Co-Clinical Imaging Resource Program (CIRP): Bridging the translational divide to advance precision medicine

Kooresh I. Shoghi  
*Washington University School of Medicine in St. Louis*

Richard Laforest  
*Washington University School of Medicine in St. Louis*

Daniel S. Marcus  
*Washington University School of Medicine in St. Louis*

et al

Follow this and additional works at: [https://digitalcommons.wustl.edu/open_access_pubs](https://digitalcommons.wustl.edu/open_access_pubs)

**Recommended Citation**

[https://digitalcommons.wustl.edu/open_access_pubs/9393](https://digitalcommons.wustl.edu/open_access_pubs/9393)

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.
Co-Clinical Imaging Resource Program (CIRP): Bridging the Translational Divide to Advance Precision Medicine

Kooresh I. Shoghi1, Cristian T. Badea2, Stephanie J. Blocker2, Thomas L. Chenevert3, Richard Laforest1, Michael T. Lewis4, Gary D. Luker3, H. Charles Manning5, Daniel S. Marcus1, Yvonne M. Mowery6, Stephen Pickup7,8, Ann Richmond9, Brian D. Ross5, Anna E. Vigeloe10, Thomas E. Yankeelov11,12, and Rong Zhou7,8

1Department of Radiology, Washington University School of Medicine, St. Louis, MO; 2Department of Radiology, Center for In Vivo Microscopy, Duke University Medical Center, Durham, NC; 3Department of Radiology, University of Michigan, Ann Arbor, MI; 4Dan L. Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX; 5Vanderbilt Center for Molecular Probes—Institute of Imaging Science, Vanderbilt University Medical Center, Nashville, TN; 6Department of Radiation Oncology, Duke University Medical Center, Durham, NC; 7Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 8Department of Pharmacology, Vanderbilt School of Medicine, Nashville, TN; 9Department of Pathology, The Ohio State University, Columbus, OH; 11Departments of Biomedical Engineering, Diagnostic Medicine, and Oncology, Oden Institute for Computational Engineering and Sciences, Austin, TX; and 12Livestrong Cancer Institutes, Dell Medical School, The University of Texas at Austin, Austin, TX

Corresponding Author:
Kooresh I. Shoghi, PhD
Department of Radiology and Biomedical Engineering, Mallinckrodt Institute of Radiology, St. Louis, MO 63110, USA; E-mail: shoghik@wustl.edu

Key Words: co-clinical trial, preclinical PET, MR, CT, quantitative imaging, informatics, precision medicine, patient-derived tumor xenograft (PDX), genetically engineered mouse model (GEMM), cell transplant model (CTM)

Abbreviations: Co-Clinical Imaging Research Resource Program (CIRP), genetically engineered mouse models (GEMMs), cell transplant models (CTMs), patient-derived tumor xenografts (PDXs), quality assurance (FLP, flippase; QA), steering committee (SC), working group (WG), hematopoietic stem cells (HSCs), quantitative imaging (QI), National Cancer Institute (NCI), magnetic resonance imaging (MRI), computed tomography (CT), positron emission tomography (PET), American College of Radiology (ACR), field of view (FOV), Bland–Altman analysis (BA)

ABSTRACT

The National Institutes of Health’s (National Cancer Institute) precision medicine initiative emphasizes the biological and molecular bases for cancer prevention and treatment. Importantly, it addresses the need for consistency in preclinical and clinical research. To overcome the translational gap in cancer treatment and prevention, the cancer research community has been transitioning toward using animal models that more faithfully recapitulate human tumor biology. There is a growing need to develop best practices in translational research, including imaging research, to better inform therapeutic choices and decision-making. Therefore, the National Cancer Institute has recently launched the Co-Clinical Imaging Research Resource Program (CIRP). Its overarching mission is to advance the practice of precision medicine by establishing consensus-based best practices for co-clinical imaging research by developing optimized state-of-the-art translational quantitative imaging methodologies to enable disease detection, risk stratification, and assessment/prediction of response to therapy. In this communication, we discuss our involvement in the CIRP, detailing key considerations including animal model selection, co-clinical study design, need for standardization of co-clinical instruments, and harmonization of preclinical and clinical quantitative imaging pipelines. An underlying emphasis in the program is to develop best practices toward reproducible, repeatable, and precise quantitative imaging biomarkers for use in translational cancer imaging and therapy. We will conclude with our thoughts on informatics needs to enable collaborative and open science research to advance precision medicine.

BACKGROUND

Co-clinical trials are an emerging area of investigation in which a clinical trial is coupled with a preclinical study to inform the corresponding clinical trial (1–7). The preclinical arm of the co-clinical trial generally uses genetically engineered mouse models (GEMMs), cell transplant models (CTMs) of human cancers or patient-derived tumor xenografts (PDXs) to aid in therapeutic efficacy assessment, patient stratification, and optimal treatment strategies designing (8, 9). The emergence of GEMMs, CTMs, and PDXs as co-clinical platforms is largely motivated by the realization that established cell lines do not recapitulate the heterogeneity of human tumors and the diversity of tumor phenotypes.
designing co-clinical imaging trials in terms of selection of animal models, considerations in designing co-clinical imaging studies, standardization of instruments, and harmonization of preclinical and clinical quantitative imaging pipelines. An underlying emphasis is to develop best practices toward reproducible, repeatable, and precise quantitative imaging biomarkers for use in translational cancer imaging and therapy. We will conclude with informatics needs to enable collaborative and open science research to advance precision medicine.

ANIMAL MODELS AND CO-CLINICAL TRIALS
NCI's precision medicine initiative emphasizes the use of translational oncology models to address the biological and molecular bases of cancer prevention and treatment. As noted above, translational oncology models considered in this context include (but are not limited to) PDXs, GEMMs, and CTMs. The advantages and disadvantages of these models are summarized in Table 1. Rapid disease progression presents a limitation for essentially all mouse models of cancer used in co-clinical trials. However, co-clinical animal models offer the opportunity for streamlined assessment of tumor sensitivity to drugs being tested in the human clinical study, as well as to evaluate mechanisms of treatment resistance and evaluate novel drug combinations (Figure 2). Initial outcomes in the human clinical trial can influence treatment strategies used in the mouse model and indicate whether new models are needed for more faithful recapitulation of the human disease process. Similarly, findings in the mouse trial can inform selection of patients most likely to benefit from an intervention, provide guidance for optimal imaging and biospecimen collection for correlative studies, and lead to adjustments in therapeutic approach. In an ideal co-clinical paradigm, the mouse trials will enable rapid transfer of information from mouse experiments to human trials to provide information to optimize treatment regimens with a focus on ultimately improving clinical care and patient outcomes (9).

Patient-Derived Tumor Xenografts
To date, PDX models perhaps come closest to addressing the co-clinical, paradigm (11). The key advantages of PDXs include the following:

- **General Advantages**:
  - Closest mimicry of human tumors and tumor microenvironment.
  - Ability to study tumor progression and response to therapy over long periods.
  - Cost-effective compared to other models.

- **Specific Considerations**:
  - Limited to human tumors and not always representative of all tumor types
  - Technical challenges in long-term serial passage
  - Ethical concerns associated with patient-derived material

---

**Figure 1.** Organizational structure of Co-Clinical Imaging Resource Program (CIRP). Members of the steering committee (SC) comprise investigators from within U24 awardees. The SC oversees the activities of 3 working groups (WGs) (animal models and co-clinical trials, image-acquisition and data processing [IADP], and informatics and outreach [IMOR]) focused on practical aspects of co-clinical imaging. Individuals not directly funded by the U24 mechanism may petition to join the CIRP network as associate members. Associate members are then affiliated with one or more of the WGs.
(1) ability to accurately reflect patients’ tumors in terms of the histomorphology, gene expression profiles, and gene copy number alterations (12–16); and (2) ability to predict therapeutic response in patients, especially when a clinically relevant drug dosage is used (17–19).

In this model, a sample of viable tumor tissue obtained from surgical resection of the tumor—or a solid (20) or liquid biopsy (21)—is implanted onto immune compromised mice, either subcutaneously or orthotopically (11, 22–24). While not depicted in Figure 2, established PDXs can be used as a renewable source of tumor cells for generation of patient-derived organoids (PDOs)
(25–27). There is also an interesting variation of the in vivo PDX method, mini-PDX, where patient-derived tumor cells are seeded into hollow fiber capsules that are implanted subcutaneously into mice. The mini-PDX-bearing mice are treated with therapeutic agents that elicit immune system infiltration and tumor regression. PDX models have been used to test the efficacy of immunotherapies, such as checkpoint inhibitors, in a preclinical setting. For example, mice bearing a PDX model of prostate cancer were treated with anti-PD1 antibodies, and the tumor regression was monitored over time.

Once the PDX model has been developed, details of the mouse strain used, engraftment procedure, rate of engraftment, tumor preparation before injection, passage number, and injection site should be recorded. PDX tumors need to be carefully validated for quality assurance (QA) based upon histology, special stains, and short tandem repeat to ensure authenticity of PDX samples. Genetic analysis should be performed to validate any genetic drift. In particular, evaluation of next-generation sequencing data generated from PDX samples requires special considerations. It is important to perform RNA sequence analysis of the primary tumor and to compare that analysis to the RNA sequence analysis of the PDX, as well as to accurately distinguish between sequencing reads that originate from the host versus those arising from the xenograft itself. Failure to correctly identify contaminating host reads can lead to incorrect mutation and expression calls (30). It is also important to note that PDX tumors do undergo some evolution as they are forced to adapt to grow in a mouse host. As a result of these mouse-host-induced changes, PDX tumors can diverge from the primary patient’s tumor from which they were derived. This genetic drift is especially evident by distinct copy number alterations in PDXs that accumulate with each PDX passage (31). Thus, it is important to perform experiments with low-passage PDXs to ensure faithful representation of the primary tumor genome.

Because PDX tumors use immune-compromised mice, one main disadvantage is the lack of an immune component to test the contribution of the immune response in therapeutic studies. To address this critical need in studies of immune-oncology, numerous mouse models have been implanted with human CD34+ cells to reconstitute the mouse with a “human immune system”; however, there are weaknesses for each model as detailed elsewhere [NSG-SGM3 (32, 33), NSG-β2m (34), MISTR (35), and NOG-EXL (36, 37)]. Despite the nuances mentioned above, PDXs hold great value for testing novel therapeutic regimens, as well as for studying mechanisms of therapeutic response and resistance in a variety of hematological and solid tumors (11, 38, 39).

**Genetically Engineered Mouse Models of Cancer**

Genetically engineered mouse models (GEMMs) typically target delivery or expression of a recombinase to trigger genetic recombination events that lead to spatially and temporally restricted tumorigenesis. The Cre-lox and Flp-frt are the most commonly used systems. Cre is a site-specific recombinase that deletes DNA flanked by loxP sites [i.e., “floxed alleles”, FL (40)] and similarly flippase (FLP) recognizes FLP recombinase target sequences [i.e., “frted alleles”, FRT] to facilitate targeted mutations (41). Numerous tissue-specific Cre drivers are available to localize mutations in particular tissues, and inducible Cre strains offer an additional layer of temporal control. For example, CreER(T2) recombinase is a tamoxifen-dependent Cre recombinase that can be activated by systemic administration of tamoxifen or localized administration of 4-hydroxytamoxifen (42). Alternatively, viruses can be used to deliver recombinase to a particular site to induce recombination of the mutant alleles. Multiple genes can be altered simultaneously in the Cre-lox system, with oncogene expression triggered by deletion of a floxed stop cassette (loxP-STOP-loxP, “LSL”) preceding an oncogene (e.g. LSL-KrasG12D) and tumor suppressor knockout from deletion of floxed exons (43). GEMMs offer several advantages, including autologous and gradual disease development in the presence of an intact immune system (44) recapitulating the inter- and intratumor heterogeneity and histopathological features of the human tumor and microenvironment (9, 45). However, disadvantages of GEMMs include their high cost, relatively long time to tumor onset, and use of genetic alterations that may not exactly mimic the heterogeneity of the individual patient’s disease. These models also lack the genetic heterogeneity typified by most human tumors (38). Furthermore, relatively large treatment groups are typically needed for GEMM experiments owing to the degree of variability among tumors. GEMMs have shown utility in co-clinical trials of immunotherapy of pancreatic cancer (46) and non-small cell lung cancer (47). However, most GEMMs have low mutational load, which can pose a challenge for immunotherapy studies that require tumors to express neoantigens to engender an immune response. This challenge has recently been overcome by combining exposure to the carcinogen 3-methylcholanthrene with Cre-mediated p53 knockout in p53R172H mice to generate a relatively high mutational load soft tissue sarcoma (48).

**Cell Transplant Models of Cancer**

Cell transplant models (CTMs) represent a distinct subset of GEMMs applied primarily to studies of hematologic cancers, such as myeloproliferative neoplasms and leukemia that arise from mutations arising in hematopoietic stem cells (HSCs) and progenitor cells. To generate a CTM model of cancer, investigators isolate bone marrow from a donor animal and transduce enriched HSCs or the total population of bone marrow cells with recombinant retroviruses or lentiviruses expressing critical oncogenes for a target disease (49). These types of viral vectors integrate into the genome of cells, ensuring stable transmission of key oncogenic mutations from HSCs to more differentiated hematopoietic cells (Figure 3). Retroviral and lentiviral vectors for CTMs commonly include a coexpressed fluorescent protein or other reporter molecules to facilitate detection of transduced cells ex vivo and in recipient mice. After conditioning with whole-
Body irradiation or high-dose chemotherapy to ablate endogenous HSCs, investigators transplant transduced HSCs intravenously into recipient mice which hone to bone marrow niches through signaling pathways analogous to stem cell transplants in humans. The hematopoietic system is subsequently reconstituted with malignant progenitor and differentiated cell lineages over several weeks. Recipient mice progressively develop features of disease that recapitulate key pathologies evident in patients, including increased cellularity in bone marrow (hypercellular marrow), splenomegaly, hepatomegaly, and/or inflammatory constitutional symptoms.

CTMs typically use syngeneic (genetically identical), immunocompetent murine models for both donors and recipients of HSC transplants, although studies have successfully transduced human HSCs with a driver oncogene and established xenograft models of hematologic cancer in immunocompromised mice. Using syngeneic mice avoids complications secondary to mismatch of donor and recipient, including graft-versus-host and graft rejection. CTMs offer several advantages in the context of co-clinical trials:

1. Generating cohorts of mice that match frequencies of driver mutations present in patients receiving the same treatment;
2. Studying disease progression and response to therapy in mice with a full range of hematopoietic cells (syngeneic all murine models);
3. Faithfully maintaining underlying genetic basis of disease present in patients; and
4. Rapidly producing large numbers of mice for treatment studies, which minimizes delays in assessing effects of therapy in studies with sufficient animals for high statistical power and rigorous validation (eg, histology).

Limitations of the model include potential adverse effects of myeloablative conditioning regimens used to facilitate engraftment of transplanted HSCs in recipient animals and accelerated course of disease relative to patients.

CO-CLINICAL IMAGING STUDY DESIGN, INSTRUMENTS, AND STANDARDIZATION

Standardization of clinical quantitative imaging (QI) has been realized to a large extent by numerous initiatives, such as the Quantitative Biomarker Alliance and NCI’s Quantitative Imaging Network to implement advanced QI methods in clinical practice (50–53). While these and other initiatives have had a great impact in advancing clinical applications of QI, preclinical imaging remains a critical component in the translational pipeline of validating QI methods and imaging agents for applications in drug discovery, cancer detection, and response to therapy assessment. An inherent challenge in preclinical imaging is lack of standardization in terms of study design and animal logistics, image acquisition, and analysis and instrument quality control/assurance. While generally applicable to other modalities, in the ensuing sections, these considerations will be discussed in relation to co-clinical magnetic resonance imaging (MRI), computed tomography (CT), and positron emission tomography (PET) in the context of co-clinical trials.

Co-clinical Imaging Study Design and Animal Logistics

Small animals are noncompliant subjects and as such the vast majority of small-animal imaging studies use general anesthesia, typically isoflurane in oxygen. As mice are not able to maintain their core body temperature while under general anesthesia, it is necessary to use vital signs’ monitoring in combination with active maintenance of core body temperature for preclinical studies. Many quantitative imaging parameters are temperature-dependent; therefore, it is critical that the animal achieves a stable core body temperature and physiological state before initiation of any quantitative imaging studies (54, 55). Numerous other factors involved in the setup for preclinical imaging have been documented to impact imaging parameters, including anesthesia, animal handling, and diet (duration of fasting), among other factors (54–61). Parameters related to animal husbandry, including housing conditions, acclimation, chow, strain of animal, and physiological stress, may also impact the outcome of
imaging studies (56). Imaging studies typically performed during day times, disrupt the animal’s circadian rhythms which modifies disease metabolism in some cases (62, 63). An important consideration in multicenter preclinical trials is institutional variability in housing. A recent analysis of data derived from Mouse Metabolic Phenotyping Centers suggests that the location (and corresponding institution) at which a study was performed contributed to differences in energy expenditure in rodents, even when the same diet was used across institutions (64). Thus, institutional differences in animal housing may also impact preclinical imaging studies. To enhance the reproducibility and the translational impact of preclinical imaging studies, these factors need to be considered and recorded to facilitate interpretation of co-clinical trials.

MRI

MRI provides superb soft tissue contrast that can be manipulated by suitable adjustment of the acquisition parameters. Image contrast can be sensitized to several properties of tissue water including nuclear relaxation rate, water diffusion, blood flow, or perfusion and chemical exchange. These properties can be quantitatively measured by magnetic resonance, resulting in numerous biometric markers of disease state (65). Typically, an MRI examination includes generation of a series of images with different contrast weighting, thereby facilitating multiparametric analysis (66–68) and improved specificity relative to single-parameter imaging techniques. MRI may also be combined with injectable contrast agents to provide additional information. The reader is referred to numerous treaties describing MRI methodologies for additional details (69–71). Although a majority of MRI techniques available clinically may be applied in preclinical studies, there are a number of significant differences between the available tools and methodologies that pose challenges to the development of co-clinical studies (65).

Motion Control in Co-clinical MRI. In the clinic, physiological motion must be addressed when imaging certain regions of the anatomy. Various techniques, including breath-holding, parallel acquisition (72), and fast acquisition methods such as echo planar imaging (73, 74) and fast imaging with steady-state precession (75), are often used alone or in combination to effectively freeze the motion. Lightly anesthetized mice have physiological motion rates that are an order of magnitude greater than those of humans (heart rates in the range of 400–600 bpm; respiratory rates of 60–90 breaths/min). The methods used in the clinic to address physiological motion are either not available or not fast enough to suppress motion artifacts in small-animal studies. It is therefore necessary to use other means of motion correction in preclinical studies of the abdomen or thorax. Prospective gating (ECG and/or respiration) (76–78), retrospective gating (79, 80), navigator echoes (81–83), and “self-gated” methods (84–86) are commonly used to minimize motion artifacts in preclinical studies. These methods are often used in combination with sampling schemes that have reduced sensitivity to motion including radial (87) and spiral (88, 89) k-space trajectories.

Differences in Preclinical and Clinical MR Instrumentation. There currently is a trend among preclinical instrument vendors to produce static fields ($B_0$) that mimic those used in the clinic (1.5, 3, and 7 T). However, the $B_0$ used in preclinical systems varies over a broader range (1–21 T) and is typically higher than those used in the clinical setting. The push to higher fields is driven by the fact that the MRI signal scales somewhere between linearly and the square of the $B_0$ (70). The additional signal strength is necessary to offset the limited signal-to-noise ratio owing to the smaller voxel size required for preclinical imaging. However, the most commonly used MRI contrast parameters (relaxation times) are known to be $B_0$-dependent. Both clinical and preclinical protocols must therefore be optimized for the target field strength. This poses many challenges when developing co-clinical trials that use significantly different field strengths. The higher $B_0$ used in preclinical studies is also problematic for methods that are sensitive to magnetic susceptibility effects, as these effects also scale with $B_0$. Gradient echo and echo planar imaging techniques often suffer geometric distortions and/or signal dropout owing to magnetic susceptibility effects. These effects make single-shot techniques extremely difficult at field strengths commonly used for preclinical studies.

The gradient system for any MRI scanner provides a magnetic field gradient that is within a few percent of linear over a prescribed spherical volume at the isocenter of the magnet. Nonlinearities in the gradient system outside of this volume lead to reproducible geometric distortions in images acquired in this region of space. On clinical scanners, the gradient nonlinearities are fully characterized during installation and suitable nonlinear deformations are applied to acquired data in order to correct for these effects, including deformation corrections for apparent diffusion coefficient maps (90). These corrections are generally not available on preclinical scanners. It is therefore critical to assure that the volume of interest is limited to the linear volume of the gradient system when performing volumetric studies on preclinical instrumentation. Alternatively, the nonlinearities could be characterized and corrected as is done in the clinical case.

Calibration and QA. The American College of Radiology (ACR) provides detailed requirements for accreditation of clinical sites (www.acraccreditation.org/modalities/). This includes detailed protocols for instrumentation calibration, QA tests and preventative maintenance. The ACR also routinely performs site visits to assure that facilities adhere to recommended best practices. No such accreditation exists for preclinical sites. Many of the best practices provided by the ACR could be extrapolated to preclinical sites. However, this is thwarted to some degree by the lack of standard phantoms to be used of QA. Tests. Though numerous phantoms have been described in the literature (91–93), there is a lack of consensus among preclinical MRI sites as to the optimal phantoms to be used for QA.

CT

CT is a fast and powerful anatomical x-ray-based imaging method in which contrast is based on x-ray attenuation which is dependent on the x-ray energy and the composition of the subject. A clinical CT system includes an x-ray tube and detector assembled in a gantry which rotates around the subject to create hundreds of radiographic projections. The projections are used by a reconstruction algorithm to produce (3D) tomographic data. The most used reconstruction algorithm for cone beam CT is filtered backprojection. The contrast is based on x-ray attenuation which is dependent on the x-ray energy and the composition of
the subject. CT is the modality of choice for bone and lung imaging. If used with contrast agents, CT can provide perfusion information and/or cardiac function. One of the major limitations of x-ray CT imaging is exposure to radiation. More information on the basics of CT imaging is provided elsewhere (69–71, 94).

**Differences in Preclinical and Clinical CT Instrumentation.** Preclinical CT also known as micro-CT is a volumetric imaging method based upon the same principles and components as clinical CT scanners, but delivering much higher spatial resolution (95). There are 2 possible system design geometries in micro-CT imaging:

(1) rotating gantry (tube and detector), and
(2) rotating specimen.

All of the current commercial systems for in vivo scanning use the rotating gantry geometry, that is, they are scaled versions of the clinical CT scanners. The micro-CT scanners with a rotating specimen geometry are mostly used for ex vivo imaging.

The higher resolution achievable with micro-CT is linked to the use of microfocus x-ray tubes with small focal spots (eg, 10 μm for 8 W power) and x-ray detectors with small voxel size (eg, 20 μm). The smaller voxels of micro-CT (relative to clinical CT) require much higher dose, as any voxel—independent of its size—needs to interact with a certain number of x-ray photons for adequate image quality (96, 97). If the photon noise (measured by its variance) is kept constant and the linear dimension of the voxel is reduced by 2 (ie, a reduction of voxel volume by 8), the dose must be increased by up to 16 times (98, 99). Thus, the acquisition protocols should be customized to minimize the x-ray dose but this comes usually as a penalty to image quality. X-ray radiation exposure can lead to biological damage and long-term health effects (100). The LD_{50/30} whole-body radiation dose in mice (the dose required to kill 50% of mice within 30 days) depends on many factors, but tends to be between 5 and 8 Gy (101, 102). The typical radiation dose for a single micro-CT scan can vary widely and reported values in the literature range from 0.017 Gy to 0.78 Gy (101). Rodents have the ability to repair damage from low doses of radiation (up to 0.3 Gy) over the course of several hours (103), so most low-dose micro-CT scans should have limited biological impact, even when the same animals are longitudinally scanned over the course of a study. But for higher-dose scans, longitudinal micro-CT imaging can potentially lead to a cumulative dose that could affect biological function (particularly immune function and tumor response) and long-term health (100). Therefore, careful consideration must be made to determine the optimal imaging protocol for each individual application to minimize the effects of radiation dose on the experiment.

For small-animal micro-CT imaging, the use of clinical contrast agents is particularly difficult. Mice have much higher renal clearance rates than humans, so injected contrast agents are rapidly excreted. To overcome the rapid clearance of traditional contrast agents, micro-CT can benefit from blood pool contrast agents, exhibiting prolonged blood residence time and stable enhancement from minutes to hours. Blood pool agents are made up of a wide variety of high-molecular-weight compounds or nanoparticles that avoid renal clearance owing to their large size. Their use for micro-CT imaging has been reviewed previously (104).

**Motion Control in Co-clinical CT Imaging.** An important aspect of in vivo imaging with micro-CT when imaging the cardio-pulmonary system is related to physiological gating (105). Unlike in clinical chest CT, which is performed in a single breath hold, preclinical projection data in micro-CT must be acquired over many breaths, requiring respiratory gating. However, both respiratory and cardiac motion can lead to artifacts and blurry appearance in reconstructed images. To compensate, gating approaches have been developed to synchronize the projection acquisition with physiological motion, ensuring that all projection images are acquired during the same phase of motion. There are 3 types of gating strategies: prospective, retrospective, and image-based gating (106). Prospective gating is being used in scanners that operate under step-and-shoot mode, in which after the gantry rotates to any new angle, the x-ray tube waits for a trigger signal before acquiring the next projection image. The trigger is based on a signal provided by a pneumatic cushion positioned on the animal’s diaphragm or an optical measurement (107). Typically, only 1 projection is acquired at each projection angle (108). However, it is also possible to acquire multiple projections (called frames) at each angle, and then sum the frames into a single projection (ie, multiframe acquisitions) to improve signal-to-noise ratio. Retrospective (109) and image-based methods of gating have also been for micro-CT (110).

**Calibration and QA.** To ensure that a CT system performs well, it is important to assess the image quality and dose routinely using phantoms (111) especially as quantitative CT image features are widely being investigated in radiomics for tissue phenotype characterization. In the animal imaging space, 1 example of a commercially available micro-CT phantom consists of 6 separate modular sections (resolution coils, slanted edge, geometric accuracy, CT number evaluation, linearity, and uniformity and noise), each designed to evaluate 1 aspect of image quality (112). This phantom has been used for performance evaluation for various scan protocols with micro-CT system (113, 114). Other custom phantoms to assess specific imaging tasks have been also reported, for example, a phantom to assess the the voxel scaling accuracy and it has been tested for a variety of micro-CT scanners covering a range of image resolutions (115). Some anatomically correct simulated phantoms have been introduced (116) (117). Moreover, a 4D digital mouse phantom (MOBY) exists, providing not only anatomical detail but also realistic motion owing to the cardiac and respiratory cycles (118).

**PET**

Over the years, numerous reports have highlighted considerations in small-animal PET imaging (54–56). Unfortunately, despite some progress, there are still gaps in standardization of small-animal imaging protocols and QI methods to produce consistent results as highlighted by recent works (119, 120). Importantly, in line with the theme of this communication, there is a need to harmonize preclinical and clinical PET QI pipelines so as to enhance the translational impact of developments in PET imaging. To harmonize clinical and small-animal PET images, instrumentation and software factors affecting spatial resolution and scanner sensitivity should be considered.
Factors Affecting Spatial Resolution and Sensitivity in PET. PET physics dictates that photon nonacolinearity and positron range negatively contribute to the spatial resolution a system can achieve (121). Photon noncolinearity effect on spatial resolution is proportional to scanner radius and thus will be of importance in human imaging system only. $^{18}$F is the most widely used nuclide in PET imaging. The positron from this nuclide has a maximum energy of 0.63 MeV and its range contributes to a loss of 0.5 mm in spatial resolution (121). This value is negligible compared with other factors in human imaging but has a small impact in small-animal PET imaging. From a camera design standpoint, the crystal size ultimately determines the intrinsic spatial resolution, and the detector technology has constantly evolved over the past few decades with progressively smaller crystals, from 6 mm in the 1990’s to ~3 mm nowadays. Absolute system sensitivity depends of the scanner diameter, axial field of view (FOV), and crystal thickness. Ultimately, the performance of a system is a compromise between sensitivity and resolution, not a simple choice for most applications.

System Design—Clinical vs Preclinical. The current generation of clinical PET/CT scanners are designed for whole-body imaging and thus have a diameter of ~80 cm, allowing for a patient port of ~70 cm, and are composed of a cylindrical configuration of PET detectors. At such radius, high sensitivity is achieved with more crystals, either thicker (2–3 cm) for increased detection efficiency and/or with the use of longer-axis scanner. The current generation of PET scanner will typically use a crystal size of ~4-mm, having a thickness of 20–30 mm and an axial FOV of ~25 cm. Recent camera designs such as the Siemens Biograph Vision pushes this limit with the use of 3.2-mm crystals and 26 cm axial FOV. The absolute sensitivity of clinical PET/CT systems ranges from ~8 to 23 cps/KBq (0.8%–2.3%) with best spatial resolution at ~4–6 mm of full-width half-max at the central portion of the FOV. Small-animal systems, on the other hand, can achieve better resolution, in part, because of the smaller diameter making the photon nonacolinearity a nonfactor, but mostly from the use of smaller crystal size. State-of-the-art systems use ~1-mm crystal or achieve ~1-mm spatial sampling and typically claim to less than ~1-mm spatial resolution with iterative image reconstruction. Because the camera radius can be kept small, mouse sensitivity of 4%–8% is typically achieved. Those values are reported for typically wide energy acceptance windows of 350–650 keV or even 250–750 keV. This acceptance energy window is much wider than that in the clinical setting. In mice, the scatter fraction is small, at least smaller than in human setting. The positron range is an additional remaining factor not to be neglected in small-animal imaging and it is unlikely that further progress can be made in small-animal PET without consideration for positron range even when imaging with $^{18}$F.

Image Reconstruction. Most systems use statistical–iterative image reconstruction and implement the 3D-OSEM (ordered subset expectation-maximization) algorithm (122), which is based on the maximum likelihood (ML–EM) algorithm (123) with a sub-division of the projection views into subsets for accelerated image reconstruction. Commonly all manufacturers will implement point spread function modeling that has the effect of improving spatial resolution and reducing imaging noise. In the clinical arena, point spread function modeling (124) is now commonly available and time of flight image reconstruction is available for most systems. In the latter, the critical parameter is the coincidence timing resolution that most systems achieve in <400 ps (125–127). The time of flight image reconstruction brings the benefit of improved signal to noise. In light of differences in system design, clinical and preclinical systems offer widely different performance levels in terms of spatial resolution and typically, small-animal PET will have a 2- to 4-fold improved system sensitivity over clinical PET systems. However, the imaging scales of the subjects to be imaged are widely different. In terms of resolution-to-scale, clinical systems have a significant advantage over preclinical systems.

Co-clinical Radiotracer Considerations. Factors related to radiotracers that could potentially affect the harmony of molecular imaging co-clinical trials are numerous and worthy of a more in-depth discussion than space allows here. However, it is worth noting a number of issues that dovetail with current trends in tracer development for cancer-specific imaging. First, specific tracer retention mechanisms should ideally be identical, spanning model system to human. Whether tracers are aimed at probing classic ligand–receptor interactions or targeting enzymes that might portend intracellular concentration, a first assumption is that similar mechanisms are in place in both the model system and humans. Parallels between the 2 systems, however, are complex. For example, the biochemical rates that might enzymatically concentrate a radiotracer within cells may be quite different between humans and models and result in different trapping rates. Even more critical, the specific targets of radiotracers might not be expressed to the same level, or even at all, in both species. Because targets of radiotracers may be expressed at different levels (concentrations), of particular consideration are mass effects and specific activity considerations (128) which may impact quantitative imaging measurement in co-clinical trial settings. Moreover, immune-oncology provides the latest examples where species-specific selectivity renders certain radiotracers active in one system but agnostic in others. Described elsewhere in this publication are humanized mouse model systems that aim to address the conundrum of species selectivity. Other factors that must be considered but are not necessarily insurmountable include species-specific physiology and metabolism. However, sophisticated imaging and parallel biochemical analyses can help in normalizing differences across species and scale. Imaging protocols can also be developed that minimize the effect of diet on results, which could be of particular importance with respect to cancer metabolism.

Phantoms for Calibration and QA

Clinical phantoms. The ACR (American College of Radiology) has developed a widely accepted phantom used for quality control and site qualification for clinical trials in PET/CT (129). The phantom consists of specially designed top flange to the widely adopted Jaszczak phantom (Peter Esser flange). The contains 4 fillable cylinders, 25 mm in length, 8, 12, 16, and 25 mm in diameter (hot lesions), in addition to 2 additional fillable 25-mm-diameter cylinders (air and nonradioactive water), and a solid polytetrafluoroethylene (PTFE) cylinder (mimicking bones) are included. A cold-rod section is inserted at the other end of the phantom to provide a means to estimate the scanner spatial resolution through visual inspection. The phantom is typically prepared in a protocol trying
to emulate a clinical FDG PET imaging scenario to define the activity levels in the hot lesions and background area. A ratio of 4:1 of activity concentration in the hot lesions relative to background is typically chosen. The phantom is then imaged as per the site clinical imaging protocol in use for FDG oncology PET/CT patients. At this set activity ratio and typical scan time, and standard image reconstruction algorithm and parameter set for clinical applications, the smallest hot lesion would be typically barely visible. Maximum values in the hot cylinders provide a measurement of count recovery as a function of object size, but these values are used for only site qualification and not for scanner performance. The uniform water section of the phantom allows the measurement of absolute scanner quantification accuracy, in plane uniformity and across plane uniformity (at least for a few centimeters of its axial section). This phantom was designed to allow easy preparation and to allow the measurement of a number of parameters useful to compare imaging performance of a scanner (and the chosen reconstruction) for the clinical task of clinical oncological FDG-PET imaging in the setting of clinical trials.

Other phantoms have been reported to investigate the dependence of PET image bias on CT-based attenuation correction (130). In the clinical setting, the NEMA IEC phantom is used routinely (131). The NEMA IEC phantom is composed of a hollow chamber, is water-fillable, and contains 6 fillable spheres (10–37 mm) in its midsection. Cylindrical insert (50-mm-diameter) filled with a mixture of Styrofoam beads and water is inserted in the phantom to represent lung material. This phantom is typically used for acceptance testing according to NEMA NU-2 standard and EANM/EARL accreditation (132). Typical filling procedure consists of a 10:1 ratio between the sphere and background at total activity commensurate to standard $^{18}$F-FDG oncologic FDG PET/CT applications. This phantom allows measuring size-dependent (contrast)-recovery curve, background variability, absolute scanner calibration, and scattering correction accuracy. More recently the SNMMI-CTN PET phantom was developed to validate scanners at sites that wish to participate in oncology clinical trials (133). The CTN oncology clinical simulator phantom is an anthropomorphic chest phantom with lung fields and 6 spherical objects with inner diameters ranging from 7 to 20 mm reproducibly secured at specific locations within the phantom. This phantom allows the measurement of contrast recovery curves, and their reproducibility, for realistic lesion and operational clinical image reconstruction settings.

**Preclinical phantoms.** For small-animal scanners, the NEMA NU-4 2008 proposes a mouse size image quality phantom that has been used to compare preclinical PET imaging systems (134). The phantom consists of a Lucite cylinder with, at one end, a 5 fillable rod pattern with diameter 1–5 mm for count recovery measurement, and at the other end, 2 small 8-mm-diameter lung inserts to evaluate scatter correction efficiency. A fillable water section in the middle allows the measurement of in-plane uniformity. The design was chosen to allow for a robust and easy-to-fill phantom and was initially designed for the purpose of scanner comparison. The fillable hot rod simplifies construction and avoids the problem of spheres with a cold wall. The hot rod-like lesions were chosen to be of diameter commensurate to organ sizes in mice; however, these are cylindrical in shape, not spherical. The recovery values are to be expected to be larger in rod-like objects relative to sphere-like objects. Importantly, the rod-like lesion sizes are of appropriate size to challenge most small-animal PET imaging systems and thus one can evaluate the quantitative performance for imaging small objects with this phantom for the purpose of comparing animal scanners. Preclinical phantoms play a critical role in harmonizing preclinical instruments across multiple sites (120).

**PRECISION AND ACCURACY IN QUANTITATIVE IMAGING**

There are a variety of considerations that must be made to acquire preclinical imaging data that best serve a study. Although the endpoints and acquisition goals will vary by study, all imaging data sets should be analyzed with a functional understanding of sources of variability and uncertainty in the data. The National Institutes of Standards and Technology cites 3 underlying sources of uncertainty in the clinical study and implementation of quantitative imaging biomarkers (135). The first is variability caused by the devices used to capture images, or instrumentation variance. The second is variability in image interpretation by clinicians/technicians, or reader variance. The third is variability owing to intrinsic properties of the biology, or biological variance. These uncertainties exist in both preclinical and clinical imaging domains, highlighting the need to define the appropriate methods by which data are measured, interpreted, and validated (136). Unfortunately, the deficit in standardized metrics for preclinical imaging and analysis is even greater than that faced by clinical imaging scientists and technicians. Although certain challenges are specific to modality, a general definition for defining and assessing the utility of imaging biomarkers is assessing their reproducibility and repeatability.

Numerous methods are used to assess reproducibility of image metrics including Lin’s concordance correlation coefficient (137) and Bland–Altman analysis (BA) (138). The Lin’s concordance correlation coefficient, is the product of the Pearson correlation coefficient and the bias correction factor and accounts for both precision and accuracy. The method outlined by Watson and Petrie (139) is typically used to calculate these metrics. The procedure used to calculate the statistical parameters for the BA plots are summarized by Galbraith (140) and Raunig (141). To assess reproducibility between image metrics derived from consecutive days, it is important to test that the “day 1” vs “day 2” absolute differences are independent of the means using Kendall tau test for correlation (140). Let $\Delta$ denote the within-mouse difference between the measurements, and $N$ denote the number of paired measurements. The standard deviation for the mean difference is calculated using the following equation:

$$\text{dsd} = \sqrt{\frac{\sum (\Delta_i)^2}{N}},$$

and the within-subject standard deviation (wSD), using the following equation:

$$wSD = \frac{\text{dsd}}{\sqrt{2}}.$$

The 95% confidence limits in the BA plots are the limits of agreement defined as the mean difference ± the repeatability coefficient (RC).
CORRELATIVE BIOLOGY IN VALIDATION OF QI BIOMARKERS

The value proposition of medical imaging is that it can interrogate human biology in vivo, noninvasively, spatially, or longitudinally, and thus provides diagnostic, predictive, and therapeutic insights to manage patient outcome. In addition to validating the precision and accuracy of QI imaging metrics (as outlined in Precision and Accuracy of Quantitative Imaging section), ideally QI metrics for a given biomarker need to be validated against the underlying biology. A “biomarker” is defined as a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or responses to a therapeutic intervention (154). Thus, an imaging biomarker is an objective QI metric derived from an in vivo image for a given biomarker. Traditionally, clinical QI biomarkers have been validated by statistical analyses against measures of outcome. More often, tissue biopsies are available and can be used to correlate in vivo QI metrics to measures pathology or OMICS (among others) measures. In contrast to biopsies, in vivo images can provide information about the spatial heterogeneity of the whole tumor, albeit at lower resolution. With the advent of radiomics (155, 156), there is an underlying effort to validate image features against pathological or genomic features of tumor heterogeneity derived from biopsies (157–165). However, there are numerous complicating nuances in correlating a QI metric derived from a clinical image to features derived from a biopsy; these nuances include mismatches in scale, discrepancies in coregistration, and importantly single point (needle) biopsies may not reflect the pathobiology of the whole tumor. The use of co-clinical models thus enables correlation and validation QI metrics against the underlying heterogeneous biology of co-clinical human tumor models when human specimen is scarce.

INFORMATICS NEEDS TO SUPPORT CO-CLINICAL RESEARCH

The preclinical imaging workflow is somewhat complex, routinely using multiple instruments to characterize the physiology and biology of a given tumor in vivo. To validate in vivo imaging measurements, in vitro or ex vivo multiscale assays such as pathology, -OMICS (genomics, proteomics, metabolomics, etc.), immunohistochemistry, multiplexed immunofluorescence are used to correlate quantitative image-derived measurements to ex vivo measures. When multiplied by the number of subjects (animals) per group in a given experiment, multitudes of interventions (e.g., drugs), descriptive data (weight, diet, tumor volume, blood, metabolic panel data, etc.) and the number of time-points in a longitudinal imaging protocol, the resulting data sets are vast and prohibitive to track and manage long term. As a consequence, nontractable data result in poor reproducibility and present obstacles for open science collaboration and data mining.

To that end, informatics solutions are needed to support co-clinical imaging research including collection of metadata to qualify co-clinical imaging studies. For example, a recent guideline on the use of PDX in preclinical research (29), lists ~45 fields (metadata) to capture to qualify research results. Importantly, there is a need for harmonization and integration of preclinical cancer imaging data, imaging acquisition protocols, and annotation. Legacy preclinical imaging databases are not equipped to support big data science and collection of metadata/annotations to support NCI’s precision medicine initiative. Although some institutions have developed databases to house preclinical imaging data, many such legacy databases are not compatible with the complexity and growing demands in preclinical cancer imaging which include big data needs and collection of metadata/annotation to support NCI’s precision medicine initiative. Importantly, the increasing prevalence of quantitative acquisition and analysis approaches depend on sophisticated computational methods that generate additional derived data. Given these “big data” challenges, informatics tools are needed that have the capacity to organize data structures, enforce QA practices, generate audit trails and provenance records, provide detailed reports and data tracking tools, and ultimately facilitate data analysis.

Lack of reproducibility in preclinical cancer research, including imaging, has been highlighted by numerous publications (119, 166). Other than promoting open science, data sharing has been suggested as one solution to address reproducibility. Similarly, sharing of quantitative imaging pipelines is expected to enhance reproducibility, as it will allow for testing of multiple analytic pipelines using a common data set for comparison and validation. The NCI has chosen to establish an open environment in which the oncology community can collaborate to tackle the sundry issues that pertain to reproducibility of animal model research as required for precision medicine. Prominent among those issues is transparency of details that document imaging experiments and their application to translational research. Thus, informatics tools and platforms are needed to enhance reproducibility in preclinical imaging, enable data mining with collection of metadata and annotations tools, and promote open science.

SUMMARY

Advances in clinical QI have been realized to a large extent by numerous initiatives such as the Quantitative Imaging Network and the Quantitative Imaging Biomarker Alliance to standardize and implement advanced QI methods in clinical practice. Although these and other initiatives have had a significant impact in advancing clinical applications of QI, preclinical imaging plays a critical role in developing in vivo translational imaging strategies to interrogate disease mechanisms, detect disease, and assess/predict response to therapy. The use of co-clinical animal models of cancer ushers-in new paradigms involving co-clinical trials where biological and molecular mechanisms of
of-the-art translational quantitative imaging methodologies to enable disease detection, risk stratification, and assessment/prediction of response to therapy. It is expected that the quest for best practices will neither result in reduced creativity nor hamper progress in translational imaging science, as some may conjecture. Rather such creativity should be viewed as investment towards progress in translational imaging and its role in guiding precision medicine into the next decade.

ACKNOWLEDGMENTS

The research agenda outlined in this communication is supported by the Washington University Co-Clinical Imaging Research Resource, NCI Grant # U24CA209837 and Siteman Cancer Center Support Grant P30CA01842; the Duke Preclinical Research Resources for Quantitative Imaging Biomarkers, NCI Grant # U24CA220245; Integrating OMICS and Quantitative Imaging Data in Co-Clinical Trials to Predict Treatment Response in Triple Negative Breast Cancer, NCI Grant # U24CA2226110 and CPRIT RR160005 (T.E.Y. is a CPRIT Scholar of Cancer Research); University of Michigan Quantitative Imaging Research, NCI Grant # U24CA237683; Vanderbilt University-PET Imaging Resource to Enhance Delivery of Individualized Cancer Therapeutics, NCI Grant # U24CA220325; and PENN Quantitative MRI Resource for Pancreatic Cancer, NCI Grant # U24CA231858.

REFERENCES


284


