

## Abstract

The use of functional imaging with positron emission tomography (PET) has increased in clinical oncology to assess response to therapy. Response assessment with PET scans is largely interpreted *qualitatively*, which results in subjective clinical evaluation. Alternatively, *quantitative* imaging can enable objective evaluation; however, the path to establish standardized response criteria of candidate quantitative imaging biomarkers (QIBs) is extremely challenging. This dissertation focused on establishing quantitative  $^{18}\text{F}$ -NaF PET-based treatment response assessment. Using  $^{18}\text{F}$ -NaF PET/CT scans of bone tumors (osseous lesions) in metastatic castration-resistant prostate cancer patients imaged in a multicenter clinical trial, we characterized test-retest repeatability of standardized uptake values (SUVs) measured from both lesions and patients, and reproducibility across imaging sites. From these studies we derived the limits of agreement, which can be interpreted as objective response criteria. To assess the generalizability of response criteria, we investigated sources of variability that may influence response assessment. Linear mixed effects models identified both differences in injected dose between scans and anatomical location of the lesion may influence repeatability. To address the need to mitigate potential variability in longitudinal imaging, we evaluated the utility of reference region normalization but found that SUVs were similarly robust without. In order to advance criteria for QIBs of response, we introduced the response-to-repeatability metric and discovered that not all candidate QIBs were able to discern statistically measurable changes at treatment follow-up. In our last study, we introduced a bootstrapping method to estimate sample size requirements needed to achieve a desired level of repeatability, a critical component in clinical trial design. Finally, we outlined statistical limitations to the generalizability of response criteria, which will guide appropriate implementation of imaging-based response assessment in the clinical routine. In summary, we present a statistical basis to enable quantitative imaging-based response criteria and methods to iteratively advance needs in both research and the clinic.