

Estimating the accuracy and precision of quantitative imaging biomarkers as endpoints for clinical trials using standard-of-care CT

Paul Kinahan¹, Darrin Byrd¹, Hao Yang², Hugo Aerts³, Binzhang Zhao², Andrey Fedorov³, Lawrence Schwartz², Tavis Allison², Chaya Moskowitz⁴

¹ Department of Radiology, University of Washington, Seattle WA

¹ Department of Radiology, Columbia University; New York Presbyterian Hospital, New York NY

¹ Department of Radiology, Harvard University, Boston MA

¹ Department of Biostatistics, Memorial Sloan Kettering Cancer Center, New York NY

ABSTRACT

Quantitative imaging biomarkers (QIBs) hold enormous potential to improve the efficiency of clinical trials that use standard-of-care CT imaging. Examples of QIBs include size, shape, intensity histogram characteristics, texture, radiomics, and more. There is, however, a well-recognized gap between discovery and the translation to practice of QIBs, which is driven in part by concerns about their repeatability and reproducibility in the diverse clinical environment. Our goal is to characterize QIB repeatability and reproducibility by using virtual imaging clinical trials (VICTs) to simulate the full data pathway. We start by estimating the probability distribution functions (PDFs) for patient-, disease-, treatment-, and imaging-related sources of variability. These are used to forward-model sinograms that are reconstructed and then analyzed by the QIB under evaluation in a virtual imaging pipeline. By repeatedly sampling from the variability PDFs, estimates of the bias, variance, repeatability and reproducibility of the QIB can be generated by comparison with the known ground truth. These estimates of QIB performance can be used as evidence of the utility of QIBs in clinical trials of new therapies.

Keywords: Imaging biomarkers, quantitative imaging, repeatability and reproducibility, clinical trials

I. INTRODUCTION

Clinical trials are a cornerstone of developing more effective cancer therapies. However, traditional clinical trials are often slow, expensive, and inefficient. Imaging of disease with standard-of-care CT plays a pivotal role in the management of patients with cancer and is used to measure endpoints in cancer drug trials to quantify efficacy in candidate compounds. There is a tremendous potential for quantitative imaging biomarkers (QIBs) to make clinical trials more efficient and informative. Examples of QIBs include size, shape, intensity histogram characteristics, and texture. Taking advantage of this potential is imperative since in the era of targeted therapies, studies will be smaller, more fractionated, with more expensive therapies. There is, however, a well-recognized gap between discovery and translation to practice for biomarkers in general and specifically for quantitative imaging biomarkers used in clinical trials. This gap arises for reasons that have been described including, among other items, a lack of data for testing and validation, a lack of rigor in the experimental design, inconsistent algorithm implementation, incomplete reporting, and a lack of appreciation for the requirements for adoption of quantitative imaging biomarkers. Addressing the lack of knowledge about the a priori distributions of random effects in imaging scenarios that should be evaluated, providing a rigorous methodology for evaluation, and ensuring pathways for adoption for all stakeholders can overcome these barriers.

To do so we propose to build a measurement error model by using virtual imaging clinical trials (VICTs) [1] to simulate the entire data pathway from patient models through image generation to QIBs. As a first step in this process, virtual imaging clinical trials (VICTs) are an emerging methodological adjunct to clinical trials using imaging. A VICT is essentially an extension to a clinical trial simulation in that the population of human subjects is replaced with a population of virtual digital subjects; imaging systems are replaced with physics-based virtual imaging simulators; and clinical interpretations are replaced with AI-derived image analyses. A VICT offers a feasible and efficient means to conduct experimentation in medical imaging by providing the practical ability to systematically assess and optimize a host of trial design factors and imaging parameters in the development and evaluation of imaging technologies, a task not possible through diagnostic clinical trials. While time, cost efficiency, and ethical feasibility are the main advantages of VICTs, VICTs offer one additional attribute; ground truth can be perfectly known and precisely controlled. As the condition of

the patient is defined a priori, a VICT makes it possible to ascertain how an image analysis metric represents the ground truth. This is a unique capability that can never be assured in clinical trials. Of course, a VICT cannot predict the impact of a novel therapy on a type of disease in a specific patient. However, VICTs can predict the range of outcomes to be expected for a pre-determined (i.e. plausible) domain of known variables, e.g. baseline tumor size and the subsequent shrinkage due to a postulated therapy. Over the last several years there has been a steady improvement of the realism of human and imaging system models. The growing maturation of VICTs as useful tools is demonstrated by multiple publications in mammography, CT, and PET, and even FDA approvals based on VICT studies of some aspects of image technology.

Our goal is to use VICTs to characterize the accuracy of QIBs using standard-of-care CT in oncology trials. From this we can develop a guide for implementation in clinical trials and also a roadmap for adoption by regulatory bodies, industry, oncologists, cooperative oncology groups and professional societies.

II. METHODS

The virtual imaging pipeline component is the computational core, which uses the XCAT patient model [2] as an input to the CT-simulator CatSim [3]. The sources of variability can be grouped into categories along the pathway of the virtual imaging pipeline: (1) patient variability, (2) tumor characteristics, (3) CT acquisition, (4) image reconstruction, and (5) the QIB algorithm.

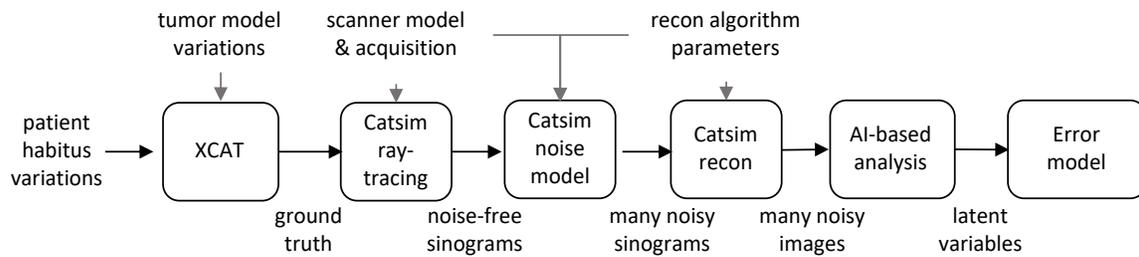


Fig. 1. Data flow in the virtual imaging pipeline.

Data available from the VELOUR clinical trial (NCT00561470) [6], one of the Vol-PACT cohorts [4], are used to define probability density functions. Some of the distributions of scanner- patient-, disease-, and imaging-related sources of variability are shown in Figs. 2 and 3.

Aquilion	Brilliance 6	Emotion 16	LightSpeed	Mx8000 Dual	Sensation 64
Aquilion ONE	Brilliance 64	Emotion 16 (2007)	LightSpeed Plus	Mx8000 IDT 10	Sensation Cardia
Asteion	Brilliance Big Bore	Emotion 6	LightSpeed Pro 16	Mx8000 IDT 16	Sensation Cardiac
AVE1	Brilliance iCT	Emotion Duo	LightSpeed Pro 32	MxTwin	Sensation Cardiac 64
Balance	Brilliance16	Esprit+	LightSpeed QX/i	null	Sensation Open
Biograph 40	CT/e	Gemini TF	LightSpeed RT	Philips CT Aura	Somaris/5 3D
Biograph 6	CT/e Dual	GEMINI TF TOF 64	LightSpeed RT16	PQ2000	SOMATOM Definition Flash
Biograph 64	Definition	GeminiGXL 16	LightSpeed Ultra	ProSpeed	SOMATOM PLUS 4
Biograph40	Definition AS	HiSpeed	LightSpeed VCT	SCT-7000TS	Spirit
BrightSpeed	Definition AS+	HiSpeed CT/i	LightSpeed Xtra	SCT-7800TX	Volume Access
BrightSpeed S	Discovery CT750 HD	HiSpeed Dual	LightSpeed16	Sensation 10	Volume Zoom
Brilliance 16	Discovery ST	HiSpeed NX/i	Manufacturer Model	Sensation 16	Xpress/GX
Brilliance 16P	Discovery STE	HiSpeed QX/i	MX 6000 Dual	Sensation 4	
Brilliance 40	Emotion	iCT 256	Mx8000	Sensation 40	

Fig. 2. List of the CT scanner models used in the VELOUR trial as recorded in the DICOM image headers

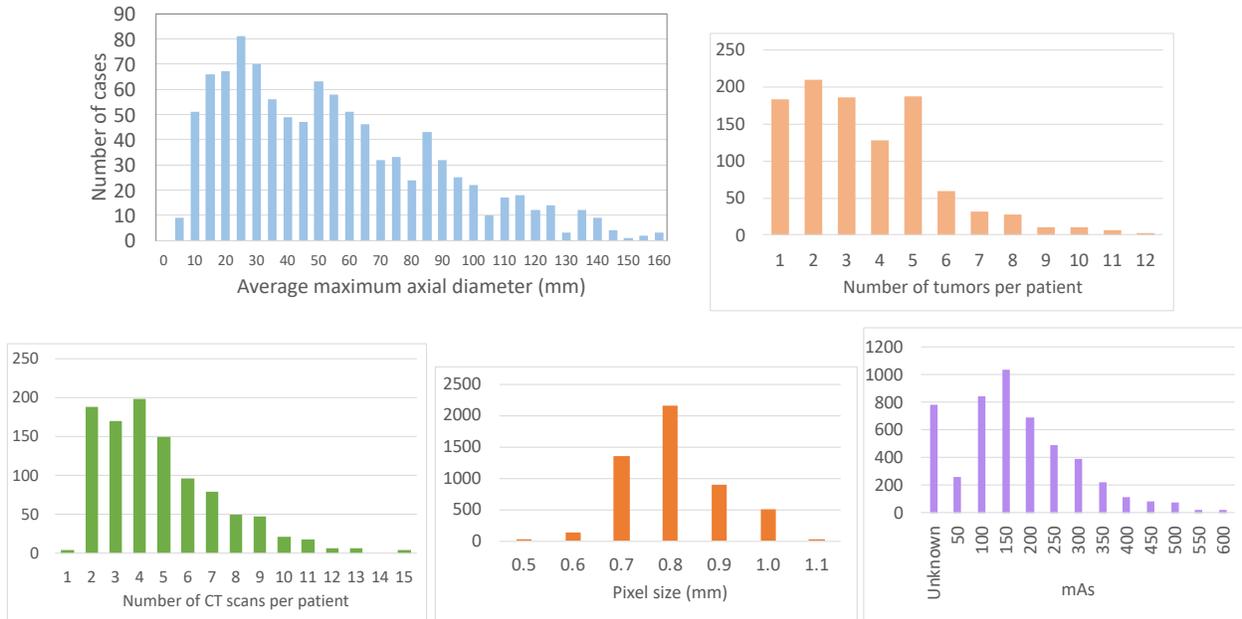


Fig. 3: Some of the sources of variability for the multicenter VELOUR trial data [6]. Shown are the average baseline tumor diameter, the number of tumors per patient at baseline, the number of standard-of-care CT scans per patient, the reconstruction pixel size and mAs per scan.

We used a VICT based on a two-arm trial (control and treatment) as shown in Fig. 4. that uses a baseline and follow-up scan to determine reduction in average tumor volume.

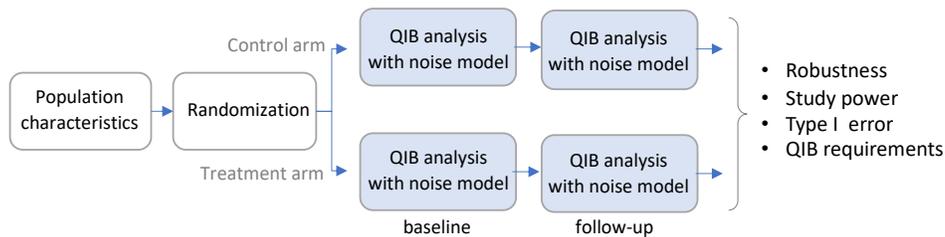


Fig. 4 Virtual imaging clinical trial (VICT) with multi-center baseline and follow-up CT scans. The impact of variability of AI-derived quantitative imaging biomarkers (AI-QIBs) on study power as a function of patient numbers, effect size, and measurement type is assessed.

For a range of effect sizes and trial sizes, we computed study power as a function of QIB variability. The error model used a generalized linear approach for bias and variance of a QIB. In this case we used prior tumor volume estimates (12.5% CoV, but over 25% has been reported). There were 1,000 simulations for each parameter combination to evaluate the QIB in terms of standard error, Type I error, and Type II error (i.e. 1 - study power).

III. RESULTS

Simulated data to be used as plausible ground truth was generated using correlated log-normal distributions modeled on the measured data (Fig. 5). Goodness of fit was checked with Q-Q plots and other statistical tests.

Initial results of study power (Fig. 6) demonstrate the impact of QIB variance in clinical trials using multicenter standard-of-care CT imaging, which features heterogeneity in imaging systems across sites.

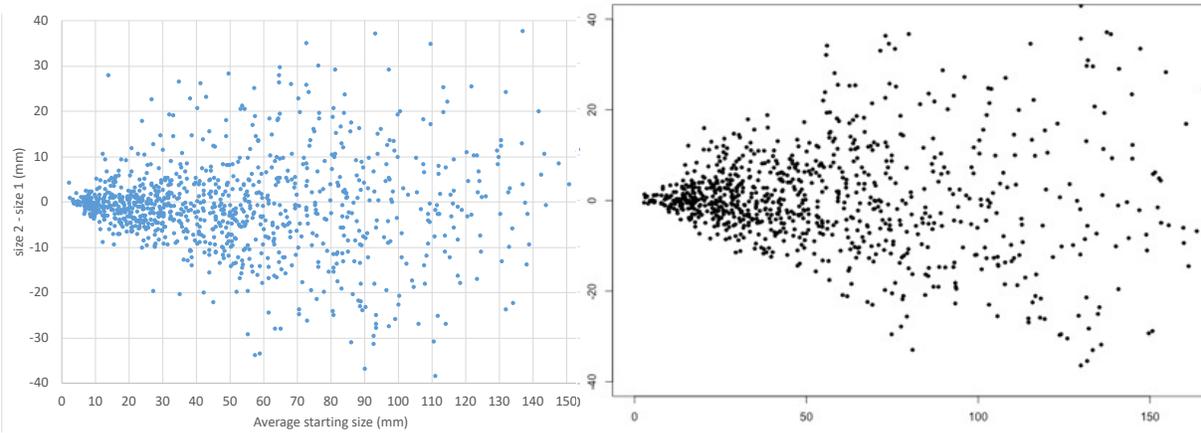


Fig. 5. Tumor size difference (i.e. follow-up - baseline) as a function of baseline tumor size. Left: Measured data from the VELOUR trial with 1,043 patients. Right: Simulated results from multivariate log-normal distributions displayed using the same scales.

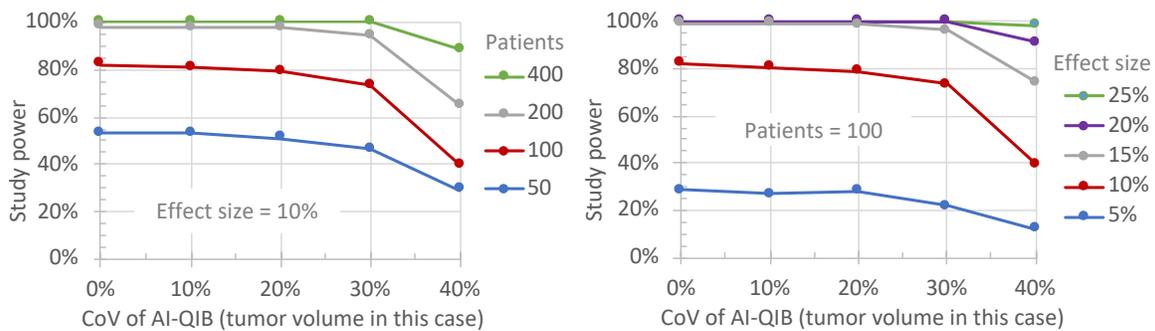


Fig. 6. Study power for the clinical trial illustrated in Fig. 4 showing the importance of understanding the variability of the QIB as a function of sample size and the true difference between arms. Top: impact of QIB coefficient of variation (CoV) and study size. Bottom: Importance of QIB CoV for small studies, i.e. targeted and/or expensive therapies. Data for 100 patients and effect size = 10% is common to both plots, showing the importance of controlling the CoV for a typical study power of 80%.

IV. DISCUSSION

Reliable smaller- n studies are imperative for clinical trials that are smaller, more fractionated, and use more expensive therapies. Understanding the application of QIBs to reduce the number of patients, while retaining study power (and knowledge of the expected study power) is important for these trials to be successful in the advancement of more effective therapies. These methods are based on data from prior clinical trials, and in turn will provide feedback on the robustness of more effective QIBs and guidance for their use in clinical trials.

REFERENCES

- [1] E. Abadi, W. P. Segars, B. M. W. Tsui, P. E. Kinahan, N. Bottenus, A. F. Frangi, A. Maidment, J. Lo, and E. Samei, "Virtual clinical trials in medical imaging: a review.," *J. Med. Imag.*, vol.7 2020.
- [2] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. W. Tsui, "4D XCAT phantom for multimodality imaging research.," *Medical Physics*, vol. 37, no. 9, pp. 4902–4915, Sep. 2010.
- [3] B. De Man, S. Basu, N. Chandra, B. Dunham, P. Edic, M. Iatrou, et al., "CatSim: a new computer assisted tomography simulation environment," *Proceedings of SPIE*, vol. 6510, p. 65102G, 2007.

- [4] L. Dercle, D. E. Connors, Y. Tang, et al., “Vol-PACT: A Foundation for the NIH Public-Private Partnership That Supports Sharing of Clinical Trial Data for the Development of Improved Imaging Biomarkers in Oncology.,” *JCO Clinical Cancer Informatics*, vol. 2, pp. 1–12, Dec. 2018.
- [5] H. Yang, L. H. Schwartz, and B. Zhao, “A Response Assessment Platform for Development and Validation of Imaging Biomarkers in Oncology.,” *Tomography*, vol. 2, pp. 406–410, Dec. 2016.
- [6] J. Tabernero, E. Van Cutsem, R. Lakomý, J. Prausová, P. Ruff, G. A. van Hazel, V. M. Moiseyenko, D. R. Ferry, J. J. McKendrick, K. Soussan-Lazard, S. Chevalier, and C. J. Allegra, “Aflibercept versus placebo in combination with fluorouracil, leucovorin and irinotecan in the treatment of previously treated metastatic colorectal cancer: prespecified subgroup analyses from the VELOUR trial.,” *Eur. J. Cancer*, vol. 50, no. 2, pp. 320–331, Jan. 2014.