WED-238

Repeatability and reproducibility assessment and its acceptable standard error of means for qFibrosis system in multi-site NASH clinical trials

Jason Pik Eu Chang[1], Claudia Filozof[2], Dean Tai[3], Kutbuddin Akbary[3], Yayun Ren[3], Elaine Chng[3], Feng Liu[4], Lai Wei[5], David E Kleiner[6], Arun Sanyal[7]

[1]Singapore General Hospital, Gastroenterology & Hepatology, Singapore, *[2]Labcorp Drug Development, Israel*, *[3]HistoIndex Pte Ltd, Singapore*, *[4]Peking University Hepatology Institute, Peking University People's Hospital, China*, *[5]Hepatopancreatobiliary Center, Beijing Tsinghua Changgung Hospital, China*, *[6]National Cancer Institute, National Institutes of Health, United States*, *[7]Stravitz-Sanyal Institute of Liver Disease and Metabolic Health, VCU School of Medicine, United States*

Email: akbary.kutbuddin@histoindex.com

Background and Aims: Accurate quantification of fibrosis is critical in clinical trials in NASH. In recent years, the application of digital pathology with artificial intelligence (AI), including qFibrosis has gained increased attention due to the potential to quantify fibrosis features from liver biopsies with better inter-/intra observer agreements as compared to conventional reads. Our group has recently reported the inter-system (repeatability) and intra-system (reproducibility) of qFibrosis. In this study, we aim to describe an acceptable standard error of means for the qFibrosis system in NASH clinical trials.

Method: The study included 41 core biopsies with confirmed NASH, of which 9, 9, 13 and 10 samples were staged F1, F2, F3 and F4, respectively. Scanning was conducted with 3 Genesis200® machines, using second harmonic generation/two-photon excitation fluorescence (SHG/TPEF) microscopy on unstained slides. 3 repeated scans were conducted for each sample by each machine (reproducibility) and by three different machines (repeatability) at different time points, and a qFibrosis continuous value (qFC) is generated based on an AI algorithm for each sample per scan. The standard error of means (SEM) was determined by 2 methods: 1) cohort based: taking the qFC for each patient (the median of 9 scans); 2) sample based: taking the 9 scans for each patient; with the same fibrosis stages.

Results: The SEM for F1, F2, F3, F4 are 0.1924, 0.1917, 0.3595, 1.1038 respectively using cohort-based method. And the SEM for F1, F2, F3, F4 are 0.1334, 0.1081, 0.1417 0.3464 respectively using sample-based method. Note that the SEM values are progressively larger for higher stage of fibrosis using the cohort-based method, this is due to the fact the qFC is generally larger for higher stage of fibrosis. This is not the case for sample-based method, as the SEM is dependent on system repeatability and reproducibility instead of the value of qFC.

Conclusion: In our recent efforts in establishing repeatability and reproducibility in the advent of AI digital pathology, there is still a gap in determining the acceptable SEM for these quantitative measurements. This study aims to summarize different approaches to determine SEMs and their results. The group intends to further investigate impact of SEM on the result of assessment in NASH clinical trials with additional approaches such as Obuchowski index and report these findings to the clinical trials community.

Figure: NA

Repeatability and reproducibility assessment and its acceptable standard error of means for qFibrosis system in multi-site NASH clinical trials

J. CHANG[1], C. M. FILOZOF[2], D. TAI[3], K. AKBARY[3], Y. REN[3], E. CHNG[3], F. LIU[4], L. WEI[5,6], D. E. KLEINER[7], A. J. SANYAL[8]

1Department of Gastroenterology & Hepatology, Singapore General Hospital, Singapore; 2Labcorp Drug Development, Tel Aviv, Israel; 3HistoIndex Pte Ltd, Singapore; 4Peking University Hepatology Institute, Beijing Key Laboratory of Hepatitis C and Immunotherapy for Liver Diseases, Peking University People's Hospital, Beijing, China; 5Hepatopancreatobiliary Center, Beijing Tsinghua Changgung Hospital, Tsinghua University, Beijing, China; 6Institute for Precision Medicine, Tsinghua University, Beijing, China; 7Laboratory of Pathology; Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA; 8Stravitz-Sanyal Institute of Liver Disease and Metabolic Health, Dept. of Internal Medicine, VCU School of Medicine, Richmond, VA

Introduction

- Accurate quantification of fibrosis is critical in clinical trials in NASH. In recent years, the application of digital pathology with artificial intelligence (AI), including qFibrosis has gained increased attention due to the potential to quantify fibrosis features from liver biopsies with better inter-/intra observer agreements as compared to conventional reads.

Aim

- Our group has recently reported the inter-system (repeatability) and intra-system (reproducibility) of qFibrosis. In this study, we aim to describe an acceptable standard error of means for the qFibrosis system in NASH clinical trials.

Method

- The study included 41 core biopsies with confirmed NASH, of which 9, 9, 13 and 10 samples were staged F1, F2, F3 and F4, respectively.
- Scanning was conducted with 3 separate Genesis200® machines, using second harmonic generation/two-photon excitation fluorescence (SHG/TPEF) microscopy on unstained slides.
- 3 repeated scans were conducted for each sample by each machine (reproducibility) and by 3 different machines (repeatability) at different time points, and a qFibrosis continuous value (qFC) was generated based on an AI algorithm for each sample per scan.
- The standard error of means (SEM) was determined by 2 methods: 1) cohort based: taking the qFC for each patient (the median of 9 scans); 2) sample based: taking the 9 scans for each patient; with the same fibrosis stages (**Figure 1**).

Conclusions

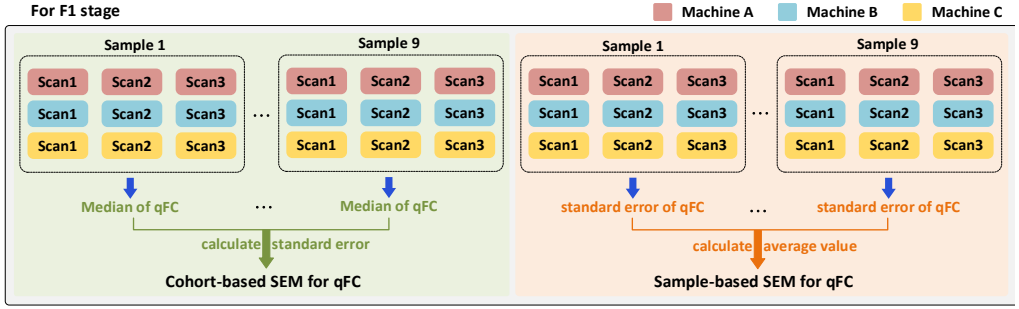
- In our recent efforts in establishing repeatability and reproducibility in the advent of AI digital pathology, there is still a gap in determining the acceptable SEM for these quantitative measurements.
- This study aims to summarize different approaches to determine SEMs and their results in clinical trials. The results described here were a part of clinical trials where cohort-based analysis was performed.
- We intend to further investigate impact of SEM on these measurements in NASH clinical trials with additional data analysis, such as Obuchowski index, and report these findings to the clinical trials community.

Contact information

jason.chang@singhealth.com.sg

Results

Figure 1: Demonstration of SEM calculation for F1 stage. Each sample was scanned 3 times at different time points by the same machine; separately, each sample was also scanned by 3 different machines. qFC was calculated for all scans and SEM of F2, F3 and F4 were comparable.



- In this study, the median qFC of all scans were 1.70, 1.83, 2.71 and 4.72 for F1, F2, F3 and F4, respectively.
- The qFC cut-off values correlating with semi-quantitative NASH-CRN scores are 1.04,1.45, 2.12 and 3.45 for each stage.
- The SEM for F1, F2, F3, F4 are 0.1924, 0.1917, 0.3595, 1.1038 respectively using cohort-based method, while the SEM were 0.1334, 0.1081, 0.1417 0.3464 using sample-based method (**Table 1**).

Table 1: The median, cohort-based SEM and sample-based SEM for qFC.

Pathologist fibrosis stage	Median of qFC	Cohort-based SEM for qFC	Sample-based SEM for qFC
F1	1.70	0.19	0.13
F2	1.83	0.19	0.11
F3	2.71	0.36	0.14
F4	4.72	1.10	0.35

- The SEM values are progressively larger for higher stage of fibrosis using the cohort-based method, since qFC is generally larger for higher stage of fibrosis. This is not the case for sample-based method, as the SEM is dependent on system repeatability and reproducibility instead of the value of qFC.
- Cohort-based SEM was used for the intra-/inter-system agreement of qFC because it is more suitable in the context of use in evaluation of cohorts from clinical trials.
- Using the cohort-based SEM, the overall intra-system agreement of qFC was 90.79% (95%; CI: 0.878-0.938), while the inter-system agreement was 81.03% (95%; CI: 0.770-0.848) (**Table 2**).

Table 2: Intra-/inter-system agreement using cohort-based SEM

Intra-system agreement (95% CI)		Inter-system agreement (95% CI)	
Machine A	93.50% (0.886, 0.976)	**Machine A vs B**	75.61% (0.675, 0.829)
Machine B	87.80% (0.821, 0.935)	**Machine A vs C**	86.18% (0.797, 0.919)
Machine C	91.06% (0.854, 0.959)	**Machine B vs C**	81.30% (0.740, 0.878)
Overall	**90.79% (0.878, 0.938)**	**Overall**	**81.03% (0.770, 0.848)**