Analysis of spleen volumetry as imaging biomarker: Precision of measurements and usability on CT and MRI

Poster No.:	C-2390
Congress:	ECR 2016
Type:	Scientific Exhibit
Authors:	H. Beaumont[1], E. Oubel[1], A. Iannessi[2], P. A. Thyss[2]; [1]Valbonne/FR, [2]Nice/FR
Keywords:	Cancer, Statistics, Segmentation, Computer Applications-3D, MR, CT-Quantitative, CT, Spleen, Hematologic, Computer applications
DOI:	10.1594/ecr2016/C-2390

Aims and objectives

Myelofibrosis is a type of chronic leukemia affecting bone marrow in impairing the normal production of blood cells. The disease manifests itself by an extensive scarring in patients bone marrow, leading to severe anemia, weakness, fatigue, and often splenomegaly and hepatomegaly (Figure 1) [1]. In the group of myeloproliferative diseases, which myelofibrosis belongs to, published epidemiology data are scarce [2].

Spleen enlargement is thus a clinical sign of blood cancers and an index of the therapeutic response. T2-weighted MRI is the preferred modality for measuring spleen volume and non-contrasted CT is an alternative. Usually, 25% volume change is considered as progressive disease and -35% a response to treatment [3] [4].

Spleen volumetry is a quantitative imaging biomarker (QIB) that is useful not only for the monitoring of myelofibrosis but that could find several other applications. For all the different applications clinicians can find, the quantification of spleen volume must go through a rigorous qualification process [5].

As a first step in the qualification of spleen volumetry as a QIB, we analyzed inter-reader precision and usability of measurement when using non-contrast CT, T2, non-contrast T1 in-phase and T1 out-of-phase MRI sequences.

Images for this section:

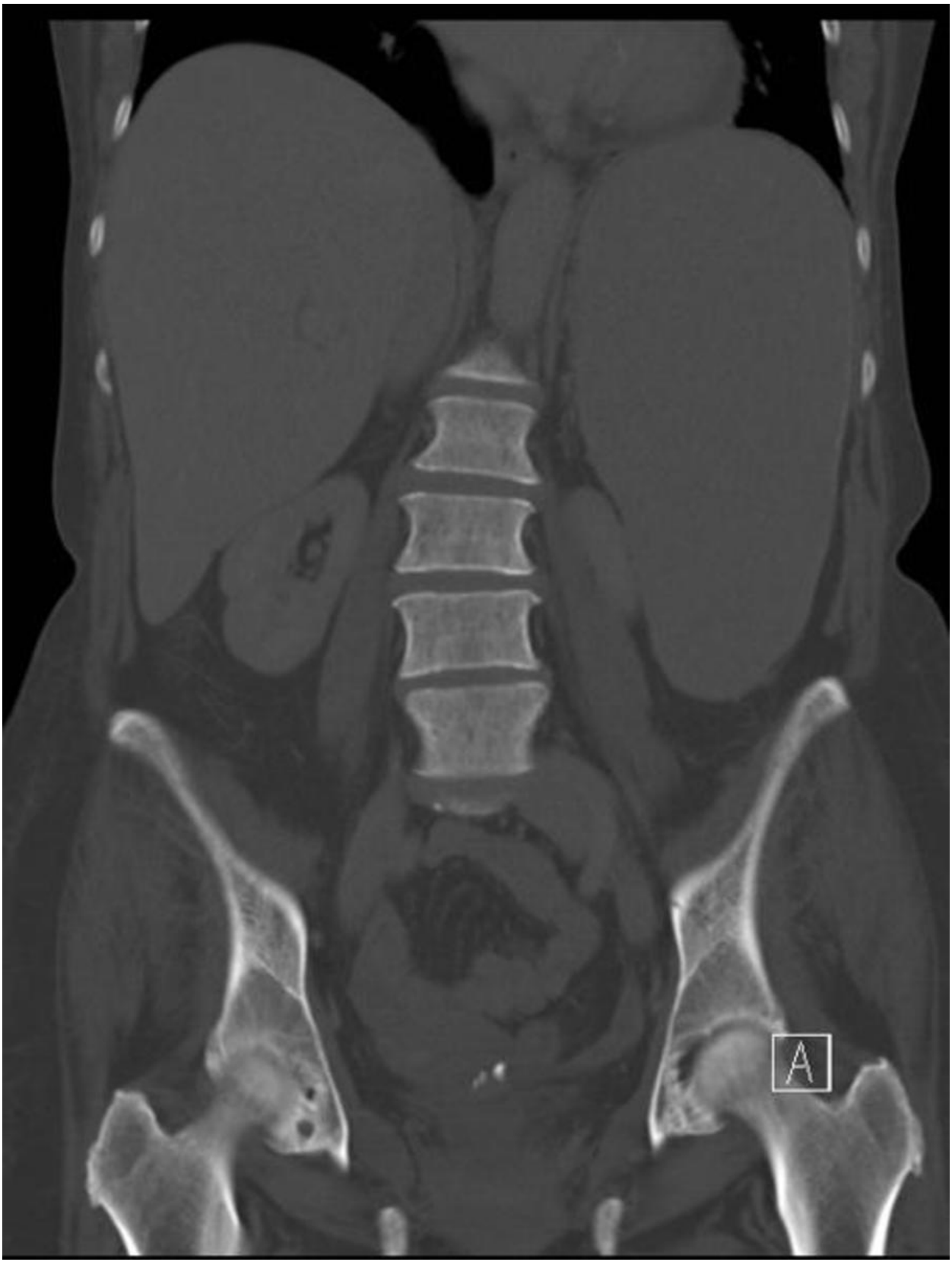
A

Fig. 1: Massive splenomegaly in a 72 year old with bony sclerosis is most likely myelofibrosis. Early hepatomegaly as suggested by slight rouding of the inferior end of the liver.

Methods and materials

We designed our study based on works previously performed by Harris et al [6] and Linguraru et al. [7].

Harrys et al documented measurement precision of an automatic tool performed on CT and relying on a cohort of 230 patients. Precision was between 6% and 10%.

Linguraru et al. reported that Volume Error (VER) also on CT, computed as the percentage of absolute volume difference between automated and manual measurements relative to the true (manual) measurement, featured inter-observer variability of 2.9% +/- 3%. The study included 14 patients and two readers performed the measurements.

As it was not documented on previous studies, we assumed that precision on CT was not highly different than MRI and that measurements involved scans thickness in the range of 1mm up to 5mm.

As a starting point, we assumed that including, at least, 15 CT and 15 MRI scans in our study and involving, at least, two readers, we would reach comparable confidence intervals than Harrys et Linguraru groups.

One expert radiologist and two imaging scientists measured the spleen volume of 39 healthy patients. Patients were retrospectively selected from 6 different European sites.

Readers reviewed 18 cases using T2, T1 in-phase and T1 out-of-phase MRI sequences. In addition, 21 cases using CT were analyzed. A semi-automatic contouring tool (MEDIAN LMS) was used(Figure 2).

For CT and the three MRI sequences, mean segmentation time and Reproducibility Coefficient (RC) [8] were computed on log-transformed bootstrapped data. Minimal Detection Change (MDC) was estimated.

Images for this section:

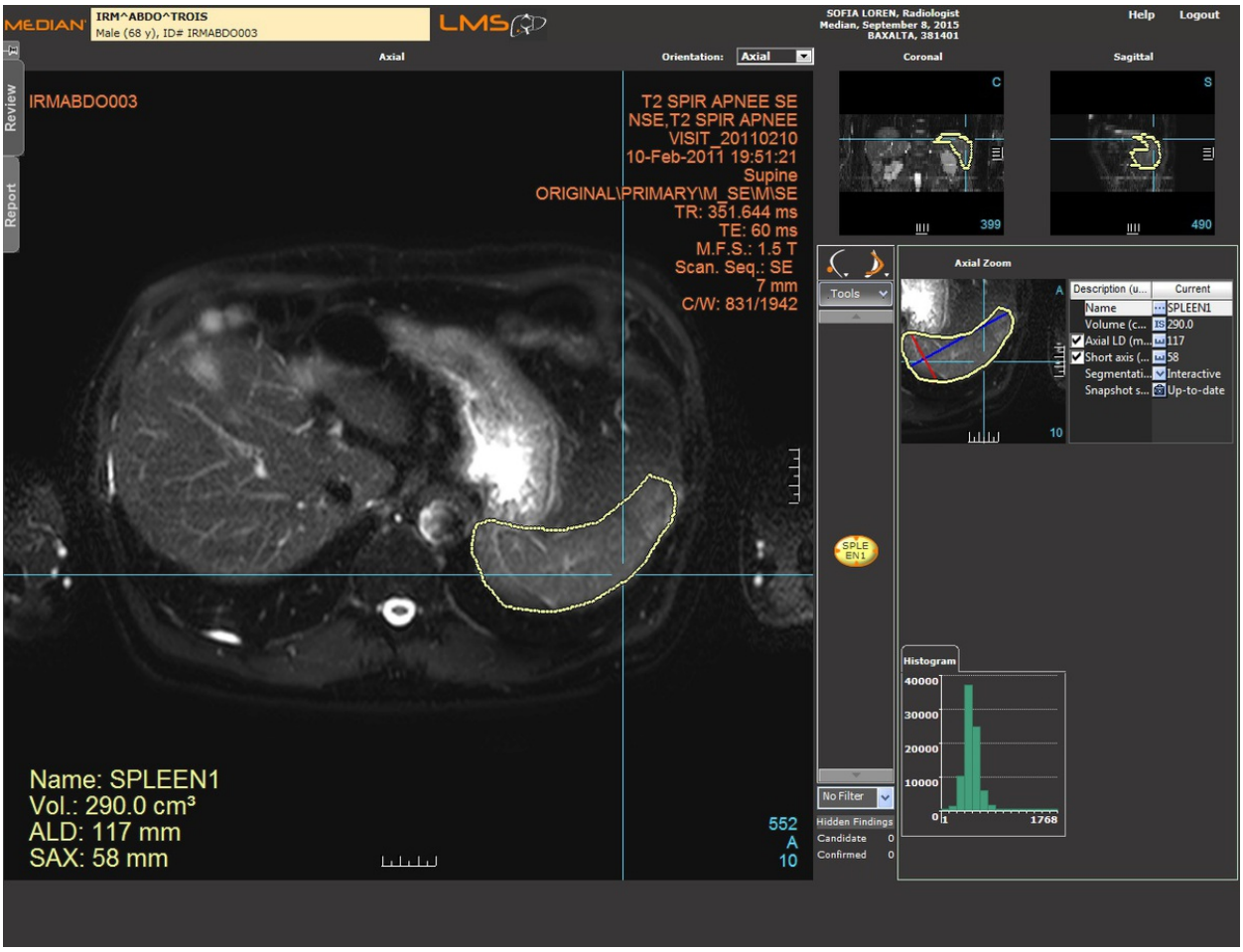

Fig. 2: Assessment of spleen volume was performed with LMS software in using a semi-automated tool enabling voxel-wise adjustments of the contours. From the 3D segmentations, contours and the automatic computing of the volume were archived for all readers. The record of contours and data, allowed retrospective and simultaneous inspection of all evaluations.

Results

RC was not significantly different for the different imaging techniques (Table 1).

As summarized on table 2 for CT and T2 respectively, RC was 0.059 [0.047; 0.070] and 0.058 [0.045; 0.072], MDC was 7.8% and 8.8%.

Readers reported the same two complex cases on the T2 sequence only.

Log-transformed mean segmentation time was not significantly different between the MRI sequences, and it was longer on CT than on MRI.

When pooling all readers together, median segmentation time was close to 4min on CT and 3min on MRI.

Images for this section:

	CT	T2**	T1 in	T1 Out*
Bias	0.15%	0.16%	1.97%	1.475%
RC (Log)	0.065 [0.039; 0.089]	0.064 [0.039; 0.089]	0.109 [0.048; 0.167]	0.075 [0.043; 0.107]
Cov	2.45%	2.46%	4.07%	2.84%
Min. detect change	6.8%	6.8%	12.5%	8.87%

Table 1: Reproducibility of spleen volume measurements. Two readers expert in image analysis were involved. From top to bottom are reported the Bias of the relative difference, the repeatability coefficient (RC) computed on log transformed data with corresponding confidence interval, coefficient of Variation (CoV) and an evaluation of the minimal detection change. The metrics was computed for left to right for CT scans, T2 MRI, T1 in phase MRI and T1 out phase MRI.

	CT	T2**
Bias	0.15%; 1.67%; 1.82%	1.02%; 0.41%; 0.61%
RC (Log)	0.059 [0.047; 0.070]	0.058 [0.045; 0.072]
Cov	2.45%; 2.88%; 3.81%	2.49%; 3.10%; 3.59%
Min Detect Change	7.78%	8.87%

Table 2: Reproducibility of spleen volume measurements. Two readers expert in image analysis and one experienced radiologist were involved. From top to bottom are reported the Bias of the relative difference, the repeatability coefficient (RC) computed on log transformed data with corresponding confidence interval, coefficient of Variation (CoV) and an evaluation of the minimal detection change. The metrics was computed for left to right for CT scans and T2 sequence MRI.

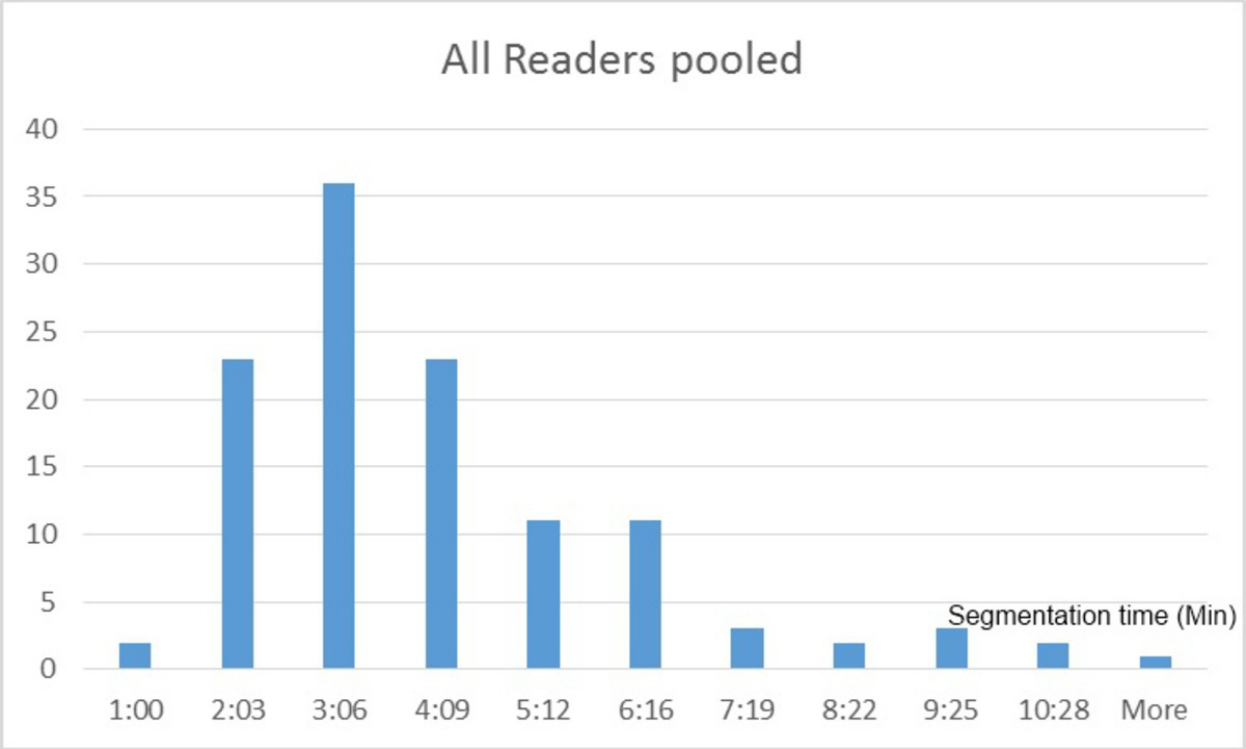

Fig. 3: Distribution of segmentation time in Minute:Seconde of all reader pooled together.

Conclusion

Precision of segmentations was independent of imaging modality and MRI sequences, therefore radiologists do not have to constrain their reads to a specific modality. MDC was compatible with thresholds of [+25%; -35%] usually used for assessing a therapeutic response, meaning that metrology does not represent a limitation in the evaluation of clinical changes. In addition to precision, similar usability was found among the MRI sequences.

As these performances were evaluated on patients with healthy spleen, a confirmation must be reached when analyzing pathological organs, notably regarding confounding factors which has been observed on one T2 scan. Another room for investigation will be the refinement of the clinical response threshold, which may potentially improve sensitivity and specificity at detecting the therapeutic response.

Personal information

Hubert Beaumont, PhD Dpt. of Science MEDIAN Technologies, Valbonne, France; hubert.beaumont@mediantechnologies.com Estanislao Oubel, PhD Dpt. of Science MEDIAN Technologies, Valbonne, France; estanislao.oubel@mediantechnologies.com Antoine Iannessi, M.D. Radiology Centre Antoine Lacassagne, Nice, France; antoine.iannessi@nice.unicancer.fr Antoine Thyss, Pr. M.D. Clinical hematology Centre Antoine Lacassagne, Nice, France; antoine.thyss@nice.unicancer.fr

References

1. Kaplan KR, Mitchell DG, Steiner RM, et al. (1992) Polycythemia vera and myelofibrosis: correlation of MR imaging, clinical, and laboratory findings. Radiology 183:329-334.

2. Moulard O, Mehta J, Fryzek J, et al. (2014) Epidemiology of myelofibrosis, essential thrombocythemia, and polycythemia vera in the European Union. Eur J Haematol 92:289-297.

3. Tefferi A, Barosi G, Mesa R a., et al. (2006) International Working Group (IWG) consensus criteria for treatment response in myelofibrosis with myeloid metaplasia, for the IWG for Myelofibrosis Research and Treatment (IWG-MRT). Blood 108:1497-1503. doi: 10.1182/blood-2006-03-009746

4. Harrison C, Kiladjian J-J, Al-Ali HK, et al. (2012) JAK inhibition with ruxolitinib versus best available therapy for myelofibrosis. N Engl J Med 366:787-798. doi: 10.1056/NEJMoa1110556

5. Raunig DL, Mcshane LM, Pennello G, et al. (2014) Quantitative imaging biomarkers: A review of statistical methods for technical performance assessment. Stat Methods Med Res. doi: 10.1177/0962280214537344

6. Harris A, Kamishima T, Hao HY, et al. (2010) Splenic volume measurements on computed tomography utilizing automatically contouring software and its relationship with age, gender, and anthropometric parameters. Eur J Radiol. doi: 10.1016/j.ejrad.2009.08.013

7. Linguraru MG, Sandberg JK, Li Z, et al. (2010) Automated segmentation and quantification of liver and spleen from CT images using normalized probabilistic atlases and enhancement estimation. Med Phys 37:771-783. doi: 10.1118/1.3284530

8. Sullivan DC, Obuchowski NA, Kessler LG, et al. (2015) Metrology Standards for Quantitative Imaging Biomarkers. Radiology 277:813-25. doi: 10.1148/radiol.2015142202