

Building effective collaborations for the study of COVID-19 across the severity spectrum

Translational patient-oriented research is essential for understanding the immunopathology of SARS-CoV-2 and the varied host responses to this virus. Successful research collaborations place the patient at the center while designing the team, setting, timing, data elements and biospecimens to address important questions and advance human health.

Roseline S. Agyekum, Benjamin A. Abramoff, E. John Wherry and Nuala J. Meyer

Since the first reports of an unusually severe viral pneumonia impacting the province of Wuhan, China, in December 2019 (ref. ¹), the research community has learned an astonishing amount about what we now recognize as the disease caused by coronavirus SARS-CoV-2 (COVID-19). The pace of research has been extraordinary, and the volumes of data reported are a testament to the power of the global biomedical research apparatus when harnessed toward a unified goal. Patient-oriented research has been transformational to our understanding of COVID-19 and has rapidly translated to patient care. Such research identified the viral genomic sequence, classified it as a coronavirus², identified each major strain of the virus, and developed tools to detect and diagnose the disease.

Beyond the molecular detection of SARS-CoV-2, patient cohorts enabled the rapid identification of clinical features that characterize this disease. Early reports from Wuhan highlighted the increased risk among older individuals and those with cardiovascular comorbidities, the often protracted and delayed hypoxic respiratory failure, and laboratory features such as a high ratio of neutrophils to lymphocytes in the peripheral blood associating with severe disease^{3,4}. However, major pathophysiological insights were possible when mechanistic scientists partnered with clinical investigators. Studies in well-curated molecular cohorts showed the marked heterogeneity in immune response among hospitalized cases with COVID-19⁵ and the potential contributions of auto-antibodies against type I interferon to worsen COVID-19 severity⁶, while multi-omic measurements in patients with specific post-acute sequelae of COVID (PASC) suggested both immunological and metabolic reprogramming⁷. In this Comment, we review some guiding principles to a successful translational

Box 1 | Key strategies for a successful translational collaboration

1. Design a patient-centered study — safe, convenient, meaningful
2. Build a team of experts
 - a. Patient representatives — true partnership
 - b. Multifaceted expertise: clinical, patient-oriented research, molecular and immunological
 - c. Create a process for new investigators to collaborate, request samples or collect new clinical data
 - d. Educate potential new collaborators about realistic sample volumes for new assays
3. Begin with the end in mind
 - a. Define the critical comparisons
 - b. Define an appropriate control population and ways to maximize similarities between populations to isolate the key differences
 - c. Collect the essential clinical variables — EHR, patient-reported symptoms (surveys), data requiring interpretation (CXR, CT, EKG)
 - d. Define biosamples, time points, populations
 - e. Logistical process mapping — ensure consistent processing across different settings
4. Reflect and refine
 - a. Whether the study is representative
 - b. Whether the most important questions are being addressed through this cohort
 - c. Define ways to adapt

patient-oriented collaboration and suggest key steps to ensure a representative, biologically informed cohort study that advances clinical, immunological and molecular understanding (Box 1). Although we focus on COVID-19 patient-oriented research, we believe our recommendations are equally relevant to the study of any human disease.

Center research around the patient

At the heart of well-designed patient-oriented research is a focus on the patient experience, and how to make the research safe, convenient, inclusive and meaningful for patients (Fig. 1). If the patient experience is not considered, there is a risk that the research will be rejected by regulatory agencies or, more commonly, that enrollment targets are not met because

patients do not participate and thus effort, resources and time are wasted. Interested research participants cite numerous barriers to engaging in research, particularly the lack of time, compensation, opportunities to participate, accessibility or transportation to the research facility and self-efficacy for facilitating their own participation⁸. Successful patient-oriented research considers and mitigates each of these barriers during study design. As investigators launch their study, they must decide on the enrollment setting, which will dictate their team's approach to the collection of both clinical data and biospecimens. Working with hospitalized patients overcomes the barriers of access, study awareness (when study and clinical staff can alert potential participants about research opportunities), transportation and time, and may facilitate

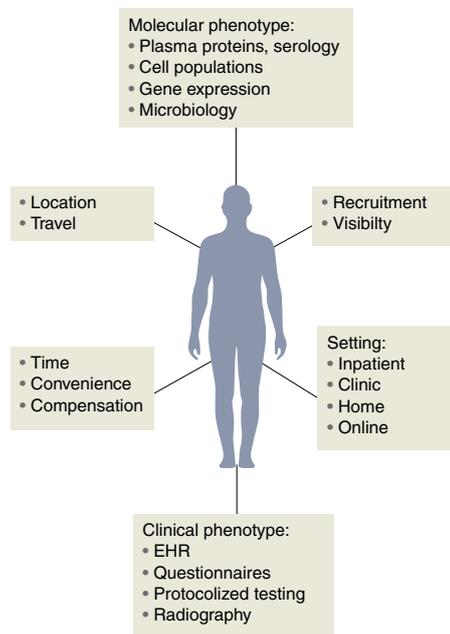


Fig. 1 | Patient-centered approach to study design. To launch a successful molecular cohort study, we recommend trying to maximize the study participant's awareness of the study and its purpose, as well as optimizing convenience and compensation. Sometimes this might include providing transportation to the study location or sending a study team to the patient. We also recommend critically evaluating the necessary clinical and molecular data to be captured, and ensuring that these data can be collected in the appropriate setting.

repeated biosampling, yet there may be more restrictions to maintaining safe blood collection volumes, particularly in critically ill patients who are prone to anemia. Special cases, such as pregnant patients or those with pre-existing illnesses, present added complexities and may warrant a revised risk-benefit consideration about participation. Severely limited personal protective equipment, as occurred early during the COVID-19 pandemic, can add complexity to working with patients that are acutely infectious, particularly when the infectivity of body fluids is uncertain⁹. By contrast, working with patients in an ambulatory setting can relieve some constraints, such as the safe volume of blood to be collected, but introduces other challenges, including securing time in the participant's schedule, organizing transportation to the study site, and navigating the logistics of biosample processing when research is performed remotely or obtained outside routine hours. Beyond the setting of the research, other patient-centered decisions include which

data elements are central to the project and how they will be collected; which biospecimens will be collected and at which time points; and how to reach potential participants and notify them about the research.

Intentional team building

Translational research is team science, and success depends on building a team with complementary expertise, insights and resources. We define the study team broadly to include all members involved in conceiving, refining and executing the research, from clinical and mechanistic investigators to staff enrolling participants, collecting data (clinical and molecular) and analyzing results. The inclusion of patient representatives on the study team ensures that the research being pursued is meaningful to those experiencing the disease. Each perspective on the team is vital; all members have input as to which research questions warrant investigation, and the team prioritizes through discussion and compromise.

Once the team of experts is built, a regular meeting schedule helps to ensure that projects meet their targets. The clinical team shares observations about how the cohort is evolving and risk factors that might influence biology, and the molecular team shares data in progress, highlighting surprises or unusual cases that might warrant a deeper phenotypic dive. Regular meetings allow the team to prioritize projects, designate leaders for different aspects of the project, clarify authorship and troubleshoot roadblocks that may hinder the research. New investigators requesting data or biospecimens to test a novel hypothesis is a frequent scenario. Cohort studies should anticipate this and have a clear process to determine how remaining biospecimens should be used after the initial study objectives are met. Setting clear priorities, assigning decisional authority, carefully tracking sample volumes, suggesting boundaries on sample volumes per assay and refining protocols on non-study samples before accessing study resources are strategies that can help to maximize the use of material from valuable patient cohorts.

Molecular cohort studies — those that prospectively enroll at-risk participants and collect carefully timed biospecimens — rely on significant human capital. The clinical team is continually screening participants, performing informed consent discussions and then collecting clinical data and biosamples from the individuals who consent. The molecular staff perform high-specialty molecular processing, cellular isolation and highly sophisticated

laboratory assays dictated by the often unpredictable enrollment. Molecular cohorts are expensive to launch and maintain. However, the COVID-19 experience has highlighted important scientific insights that were achieved by well-designed cohorts. Early in the pandemic, the need for patient-oriented research outpaced the ability of traditional funding mechanisms to respond. In the USA, the National Institutes of Health (NIH) allowed redeployment of resources from funded works in progress and provided some supplemental awards, but a more proactive solution might be to create designated 'disease response' cohort-building resources that could be accessed in future pandemics, taking cues from successful trials such as the RECOVERY and REMAP-CAP platforms¹⁰. Investigators should also seek institutional investment, potentially through pilot awards that can help to establish new collaborations, while awaiting new grant opportunities.

Cohort assembly

Successful research teams balance a goal-oriented approach — designing the study to answer the desired biological questions — with the flexibility to adapt as new knowledge emerges. The research question will determine the patient population to study, the key variables necessary for phenotyping and the biosamples needed to make the biological comparisons. Selecting the optimal comparison population is crucial. In the case of patients with COVID-19, comparisons can be made with never-exposed (SARS-CoV-2-naive) patients, with patients infected with another respiratory virus or between patients with different COVID-19 severity or different phases of COVID-19 disease. Each research question prompts a different recruitment strategy. Some investigations attempt to make several comparisons (for example, healthy, SARS-CoV-2-infected, infected with pneumonia and infected with both pneumonia and respiratory failure). For such complex designs, each category of participants needs a careful consideration of the issues raised above: how to identify patients, where to enroll, how to collect clinical information and how to collect biospecimens.

Biosample collection can be a complicated process. In our experience with a COVID-19 molecular cohort, it became apparent that the availability of processing space and personnel compatible with biosafety level 2-enhanced (BSL-2+) practices was a major bottleneck. By creating a flow chart for the patient-derived samples from acquisition to storage, we were able to identify the appropriate

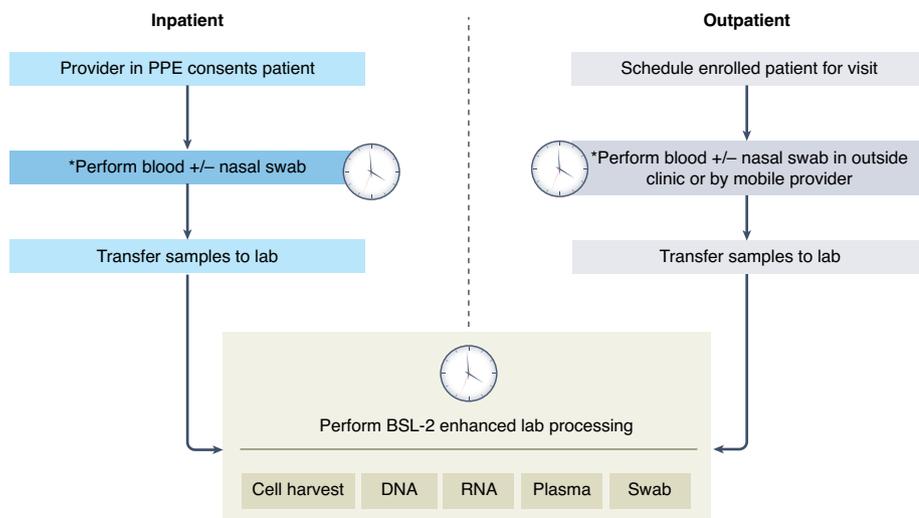


Fig. 2 | Setting-specific COVID-19 biospecimen process map. To ensure appropriate collection and processing of biosamples, trace the steps from sample collection to assay, and confirm a consistent process whether collecting from hospitalized inpatients or healthy volunteers in the community. Once the process is mapped, confirm the appropriate personnel, supplies and equipment are available to conduct the work safely. Asterisks indicate steps initially limited to nurse, physician or staff phlebotomy when personal protective equipment (PPE) was limited. The clock signal indicates that a time stamp is recorded.

personnel, protective equipment, collection and transport materials and laboratory space to accomplish each step safely and in accordance with all institutional and national guidelines. In turn, the detailed flow chart (Fig. 2) increased the confidence of the team members who were volunteering to work with potential biohazards during a time of great uncertainty⁹. Wherever possible, identical sample processing, including the time between acquisition and processing, should be conducted across all patient subgroups to prevent flawed conclusions based on process variation. Recording time stamps for each processing step allows researchers to statistically test whether processing variables influenced the results. In addition, we recommend that investigators clearly define their primary study objectives and apply stringent methodology to declare these objectives significant¹¹, and explicitly state when secondary outcomes do not survive multiple comparison adjustment.

Equally important as the biosample design is the design for the capture of clinical data. This includes defining the data elements that are crucial to the scientific question and those that are helpful to gauge generalizability by describing the study population. Consider the factors that might confound associations, and collect them prospectively. Numerous options for clinical data collection are possible, and each has advantages and challenges. Some

studies will rely on dedicated data capture through trained research personnel that can extract data from the electronic health record (EHR). Although this practice often yields high quality data completeness and reliability, it can be slow and inefficient, and more difficult to adjust once the study has launched. A data item that was overlooked, or one added as the landscape changed — such as vaccination status, SARS-CoV-2 strain¹² or specific treatments received — may require a return to the EHR of every participant. The importance of the additional information is judged against the extra work to obtain it.

Harnessing the power of the EHR

Notable strides have been made to phenotype aspects of COVID and potentially PASC through the EHR directly. An international effort (the Consortium for Clinical Characterization of COVID-19, or 4CE, by EHR), which was mobilized to consolidate, share and interpret data from hospitalized patients with polymerase chain reaction (PCR)-confirmed COVID-19 diagnosis, rapidly disseminated data about national case rates, patient characteristics and laboratory test trajectories, as well as variations in practice across different sites¹³. The US NIH has launched the National COVID Cohort Collaborative (N3C) EHR repository, which currently has more than 1.7 million participants and has developed machine learning models that identify

patients with ‘potential long COVID’ based on symptom report and health care utilization¹⁴. Such models have tremendous power to identify patients who might participate in clinical research and to test associations in a well-powered pragmatic design. However, cohort building by EHR definition is not straightforward. One large effort to validate insurance-coding-based methods to identify patients found that defining cases using the COVID-19-dedicated international classification of diseases claims code (ICD-10) within a patient’s record had high sensitivity to identify patients, but poor specificity (below 50%)¹⁵. EHR algorithms based on ICD-10 with symptom codes achieved higher specificity, but sensitivity fell below 50%, leading the authors to conclude that none of the 11 algorithms built on ICD-10 codes alone exhibited a satisfactory combination of high sensitivity and specificity compared to using PCR positivity, which was their gold standard¹⁵. However, as home antigen testing is now widely available, the sensitivity of algorithms requiring PCR testing will suffer, because positive antigen tests might not be captured by the EHR and true positive cases will be missed by this definition. Finally, some clinical phenotyping may not exist in the EHR, and may require a dedicated study effort to collect certain features. For COVID-19, although artificial algorithms are being developed for chest radiology interpretation, by and large the chest X-ray or computed tomography interpretation still requires a trained interpreter to determine whether the patient has radiographic pneumonia or acute respiratory distress syndrome¹⁶. Some features may require patient reporting of symptoms, and validated instruments are strongly encouraged to ensure reproducibility^{17–20}.

Another factor with a major effect on the study design is the phase of disease being studied. As more and more of the global population enters a ‘COVID-resolved’ state, there is strong recognition of the breadth and depth of PASC and the multiple life and health domains impacted by this virus²¹. The patients seeking outpatient care for PASC seem to be different from those hospitalized during acute disease, and also from those discharged from the hospital to acute rehabilitation after COVID-19 infection^{5,22,23}, causing investigators to think specifically about how best to reach, enroll and phenotype these patients. The American Academy of Physical Medicine and Rehabilitation has convened expert panels that issued consensus guideline statements on how to assess and treat the symptoms of PASC²⁴. These guidelines are a helpful resource to encourage more

standardized data collection and which symptom survey instruments might facilitate rigorous, reproducible research. As an added complexity, many patients receive post-COVID care through novel platforms, including telehealth, which has evolved during the periods of social distancing and to expand care to patients unable to travel to appointments. New methods to link biosampling –through patient-collected samples that are shipped to the study team, dedicated phlebotomy visits, or even a study team with phlebotomy who travels to the patient — should be considered to facilitate this work while developing solutions to process biosamples effectively.

One final aspect of successful translational research is to continuously reflect and refine the cohort design. It is imperative that the enrolled population is representative of the patients experiencing the illness and requiring treatment. The COVID-19 pandemic has further exposed the disproportionate burden on historically disadvantaged and underrepresented communities, and research should help to decrease these disparities and avoid perpetuating them by failing to enroll a representative sample²⁵. There are many reasons for communities that historically have been treated unequally to choose not to participate in observational research. At the same time, it is our responsibility as investigators to make the science meaningful and impactful for all communities, and to communicate this imperative effectively. We must engage with patients in such a way that they understand the risks and benefits of a study and know the effect of their involvement. When studies identify an underrepresentation problem, we recommend revisiting each step in the recruitment process, asking whether unintended barriers to participation are present and engaging community members to advise the study team. Ultimately, we need improved methods to disseminate the findings of the study back to the patients and communities who made this research possible, and better understanding of what further research is best aligned with and can improve the health of the patients we serve. We believe the patient-centered approach is what

defines a successful translational research collaboration and advances health while growing knowledge. □

Roseline S. Agyekum^{1,2},
Benjamin A. Abramoff³, E. John Wherry^{4,5,6}
and Nuala J. Meyer^{1,2,4}✉

¹Pulmonary, Allergy, and Critical Care Medicine, University of Pennsylvania Perelman School of Medicine, Pennsylvania, PA, USA. ²Center for Translational Lung Biology, Lung Biology Institute, University of Pennsylvania Perelman School of Medicine, Pennsylvania, PA, USA. ³Department of Physical Medicine and Rehabilitation, University of Pennsylvania Perelman School of Medicine, Pennsylvania, PA, USA. ⁴Institute for Immunology, University of Pennsylvania Perelman School of Medicine, Pennsylvania, PA, USA. ⁵Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania Perelman School of Medicine, Pennsylvania, PA, USA. ⁶Parker Institute for Cancer Immunotherapy, University of Pennsylvania Perelman School of Medicine, Pennsylvania, PA, USA.

✉e-mail: nuala.meyer@penmedicine.upenn.edu

Published online: 23 August 2022
<https://doi.org/10.1038/s41590-022-01283-2>

References

- World Health Organization. Pneumonia of unknown cause — China; <https://www.who.int/csr/don/05-january-2020-pneumonia-of-unknown-cause-china/en/> (2020).
- Wu, A. et al. Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell Host Microbe* **27**, 325–328 (2020).
- Wang, D. et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. *J. Am. Med. Assoc.* **323**, 1061–1069 (2020).
- Yang, X. et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respir. Med.* **8**, 475–481 (2020).
- Mathew, D. et al. Deep immune profiling of COVID-19 patients reveals distinct immunotypes with therapeutic implications. *Science* **369**, eabc8511 (2020).
- Bastard, P. et al. Autoantibodies against type I IFNs in patients with life-threatening COVID-19. *Science* **370**, eabd4585 (2020).
- Su, Y. et al. Multiple early factors anticipate post-acute COVID-19 sequelae. *Cell* **185**, 881–895.e820 (2022).
- Courvoisier, M. et al. Evaluation of the partners in research course: a patient and researcher co-created course to build capacity in patient-oriented research. *Res. Involv. Engagem.* **7**, 76 (2021).
- Chou, R. et al. Epidemiology of and risk factors for coronavirus infection in health care workers: a living rapid review. *Ann. Intern. Med.* **173**, 120–136 (2020).
- Goossens, H., Derde, L., Horby, P. & Bonten, M. The European clinical research response to optimise treatment of patients with COVID-19: lessons learned, future perspective, and recommendations. *Lancet Infect. Dis.* **22**, e153–e158 (2022).
- Midway, S., Robertson, M., Flinn, S. & Kaller, M. Comparing multiple comparisons: practical guidance for choosing the best multiple comparisons test. *PeerJ* **8**, e10387 (2020).
- Goel, R. R. et al. mRNA vaccines induce durable immune memory to SARS-CoV-2 and variants of concern. *Science* **374**, eabm0829 (2021).
- Schnellinger, E. M., Yang, W., Harhay, M. O. & Kimmel, S. E. A comparison of methods to detect changes in prediction models. *Methods Inf. Med.* <https://doi.org/10.1055/s-0042-1742672> (2022).
- Pfaff, E. R. et al. Identifying who has long COVID in the USA: a machine learning approach using N3C data. *Lancet Digit. Health* **4**, e532–e541 (2022).
- Brown, C. A. et al. Development and validation of algorithms to identify COVID-19 patients using a US electronic health records database: a retrospective cohort study. *Clin. Epidemiol.* **14**, 699–709 (2022).
- Sjoding, M. W. et al. Deep learning to detect acute respiratory distress syndrome on chest radiographs: a retrospective study with external validation. *Lancet Digit. Health* **3**, e340–e348 (2021).
- Freitas, S., Simoes, M. R., Maroco, J., Alves, L. & Santana, I. Construct validity of the Montreal Cognitive Assessment (MoCA). *J. Int. Neuropsychol. Soc.* **18**, 242–250 (2012).
- Nasreddine, Z. S. et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* **53**, 695–699 (2005).
- Garin, O. et al. Validation of the “World Health Organization disability assessment schedule, WHODAS-2” in patients with chronic diseases. *Health Qual. Life Outcomes* **8**, 51 (2010).
- Meachen, S. J., Hanks, R. A., Millis, S. R. & Rapport, L. J. The reliability and validity of the brief symptom inventory-18 in persons with traumatic brain injury. *Arch. Phys. Med. Rehabil.* **89**, 958–965 (2008).
- Sudre, C. H. et al. Attributes and predictors of long COVID. *Nat. Med.* **27**, 626–631 (2021).
- Brown, L. A. et al. The unique contribution of depression to cognitive impairment in post-acute sequelae of SARS-CoV-2 infection. *Brain Behav. Immun. Health* **22**, 100460 (2022).
- Abramoff, B. A., Dillingham, T. R., Caldera, F. E., Ritchie, M. D. & Pezzin, L. E. Inpatient rehabilitation outcomes after severe COVID-19 infections: a retrospective cohort study. *Am. J. Phys. Med. Rehabil.* **100**, 1109–1114 (2021).
- Herrera, J. E. et al. Multidisciplinary collaborative consensus guidance statement on the assessment and treatment of fatigue in postacute sequelae of SARS-CoV-2 infection (PASC) patients. *PM. R.* **13**, 1027–1043 (2021).
- Witham, M. D. et al. Ensuring that COVID-19 research is inclusive: guidance from the NIHR INCLUDE project. *BMJ Open* **10**, e043634 (2020).

Acknowledgements

We thank the patients and volunteers who participated in our research, the many research and clinical staff who contributed their services, particularly the Penn Immune Health and COVID-19 Processing Unit, and our institution for sponsoring impactful collaborations.

Author contributions

E.J.W. is a founder of Surface Oncology, Arsenal Biosciences, and Danger Bio and an inventor on a patent submitted by Emory University that covers the use of PD-1 blockade to treat infections and cancer. E.J.W. is a consultant or advisor for Merck, Marengo, Janssen, Related Sciences, Synthekine and Surface Oncology. N.J.M. is an advisor for Endpoint Health, Inc. and her institution receives funding from Quantum Leap Healthcare Collaborative for work unrelated to this manuscript. R.S.A. and B.A. report no competing interests.