***Please add your comments directly in this document for others to see and comment.***

***ALL ideas, references, suggestions and critiques are welcome. Thank you!***

**CONTENTS**

**I.** **Moderator(s)**

**II.** **Summary of Action Ideas**

**III.** **Additional Resources from Participants**

**IV.** **Discussion Summary**

**V.** **List of Participants**

**VI.** **Digital Twin Chat Log**

**I.** **Moderators**:

* Tina Hernandez-Boussard *– Stanford University, computational biology, and health sciences research; accelerate* patient centered care
* Paul Macklin *– Indiana University*
* Tanveer Syeda-Mahmood *– IBM Research, multimodal imaging and informatics, analytics perspective*

**II.** **Summary of Action Ideas**

* Concrete definition needed for “twin,” i.e. what might be a possible quantitative metric to define similarity

* Primary goal should be eventual digital testing on the twin for genuine treatment decision enabling.

* Oncology physicians and researchers – pull from the NCI community – want to engage them from the beginning to get a physician’s voice.
  + Reach out to community meetings such as [ASCO](https://www.asco.org/) (great place to engage clinicians and get them involved)

* Potential funding resource: Philanthropic groups – [Breast Cancer Research Foundation](https://www.bcrf.org/)

* Industry could build a data commons or sandbox and collect data across universities and begin modeling process
* Defining a tractable problem would give those interested in participating a way of having something concrete to explore, and a way to get some footing in some actual work.
  + A tractable problem that could be modularly expanded would be helpful. Current efforts include building tumor metastasis dissemination models, based on longitudinal mouse models. Would be interesting to improve and scale that to human scales, with joint expertise, then start plugging in better details.
* Envision a stream of data where there would be interest in projecting treatment effects based on considerations like PKPD, tumor localization, target protein interactions with small molecules, etc. There is tremendous scope in this group for really working on phenotype prediction.
  + Propose circulating to this interest group some ways that we envision the data needs and what could be used for development of digital twin in a clinical sense, and then see what others in the group envision, and hone in on a set of working questions.
    - Selection of use case is important because each cancer site requires specialty treatment and has different clinical questions.
  + Propose picking a clinical target area that already includes some “success stories” retrospectively, such as melanoma or non-small-cell lung cancer (immunotherapy innovations) and/or glioblastoma?

**III.** **Additional Resources Recommended by Participants**

* Roswell Park, experience in phenotypic projection<https://www.roswellpark.org/>
* American Society of Clinical Oncology <https://www.asco.org/>
* NIH Science and Technology Research Infrastructure for Discovery, Experimentation, and Sustainability (STRIDES) Initiative<https://datascience.nih.gov/strides>
* Melanoma Research Foundation – Accelerating Research and Treatment Development
  + <https://melanoma.org/research-science/scientific-initiatives/mrf-breakthrough-consortium-mrfbc/>
* Potential funding resources:
  + <https://grants.nih.gov/grants/guide/pa-files/par-16-349.html>
  + NCI Informatics Technology for Cancer Research (ITCR)<https://itcr.cancer.gov/funding-opportunities>
* Society for Simulation in Healthcare - Healthcare Systems Modeling & Simulation Affinity Group
  + <https://www.ssih.org/Interest-Groups/Healthcare-Systems-Modeling-Simulation>
* Healthcare Systems Modeling and Simulation Webinar<https://www.youtube.com/channel/UCMWVW9plawga7UtWnISrszg/videos>
* Cancer Intervention and Surveillance Modeling Network <https://cisnet.cancer.gov/>
* Breast Cancer Research Foundation <https://www.bcrf.org/>
* Interagency Modeling and Analysis Group <https://www.imagwiki.nibib.nih.gov/>
* Presentation on Industrial Digital Twins
  + <https://www.imagwiki.nibib.nih.gov/sites/default/files/GE%20Digital%20Twin%20Overview%20and%20Tutorial_RRI%20v2.pdf>

**IV.** **Discussion Summary**

**Q – Digital Twin is an overloaded term**. Treatments are not so personalized and not looking at all scales. Needs to be more personalized and individualized. **Do you think this vision is realistic and feasible in our time frames?**

* There seems to be everything for everyone.
* This is a big area – Digital Twin in “focused conditions” – “disease focus”
* I also wonder whether the term “twin” needs to encompass the organismal scale or whether it can stop at a lower scale, such as organ/tissue, which can then be linked to phenotype (perhaps not mechanistically, but statistically)

**Q –** **What type of cancer would be important to study in this context? Which community is open to trying this?**

* Breast cancer Foundation and community may be an area to start.

* Melanoma area: Open to Immunotherapy options.

* Many single lab options – common software ecosystem instead of multiple labs doing things individually.

* May require SOFTWARE ENGINEERS and DATA STANDARDS
  + This will take a lot to put together a DATA DASHBOARD and INTERFACE EXPERTS; mock-ups and patient advocates asking for what they wanted (be a part of the design process)

* Prostate and Lung? Imaging modalities for lung cancer, for example, could be an option.
  + Phenotypic projection – Roswell Park (Miranda Lynch)
  + Major lung cancer data is publicly available

**Q –** **VA Data Availability? What about Prostate cancer?**

* Cancer Registries have info at time of diagnosis and recurrence, but not necessarily trajectory; doing a lot of large-scale data linkages.
* would be great to have data from full body CT scanner; e.g. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5750522/>

**Challenges**

* How do we get new sources of data to enable trajectory modeling and bring in all points – genomics, social factors, etc.? How do we find and aggregate data?

* One-off versus longitudinal data – how do we collect and automatically annotate?
  + Lung cancer may be a good start – patient trajectory is a lot shorter

* How do we ensure linkages with sensors and that data becomes available in real-time in daily, continuous feeds?

* How do we incorporate novel therapies and predict outcomes?

* How do we represent patients, their clinical history, contextual factors, and outcomes of interest?

* How do we use the Electronic health record at scale to do this kind of work?

* How do we identify and measures treatment-related adverse responses?

**Specialty Areas and Collaborators**

* How do clinicians see this as useful?
* HCI: Human-computer interaction
* User interface people
* Data visualization
* Uncertainty quantification
* Bring mechanistic and ML modeling – nucleate work
* Oncology physicians and researchers – pull from the NCI community – want to engage them from the beginning to get physician voice.
* Reach out to community meetings – ASCO (great place to engage clinicians and get them involved)
* Identify champions from your own collaborations.

**Funding Sources**

* Pilot funding may be helpful. If this community developed a specific use case, would there be opportunities? What about through DOE and NCI?
  + NIH, NSF, Big science questions, individual grants?

* Philanthropic groups – Breast Cancer Research Foundation, Prostate (not sure of groups), brain tumor groups, - PICK A USE CASE (cancer topic) and foundations related to a particular disease may fund

**Academia VERSUS Industry**

* Industry – Imaging, modeling – Getting behind a platform; industry could build a data commons or sandbox and collect data across universities and begin modeling process (industry and university partner). AI and ML algorithms and causal reasoning; bring algorithms to data models. May require joint exploration. Maybe industry can take to scale to explore.

* Industry – GE point of view – medical imaging. Interest in AI, clinical side, life sciences side. More expertise in the imaging side.

**Next Steps**

* What would be your ideal cancer use type? Most progress? Where cancer has treatment choices/options?

* What cancer type does this group want to focus on? Is data accessible and available?

* Patient Advocacy groups?

* Foundations / Funding opportunities?

* Breast cancer has so much funding, data, genomic information – one to think about
* Shorter care-trajectory à Lung cancer? Prostate Cancer? Glioma / Glioblastoma? Virtual controls? Bring in UQ?

**V.** **Participants**: ***Please add/correct your name & affiliation if you participated in this breakout group!***

* Emily Greenspan *– NCI CBIIT*
* Ernesto Lima – *The University of Texas at Austin*
* Roxanne Jensen *– NCI DCCPS*
* KJ Wilkins – *National Institute of Diabetes and Digestive and Kidney Diseases*
* Tom Bartol
* Amy Gryshuk – *Lawrence Livermore National Lab*
* Julio Perez - *Instituto Nacional de Astrofisica, Optica y Electronica*, Mexico.
* Nadia *– Cancer bioinformatics*
* Brian *– GE research; computer engineering; computational biology; Co-PI on ITCR for tumor heterogeneity*
* Miranda Lynch – *Hauptman-Woodward Medical Research Institute*
* Berdan Meyers *– GE Research, software engineer*
* Hildur Knutsdottir *– Johns Hopkins, mathematical biology, bioinformatics*
* Jonathan Ozik *– ANL, agent-based modeling, worked with Paul Macklin, high-throughput verification*
* Matthew McCoy *– Georgetown University, MD simulations, cell scale simulations*
* Veera Baladandayuthapani *– University of Michigan, Prof. of Biostatistics*
* John Rice *– Dublin Ireland, observer, associated with modeling group after retiring, interested in digital twin*
* [UNKNOWN] *– University of Texas Austin*
* Tiphaine Martin *– Post-doc at Icahn School of Medicine at Mount Sinai, NY. Bioinformatics and biostatistics*
* Donna Rivera *– NCI, SEER program*
* Yue Dong – *Mayo Clinic, interested in OR/ICU patient care delivery process and provider information process*

**Group Contact Info**

* Tina Hernandez-Boussard, [boussard@stanford.edu](mailto:boussard@stanford.edu). Use case ideas: prostate and breast cancer
* Tanveer Syeda-Mahmood, IBM Research, [stf@us.ibm.com](mailto:stf@us.ibm.com)
* Jonathan Ozik, Argonne National Laboratory, [jozik@anl.gov](mailto:jozik@anl.gov)
* Miranda Lynch [mlynch@hwi.buffalo.edu](mailto:mlynch@hwi.buffalo.edu) Hauptman-Woodward Medical Research Institute
* Veera Baladandayuthapani, University of Michigan; [veerab@umich.edu](mailto:veerab@umich.edu); bayesrx.com
* Ernesto Lima - The University of Texas at Austin, [lima@ices.utexas.edu](mailto:lima@ices.utexas.edu)
* Matthew McCoy - Georgetown University - [mdm299@georgetown.edu](mailto:mdm299@georgetown.edu)
* KJ Wilkins: [kenneth.wilkins@nih.gov](mailto:kenneth.wilkins@nih.gov) - I like the choice of differing length of care / progression
* Donna Rivera, Surveillance Research Program, National Cancer Institute [donna.rivera@nih.gov](mailto:donna.rivera@nih.gov)
* Yue Dong, Mayo Clinic [dong.yue@mayo.edu](mailto:dong.yue@mayo.edu)
* Julio Perez [jcp.sansalvador@inaoep.mx](mailto:jcp.sansalvador@inaoep.mx)

**VI.** **Digital Twin Chat Log**

*Miranda Lynch*: A very interesting aspect here is that there is a ‘built in’ longitudinal nature to the evolution of the digital twin, and that evolution (learning over time) will rely intimately on the ability to get clinical patient info over time.

*KJ Wilkins*: could offer “counterfactual digital twin” as more precise distinction from other uses, one that gets at goal of precision medicine. Given an individual’s data to date, which (set of) treatment(s) are “right” for that individual at that time? Whether ML, or ODE dynamical models, or other modeling approaches are used it seems this involves the goal of what this aimed at estimating...

*Paul Macklin*: I agree with this perspective. We also want more than current "precision medicine" (which tends to focus on stratification) and more on "predictive medicine" that looks at the dynamics of individual patients

*Veera B*: From an analytical perspective it might be critical how we define a “twin,” i.e. what might be a possible quantitative metric to define similarity

*KJ Wilkins*: Thanks for the emphasis on Common Data Standards, and architectures that ease use of data interactions, whether longitudinal or multi scale; the tech companies offering cloud computing through STRIDES initiative have the novice-ready architecture part up and running, it’s the human-interactive aspect that’s disease focused that could use the sort of expertise you mentioned.

*KJ Wilkins*: @Donna: any sources for integrating #melanoma data re: novel immunotherapies outside of SEER? I like your mention of dual needs considerations for clinicians AND patients. Please also explicitly mention WG I.

*Paul Macklin*: There have been data commons efforts here and there. Has been way too much lock in so far, since many are based on a big company trying to get everybody on their platform. Has been a challenge, but also good prototyping / early results. Agreed we need to leverage wherever we can, and then plug dynamical and ML models into whatever best ecosystem.

*KJ Wilkins*: ...which therapy was recently approved by FDA? off n=55 across 12 sites.

*Donna Rivera*: Data Sources: SEER can provide incidence and prevalence data for melanoma, as far as specific sources data can come from claims or EHR, also other groups may want be interested:<https://melanoma.org/research-science/scientific-initiatives/mrf-breakthrough-consortium-mrfbc/>

<https://grants.nih.gov/grants/guide/pa-files/par-16-349.html>

Potential Funding:<https://grants.nih.gov/grants/guide/pa-files/par-16-349.html>

<https://itcr.cancer.gov/funding-opportunities>

*KJ Wilkins*: I would ask Amy Gryshuk to include patients/patient advocates that are associated with disease areas mentioned.

*Miranda Lynch*: Defining a tractable problem would give all of us who are interested in participating a way of having something concrete to explore, and a way to get some footing in some actual work.

Paul Macklin: A tractable problem that could be modularly expanded would be helpful. We've been working at building tumor metastasis dissemination models, based on longitudinal mouse models. Would be interesting to improve and scale that to human scales, with joint expertise. Then start plugging in better details. But as an early platform for treatment response for circulating drugs. I would think that building that as a scaffolding would be tractable and attract help.

*Miranda Lynch*: I come at many clinical questions in terms of structural biology and use of imaging to provide actual ‘local’ information (of the what’s where and what is it doing, at molecular and cellular level). I would like to envision a stream of data where there would be interest in projecting treatment effects based on considerations like PKPD, tumor localization, target protein interactions with small molecules, etc. I think there is tremendous scope in this group for really working on phenotype prediction.

*Paul Macklin*: That's a project that would get me fired up, Miranda.

*Miranda Lynch*: Could we circulate to this interest group some ways that we envision the data needs and what could be used for development of digital twin in a clinical sense, and then see what others in the group envision, and hone in on a set of working questions?

*Donna Rivera*: To engage clinicians effectively, the selection of use case is important because each cancer site requires specialty treatment and has different clinical questions.

*Miranda Lynch*: I agree with Donna’s comment, that finding a focused clinical target area would be one of the key parts of homing in on a tractable problem.

*Paul Macklin*: That's a good point. Being more concrete will be very helpful in driving advances.

*KJ Wilkins*: Why not pick a clinical target area that already includes some “success stories” retrospectively, such as melanoma or non-small-cell lung cancer (immunotherapy innovations) and/or glioblastoma?

*Yue Dong*:

· We have a group try to reach out outside our professional to partnership.<https://www.ssih.org/Interest-Groups/Healthcare-Systems-Modeling-Simulation>

· We have some webinars online to educate our clinicians<https://www.youtube.com/channel/UCMWVW9plawga7UtWnISrszg/videos> . Please contact me if you would like to present a webinar.

*Tanveer* Syeda-Mahmood: Lung cancer particularly the use of low dosage screening. Prostate and breast is ok too.

*Paul Macklin*: if concrete with prior existence of some tools, I'd suggest GBM, using Kristin Swanson's tools. I bet she'd be very amenable. But the idea of there being true choices is a great point. prostate: watchful waiting vs. treatment, etc.

*KJ Wilki*ns: Patient goals would be a key feature with UQ an accompanying toolkit. We need to be realistic about the volume and granularity/complexity of data required to make predictions at a patient-tolerated level of uncertainty; great discussion, all!

*Jonathan Ozik*: Thanks for facilitating the interesting discussion in the micro lab today. I thought the topics of data access and expert inclusion (from as early as possible) were great. Even with the assumption of the future existence of digital twins, I was a little surprised how little discussion was dedicated to the computational experimentation with digital twins, which I assumed would be one of the main reasons for using them. I do also think that there is a lot of expertise (e.g., CISNET) and work to do (e.g., large-scale uncertainty quantification) to develop the calibrated digital twins in the first place. I'd be curious to get your take on this.

I'd also be curious to understand what types of models would be considered digital twins. With your inclusion I'd assumed that mechanistic ones were the focus, but it sounded like microsimulation, natural history models could also qualify?

*Miranda Lynch*: Thanks for sharing the very helpful comments. I also think that the primary goal should be eventual digital testing on the twin for the purpose of genuine treatment decision enabling.

###