580 [PubMed: 22025886]
4. Reeder SB, Hu HH and Sirlin CB (2012) Proton density fat-fraction: a standardized MR-based biomarker of tissue fat concentration. *J Magn Reson Imaging* 36:1011–4. 10.1002/jmri.23741 [PubMed: 22777847]
5. Reeder SB, Pineda AR, Wen Z et al. (2005) Iterative decomposition of water and fat with echo asymmetry and least-squares estimation (IDEAL): Application with fast spin-echo imaging. *Magn Reson Med* 54:636–44. 10.1002/mrm.20624 [PubMed: 16092103]
6. Hong CW, Fazeli Dehkordy S, Hooker JC, Hamilton G and Sirlin CB (2017) Fat Quantification in the Abdomen. *Top Magn Reson Imaging* 26:221–7. 10.1097/RMR.0000000000000141 [PubMed: 29176468]
7. Yokoo T and Browning JD (2014) Fat and Iron Quantification in the Liver. *Top Magn Reson Imaging* 23:129–50. 10.1097/RMR.0000000000000016 [PubMed: 24690617]
8. Cassidy FH, Yokoo T, Aganovic L et al. (2009) Fatty Liver Disease: MR Imaging Techniques for the Detection and Quantification of Liver Steatosis. *RadioGraphics Radiological Society of North America*. 29:231–60. 10.1148/rg.291075123
9. dilman S, Halilo lu M, Gümrük F and Karçaaltuncaba M (2016) The Feasibility of Magnetic Resonance Imaging for Quantification of Liver, Pancreas, Spleen, Vertebral Bone Marrow, and Renal Cortex R2\* and Proton Density Fat Fraction in Transfusion-Related Iron Overload. *Turkish J Hematol* 33:21–7. 10.4274/tjh.2015.0142
10. Tang A, Desai A, Hamilton G et al. (2015) Accuracy of MR imaging-estimated proton density fat fraction for classification of dichotomized histologic steatosis grades in nonalcoholic fatty liver disease. *Radiology* 274:416–25. 10.1148/radiol.14140754 [PubMed: 25247408]

11. Tyagi A, Yeganeh O, Levin Y et al. (2015) Intra- and inter-examination repeatability of magnetic resonance spectroscopy, magnitude-based MRI, and complex-based MRI for estimation of hepatic proton density fat fraction in overweight and obese children and adults. *Abdom Imaging* 40:3070–7. 10.1007/s00261-015-0542-5 [PubMed: 26350282]
12. Bonekamp S, Tang A, Mashhood A et al. (2014) Spatial distribution of MRI-Determined hepatic proton density fat fraction in adults with nonalcoholic fatty liver disease. *J Magn Reson Imaging* 39:1525–32. [PubMed: 24987758]
13. Larson SP, Bowers SP, Palekar NA et al. (2007) Histopathologic variability between the right and left lobes of the liver in morbidly obese patients undergoing Roux-en-Y bypass. *Clin Gastroenterol Hepatol* 5:1329–32. 10.1016/j.cgh.2007.06.005 [PubMed: 17702661]
14. Merriman RB, Ferrell LD, Patti MG et al. (2006) Correlation of paired liver biopsies in morbidly obese patients with suspected nonalcoholic fatty liver disease. *Hepatology* 44:874–80. 10.1002/hep.21346 [PubMed: 17006934]
15. Vu K-N, Gilbert G, Chalut M et al. (2016) MRI-determined liver proton density fat fraction, with MRS validation: Comparison of regions of interest sampling methods in patients with type 2 diabetes. *J Magn Reson Imaging* 43:1090–9. 10.1002/jmri.25083 [PubMed: 26536609]
16. Cui J, Philo L, Nguyen P et al. (2016) Sitagliptin vs. placebo for non-alcoholic fatty liver disease: A randomized controlled trial. *J Hepatol* 65:369–76. 10.1016/j.jhep.2016.04.021 [PubMed: 27151177]
17. Middleton MS, Heba ER, Hooker CA et al. (2017) Agreement Between Magnetic Resonance Imaging Proton Density Fat Fraction Measurements and Pathologist-Assigned Steatosis Grades of Liver Biopsies From Adults With Nonalcoholic Steatohepatitis. *Gastroenterology* W.B. Saunders. 153:753–61. 10.1053/j.gastro.2017.06.005
18. Loomba R, Sirlin CB, Ang B et al. (2015) Ezetimibe for the treatment of nonalcoholic steatohepatitis: assessment by novel magnetic resonance imaging and magnetic resonance elastography in a randomized trial (MOZART trial). *Hepatology* 61:1239–50. 10.1002/hep.27647 [PubMed: 25482832]
19. Fazeli Dehkordy S, Fowler KJ, Mamidipalli A et al. (2018) Hepatic steatosis and reduction in steatosis following bariatric weight loss surgery differs between segments and lobes. *Eur Radiol*. 10.1007/s00330-018-5894-0
20. Le T-A, Chen J, Changchien C et al. (2012) Effect of colesvelam on liver fat quantified by magnetic resonance in nonalcoholic steatohepatitis: a randomized controlled trial. *Hepatology* 56:922–32. 10.1002/hep.25731 [PubMed: 22431131]
21. Permutt Z, Le TA, Peterson MR et al. (2012) Correlation between liver histology and novel magnetic resonance imaging in adult patients with non-alcoholic fatty liver disease - MRI accurately quantifies hepatic steatosis in NAFLD. *Aliment Pharmacol Ther*. 10.1111/j.1365-2036.2012.05121.x
22. Tang A, Tan J, Sun M et al. (2013) Nonalcoholic fatty liver disease: MR imaging of liver proton density fat fraction to assess hepatic steatosis. *Radiology* 267:422–31. 10.1148/radiol.12120896 [PubMed: 23382291]
23. Hong CW, Wolfson T, Sy EZ et al. (2018) Optimization of region-of-interest sampling strategies for hepatic MRI proton density fat fraction quantification. *J Magn Reson Imaging* 47:988–94. 10.1002/jmri.25843 [PubMed: 28842937]
24. Negrete LM, Middleton MS, Clark L et al. (2014) Inter-examination precision of magnitude-based MRI for estimation of segmental hepatic proton density fat fraction in obese subjects. *J Magn Reson Imaging* 39:1265–71. 10.1002/jmri.24284 [PubMed: 24136736]
25. Yokoo T, Bydder M, Hamilton G et al. (2009) Nonalcoholic fatty liver disease: diagnostic and fat-grading accuracy of low-flip-angle multiecho gradient-recalled-echo MR imaging at 1.5 T. *Radiology* 251:67–76. 10.1148/radiol.2511080666 [PubMed: 19221054]
26. Yokoo T, Shieh-morteza M, Hamilton G et al. (2011) Estimation of hepatic proton-density fat fraction by using MR imaging at 3.0 T. *Radiology* 258:749–59. 10.1148/radiol.10100659 [PubMed: 21212366]

27. Kühn J-P, Hernando D, Mensel B et al. (2014) Quantitative chemical shift-encoded MRI is an accurate method to quantify hepatic steatosis. *J Magn Reson Imaging* 39:1494–501. 10.1002/jmri.24289 [PubMed: 24123655]
28. Bydder M, Yokoo T, Hamilton G et al. (2008) Relaxation effects in the quantification of fat using gradient echo imaging. *Magn Reson Imaging* 26:347–59. 10.1016/j.mri.2007.08.012 [PubMed: 18093781]
29. Hamilton G, Schlein AN, Middleton MS et al. (2016) In vivo triglyceride composition of abdominal adipose tissue measured by (1) H MRS at 3T. *J Magn Reson Imaging*. 10.1002/jmri.25453
30. Hernando D, Hines CDG, Yu H and Reeder SB (2012) Addressing phase errors in fat-water imaging using a mixed magnitude/complex fitting method. *Magn Reson Med NIH Public Access*. 67:638–44. 10.1002/mrm.23044
31. Yu H, Shimakawa A, Hines CDG et al. (2011) Combination of complex-based and magnitude-based multiecho water-fat separation for accurate quantification of fat-fraction. *Magn Reson Med* 66:199–206. 10.1002/mrm.22840 [PubMed: 21695724]
32. Raunig DL, McShane LM, Pennello G et al. (2015) Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. *Stat Methods Med Res* 24:27–67. 10.1177/0962280214537344 [PubMed: 24919831]
33. Neyman J and Scott EL (1948) Consistent Estimates Based on Partially Consistent Observations. *Econometrica JSTOR*. 16:1. 10.2307/1914288
34. Campo CA, Hernando D, Schubert T et al. (2017) Standardized Approach for ROI-Based Measurements of Proton Density Fat Fraction and R2\* in the Liver. *Am J Roentgenol* 209:592–603. 10.2214/AJR.17.17812 [PubMed: 28705058]
35. Hooker JC, Hamilton G, Park CC et al. (2018) Inter-reader agreement of magnetic resonance imaging proton density fat fraction and its longitudinal change in a clinical trial of adults with nonalcoholic steatohepatitis. *Abdom Radiol* 10.1007/s00261-018-1745-3
36. Wang K, Mamidipalli A, Retson T et al. (2019) Automated CT and MRI Liver Segmentation and Biometry Using a Generalized Convolutional Neural Network. *Radiol Artif Intell Radiological Society of North America (RSNA)*. 1:180022. 10.1148/ryai.2019180022



**Figure 1:** Sampling strategy with 9 ROIs (yellow circles). An ROI is propagated onto each of the 9 hepatic segments on multiple slices through the liver on the PDFF map. Scale bar denotes a PDFF dynamic range of 0 – 50% for magnitude-based MRI due to fat-water ambiguity and the assumption that water is the dominant signal.



**Figure 2:** Schematic of the imaging protocol. Subjects were scanned in three separate exams, each of which had three acquisitions (nine acquisitions total). Intra-exam repeatability was computed from the three acquisitions of the first exam (dotted rectangle). Inter-exam repeatability was computed from the first acquisition of each exam (dashed rectangle). Accuracy was computed using all acquisitions (solid rectangle).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

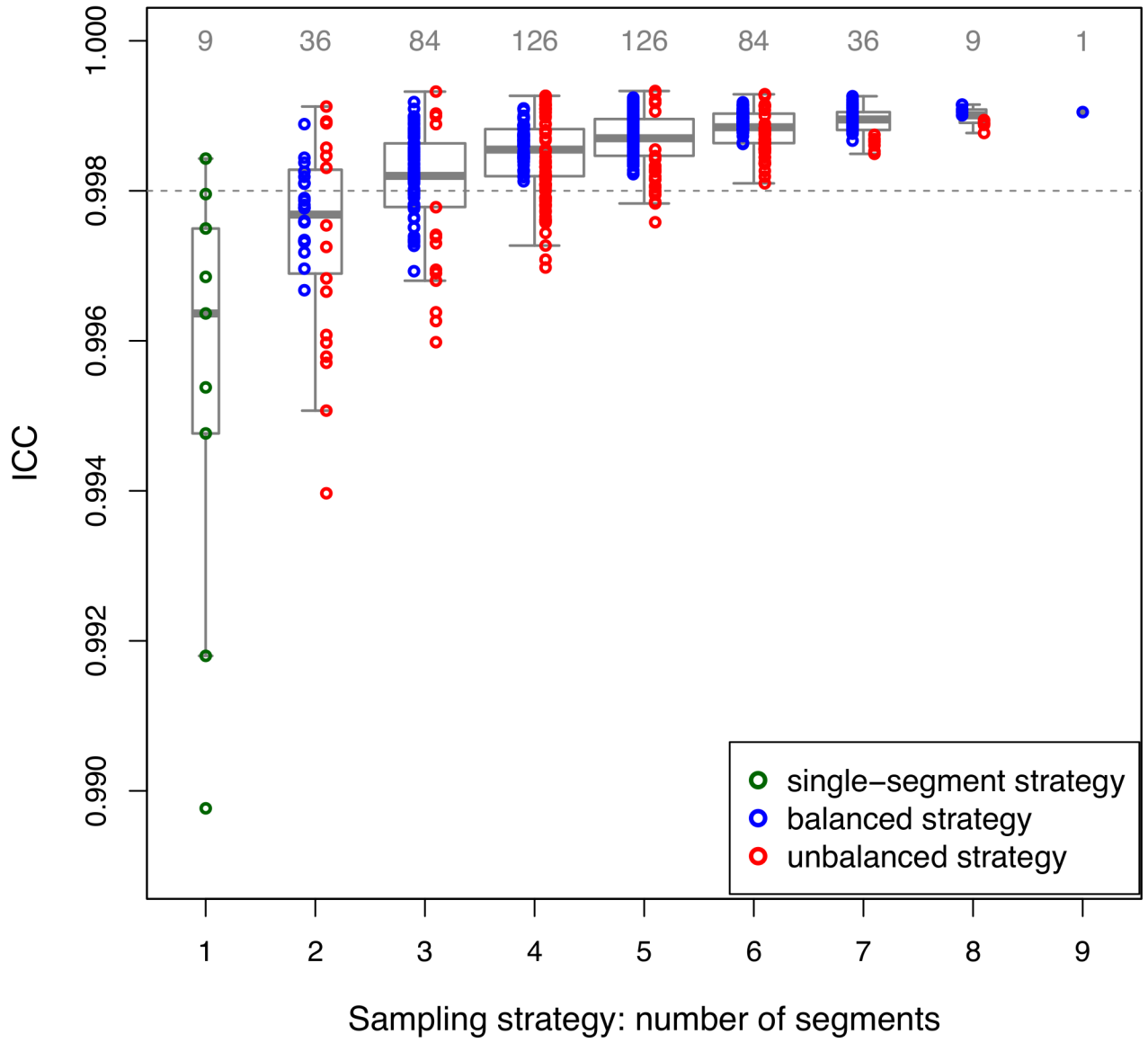
|                   | Strategy                         | Reference                      |
|-------------------|----------------------------------|--------------------------------|
| Exam 1; Repeat 1: | 9.9                              | 9.8                            |
| Exam 1; Repeat 2: | 10.3                             | 9.9                            |
| Exam 1; Repeat 3: | 9.9                              | 9.6                            |
| ←————→            |                                  |                                |
| Exam 2; Repeat 1: | 10.4                             | 9.9                            |
| Exam 2; Repeat 2: | 10.5                             | 10.2                           |
| Exam 2; Repeat 3: | 10.4                             | 9.9                            |
| ←————→            |                                  |                                |
| Exam 3; Repeat 1: | 10.6                             | 10.3                           |
| Exam 3; Repeat 2: | 10.3                             | 9.8                            |
| Exam 3; Repeat 3: | 9.8                              | 9.7                            |
| ←————→            |                                  |                                |
|                   | Segments 3, 5, 8<br>Average PDFF | All 9 Segments<br>Average PDFF |

|                   | Strategy                         |   | Reference                      |
|-------------------|----------------------------------|---|--------------------------------|
| Exam 1; Repeat 1: | 10.2                             | ↔ | 10                             |
| Exam 1; Repeat 2: | 9.7                              | ↔ | 9.8                            |
| Exam 1; Repeat 3: | 10.1                             | ↔ | 10.1                           |
| Exam 2; Repeat 1: | 9.8                              | ↔ | 9.9                            |
| Exam 2; Repeat 2: | 9.5                              | ↔ | 9.9                            |
| Exam 2; Repeat 3: | 8.9                              | X | missing                        |
| Exam 3; Repeat 1: | 9.5                              | ↔ | 9.7                            |
| Exam 3; Repeat 2: | missing                          | X | missing                        |
| Exam 3; Repeat 3: | missing                          | X | missing                        |
|                   | Segments 3, 5, 8<br>Average PDFF |   | All 9 Segments<br>Average PDFF |

**Figure 3:**

Illustration of the selection of paired data for the accuracy analysis, using the data from two study subjects and a sampling strategy which combines segments 3, 5, and 8. The subject in panel A has all 9 valid acquisitions. All nine strategy-reference data pairs are included in the computation of bias and MAE. The subject in panel B has missing data on multiple segments for acquisitions 2 and 3 of exam 3. Those acquisitions were excluded. Additionally, although this subject has valid measurements for segments 3, 5, and 8 on exam 2 acquisition 3, there is missing value for segment 7 and the reference average is incomplete. That acquisition is also excluded. This subject contributes 6 data pairs to the computation of bias and MAE.

### Intra-Exam ICC



Author Manuscript

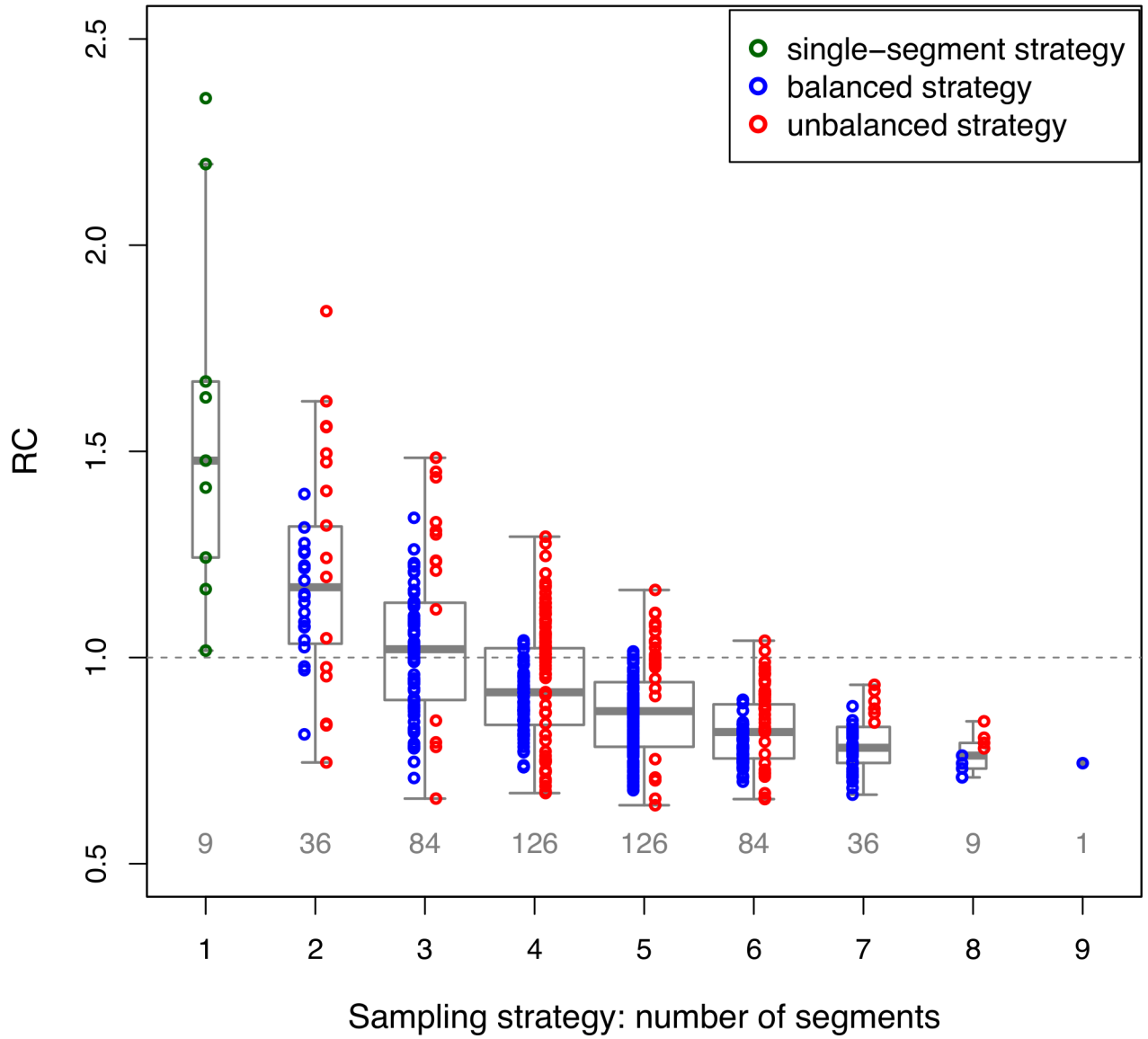
Author Manuscript

Author Manuscript

Author Manuscript



### Intra-Exam RC



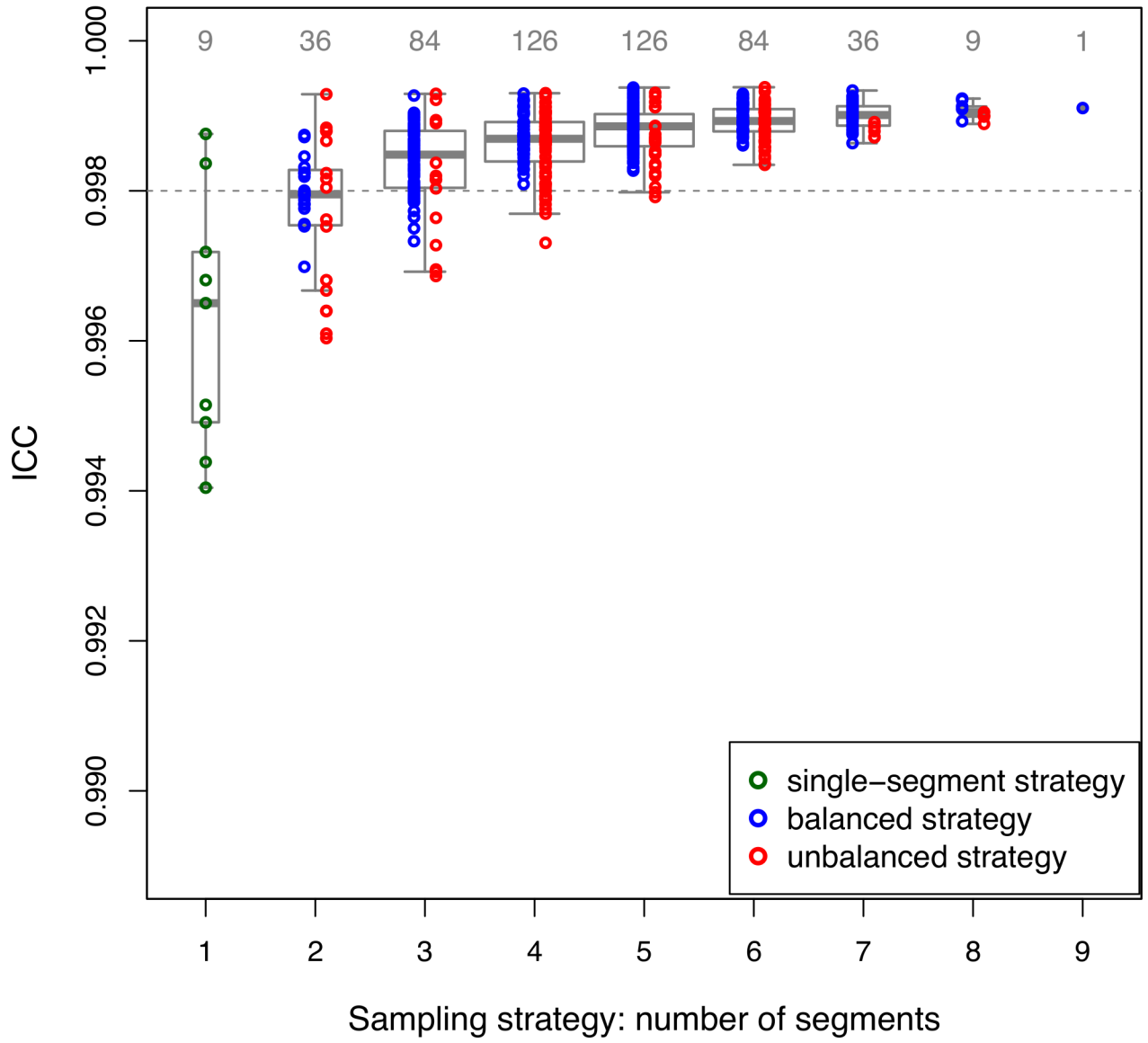
Author Manuscript

Author Manuscript

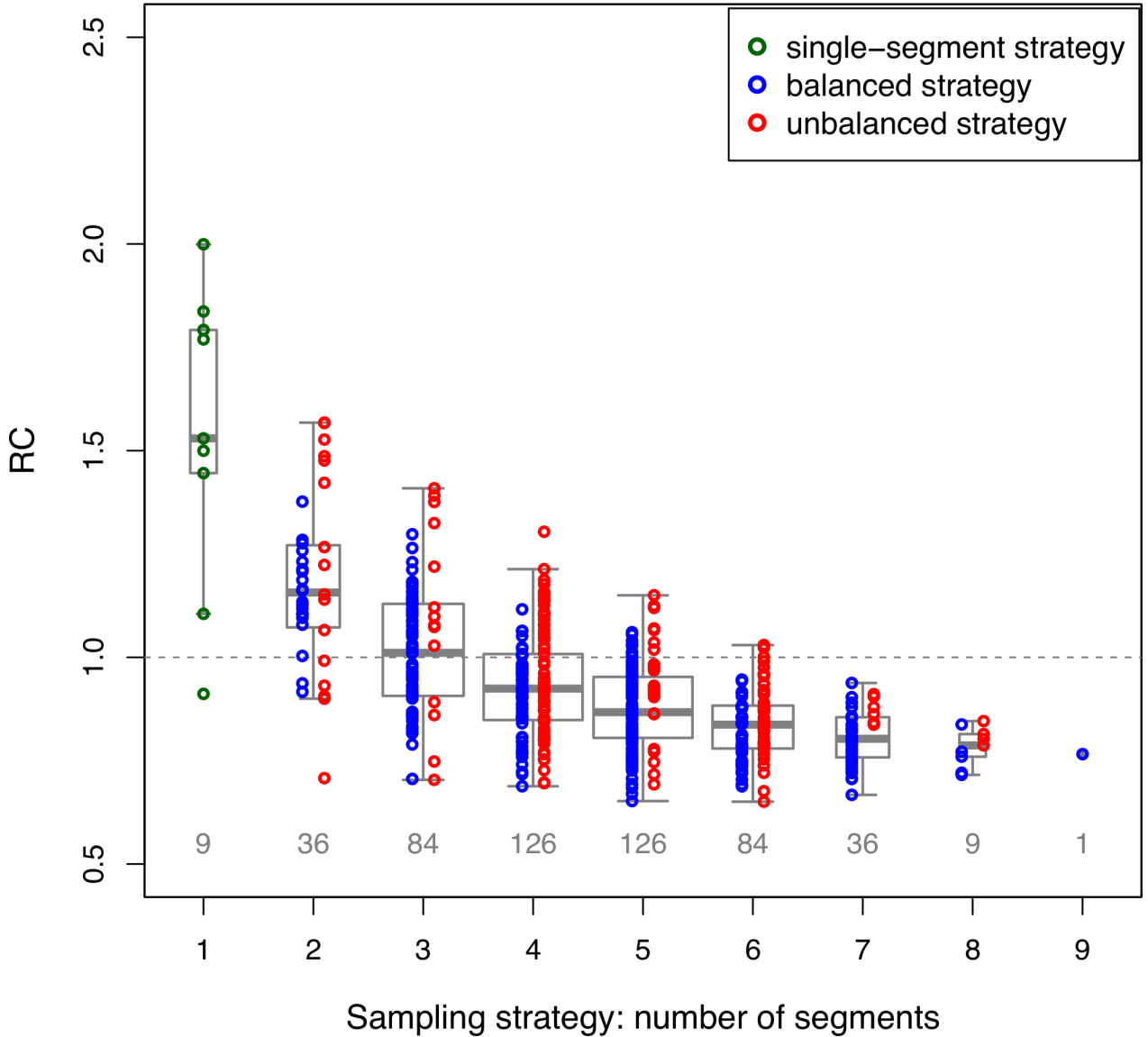
Author Manuscript

Author Manuscript

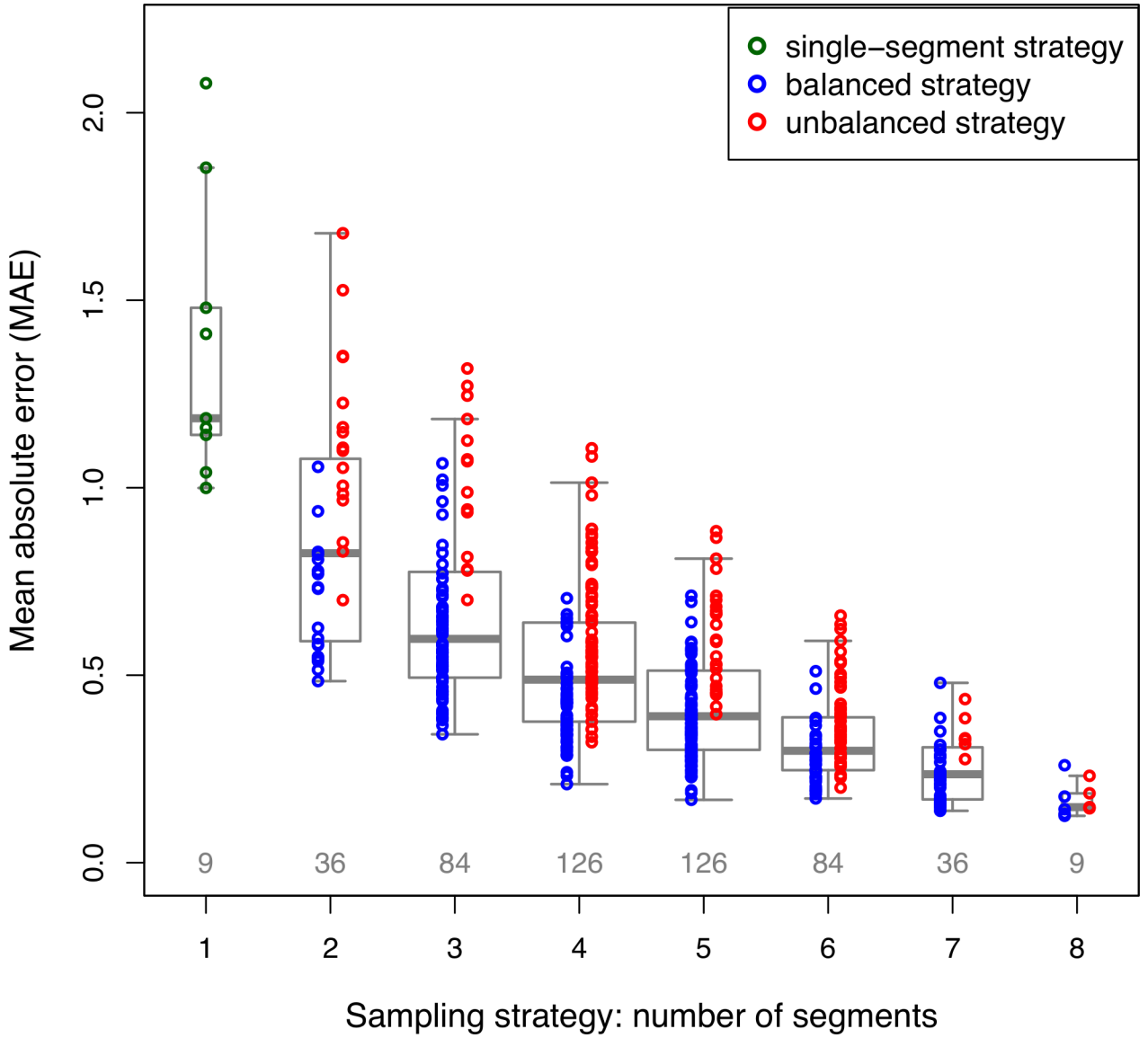
### Inter-Exam ICC



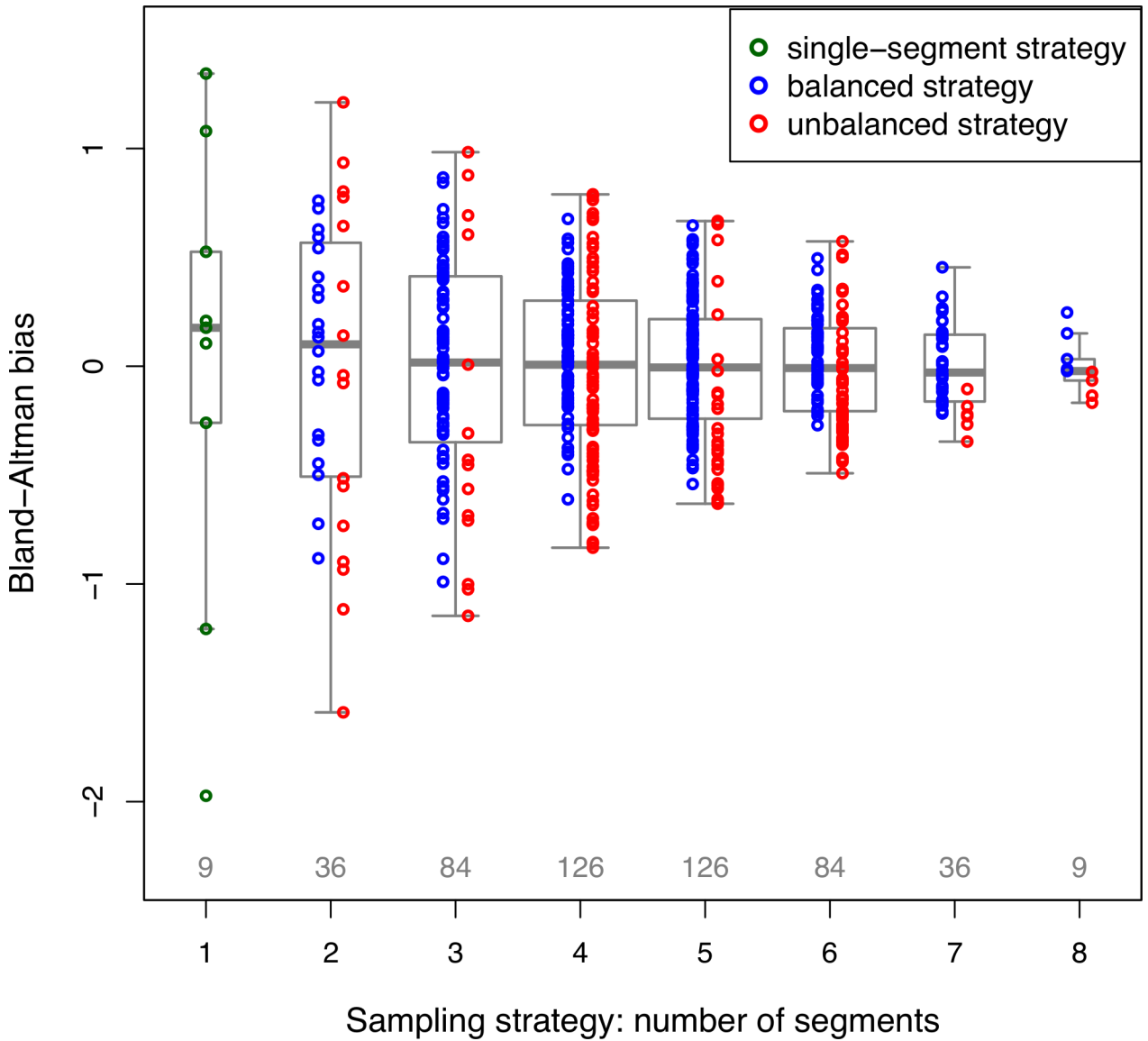
### Inter-Exam RC



**Figure 4:** Box plots showing repeatability (y-axis) shown for sampling strategies by the number of ROIs used. Intra-exam repeatability assessed by ICC (A) and RC (B), as well as inter-exam repeatability assessed by ICC (C) and RC (D) are shown. In each subfigure, each dot represents a particular sampling strategy (510 subsets and the 9-ROI strategy). Strategies where the number of ROIs in the left and right hepatic lobes differed by no more than 1 (i.e. balanced) are color-coded in blue, strategies where the number of ROIs in the left and right hepatic lobes differed by 2 or more (i.e. unbalanced) are color-coded in red. The special case of strategies with a single ROI is color-coded in green. The thresholds of 0.998 for ICC and 1% for RC are illustrated by the dashed horizontal lines. Balanced strategies tended to achieve higher repeatability than unbalanced strategies.



**Figure 5:** Box plot showing mean absolute error (y-axis) of each sampling strategy relative to the 9-ROI PDFFF shown for sampling strategies by the number of ROIs used (x-axis). Each dot represents a particular sampling strategy (510 total). Strategies where the number of ROIs in the left and right hepatic lobes differed by no more than 1 (i.e. balanced) are color-coded in blue, strategies where the number of ROIs in the left and right hepatic lobes differed by 2 or more (i.e. unbalanced) are color-coded in red. The special case of strategies with a single ROI is color-coded in green. Balanced strategies tended to have lower mean absolute error than unbalanced strategies.



**Figure 6:** Box plot showing Bland-Altman bias (y-axis) of each sampling strategy relative to the 9-ROI PDFF shown for sampling strategies by the number of ROIs used (x-axis). Each dot represents a particular sampling strategy (510 total). Strategies where the number of ROIs in the left and right hepatic lobes differed by no more than 1 (i.e. balanced) are color-coded in blue, strategies where the number of ROIs in the left and right hepatic lobes differed by 2 or more (i.e. unbalanced) are color-coded in red. The special case of strategies with a single ROI is color-coded in green. Balanced strategies tended to bias closer to zero than unbalanced strategies.

**Table 1:**

Table of MR imaging parameters for confounder-corrected chemical-shift-encoded magnitude-based proton density fat fraction multi-echo GRE sequence.

| Parameter                | Values                                |
|--------------------------|---------------------------------------|
| Pulse sequence           | 2D Spoiled GRE                        |
| Slice thickness          | 8 mm                                  |
| Flip angle               | 10°                                   |
| Interslice gap           | 0 mm                                  |
| TE                       | 1.15, 2.30, 3.45, 4.60, 5.75, 6.90 ms |
| TR                       | 125 ms                                |
| Image matrix             | 192 × 192 base matrix                 |
| FOV                      | 44 × 44 cm base                       |
| NEX                      | 1                                     |
| Parallel imaging         | Off                                   |
| BW                       | ± 142 kHz                             |
| Fractional echo sampling | 0.8                                   |

Abbreviations: GRE = gradient recalled echo; TR = repetition time; TE = echo time; FOV = field of view; NEX = number of excitations; BW = bandwidth (units are kHz= kilo Hertz)

**Table 2:**

Table of intra-exam and inter-exam repeatability (ICC) by the number of ROIs used. Means and ranges are reported overall and separately for the subsets of balanced and unbalanced sampling strategies.

| Number of ROIs | Number of balanced strategies | Number of unbalanced strategies | Intra-exam ICC              |                             |                          | Inter-exam ICC              |                             |                          |
|----------------|-------------------------------|---------------------------------|-----------------------------|-----------------------------|--------------------------|-----------------------------|-----------------------------|--------------------------|
|                |                               |                                 | All                         | Balanced                    | Unbalanced               | All                         | Balanced                    | Unbalanced               |
| 1 (n = 9)      | -                             | -                               | 0.9954<br>(0.9898 – 0.9984) | -                           | -                        | 0.9962<br>(0.994 – 0.9988)  | -                           | -                        |
| 2 (n = 36)     | 20                            | 16                              | 0.9975<br>(0.994 – 0.9991)  | 0.9978<br>(0.9967 – 0.9989) | 0.9971 (0.994 – 0.9991)  | 0.9978<br>(0.996 – 0.9993)  | 0.998 (0.997 – 0.9987)      | 0.9976 (0.996 – 0.9993)  |
| 3 (n = 84)     | 70                            | 14                              | 0.9981<br>(0.996 – 0.9993)  | 0.9983<br>(0.9969 – 0.9992) | 0.9975 (0.996 – 0.9993)  | 0.9984<br>(0.9969 – 0.9993) | 0.9984<br>(0.9973 – 0.9993) | 0.9981 (0.9969 – 0.9993) |
| 4 (n = 126)    | 60                            | 66                              | 0.9985<br>(0.997 – 0.9993)  | 0.9987<br>(0.9981 – 0.9991) | 0.9983 (0.997 – 0.9993)  | 0.9986<br>(0.9973 – 0.9993) | 0.9987<br>(0.9981 – 0.9993) | 0.9986 (0.9973 – 0.9993) |
| 5 (n = 126)    | 100                           | 26                              | 0.9987<br>(0.9976 – 0.9993) | 0.9988<br>(0.9982 – 0.9992) | 0.9984 (0.9976 – 0.9993) | 0.9988<br>(0.9979 – 0.9994) | 0.9989<br>(0.9983 – 0.9994) | 0.9986 (0.9979 – 0.9993) |
| 6 (n = 84)     | 40                            | 44                              | 0.9988<br>(0.9981 – 0.9993) | 0.9989<br>(0.9986 – 0.9992) | 0.9987 (0.9981 – 0.9993) | 0.9989<br>(0.9983 – 0.9994) | 0.999<br>(0.9986 – 0.9993)  | 0.9989 (0.9983 – 0.9994) |
| 7 (n = 36)     | 30                            | 6                               | 0.9989<br>(0.9985 – 0.9993) | 0.999<br>(0.9987 – 0.9993)  | 0.9986 (0.9985 – 0.9987) | 0.999<br>(0.9986 – 0.9993)  | 0.999<br>(0.9986 – 0.9993)  | 0.9988 (0.9987 – 0.9989) |
| 8 (n = 9)      | 5                             | 4                               | 0.999<br>(0.9988 – 0.9991)  | 0.9991<br>(0.999 – 0.9991)  | 0.9989 (0.9988 – 0.9989) | 0.999<br>(0.9989 – 0.9992)  | 0.9991<br>(0.9989 – 0.9992) | 0.999 (0.9989 – 0.999)   |
| 9 (n = 1)      | 1                             | -                               | 0.999<br>(0.999 – 0.999)    | 0.999 (0.999 – 0.999)       | -                        | 0.9991<br>(0.9991 – 0.9991) | 0.9991<br>(0.9991 – 0.9991) | -                        |

**Table 3:**

Table of intra-exam and inter-exam repeatability (RC) by the number of ROIs used. Means and ranges are reported overall and separately for the subsets of balanced and unbalanced sampling strategies.

| Number of ROIs | Number of balanced strategies | Number of unbalanced strategies | Intra-exam RC         |                    |                    | Inter-exam RC         |                    |                    |
|----------------|-------------------------------|---------------------------------|-----------------------|--------------------|--------------------|-----------------------|--------------------|--------------------|
|                |                               |                                 | All                   | Balanced           | Unbalanced         | All                   | Balanced           | Unbalanced         |
| 1 (n = 9)      | -                             | -                               | 1.57<br>(1.02 – 2.36) | -                  | -                  | 1.54<br>(0.91 – 2.00) | -                  | -                  |
| 2 (n = 36)     | 20                            | 16                              | 1.19<br>(0.75 – 1.84) | 1.14 (0.81 – 1.40) | 1.26 (0.75 – 1.84) | 1.18<br>(0.71 – 1.57) | 1.15 (0.92 – 1.38) | 1.21 (0.71 – 1.57) |
| 3 (n = 84)     | 70                            | 14                              | 1.02<br>(0.66 – 1.48) | 1.00 (0.71 – 1.34) | 1.16 (0.66 – 1.48) | 1.02<br>(0.70 – 1.41) | 1.00 (0.71 – 1.30) | 1.09 (0.70 – 1.41) |
| 4 (n = 126)    | 60                            | 66                              | 0.93<br>(0.67 – 1.29) | 0.89 (0.73 – 1.04) | 0.97 (0.67 – 1.29) | 0.93<br>(0.69 – 1.30) | 0.91 (0.69 – 1.12) | 0.96 (0.70 – 1.30) |
| 5 (n = 126)    | 100                           | 26                              | 0.87<br>(0.64 – 1.16) | 0.84 (0.68 – 1.01) | 0.96 (0.64 – 1.16) | 0.88<br>(0.65 – 1.15) | 0.86 (0.65 – 1.06) | 0.94 (0.69 – 1.15) |
| 6 (n = 84)     | 40                            | 44                              | 0.82<br>(0.66 – 1.04) | 0.79 (0.70 – 0.90) | 0.86 (0.66 – 1.04) | 0.84<br>(0.65 – 1.03) | 0.81 (0.69 – 0.95) | 0.86 (0.65 – 1.03) |
| 7 (n = 36)     | 30                            | 6                               | 0.79<br>(0.67 – 0.93) | 0.77 (0.67 – 0.88) | 0.89 (0.84 – 0.93) | 0.81<br>(0.67 – 0.94) | 0.79 (0.67 – 0.94) | 0.87 (0.84 – 0.91) |
| 8 (n = 9)      | 5                             | 4                               | 0.76<br>(0.71 – 0.85) | 0.73 (0.71 – 0.76) | 0.81 (0.78 – 0.85) | 0.78<br>(0.72 – 0.85) | 0.76 (0.72 – 0.84) | 0.81 (0.79 – 0.85) |
| 9 (n = 1)      | 1                             | -                               | 0.74                  | 0.74               | -                  | 0.77                  | 0.77               | -                  |



**Table 4:**

Table of proportions of sampling strategies that achieved the intra-exam and inter-exam repeatability threshold for ICC and RC by number of ROIs used, computed overall and separately for balanced and unbalanced subsets of sampling strategies.

| Number of ROIs | Intra-exam ICC |                |              | Inter-exam ICC |                |              | Intra-exam RC |              |             | Inter-exam RC |              |             |
|----------------|----------------|----------------|--------------|----------------|----------------|--------------|---------------|--------------|-------------|---------------|--------------|-------------|
|                | All            | Balanced       | Unbalanced   | All            | Balanced       | Unbalanced   | All           | Balanced     | Unbalanced  | All           | Balanced     | Unbalanced  |
| 1 (n = 9)      | 11% (1/9)      | -              | -            | 22% (2/9)      | -              | -            | 0% (0/9)      | 0% (0/9)     | -           | 11% (1/9)     | -            | -           |
| 2 (n = 36)     | 36% (13/36)    | 35% (7/20)     | 38% (6/16)   | 39% (14/36)    | 35% (7/20)     | 44% (7/16)   | 22% (8/36)    | 15% (3/20)   | 31% (5/16)  | 19% (7/36)    | 10% (2/20)   | 31% (5/16)  |
| 3 (n = 84)     | 63% (53/84)    | 70% (49/70)    | 29% (4/14)   | 74% (62/84)    | 77% (54/70)    | 57% (8/14)   | 46% (39/84)   | 50% (35/70)  | 29% (4/14)  | 49% (41/84)   | 53% (37/70)  | 29% (4/14)  |
| 4 (n = 126)    | 83% (105/126)  | 100% (60/60)   | 68% (45/66)  | 91% (115/126)  | 100% (60/60)   | 83% (55/66)  | 71% (89/126)  | 92% (55/60)  | 52% (34/66) | 72% (91/126)  | 83% (50/60)  | 62% (41/66) |
| 5 (n = 126)    | 94% (118/126)  | 100% (100/100) | 69% (18/26)  | 98% (123/126)  | 100% (100/100) | 88% (23/26)  | 88% (111/126) | 98% (98/100) | 50% (13/26) | 88% (111/126) | 92% (92/100) | 73% (19/26) |
| 6 (n = 84)     | 100% (84/84)   | 100% (40/40)   | 100% (44/44) | 100% (84/84)   | 100% (40/40)   | 100% (44/44) | 98% (82/84)   | 100% (40/40) | 95% (42/44) | 96% (81/84)   | 100% (40/40) | 93% (41/44) |
| 7 (n = 36)     | 100% (36/36)   | 100% (30/30)   | 100% (6/6)   | 100% (36/36)   | 100% (30/30)   | 100% (6/6)   | 100% (36/36)  | 100% (30/30) | 100% (6/6)  | 100% (36/36)  | 100% (30/30) | 100% (6/6)  |
| 8 (n = 9)      | 100% (9/9)     | 100% (5/5)     | 100% (4/4)   | 100% (9/9)     | 100% (5/5)     | 100% (4/4)   | 100% (9/9)    | 100% (5/5)   | 100% (4/4)  | 100% (9/9)    | 100% (5/5)   | 100% (4/4)  |
| 9 (n = 1)      | 100% (1/1)     | 100% (1/1)     | -            | 100% (1/1)     | 100% (1/1)     | -            | 100% (1/1)    | 100% (1/1)   | -           | 100% (1/1)    | 100% (1/1)   | -           |

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 5:**

Mean absolute error and bias by number of ROIs, reported overall and separately for balanced and unbalanced strategies. Mean  $\pm$  standard deviation is reported for mean absolute error and range is reported for bias.

| Number of ROIs | Number of balanced strategies | Number of unbalanced strategies | Mean Absolute Error (mean $\pm$ standard deviation) |                 |                 | Bias (range) |              |               |
|----------------|-------------------------------|---------------------------------|---|-----------------|-----------------|--------------|--------------|---------------|
|                |                               |                                 | All   | Balanced        | Unbalanced      | All          | Balanced     | Unbalanced    |
| 1              | -                             | -                               | 1.37 $\pm$ 1.05                                     | -               | -               | -1.97 – 1.34 | -            | -             |
| 2              | 20                            | 16                              | 0.88 $\pm$ 0.66                                     | 0.68 $\pm$ 0.48 | 1.13 $\pm$ 0.84 | -1.59 – 1.21 | -0.88 – 0.76 | -1.59 – 1.21  |
| 3              | 70                            | 14                              | 0.67 $\pm$ 0.50                                     | 0.60 $\pm$ 0.43 | 1.02 $\pm$ 0.73 | -1.15 – 0.98 | -0.99 – 0.87 | -1.15 – 0.98  |
| 4              | 60                            | 66                              | 0.52 $\pm$ 0.39                                     | 0.41 $\pm$ 0.29 | 0.62 $\pm$ 0.46 | -0.83 – 0.79 | -0.61 – 0.68 | -0.83 – 0.79  |
| 5              | 100                           | 26                              | 0.42 $\pm$ 0.31                                     | 0.37 $\pm$ 0.27 | 0.60 $\pm$ 0.40 | -0.63 – 0.67 | -0.54 – 0.65 | -0.63 – 0.67  |
| 6              | 40                            | 44                              | 0.33 $\pm$ 0.25                                     | 0.27 $\pm$ 0.19 | 0.39 $\pm$ 0.27 | -0.49 – 0.57 | -0.27 – 0.50 | -0.49 – 0.57  |
| 7              | 30                            | 6                               | 0.25 $\pm$ 0.19                                     | 0.23 $\pm$ 0.17 | 0.35 $\pm$ 0.08 | -0.35 – 0.45 | -0.22 – 0.45 | -0.35 – 0.10  |
| 8              | 5                             | 4                               | 0.17 $\pm$ 0.13                                     | 0.17 $\pm$ 0.12 | 0.18 $\pm$ 0.06 | -0.17 – 0.25 | -0.02 – 0.25 | -0.17 – -0.03 |