DATA USE AGREEMENT FOR:
CHOOSING IMMUNE SUPPRESSION IN RENAL TRANSPLANTATION BY EFFICACY AND MORBIDITY 2 – CISTEM2

This DATA USE AGREEMENT FOR: CHOOSING IMMUNE SUPPRESSION IN RENAL TRANSPLANTATION BY EFFICACY AND MORBIDITY 2 – CISTEM2 ("**Agreement**"), which is effective as of December 15, 2024 ("**Effective Date**"), is entered into by and between THE CURATORS OF THE UNIVERSITY OF MISSOURI ("**MU**") and _____ ("**Site Participant**", each a "**Party**" and collectively, the "**Parties**").

WHEREAS, Rutgers University will serve as the Institutional Review Board (IRB) of record and the primary recipient of the NIH award for CISTEM2;

WHEREAS, Any funding for Site Participants to participate in the CISTEM2 program shall be pursuant to a separate contract between Site Participants and Rutgers University;

WHEREAS, MU is the recipient of Patient Centered Outcomes Research Institute ("**PCORI**") funding and established the Greater Plain Collaborative Network ("**GPC**").

WHEREAS, MU is responsible for operation of GPC and thus serves as the GPC coordinating center ("**GCC**").

WHEREAS, pursuant to the PCORnet Master Data Sharing and Use Agreement ("**PCORnet DSA**") effective _____, GPC Data Sharing Agreement ("**GPC DSA**") effective _____, and as subsequently amended, GPC participating health systems have contributed a Limited Data Set to GCC for various purposes including research.

WHEREAS, MU will serve as the Data Coordinating Cetner ("**DCC**") of CISTEM2 study.

WHEREAS, CISTEM2 Study Team may request to use such Curated Data, as defined below, for the purposes of conducting Research (as defined herein) relating to transplant, as set forth in the Research Purpose and Study Protocol (attached hereto as Exhibit A) (collectively, "**Research Protocol**").

NOW, THEREFORE, in consideration of the respective representations and agreements contained herein, the receipt and sufficiency of which is hereby acknowledged and agreed, the Parties, intending to be legally bound, agree as follows:

1. **Definitions.** Except as otherwise expressly provided herein, terms used in this Agreement shall be defined as follows:

   a. **Aggregate Data:** Aggregated, De-identified, non-Individual Level Data across specified strata of individuals. For example, counts of patients within a stratum that includes a particular age group, gender and diagnosis.

   b. **Authorized Users**: Individuals associated with and selected by Site Participant who have been granted access to the Secure File Transfer Method in accordance with minimum standards developed by the DCC.

c. **Clinical Research Network ("CRN")**: A clinical research network that participates in PCORnet.

d. **Common Data Model ("CDM"):** This term is defined as the PCORnet common data model defined in accordance with PCORnet and GPC DSA.

e. **Contributed Data:** This term is defined as any type of data in its original or raw format submitted by Participating Sites.

f. **Curated Data:** This term is defined as a curated or cleaned version of Contributed Data which deemed sufficient to serve the Research Purpose.

g. **Data Coordinating Center**: or DCC, means MU in its respective roles of (i) receiving Contributed Data; (ii) processing Contributed Data to generate Curated Data; and (iii) receiving, maintaining, and disclosing the Curated Data according to the terms of this Agreement, the Research Protocol, and the PCORnet and GPC policies. All references to Site Participants include MU when acting in their capacity as a Site Participant and not in their capacity as the Coordinating Center.

h. **De-identified Data**: This term has the meaning ascribed to it in the HIPAA Privacy Rule at 45 CFR Section 164.514(a). Processes for de-identifying data are set forth in 45 CFR Section 164.514(b) of the HIPAA Privacy Rule.

i. **HIPAA**: The Health Insurance Portability and Accountability Act of 1996, the Health Information Technology for Economic and Clinical Health Act, and all implementing regulations, as may be amended from time to time.

j. **HIPAA Privacy Rule**: The HIPAA Privacy Rule (45 CFR Part 160 and Subparts A and E of Part 164), as may be amended from time to time.

k. **HIPAA Security Rule**: The HIPAA Security Rule (45 CFR Part 160 and Subparts A and C of Part 164), as may be amended from time to time.

l. **Individual Level Data**: Data that are not Aggregate Data. Individual Level Data contain information that is specific to individual patients. Individual Level Data may or may not be De-Identified Data.

m. **Legal Requirements**: includes all applicable federal, state and local laws, guidelines, guidance, rules, regulations, ordinances, standards, and policy requirements, including, but not limited to: (i) the Common Rule and Legal Requirements relating to human subjects research, including applicable Legal Requirements of the U.S. Food and Drug Administration and the U.S. Department of Health and Human Services Office for Human Research Protections; (ii) HIPAA; (iii) all other applicable Legal Requirements governing the privacy, security, or protection of Personal Information; and (iv) all Legal Requirements to which each Party has agreed to comply in the PCORI subawards.

n. **Limited Data Set ("LDS")**: This term has the meaning ascribed to it in the HIPAA Privacy Rule at 45 CFR Section 164.514 (e).

o. **Minimum Necessary**: This term has the meaning ascribed to it in the HIPAA Privacy Rule at 45 CFR Section 164.514(d).

p. **PCORnet Policies**: Policies adopted by PCORnet leadership and made available to Site Participants regarding the operation and governance of PCORnet, which may be amended from time to time.

q. **PPRL Hash:** PPRL stands for Privacy-Preserved Record Linkage, which describes a set of methods and/or software for performing linkage across multiple data sources without exposing or exchanging PHI information.

r. **Protected Health Information ("PHI")**: This term has the meaning ascribed to it in the HIPAA Privacy Rule at 45 CFR Section 160.103.

s. "**Research**" or "**research**": shall have the meaning given to that term by HIPAA at 45 C.F.R. § 164.501, and may also include related public health analyses.

t. **Secure File Transfer Method**: A method to securely transfer Data between Site Participants and DCC. This method will be specified for each Query and Data transfer by the DCC, subject to Site Participant's approval.

u. **Structured Data**: This term is defined as the portion of Contributed Data that can be organized into a standardized format.

v. **Site Participant Data ("Data")**: Data generated, collected, processed, maintained, held or stored by Site Participant locally in connection with its participation in CISTEM2, which may be transferred to the DCC in response to a study-specific Query.

w. **Unsecured PHI**: PHI that is not rendered unusable, unreadable or indecipherable to unauthorized persons through the use of a technology or methodology specified by the Secretary in the guidance issued under section 13402(h)(2) of Public Law 111-5, as the term is defined in 45 CFR 164.402.

x. **Unstructured Data:** This term is defined as the portion of Contributed Data that cannot be organized into a standardized format.

y. **Study Team:** Study Team consists of authorized members who are permitted to access either Contributed Data or Curated Data generated for the CISTEM2 study. It is mandatory for the approved study team members to be included in the Institutional Review Board (IRB) documentation or to be recorded in a comparable manner that demonstrates their compliance with training requirements for conducting research involving human subjects.

z. **Data Use Approval:** Data Use Approval shall refer to the approval process established by the GPC Data Request Oversight Committee (DROC) or any equivalent process mutually agreed upon by all Parties involved.

2. Contributed Data.

a. **DCC Receipt, Use and Disclosure of Contributed Data**.

i. Each Site Participant agrees to create, organize, transmit, and disclose Contributed Data to the DCC from time to time in accordance with PCORnet and GPC Policies and Research Protocol. Each Site Participant hereby grants (or has caused the applicable licensor or rights-holder to grant) to the DCC a worldwide, royalty-free, nonexclusive, limited license, with the right to grant sublicenses through multiple tiers to access, use, and disclose the Contributed Data for curation purposes, solely to the extent such purposes are in accordance with the PCORnet, GPC Policies and the Research Protocol, and applicable Legal Requirements, and as described in this Agreement, which may include:

- the creation of Limited Data Sets from the Contributed Data (yielding Curated Data) created by the DCC;
- the De-Identification of Contributed Data (yielding Curated Data), whether by the DCC;
- the creation and implementation of PPRL Hashes using a software tool, such as Datavant, selected by the Study Team and the use of such PPRL Hashes to deduplicate Contributed Data or to link to death and transplant registries.

ii. Notwithstanding anything to the contrary in this Agreement, in the event the DCC determines any of the Contributed Data provided by any Site Participant is not permitted by this Agreement or otherwise inappropriate, the DCC shall promptly notify the relevant Site Participant and, pursuant to the DCC's policy and practice, may immediately remove or destroy any impermissible or inappropriate data.

b. **Site Participant Responsibility for Contributed Data.** Each Site Participant agrees that:

i. All Contributed Data shall be disclosed by such Site Participant to the DCC for inclusion in the CISTEM2 Database in accordance with PCORnet and GPC Policies and Research Protocol. All Structured Data shall be disclosed by Site Participant in format that is in accordance with the Common Data Model, or as otherwise determined by the DCC.

ii. Site Participant may, upon mutual agreement with the Study Team, be selected to provide Unstructured Data on an optional basis. All Unstructured Data shall be provided in accordance with the standards set forth in the Research Protocol and its supplements, which shall require Site Participants to perform best-practice De-Identification techniques locally, using known patient identifiers and pattern recognition and to transmit Unstructured Data in flat file versions of obfuscated text to the DCC, who will convert such Unstructured Data into Limited Data Sets or De-Identified Data.

iii. Site Participant has collected, used, secured and disclosed, and shall continue to collect, use, secure and disclose, the Contributed Data it provides to the DCC in full compliance with applicable Legal Requirements, including, without limitation, HIPAA.

iv. Site Participant has obtained all necessary consents and authorizations to disclose any Contributed Data provided to the DCC for the purposes set forth in this Agreement, to the extent applicable. The Parties understand and agree that Site Participants may provide Contributed Data that is covered by (i) an Informed Consent and HIPAA

Authorization or (ii) an IRB Waiver, and that such Contributed Data does not constitute Protected Health Information under HIPAA.

v. Site Participant will not knowingly provide Contributed Data to the DCC that misappropriates, violates or infringes any rights of the DCC or any third party.

vi. As between Site Participant and DCC, Site Participant is solely responsible for ensuring that any Contributed Data that such Site Participant provides to the DCC for curation and inclusion in the CISTEM2 Database accurately reflects the information held by the Site Participant (in its EHR or otherwise), and, at the time of the provision of the Contributed Data to the DCC, Contributed Data will be transmitted securely and free from any viruses or any other contaminants or disabling devices and, to the best of its knowledge, any material that may violate or infringe any patent, copyright, trademark, service mark, trade secret or any other right of any third party or any material that is otherwise unlawful, obscene or objectionable.

vii. The DCC shall curate Contributed Data and make available Curated Data to other Site Participants and approved study team members as permitted under this Agreement only in reliance on such Site Participant's representations set forth in this Agreement. Site Participant shall assume and be solely responsible for any reporting obligations under any Legal Requirements or contract arising from such Site Participant's breach of this Section 2(b).

3. <u>Curated Data.</u>

a. **Data Coordinating Center Use and Disclosure of Curated Data.**

i. Each Site Participant hereby grants (or has caused the applicable licensor or rights-holder to grant) to the DCC a worldwide, royalty-free, non-exclusive, limited license, with the right to grant sublicenses through multiple tiers, to access, use, and disclose the Curated Data only to support, maintain, provide and improve the CISTEM2 Database to support the research mission of CISTEM2 in a manner consistent with the Research Protocol, applicable Legal Requirements, and PCORnet and GPC Policies, and as set forth in this Agreement.

ii. The DCC agrees to use, secure, disclose, and maintain the Curated Data, including Limited Data Set and De-Identified Data, in the CISTEM2 Database in full compliance with applicable Legal Requirements, including without limitation, HIPAA. To the extent such Curated Data constitutes a Limited Data Set under HIPAA, the DCC will ensure the approved use of the data complies with the requirements set forth in 45 C.F.R. § 164.514(e).

4. <u>PPRL Hashes.</u> The obligations set forth in this Section 4 apply only to Contributed Data for which the DCC will implement PPRL Hashes and are in addition to other obligations established by this Agreement.

a. Any PPRL Hashes are for the sole purpose of de-duplicating Contributed Data and linking Contributed Data with other data that also includes PPRL Hashes, including death registry data, transplant registry data and other administrative databases, in accordance with the PCORnet and GPC policies and the Research Protocol. The DCC will not use any PPRL Hashes for any other purpose and shall not use such hashes to identify the subject of any Contributed Data or Curated Data provided by Site Participant, nor shall the DCC disclose any PPRL Hashes to any other Party who does not need access to the PPRL Hashes for purposes of linking Contributed Data or Curated Data. The DCC specifically agrees not to

use the PPRL Hashes to identify, re-identify, attempt to identify or to contact the subject of any Contributed Data or Curated Data provided under this Agreement.

b. The DCC shall use appropriate safeguards to prevent the use or disclosure of any PPRL Hashes other than as strictly necessary to link the Contributed Data or Curated Data in connection with the CISTEM2 project.

5. **Compliance Obligations.** The Parties agree to comply with, and take all reasonable actions to maintain compliance with, the PCORnet and GPC Policies and the Research Protocol, and to comply with all applicable Legal Requirements with respect to such Party's activities hereunder. Site Participants shall cause each of its representatives that may have access to the DCC's systems to comply with the DCC's applicable policies and procedures.

6. **Governance.** Notwithstanding the term and termination provisions set forth at Section 10 and 11 herein, the core study team member at its discretion may choose not to renew a Site Participant's participation in the CISTEM2 project if the Site Participant is found to be not in good standing in accordance with the PCORnet and GPC policies, Research Protocol and this Agreement.

7. **Expenses**. Each party shall be solely responsible for all costs and expenses incurred by such Party in the performance of this Agreement except as otherwise provided in this Agreement. For the avoidance of doubt, each Party understands it shall receive certain funding relating to the CISTEM2 project pursuant to the NIH Subawards.

8. **Acknowledgements; Disclaimer; Limitation of Liability; Insurance**.

a. **Acknowledgements**. The Parties acknowledges that the DCC is creating and managing the CISTEM2 Database solely in furtherance of the Research Protocol.

b. **Disclaimer**. The CISTEM2 Database, and access thereto, is provided on an "AS IS" and "AS AVAILABLE" basis. To the maximum extent permitted by applicable law, the DCC expressly disclaims all warranties and conditions, either express or implied, with respect to the CISTEM2 Database, including, but not limited to: all implied warranties and conditions of merchantability, title, fitness for a particular purpose, adequacy, or suitability for a particular purpose, use, or result, or arising from a course of dealing, usage, or trade practice; and any warranties of freedom from infringement of any domestic or foreign patents, copyrights, trademarks, trade secrets or other proprietary rights of any Party. The DCC specifically disclaims any warranty that the CISTEM2 Database will meet any Site Participant's requirements or will operate in combinations or in a manner selected for use by such Site Participant, or that the CISTEM2 Database will be provided without error or interruption. The DCC shall not be liable for any losses or damages resulting from or due to (i) mistakes or errors in Curated Data or (ii) service interruptions or downtime with respect to Site Participants' access to or use of the CISTEM2 Database, even if the Coordinating Center has been advised of the possibility of such damages. The Parties agree not to seek to hold the DCC or any of the DCC's directors, officers, employees, medical staff, agents, contractors liable for the consequences of any action or inaction of any other Site Participant and/or External Collaborator.

c. **Limitation of Liability.** In no event will any Party be liable for any indirect, special, consequential, incidental, punitive or non-contractual damages or lost profit or income arising out of or related to this Agreement, even if a Party has been advised of the possibility thereof. Each Party will be solely responsible for obtaining, and hereby represents that it has obtained, any necessary approvals and authorizations from individuals and Data contributors, regarding the inclusion of Contributed Data in the CISTEM2 Database. Each Party agrees to be responsible for the negligence, of any of its Authorized Users, and for assuring the Authorized User's compliance with the agreement(s) governing the Authorized User's submission to and receipt of Contributed Data. No Party will be responsible for claims, expenses, damages or liabilities arising out of the negligence or wrongful act or omission of another Party or that other Party's agents, servants or employees in connection with this Agreement. Liability for Parties that are state institutions is limited to the extent of liability incurred under the Party's applicable state tort claims act. Each Party shall be responsible for its own negligent acts and omissions under this Agreement and the negligent acts or omissions of its employees, officers, or directors, to the extent allowed by law. Under no circumstances will any Party be liable to another Party for any indirect or consequential damages of any kind, including lost profits (whether or not the Parties have been advised of such loss or damage) arising in any way in connection with this Agreement.

d. **Insurance.** Each Party shall maintain in force at its sole cost and expense with reputable insurance companies, insurance of a type and in an amount reasonably sufficient to protect against liability hereunder. In addition to such insurance and/or in the alternative, a Party may maintain a program of self-insurance to protect against the same. Each Party shall have the right to request the appropriate certificates of insurance from the other Parties for the purpose of ascertaining the sufficiency of such coverage. Notwithstanding any other terms or conditions of this Agreement, no state/federal public institution that is an instrumentality of a state/federal government shall be required to comply with the insurance requirements of this Section so long as such institution relies on the applicable law of its state/federal jurisdiction to protect and limit its liability as an instrumentality of such state/federal government.

e. THE FOREGOING LIMITATION OF LIABILITY WILL COVER, WITHOUT LIMITATION, ANY OTHER INJURY, ARISING, DIRECTLY OR INDIRECTLY, FROM A PARTY'S USE OF THE DATA AS PERMITTED HEREUNDER.

9. **Term.** This Agreement shall be effective as of the Effective Date hereof and shall continue until the completion of the study (March 31, 2029) ("**Initial Term**"). This Agreement will thereafter be automatically renewed for three days (3) successive one-year terms (each a "**Renewal Term**" and together with the Initial Term, the "**Term**"), unless terminated in accordance with Section 11.

10. Termination.

a. Any Party may terminate its participation in this Agreement immediately for any or no reason after providing written notice to the DCC, upon at least ninety (90) days' prior written notice.

b. The CISTEM2 study team and DCC may terminate this Agreement pursuant to Section 6 above.

c. **Effect of Termination of Entire Agreement.**
   i. All licenses granted in this Agreement with respect to Curated Data that have not already been approved in a Data Use Approval shall immediately terminate; provided, that any licenses granted in this Agreement to Curated Data that are already subject to a Data Use Approval shall continue solely for the purposes set forth in that Data Use Approval;
   ii. Any continued use of the Contributed Data and Curated Data after the termination date shall be in full compliance with the Data Use Approval and this Agreement;
   iii. The following Sections will survive termination of the entire Agreement: Sections 1, 7, 8, 9, 10, 11 and 13. Sections 2 and 3 shall survive solely to the extent necessary to permit the DCC to destroy Contributed Data and Curated Data and to permit the use of Curated Data to the extent permitted by a Data Use Approval issued prior to the date of termination of this Agreement.

11. <u>Confidentiality</u>. "**Confidential Information**" means any information i) disclosed by a Party to this Agreement to the other Party of this Agreement, ii) disclosed during the term of this Agreement, iii) that was reasonably disclosed in furtherance of the Research Protocol, and iv) designated confidential in writing by a Party, or if given orally, is confirmed in writing as having been disclosed as confidential or proprietary within a reasonable time (not to exceed thirty (30) days) after the oral disclosure and disclosed to another Party hereunder. Confidential Information shall not include the Curated Data, which shall be subject to the applicable restrictions on use and disclosure set forth in Sections 2, 3, and 4 hereof. Confidential Information shall also not include information that the receiving Party can demonstrate: (a) is, as of the time of its disclosure, or thereafter becomes, publicly available through a source other than the receiving Party; (b) was known to the receiving Party as of the time of its disclosure; (c) is independently developed by the receiving Party without use of the Confidential Information; (d) is subsequently learned from a third party that is not under a confidentiality obligation to the disclosing Party. Each Party agrees that, during the Term and for six (6) years thereafter: (i) it will not disclose to any third party any Confidential Information disclosed to it by the other Party except as expressly permitted in this Agreement; (ii) it will not use any Confidential Information disclosed to it by the other Party for its own purposes (except as necessary to perform its obligations or exercise its rights under this Agreement); and (iii) it will maintain the confidentiality of all Confidential Information of the other Party in its possession or control, which obligation shall be satisfied by the receiving Party using the same effort it uses to protect its own confidential information of a similar nature, but in no event less than reasonable care. Notwithstanding the foregoing, any Party may disclose Confidential Information (x) to the extent required by a court of competent jurisdiction or other governmental authority or otherwise as required by law, provided that such Party notifies the disclosing Party so that the disclosing Party may seek a protective order before such disclosure; (y) on a "need-to-know" basis under an obligation of confidentiality to its and its affiliates' legal counsel, accountants, employees, agents, contractors and consultants, provided that such individuals are bound by obligations of confidentiality at least as restrictive as those required for a Receiving Party under this Agreement; or (z) to the other Parties in compliance with this Agreement. Upon termination of this Agreement, or at any time the disclosing Party may so request, the receiving Party will deliver promptly to the disclosing Party, or, at the disclosing Party's option, the receiving

Party will destroy, all Confidential Information of the disclosing Party obtained hereunder (and all copies thereof) that the receiving Party may then possess or have under its control. Notwithstanding the foregoing, the receiving Party may retain one (1) copy of Confidential Information for archival purposes, solely for verification of compliance hereunder and/or in accordance with applicable laws or regulations; provided, however, that any Confidential Information so retained shall continue to be subject to the confidentiality obligations of this Agreement.

12. Change in Law. Upon the enactment of any law or regulation affecting the use or disclosure of Data, or the publication of any decision of a court of the United States or applicable state relating to any such law, the publication of any interpretive policy or opinion of any governmental agency charged with the enforcement of any such law or regulation, or the opinion of counsel, the Parties may, amend this Agreement in such manner as the Parties determine necessary to comply with such law or regulation. If the Parties are unable to agree on an amendment within thirty (30) days thereafter, one of them may terminate this Agreement on written notice to the other.

13. Miscellaneous.

a. **Amendments**. From time to time, the DCC may need to amend this Agreement to respond to changes in applicable Legal Requirements, PCORnet and GPC policies, or other operations deemed necessary for CISTEM2 project. Any such amendments shall take effect upon the sooner of the effective date of the change precipitating the amendment or thirty (30) days after notice by the DCC to the Site Participants of the need for the amendment, and unless otherwise required by such change, shall only apply to Contributed Data provided and Curated Data created after the time the change becomes effective. In the event that Site Participant disagrees with any such amendment, it may terminate this Agreement upon written notice to the other Parties in accordance with Section 10.

b. **Reporting; Breach.** Each Party agrees to report, within five (5) days of discovery, any use or disclosure of Data not provided for by this Agreement, of which it becomes aware, to the Privacy Officer of the DCC and the Privacy Officer of the Site Participants that supplied the Data. The Parties agree to cooperate in the handling and mitigation of any unauthorized use, disclosure or breach of Data in accordance with the requirements of HIPAA and any other applicable laws.

c. **Successors/Assigns**. The provisions of this Agreement shall be binding upon and shall inure to the benefit of the Parties hereto and to each of their respective successors and assigns, if any.

d. **Force Majeure**. No Party shall be liable to any other Party for failure or delay in fulfilling its obligations under this Agreement to the extent such failure or delay is due to causes beyond its control, including, without limitation, acts of God or government, labor disputes (other than labor disputes involving a Party's employees, which shall not excuse that Party's non-performance), inability to secure labor, acts of war or terrorism, riots, natural disasters, pandemics, power surges or power failures, malfunctioning data transport and telecommunication lines, utilities, Internet connectivity and computer

problems.

e. **Relationship of the Parties**. This Agreement will not be deemed or construed to create any partnership, employer/employee, joint venture or agency relationship among any of the Parties. Nothing in this Agreement will be construed to constitute or appoint any Party as the agent or representative of any other Party for any purpose whatsoever, or to grant to any Party any right or authority to assume or create any obligation or responsibility, express or implied, for or on behalf of or in the name of the other, or to bind the other in any way or manner whatsoever.

f. **Entire Agreement**. This Agreement, including exhibits incorporated herein, constitutes the entire agreement of the Parties relating to the subject matter of this Agreement.

g. **Severability**. If any provision of this Agreement shall be held invalid, illegal or unenforceable for any reason, the validity, legality and enforceability of the remaining provisions shall not in any way be affected or impaired thereby and such remaining provisions shall be severable and remain enforceable to the full extent permitted by law.

h. **Counterparts**. This Agreement, including any amendment, may be executed via original signature or mutually acceptable electronic signature verified by a commercially acceptable third-party verification (i.e., Adobe Sign, Conga Sign, DocuSign), and delivered in one or more counterparts, each of which shall be deemed an original, and all of which together shall constitute one and the same instrument. A facsimile, electronic copy, or other reproduction of this Agreement shall be deemed an original.

i. **Notices**. Any notice to be given to a Party shall be given in writing and delivered to the following addresses by certified or registered mail, return receipt requested, or in person with proof of delivery. Such notice shall have been deemed received upon the date of mailing if by certified or registered mail or electronic mail and upon the date of delivery if by private courier or hand delivery:

Site Participant:	Name: Title: Institution: Address: Email:
MU:	Name: Title: Institution Address: Email:

[REMAINDER OF PAGE INTENTIONALLY BLANK]

IN WITNESS WHEREOF, the Parties have caused this Agreement to be executed by their duly authorized representatives as of the day and year first written above.

THE CURATORS OF UNIVERSITY OF MISSOURI

By: _____
Signature

Name:

Title:

Date:

Site Participant

[Site's Full Legal Name]

By: _____
Signature

Name:

Title:

Date:

EXHIBIT A
Research Purpose and Study Protocol

This Research Purpose and Study Protocol ("**Protocol**") defines the scope of the study titled Choosing IS regimens in kidney Transplant by Efficacy and Morbidity 2 (**CISTEM2**). In this study, we will establish a novel, robust and curated database (CISTEM2 database) integrating transplant registry data with multi-site electronic medical records (EMRs), administrative claims, and social determinants of health data for renal transplant patients leveraging the PCORnet infrastructure. To inform patient management in the current complex transplant ecosystem, we propose to enhance our prior work with the following **Specific Aims**:

Aim 1: Recognizing the data continuum gaps in previous studies, including the lack of longitudinal and granular clinical observations and outcome measures, such as serum creatinine levels (for computation of estimated glomerular filtration rate, eGFR), tacrolimus drug levels, measures of viremia and viruria, malignancy diagnoses, and rehospitalization events, we will establish a novel, robust and curated dataset (CISTEM2 Dataset) integrating transplant registry data with multi-site EMRs, administrative claims, and social determinants of health data for KT recipients, leveraging the PCORnet infrastructure.

Aim 2: We will develop longitudinal machine learning algorithms to dynamically suggest IS strategies that optimize renal allograft function (at 1-, 3- and -5-years post-KT), reduce cost, and limit IS comorbidities identified by patients as contributors to diminished quality of life.

Aim 3: We will validate the predictive models refined in Aim 1 and the ML models developed in Aim 2, using two additional PCORnet sites with 12 more KT centers (Total Sample: 40,535 KTs). Using computable phenotypes developed using the CDM model, our temporal-aware ML models will be evaluated to determine reliability in predicting long-term graft function in independent populations. This aim will additionally incorporate data from all three PCORI networks in a distributed learning model to refine the first dynamic clinical decision tool for longitudinal IS management after KT.

A. BACKGROUND and SIGNIFICANCE

A.1 This grant builds on our previously R01-funded study "**Choosing Immune Suppression in renal Transplantation by Efficacy and Morbidity (CISTEM)**". This proposal was funded under NIH PA-12-299 "Ancillary studies of End Stage Renal Disease, Accessing Information from … and Databases". The goal of this grant (funding period 8/1/14 to 7/31/19) was to develop a unique integration of three large national datasets (transplant registry, Medicare claims, and pharmacy fill records), to accurately assess the efficacy and safety of immunosuppression (IS) regimens in kidney transplantation (KT), leading to precision IS selection considering patient and donor factors. The CISTEM results, disseminated via multiple peer-reviewed journal publications, quantified the tradeoffs between effective prevention of acute rejection (AR) and the development of IS related complications (e.g. infection, malignancy, and new onset diabetes mellitus (NODAT)).[3-13] Results of these analyses are accessible to providers via a real-time web-based interface to facilitate personalized IS choice and patient-centered care (https://cistem.wustl.edu). While the initial CISTEM study focused on *de novo* immunosuppression, contemporary patient requires a dynamic, patient-centered

approach which adjusts IS management based on clinically significant events (e.g. AR, infection, malignancy). The current proposal will improve longitudinal care through three **Aims**: First, we will employ CISTEM methods to better assess patient-centered outcomes including allograft function. This investigation is possible through the availability of integrated clinical registry, healthcare claims, and transplant registry data within the Great Plains Collaborative (GPC) Patient Centered Outcome Research Network (PCORnet). Second, we will incorporate machine learning (ML) algorithms to create a dynamic model which better reflects clinical events. Finally, we will apply the statistical assessments and ML algorithms developed in Aim 1 and the PCORnet methods in Aim 2, to 12 additional sites in two additional PCORnet clinical research networks (CRNs), STAR and OneFlorida to validate our dynamic, patient-centered transplant management decision support algorithm (CISTEM2) in a large multiethnic population.

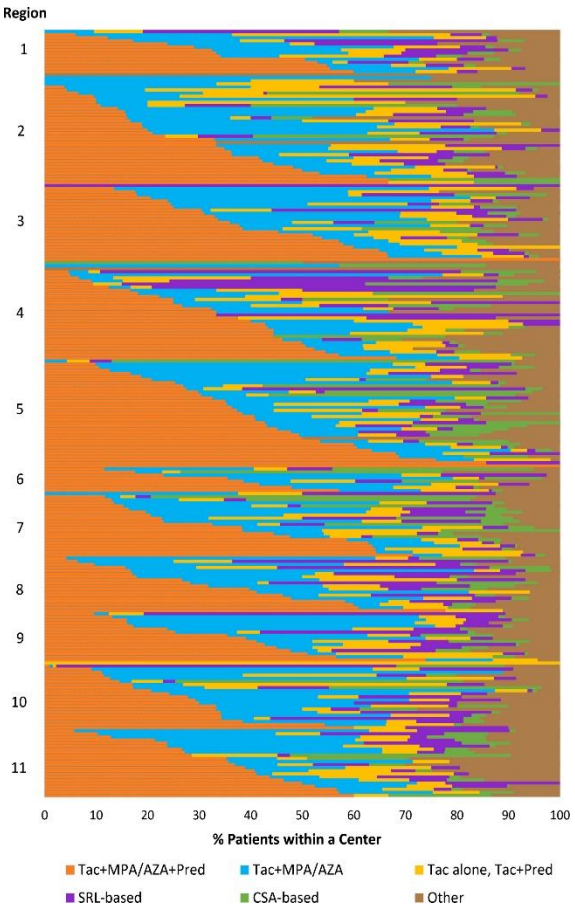


Fig. 1 Proportion of patients receiving one of six mutually exclusive IS regimens during months 6–12 after **KT**. Each horizontal bar represents an individual center within U.S. regions ordered by the proportion of patients that received triple ISx. AZA, azathioprine; CsA, cyclosporine; IS, immunosuppression; MPA, mycophenolic acid; Other, other regimens including CsA withdrawal or other trial medications; Pred, prednisone; Tac, tacrolimus.

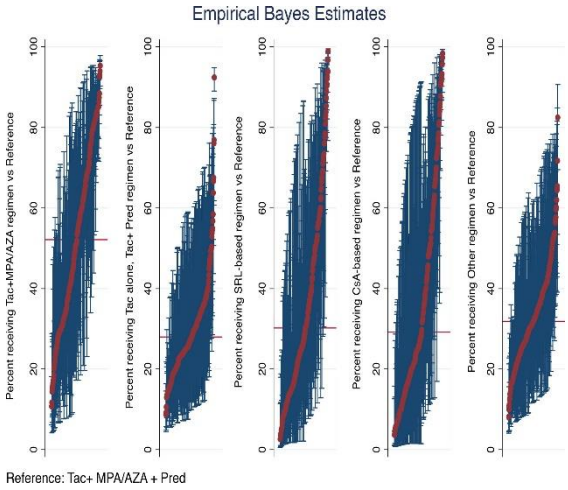


Fig. 2. Empirical Bayes estimates for likelihood of regimen use compared with reference regimen. Red bar demonstrates national average rate of use of each regimen (within pairwise regimen comparisons). Each red dot represents adjusted use at one center, and the blue bars reflect 95% CI for use at the center determined by empirical Bayes estimates, adjusting for case factors of recipients at the center; exclusion of the national average by a 95% CI reflects adjusted center use significantly above or below the national average.

A.1.1. The need for evidence-based IS regimen selection. The US Food and Drug Administration (FDA) highlighted work from the CISTEM study in its workshop, 'Evidence-based Treatment Decisions in Transplantation' in Sept 2018. CISTEM investigators presented our data documenting the substantial and persistent variation in kidney transplant (KT) practice nationally.[1, 7] The choice of IS varies markedly across centers, even those serving nearly identical populations in the same region in the U.S. Use of triple immunosuppression comprising tacrolimus (Tac), mycophenolic acid or azathioprine (MPA/AZA), and steroids (Pred) varied widely (0-100% of patients per program), as did use of steroid-sparing regimens (0-77%), mammalian target of rapamycin (mTORIi)

regimens (0-100%) and cyclosporine (CsA) based regimens (0-78%) (**Fig. 1**). Further analysis confirmed that center specific practice accounted for >40% of these differences.[1] (**Fig 2**). Variation was also demonstrated in the choice of induction medications at the time of transplant.[7] As noted by members of the FDA panel, there is an urgent need to identify relevant clinical markers to reduce IS related toxicity by individualize therapy, identifying graft injury early using noninvasive methods, and developing new data sets to facilitate drug development. We have published data highlighting the association of within-first year IS choice to 3-year clinical outcomes.[10, 14]

A.1.2. CISTEM demonstrates initial IS regimen selection impacts post-transplant morbidity.

Statistically and clinically significant variation in the incidence of complications by IS regimen (**Fig. 3**) was seen in propensity- and risk-adjusted multivariate regression analysis.[2] Compared with the reference IS (Thymoglobulin [TMG] induction with Tac+MPA/AZA+Pred), SRL-based IS was associated with significantly higher 3-year risks of pneumonia, sepsis, diabetes, AR, graft failure, and patient death, but reduced skin cancer risk (P<.001 for all) after KT. CsA-based IS increased the risk of pneumonia, sepsis, AR, and graft failure but reduced the risk of new onset diabetes mellitus after transplant (NODAT). Prednisone-free IS appeared to reduce the risk of pneumonia and sepsis. These data can be utilized to inform IS choice by incorporating patient preferences and medical risk factors such as prior malignancy and known glucose intolerance into IS regimen selection.[15] The impact of each regimen is also modified by the use of induction medications which continue to be largely driven by center preference and partly by drug availability. In yet unpublished data, we analyzed IS regimen through the first year and outcomes out to 5-years after KT. These data were striking in the temporal variation in risks and the impact of early events on later outcomes. For major infections such as urinary tract infections (UTI), early infection (months 0-11) dramatically increased the risk of late UTI (months 12-60). The risk of pneumonia or sepsis in months 12-60 increased significantly (P < 0.001) if the patient had experienced either a UTI, pneumonia or sepsis in months 0-11. AR within months 0-11 increased the risk only for sepsis. Induction antibody use did not affect the risk of any of these three infections at months 12-60, nor did use of mTORi-based or CsA-based regimens. Collectively these results demonstrate the need to adjust long-term management strategies, once key clinical events occur, as subsequent risk mitigation requires dynamic clinical management. Notably, developing evidence-based practice

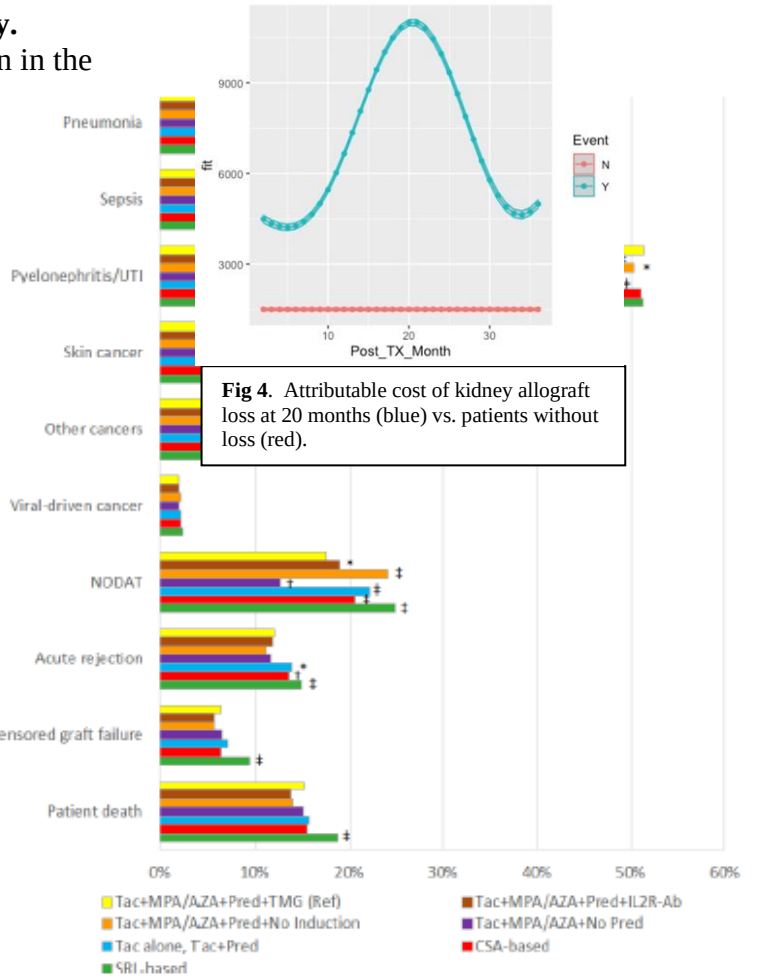


Fig 3. Incidence of complications at 3-years post KT analyzed by selected IS regimen.

Fig 4. Attributable cost of kidney allograft loss at 20 months (blue) vs. patients without loss (red).

requires a robust analysis of a large, multicenter dynamic, longitudinal clinical dataset, which has not been previously available for transplant care.

A.1.3. Clinical Events Impact the Cost of Care and the Cost Effectiveness of Kidney Transplant. Through CISTEM, we also quantified the marginal cost associated with specific post-transplant complications ranging from $17,691 (standard error (SE) $591) for UTI alone, to approximately $134,773 (SE $1876) for those with UTI, pneumonia, and sepsis. Clinical and economic impacts persisted in years 2-3 post-transplant.[10] Among malignancies studied, viral-linked cancer had the largest inpatient and outpatient cost impacts per case, followed by "other" cancer. Non-melanoma skin cancer impacted only outpatient costs. Cancer accounted for 3% to 5.5% of total inpatient Medicare expenditures and 1.5% to 3.3% of outpatient expenditures in the first 3 years posttransplant.[14] Premature allograft failure remains the most expensive event after KT. Our analyses demonstrate that a single graft failure is associated with > $100,000 of increase Medicare expenditures (Fig 4).[16] In addition, patients return initially to dialysis after allograft failure and many subsequently require re-transplantation, creating tremendous economic burdens on the transplant system. These data support the economic benefit of optimal IS management to reduce the burden of complications. Recent shifts in kidney care reimbursement by Medicare to enhance value based purchasing have increased the importance of cost minimization after transplantation.[17]

A.1.4 Successful Kidney Transplantation Improves Patient Report QOL. KT has been demonstrated to result in dramatic improvement in all domains of patient reported quality of life (Physical, Mental, and disease specific). In a meta-analysis of 44 studies (n=6929 KTs), KT resulted in improvement in both physical HRQOL and mental HRQOL results on the SF-36.[18] Similarly, KT improved kidney specific health in the disease specific HRQOL questionnaires.[18] This improvement was consistent across young patients and older adults. However, the impact of IS complications on outcomes correlated with HRQOL (e.g. percent of days free of hospitalization) has been poorly studied to date.[18] Furthermore, there has been no assessment of patients' perceptions of the impact of key posttransplant complications on their HRQOL and tradeoffs between complications of more intense IS versus likelihood of graft loss from under-immunosuppression.

A.2. High-fidelity, large, real world data sets are crucial to transplant management. While historically early threats to transplant graft survival were largely cellular and antibody-mediated AR, contemporary post-transplant patient survival is reduced predominantly by cardiovascular, infectious, and malignant complications.[19] The frequency of these complications varies markedly across recipient populations. NODAT rates are higher in African American (AA) kidney recipients treated with Tac than in either Caucasian patients treated with the same IS or AAs on CsA.[20, 21] Skin cancers are more common in Caucasians maintained on MPA/AZA[22], than in patients on Tac alone (with or without Pred). However, studies of these late transplant complications are generally limited as malignancy data are not well captured in the transplant registry. While alternatives datasets including the SEER cancer database provide some insight, missing data and lack of IS exposure records, preclude firm conclusions[23]. Similarly, estimated glomerular filtration rate (eGFR) after KT has been correlated with long-term allograft survival, IS selection, and patient quality of life. Unfortunately, data on eGFR beyond 3 years is limited by missingness of measurements in the transplant registry.

A.3. **Limitations of clinical trials for evaluating the long-term risks of IS regimens.** Current clinical trials in IS provide limited insight to optimize IS use in clinical practice. Existing trials are constrained due to: short follow-up duration; enrollment of small, select, low-risk populations; and suboptimal reporting of serious infections[24] From the Orion, Symphony, Caesar, Benefit, Freedom, Spiesser, Smart and Ascertain transplant trials,[25-33] the sepsis rate was reported only in the Benefit study, with a frequency of just 1% [29]. Pneumonia frequency was

reported only in the Benefit, Spiesser, and Smart studies,[29, 31, 33] and ranged from 2-16% over 1-3 years.[29, 31, 33] The incidence rates are markedly lower than data from United States Renal Data System (USRDS) and our CISTEM published data. Prospective trials report low cumulative cancer incidences[25-28, 30, 32, 33] and have inadequate power to detect endpoints with low clinical frequency but high impact. [21,36-39,41] Larger, robust sets such as CISTEM2 will provide more accurate assessment of infection and cancer risk, stratified by recipient demographics, clinical characteristics, and donor factors.

A.4. Novel multicenter data sets. CISTEM data suggest that factors beyond traditional selection to drive clinical outcomes. However, certain key clinical outcomes are not identifiable even with registry or national claims data (e.g. viral infections, chronic rejection, severity of cardiac disease, degree of diabetic control). Accurate cardiovascular risk stratification, lab data (e.g. IS drug levels, HbA1c), improved definition of infection type, and complete staging of malignancies are needed to provide evidence-based recommendations. Fortunately, EMR data are increasingly collected using field-defined variables which can be extracted and collated from multiple centers using a common data model. When combined with natural language processing (NLP) and ML methodologies, these robust datasets can be used to validate and refine CISTEM models and identify novel associations. Prior attempts have been limited by the lack of a robust and consistent data platform. PCORnet's common data model (CDM) provides robust, consistent data mapping and quality assurance. Because CDM-compliant datasets have been incorporated into health system data warehouse infrastructure, it permits the development of easily deployable models for clinical decision making. PCORnet data have been successfully used in the development of predictive models which quantify the risk of acute kidney injury and estimate survival from COVID-19 in diverse populations.[34, 35] PCORNet data offer particulary advantage for transplant populations as this patients have longitudinal follow-up within their transplant center.

A.4.1. Data linkage. PCORnet data provide the opportunity for linkage to additional datasets through tokenized protected health information (PHI). Accurate assessment of social deprivation index can be accomplished though geolocation. Tokenized linkage is also possible with transplant registry data and via linkage to the United Stated Renal Data Service, to Medicare claims for post-transplant patients.

A.4.2. Data linkage experience. Our proposed research team has extensive experience in the linkage of healthcare claims, transplant registry, and additional datasets using tokenized data. This has led to multiple high-impact publications demonstrating variation in clinical practice and outcomes in kidney, liver, and heart transplant.[1, 7] We are pleased to add researchers from the University of Missouri who have NIH funding to utilize linked PCORNet datasets to evaluate Amyotrophic Lateral Sclerosis (ALS). The ALS projects uses the tokenized data linkage, EMR data, Medicare, and Medicaid claims as proposed in this project.

A.5. Addition of Machine Learning (ML) analyses. Traditional analytic approaches using multivariable regression analyses, which form the core of CISTEM1, are limited by the need to determine both the variables of interest and their functional form. This precludes the incorporation of all data elements and identification of heretofore unidentified associations. This analytic limitation is particularly pronounced when analyzing complex longitudinal data sets as proposed in the current study. ML techniques including artificial Neural Networks, Random Forest and Reinforcement Learning can analyze larger data sets without pre-selection of important variables. ML successfully applied in other areas of transplant care, to predict graft outcomes and improve organ allocation.[36, 37] Thus far, ML techniques have not been applied in a robust analysis of IS practice. We propose to utilize both traditional modelling with time-varying covariates (AIM 1) and ML models (AIM 2) to assess the impact of IS selection on outcomes.

A.5.1 Models based only on data acquired prior to or at the time of transplant do not permit dynamic management of IS to optimize patient care. While effective in predicting graft outcomes for patient populations, to permit risk adjustment in performance assessment, data from the transplant registry collected at the time of transplant are not sufficiently precise to inform ongoing clinical care. In addition, prior models have focused on binary outcomes (death, graft loss) rather than dynamic measures of allograft function (e.g., eGFR) which impact patient HRQOL. Assessment of these outcomes is critical given widespread efforts to increase use of higher-risk donor organs and reduce organ non-utilization ('discard'). Furthermore, early posttransplant events, particularly delayed graft function (DGF), have long term effects on allograft survival which should inform immunosuppression management decisions.

A.6 Maximization of organ function is vital to improve the health among patients negatively influence by Social Determinants of Health (SDOH). Patients of lower socioeconomic status, racial/ethnic minorities, and other at-risk populations have historically experienced higher rates of allograft loss.[38] Explanations of this differential survival include assertions of non-adherence. Yet, there are well established pharmacogenomic differences that impact IS effectiveness. While detailed pharmacogenomic data are not available for this study, proxy values including mg of calcineurin per kg of body weight can be calculated from EHR data and incorporated into CISTEM-based predictive models. In addition, we will include both individual characteristics available in the registry (employment, educational achievement, and insurance type) and community characteristics derived from geolocation.

A.7. Summary. ML analysis of multicenter data will utilize 1) real world health data to assess the impact of IS regimen, without subject selection bias inherent in claims-based analyses which rely exclusively on analyses of Medicare recipients; 2) describe the association of IS with long-term outcomes not captured in the transplant registry or trials, and 3) provide insight into relatively uncommon but important safety outcomes that are challenging to study with adequate statistical power using standard techniques. Through improved prediction models guided by patient insights, which we will develop in Aims 1 and 2 and validate in Aim 3, we will make an immediate improvement in KT care, by allowing transplant specialists to implement precision medicine-based IS selection using a revised webtool (CISTEM2).

B. INNOVATION

B.1. Health informatics and data integration: The promise of real-world data to evaluate clinical outcomes with ML has been limited by lack of data consistency and HIPAA-compliant data linkages across health systems. To address this issue, PCORI funded the development of the PCORnet infrastructure to support multi-institutional retrospective and prospective pragmatic trials using data sets with limited PHI. The CDM allows high fidelity data linkages and has been used to investigate high prevalence conditions (e.g., acute kidney injury[35], coronary artery disease[39], COVID-19 sub-phenotyping[34]) and rare conditions such as ALS[40].

B.1.1. Greater Plains Collaborative (GPC): GPC is a network of 13 large academic and community health systems which serve more than 30 million patients. GPC extracts data from hospital EMRs into the PCORnet Common Data Platform allowing cross-institutional validated data analysis. Data is linked through unidirectional tokenization of personal identifiable information to preserve privacy. GPC data have also been successfully linked to claims data using tokenization (Medicare and Medicaid) to supplement EMR records. This linkage is vital as EMR data may miss key clinical endpoints. In the national Aspirin Dosing: A Patient-Centric Trial Assessing Benefits and Long-term Effectiveness (ADAPTABLE) trial conducted at PCORnet sites, up to 35% of hospitalizations for stroke or myocardial infarction occurred at health systems outside the enrolling center.[39, 41, 42] Use of claims from the USRDS which includes transplant registry data from Scientific Registry of Transplant Recipients (SRTR) linked

to PCORnet data will allow for construction of the most robust and representative transplant cohort ever assembled.

B.1.2. Computable phenotypes: The PCORnet CDM includes diagnoses, laboratory data, and procedures which can be combined into "computable phenotypes" to identify specific clinical outcomes. These computable phenotypes can be deployed across centers to accurately identify clinical complications without the need for detailed clinical chart review. This study is unique as the integration of SRTR data will allow for robust capture of key outcomes (death and graft failure) to supplement the computable phenotypes. Prior investigations have noted that the recall and precision of computed phenotypes differ significantly based on the inclusion of specific characteristics (e.g., laboratory values).[43-45] This project will improve the posttransplant computable phenotypes developed using health claims data in CISTEM1 (e.g. infection, malignancy, rejection) by incorporating laboratory values, problem lists, and inpatient pharmaceutical administration data available through PCORNet CDM tables supplemented with additional analyses of field defined data as necessary.

B.1.3 Validation across PCORnet sites. Although GPC provides a robust, geographically, and ethnically diverse population, the analytic population is limited to 11 transplant programs. The robust PCORNet infrastructure allows models developed in the GPC to be easily validated through data collected in the OneFlorida and STAR networks. Both networks use a consistent tokenization platform for linkage, allowing EMR data to be combined with transplant registry/USRDS datasets. The CDM allows this process to be rapidly deployed, tested, and refined using data from any participating PCORnet site. In addition, once developed, any health system which utilizes the freely available PCORNet CDM within its clinical data warehouse infrastructure can utilize the CISTEM2 tools for their patients regardless of their membership in PCORnet.

B.2. Machine Learning for Dynamic Modeling. We will leverage a variety of state-of-art statistical and ML models designed to analyze longitudinal data (i.e., temporal-aware machine learning models) to not only dynamically incorporate time-varying covariates and exposures so as to make robust rolling predictions for outcomes of interest, but also identify optimal treatment sequences with respect to health or condition states, latent or manifested.

B.2.1. Dynamic Prediction. In CISTEM1 as well as majority of the existing literature, predictions are often made at a single time point (i.e., at the time of transplantation) based on data at baseline, without continuously updating over time as more information about patient's clinical course is collected and made available. Such models are also called "offline learning" models, which often fail to capture important time-varying information (e.g. rejection, infection) resulting in less accurate outcome predictions. In contrast, ML algorithms inherently support "online learning" or "incremental learning" models, which dynamically adapt to new data without forgetting existing knowledge. In particular, we will leverage two state-of-the-art learning techniques: a) Landmark Boosted Tree Model (LMBT)[46]; and b) Functional Joint Model (FJM) [47-49]. FJM semi-parametrically modeled joint likelihood of longitudinal prognostic factors with time-to-event survival outcome, showing promise in predicting kidney allograft failure, chronic kidney disease progression, and Alzheimer disease cognitive function decline.[47-51] LMBT is a non-parametric tree-based boosting model which can enhance prediction over time by introducing a robust way to incorporate high-dimensional and time-varying factors.[46]

B.2.2. Dynamic Treatment Regimes. Dynamic treatment regimens (DTRs), alternatively named as dynamic treatment policies,[52] adaptive interventions,[53] or adaptive treatment strategies,[54] provide a new paradigm to automate the process of developing new effective treatment regimens for individual patients with long-term care. [45, 55] A DTR is composed of a sequence of decision rules to determine the course of actions (e.g., treatment type, drug dosage, or re-examination timing) at a time point according to the current health status and prior treatment history of an

individual patient. Unlike traditional randomized controlled trials that are mainly used as an evaluative tool for confirming the efficacy of a newly developed treatment, DTRs are tailored for generating new scientific hypotheses and developing optimal treatments across or within groups of patients. We will adopt two types of well-established methods for modeling DTR: a) Reinforcement Learning (RL)[56] or Markov Decision Process (MDP)[57]; and b) marginal structural model (MSM).[58] The parametric G-formula and MSM are both statistical methods widely used in causal inference for effectively controlling for time-varying confounding and uncovering direct treatment effects, which are the dominating methods used for DTR applications for the past decade.[59] However, MSM is restricted to pre-defined treatment sequences, while RL/MDP could allow more exploratory search for generating new hypotheses on optimal treatment strategies over time, an approach gaining increasing attention, especially for cancer[60, 61] and sepsis treatments.[62] Nonetheless, to the best of our knowledge, both methods have yet to be effectively utilized to address DTR problem for KT patients such as the optimal strategy for management of patients with DGF who remain AR-free at one year.

C. APPROACH

C1. Integration of national transplant registry, electronic medical records and claims data.

C.1.1. The Scientific Registry of Transplant Recipients (SRTR) Registry. The SRTR registry collects clinical and demographic data for all patients listed for and receiving organ transplants in the United States. In addition, the SRTR collects deceased and living donor data. Post-transplant follow-up regarding AR, graft function and patient survival is collected at 6 and 12 months after transplant and then annually.[63]

C.1.2. Medicare Claims. Medicare claims include diagnosis codes, procedural codes, location of care, as well as associated charges and payments. Medicare served as the primary insurer for ~57% of KT from 2008-17. The CISTEM team has already linked Medicare to SRTR for transplant recipients, thereby overcoming major administrative and technical hurdles in supplying a valuable resource for the current project.[2, 13, 64] The investigative team possesses 2008–2020 Medicare claims data and will purchase 2020-2024 data when available. In addition, the investigative team will also purchase 2015–2020 Medicare Advantage (MA), or Medicare "Encounter" data as more than 30% Medicare beneficiaries are enrolled in MA program with an increasing rate over time, whose claims could be sparse in the traditional fee-for-service Medicare data files.

Table 1. Number of KT Recipients across Participating Sites

Transplant Hospital/ Health System	PCORnet CRN	#KTx 2012-22
University of Missouri (GCC)	GPC	179
Allina Health System	GPC	290
Intermountain Healthcare	GPC	1078
Medical College of Wisconsin	GPC	728
University of Iowa	GPC	836
University of Kansas Medical Center	GPC	1383
University of Nebraska Medical Center	GPC	1448
University of Utah	GPC	1276
UT Health Science Center at San Antonio	GPC	1078
UT Southwestern Medical Center	GPC	989
Washington University in St. Louis	GPC	2526
UT Health Science Center at Houston	GPC	637
GPC Total		**12448**
Vanderbilt University Medical Center	STAR	2286
Duke University Medical Center	STAR	1497
Health Sciences of South Carolina	STAR	2630
Mayo Clinic Arizona	STAR	3407
Mayo Clinic Rochester	STAR	2171
Mayo Clinic Florida	STAR	1673
University of Florida	OneFlorida+	970
University of Miami	OneFlorida+	3726
Advent Health	OneFlorida+	1592
Emory University	OneFlorida+	2766
Tampa General Hospital	OneFlorida+	2749
University of Alabama at Birmingham	OneFlorida+	2620
Total (GPC/STAR/One Florida)		**40535**

C.1.3.PCORnet Clinical Data and Research Network and Greater Plains Collaborative.

PCORnet is funded by the Patient-Centered Outcomes Research Institute (PCORI) to harnessing the power of "real-world" EHR data through a harmonized CDM [65-68] PCORnet CDM is organized in "tables" that reflect type of data (e.g. diagnosis, vital signs, lab results) and include modifier codes for encounter type (e.g. ambulatory visit, telehealth) or source (e.g. order, billing claim). Although data follow PCORnet CDM standards and guidance, the allowable medical ontologies and coding vocabularies are more extensive (e.g. HCPCS, ICD, SNOMED). The Greater Plains Collaborative (GPC) is a founding PCORnet CRN.[69] The GPC network currently includes 13 partner sites in 8 states with a potential observational study pool exceeding 15 million.[69] The project will also collaborate with two addition CRNs: STAR and OneFlorida+ within 12 additional sites (**Table 1)** capturing > 40,000 KT patients.

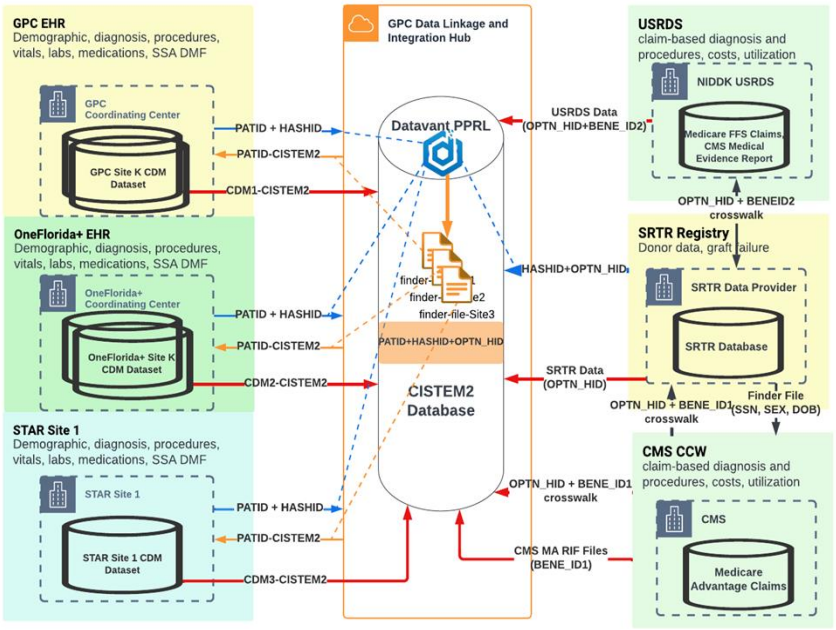


Figure 5. CISTEM2 Data Linkage and Integration.

<u>C.1.4. Privacy-Preserving Database Linkage approach</u>. To prepare the analytical patient cohort, we will leverage PCORnet's selected privacy-protecting linkage solution by Datavant to securely link patient data from all study sites and data sets (Figure 5). [70, 71] All PCORnet sites have an existing Site License Agreement (SLA) in place with Datavant and already have access to the Datavant De-ID software. All HIPAA identifiers will be masked and all dates will be reported as the number of days since the patient's transplant date. Each patient record will include an encrypted token (a hash), a 44-character unique patient identifier that can be reproducibly generated from identifiable patient information from different sites but cannot be reverse-engineered to reveal the original information. Year of birth will be included to adjust for secular trends over time. This approach was previously successfully used to link patient data from different sources in PCORnet-based and other studies.[70-75] As entailed in Figure 5, GPC coordinating center (GCC) will first leverage Datavant hash tokens to identify the crosswalk population between PCORnet EHR cohorts and SRTR registry to generate *site-specific finder files*. The finder files will then be disseminated to participating sites for clinical data extraction (on the crosswalk KT patients. Upon receiving data from participating sites, plus SRTR, USRDS and CMS, GCC will perform data linkage and integration using the mapping among four source-specific, masked identifiers: PATID, HASH_ID, OPTN_HID, and BENE_ID. HASH_ID is the study-specific transit token generated by Datavant De-ID software.

C2. Exposure and Covariates.

C.2.1. Exposure. Recognizing that multiple IS medication changes happen frequently to a given patient, we have developed methods of quantifying predominant IS regimen exposure (Table 2). In CISTEM1, we found that the IS regimens were most optimally classified into 7 groupings shown in Fig 3. Tac + MPA/AZA + Pred + TMG was the most common regimen and was used as reference. Mycophenolic acid (MPA) includes both mycophenolate mofetil (MMF) and mycophenolate sodium. IL-2 receptor antibodies (IL2rAb) includes basiliximab and daclizumab. mTORi-based IS was classified before CsA-based to enable assignment of mutually exclusive regimens, as per previous methods[1]. mTORi and CsA-based regimens were not further sub-classified due to low frequencies of patients treated with these regimens. We will supplement these data with EMR data from the CDM, specific additional EMR queries including laboratory measurements of Tac levels. For patients with CDM and Medicare Part B claims, we will correlate medication fill and administration data to assess accuracy of the EMR data.

Table 2 – Outcome, Exposure and Covariates

Patient/Transplant Characteristics	Exposure	Outcome
Demographic (e.g. age, race, gender)	Induction treatment	Global: death, graft failure, eGFR
Clinical (e.g. diagnosis, years on dialysis, diabetes, vascular disease, prior transplant	Immunosuppression maintenance meds: tacrolimus w/level corticosteroid	Complications/Cost: infection, malignancy, sepsis, hospitalization,
Transplant (Living vs. deceased donor, donor age, race, cold ischemic time, cause of death, eGFR)	Key early clinical events: infection, rejection, delayed graft function	Quality of life: days in the hospital, skilled nursing facility

C.2.2. Covariates. Adjustment covariates available in the SRTR registry include: recipient age, gender, race, ethnicity, blood type, serum creatinine and eGFR (kidney function), comorbid conditions (e.g. diabetes, peripheral vascular occlusive disease, hypertension, prior surgeries); donor type (living, deceased after brain death, deceased after circulatory death), donor age, gender, race and cause of death; cold ischemia time; degree of human leukocyte antigen (HLA) matching; CMV sero-pairing; year of transplant, peak panel of reactive antibodies (PRA), duration of dialysis prior to transplant. Additional covariates will be determined from the EMR data as possible including serial viral loads for CMV, EBV and BK virus, baseline cardiovascular evaluation (ejection fraction) and diabetic control (HbA1c).
*Further, we will adjust for **social determinants of health** covariates, using the Area Deprivation Index or geocoding other social determinants of health by the KT recipient zip code, available in the SRTR registry.*

C3. Outcomes.

C.3.1. Clinical outcomes measures and ICD codes. In Aim 1, we will assess clinical outcomes using the CDM data and SRTR registry. Given the 2008-2020 study period, the claims will contain ICD-9 and ICD-10 codes. Outcomes study include: a) primary outcome measures: AR, graft survival and patient death, assessed from SRTR follow-up form data. We define AR as a center report in the registry SRTR that an AR event occurred in a reporting interval. Death and graft failure (return to dialysis or re-transplantation) are identified in the registry. b) NODAT, infection outcomes (pneumonia, sepsis, UTI/pyelonephritis, cytomegalovirus disease and viremia, hepatitis virus disease or viremia pre- and post-transplant, BK virus nephropathy), malignancy outcomes (viral-driven cancers, skin cancer, other cancers) and cardiovascular outcomes (cardiac failure, myocardial infarction, stroke) will be ascertained from GPC data (in Aim 1) and EMR merged synthetic data (in Aim 3). Diagnosis based on billing claims in the CDM on at least two outpatient or one inpatient claim associated with condition specific ICD-9 (and if available, ICD-10) diagnostic codes. The ICD-9/10 codes used in Aim 1 include: cardiac failure (428.xx/I50.x, I11.0, I13.0, I 13.2), myocardial infarction (410.xx/I21,I22), stroke

(433.xx-435.xx/I60-I63), NODAT (250.0 to 250.93/E08-E13 with first claim after date of transplant, I those without pre-KT diabetes), pneumonia (486.xx/J12-15,J17), sepsis (038.9, 995.xx/R65.21, T81.12), hospitalized pyelonephritis (590.10, 590.15/N10,N11.0,N11.1, N13.6), viral-driven cancers, including lymphoma, Hodgkin's, myeloma, leukemia, Kaposi's, vulvovaginal, cervix (176.x, 184.x, 200.x to 208.x/C81-C96), skin cancer (140.x, 173.x, 187.x/C.44) and other cancers, including mouth, esophagus, stomach, small intestine, colon, hepatobiliary, pancreas, larynx, lung, bone, breast, melanoma, uterus, ovary, prostate, testes, bladder, kidney, CNS, endocrine (141.x-194.x/C00-C80). ICD9/10 diagnoses will be suplemented with additional data drawn from laboratory records in the CDM. Additional data will be considered based on the success of supplemental queries including viral serologies, pathology reports, and IS drug levels. Diagnoses will be restricted to those that appear either in an inpatient record or 2 or more outpatient records. Problem lists will not be used due to inaccuracy.

C.3.2. Complications, graft failure and death. Time to post-transplant cardiovascular, infectious and cancer complications will be computed as the time from KT to the date of clinical-reported diagnoses as defined by billing claims. We will define graft failure as the earliest reported date of return to maintenance dialysis or re-transplantation. Separate ML models will be developed for AR, graft failure, death, and progression of renal function. These models will be dynamic and incorporate the development of complications after transplant (e.g. AR, sepsis, DGF) on the estimates of allograft function.

C.3.3. Associations of IS regimen with post-transplant complications. Separate prediction models will be constructed for each of the infectious, malignant, and metabolic complications using the integrated CDM and transplant registry data. We will also examine a composite outcome comprising any of the infectious, malignant, or metabolic complications. We will supplement our prior analyses, which were based solely on administrative claims,[10, 14] with integrated data to allow detailed validation of claims-based complications with clinical outcomes recorded in the EMR, including chronic allograft damage as assessed by eGFR and development of specific viral infections requiring medical interventions (e.g. CMV treatment with antivirals).

C.3.5. Serum Creatinine and Estimated Glomerular Filtration Rate (eGFR). Renal function will be determined using the race-neutral 2021 CKD-EPI equation. Recipient baseline renal function will be defined as the nadir serum creatinine record 30-60 days after transplant. Subsequent recorded values will be used to determine the trajectory of renal function at prespecified time points (e.g. 3, 6, 9, 12, 18, 24, 36 months) with a window of +/- 30 days.

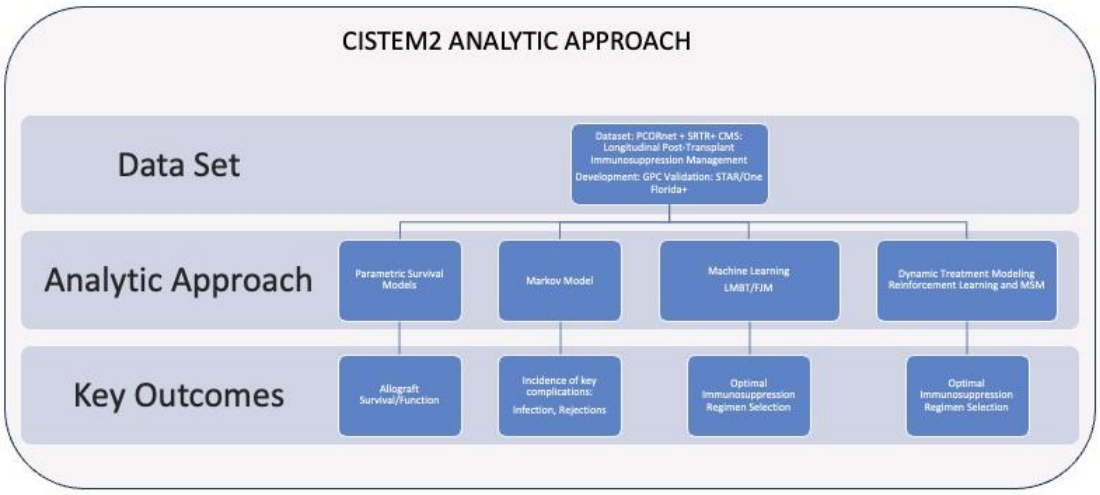
C.3.6. Cost. Resource utilization will be calculated using Medicare payments for patients with Medicare as the primary payer for their transplant including patients with Medicare Advantage plans for who paid claims data are available. Effect estimates will be obtained with general linear modeling following the general structure described above (section **C.2.5**). Reported costs will be Medicare payments for all services provided to a patient. Medicare payments for organ acquisition at the time of transplant are unavailable. All costs are reported in $US adjusted for inflation using the medical component of the CPI with 2020 as the base year.

C.4. Patient Directed Outcomes Assessment and Data Presentation.

C.4.1 Patient Complication Prioritization: With the assistance with Amy Waterman, PhD an experienced collaborator with expertise in patient outcomes assessment, we will assess patient preferences for outcomes/complications. A convenience group of KT recipients will be recruited at two institutions (Washington University and Houston Methodist Medical Center) in year 1 to participate in focus groups. The KT recipients will be asked to assess the importance of key posttransplant complications. We will describe the computed phenotypes in CISTEM and identify additional potential outcomes. This methodology is similar to a "talk aloud" session,

often used in usability studies, but the focus for these initial interviews is to identify outcomes of greatest importance to patients. The total N for the focus groups is 20 (10 per institution). Patients will be offered the opportunity for updates every 6 months on study progress and refine outcome selection. The focus groups will take place in person or by Zoom to provide an online option for participants' convenience and will build on prior experience at Methodist and at Washington University. The video and transcription functions on Zoom will be used to capture the online interviews. Each focus group is estimated to last 45-60 minutes. Participants will receive a $25 incentive for their participation, which will be processed as a check. Staff will gather the information needed to process payment with participants prior to the focus group.
C.4.2 Data Presentation Refinement: A second set of focus groups will be held in year 4 to refine the presentation of the results of the ML model. We will identify which of the outcomes identified in focus group 1 could be defined and modelled. Patient accessible graphical displays will be developed to allow patients and providers to collaboratively choose IS regimens and assess the incidence of key complications.

C.5 Baseline Statistical Modeling.
C.5.1. Statistical Analyses. The first step in these analyses will be to characterize each subject according to which of seven IS regimens they used initially and then at 3,6,12,24, and 36 months after transplant (**Fig. 3)**. For each outcome measure (also depicted in **Fig. 3**), the primary analysis will involve the use of Cox regression models with time to the event as the outcome measure as defined by the new computed phenotypes derived from the PCORI CDM. The predictors in these analyses will include IS regimen, a pre-defined set of covariates that will be applied to all outcome measures, covariates that are specific to particular outcome measures, and a propensity score that measures the propensity for the subject to be on the particular IS regimen. In all of these analyses, the reference IS regimen (**C.1.3**) will be TMG induction with maintenance Tac, MPA, and prednisone. This means that analyses will generate hazard ratios comparing each IS regimen with the reference regimen after adjusting for covariates. The proportional hazards assumption will be evaluated using Schoenfeld residuals. If it is violated, we will pursue remedial approaches such as stratifying by the level of categorical variables that violate the assumption, using the extended Cox model that permits time-varying covariates, and fitting separate Cox models for different time intervals. Pre-defined covariates to be used in all of the Cox models are listed in section **C.2.3**. Multivariate linear regression models be used to



predict the impact of IS selection on eGFR at 1, 3, and 5 years post-transplant after adjustment for the same variables. To account for dynamic changes in treat algorithms, appropriate analytic techniques will be applied for each key outcome (**Fig. 6**)

C.5.2. Parametric Survival Models. For all probability estimates where observed transplant patient data exist, parametric survival (or time-to-event) models will be constructed. Several classes of distributions will be explored, including but not limited to exponential, Weibull, Gompertz, and generalized gamma functions. Parametric estimates will be carefully compared with observed data in terms of applicability to various patient subgroups (stratified by important covariates). Standard censoring models will be compared with competing risk (sub-hazard/sub-distribution-based) models [56-59] to account for potential informative censoring. Those functions that best fit the observed data will be converted to state transition probabilities *p(t),* where t is the cycle number, using the property *p(t) = 1 - S(t+1)/S(t)*.

C.5.3. Markov Analyses. Parametric survival functions and competing risk models will be estimated using SAS 9.4 or higher and/or R. Markov decision process models will be implemented in TreeAge Pro (TreeAge Software, Williamstown, MA) using the Monte Carlo micro-simulation and discrete event simulation options to allow for per-trial sampling distributions to assess the impact of regimen selection after initial clinical events (e.g. the resumption of corticosteroids after AR). Probabilistic sensitivity analysis will also be performed using the same software[60, 61].

C.5.4 Statistical power. The power calculation is mainly based on one of the primary outcomes of allograft failure at 36 months. National data suggests a 20% incidence of allograft failure at this time point. We will have approximately 12500 KT recipients in the GPC data set alone, with an expected 2,500 graft failures.[76] In a previous study, we observed the two most common immunosuppression regimens to be used in 54% and 26% of KT recipients.[2] (As an example of the power available in our expected sample (Table 1), if the graft failure rate in the most common regimen was 20% we would be able to detect a rate of 22.5% in the second most common regimen, a risk ratio of 1.125, with alpha at 0.05, and power achieving 80%. Importantly, if we are not able to detect a statistically significant association between key outcomes and immunosuppression using only the GPC data, we will incorporate data from STAR and OneFlorida+ earlier to increase our power. In this case, alternative methods of validation (e.g. bootstrapped confidence intervals) will be used. Therefore, with power = 80% and alpha = 0.05, we could detect a difference in AR at one-year post-transplant of 10% versus 12.5% between two equal sized groups, an odds ratio = 1.25. Similarly, a difference of between an eGFR of 62 and 63.4 is detectable, a 1.4% difference. With equal groups a difference between 5% and 6.9% death-censored graft failure is detectable, an odds ratio of 1.38. We will thus have vast power to detect statistical significance for many covariates to many outcomes, even with adjusting for multiple testing. We will show point estimates and 95% confidence intervals to quantify the magnitude of the difference, through which we will be able to judge which associations are clinically meaningful.

C.6. Machine Learning for Dynamic Predictions. Each longitudinal patient record can be partitioned into a temporal representation, i.e., a sequence of observations $\{(\boldsymbol{X}_{i_t}, Y_i)|\boldsymbol{X}_{i_t} \in \mathbb{R}^p, y_i \in (0,1)\}_{i=1}^{N_t}$ at each time window t, where Y_i represent the outcome of interests for individual i and $\boldsymbol{X}_{i_t}$ denote time-varying covariates aggregated at time t. Let T_i be the ith subject's survival time to the event of interest or censoring, and (T_i, Δ_i) denote the competing risk data on the subjects, where $\Delta_i = 0$ indicates a censored event and $\Delta_i = d$ that this subject fails from the dth type of time-to events ($d = 1, \dots, g$; e.g., death, graft failure).

C.6.1. Landmark Boosted Tree Model (LMBT). LMBT is an ensemble model adaptively built upon individual boosted tree models trained at each intermediate endpoint, or landmark time.[46] We have developed an efficient and cost-effective algorithm to solve the optimization problem sequentially for all $1 \leq t \leq T$ with increment Δt, $\min E_{t|t-1}[L(y, F_t(\boldsymbol{x}_t, F_{t-1}(\boldsymbol{x}_{t-1}, y_{t-1}))\,)]$. In other

words, we can use the predicted probability from time $t-1$ as the baseline risk and ensemble new boosting models based on features updated at time t. Algorithm details in pseudo-codes can be found in our published work and the corresponding R package for implementation is publicly available.[77] The increment, Δt, will be tuned as hyperparameter using cross-validation starting from 6 months to more refined time scale. In previous studies, we have shown prediction benefits of landmark boosting model over other longitudinal models, especially during the later years. For this proposal, we will take one step further to extract patterns from model post hoc by using SHapley Additive exPlanations (SHAP) [78, 79] values over each of $\{F_t\}_{t=1}^{T}$ to generate a matrix of explanations for each individual $\{\boldsymbol{x}_i, y_i\}$ over time as $g_t(\boldsymbol{x}_i, y_i) = \phi_{i,0}^{t} + \sum_{j=1}^{p_t} \phi_{i,j}^{t}(x_{i,j})$, where $\phi_{t,j} \in \mathbb{R}$ is the attribution for feature j selected at time t. The values of $\phi_{i,j}^{t}$ can be used to reveal temporal patterns of time-varying features as well as identify high-risk event sequences, i.e., prognostic phenotypes.

C.6.2. Alternative Approaches. Alternatively, we will adopt two different approaches for dynamic predictions: a) a traditional parametric statistical model, Functional Joint Modeling (FJM), which has been applied to characterizing the individual dynamic trajectory of longitudinal progression of eGFR in conjunction with corresponding associated time-to-event outcomes.[47, 49] Instead of specifying a fixed form of hazard function being linear over time, FJM built in a flexible Bayesian B-spline model for longitudinal process which allows for nonlinear trajectories[50]; b) a more flexible deep learning model, recurrent neural network (RNN), which has been adopted for dynamic predictions of disease onset[80] and pregrssion[81], as well as drug effects.[82]

C.7. Machine Learning for Dynamic Treatment Regimes (DTRs). We will design a decision process model that answers the clinically relevant question: which IS regimen and dose offers the most health benefits in the most cost-effective fashion to a given patient with a specific recipient and donor and transplant profile. State transition probabilities will be derived from observed transplant populations. The result is a clinically applicable tool that providers can use for IS selection, and risk/benefit counseling. The models consider both initial IS choice and decisions to change or maintain IS regimen and dosing plans following transplant. The goal of this Aim 1 model is to produce tailored, patient-specific comparisons of the relative outcomes and cost-effectiveness between alternative IS regimens.

C.7.1. Reinforcement Learning. The design of DTRs can be viewed as a sequential decision-making problem. A natural algorithmic selection for learning DTRs is Reinforcement Learning (RL). In RL, an *agent* chooses an *action*, or treatment option, at *each time step* based on its current *state*, in this case a set of clinical observations and assessments of patients. The agent receives a reward, in this case treatment outcomes, and the new state from the *environment*. The goal of the *agent* is to learn an *optimal policy* that maximizes the expected accumulated reward it receives over time. The learned policy is a DTR. Modern RL approaches rely on deep neural nets and what are known as policy gradient algorithms to guide the training of the neural nets. There are often two key challenges. First, the neural nets operate on a set of features, often including the state information. However, the most effective RL methods benefit from "feature engineering" that introduces problem-specific features. The second challenge is that RL methods relying on neural nets suffer from the "black box syndrome." That is, the decision making is not explainable or interpretable. To overcome this issue, the learning process must be tailored. Similar approaches have been used by our collaborator Dr. Street and colleagues to optimize warfarin dosing for anticoagulation.[56]

C.7.2. Alternative Approaches. Alternatively, we will adopt two different approaches to evaluate time-varying treatment effects: a) a more conventional statistical method, Marginal

Structural Modeling (MSM), with inverse probability weighting.[83] MSM was introduced in 2000 and has been widely applied to longitudinal data to effectively account for time-varying confounding and identify causal associations of time-varying exposures with interested outcomes under the same standard assumptions.[58, 83-89]; b) a recently introduced RNN-based G-computation approach, or G-Net, developed to generate counterfactual estimates for DTRs.[90]

C.8. Experimental Pipeline. For all the three aims, we will follow a consistent experimental pipeline as described in sections below from data preprocessing to model validation and post-hoc analyses.

C.8.1. Data preprocessing. For each individual i, the raw data will be collected in a longitudinal fashion with pre-defined, consecutive observation windows following routine visit schedules. There will be multiple features within each sub-domain and multiple observations at different timepoints for each feature, consisting of a mixture of categorical and numerical data types. Categorical features will be presented as binary values (1 for present and 0 for absence of the feature) using the "one-hot-encoding" strategy. For continuous features, to account for informative missingness and random outliers among numerical features, we will keep both the raw values as well as perform value discretization with a two-pronged strategy: i) rule-based - group values into "low", "normal", "high" categories using existing medical knowledge base (e.g., BMI of 32 can be categorized as "overweight" based on CDC definition); ii) distribution-based - transform values into percentile ranks to abstract the underlying linear structure (e.g., BMI of 32 can be can also be mapped to a percentile rank of 70, meaning only 30% of the entire population has a BMI above 32).

C.8.2. Data bias. Bias may be present in any real-world datasets, likely resulting from unintentional over- or under-represented populations, correlations in missing data causing important classes to be decimated before training, or natural distributions of data causing rare subgroups to be treated as noise in the inputs that the ML model eliminates. For perceivable under-represented subgroups, we will use oversampling[91] and optimization weighting to inflate their impact to prevent the ML training from ignoring their contributions. We have complete data capture for all patients initially and will ensure there is not disproportionate drop out. For all the aims, we will constantly check for algorithmic biases especially for those associated with the "protected attributes" (e.g., sex, ethnicity, race, veteran status, disability status) by using model fairness metrics such as predictive parity, error rate balance, equalized odds, and overall accurate equality.[92-94]

C.8.3. Data partitioning. The overall eligible study population will first be partitioned cross-sectionally with 12 GPC sites as the master training set and 1 site as external testing site, where each site will rotate to be the external testing set for 12 iterations (i.e., cross-site validation). At each iteration, the master training set will be further partitioned into 80% training and 10% internal testing and 10% temporal testing set. All the model development activities (e.g., bootstrapping, cross-validation) at the training stage will occur within the 80% training data and we will ensure that no information about the testing data is leaked into training, to minimize the chance of overfitting. The cross-site validation will serve as a common ground for fair comparisons of different models and fair reporting of sensitivity analysis. All hyper-parameter tuning for different dynamic models will be performed within the training set using 5-fold cross-validation. Further cross-validation will occur in Aim 3 with the incorporation of PCORnet CDM data from STAR and One-Florida+ centers.

C.8.4. Hybrid Validation Approach with Expert Guidance. We will take an expert-guided hybrid validation approach with a qualitative component and a quantitative component throughout the entire study. For the expert-guided qualitative validation: a) we will start with comparing our results with existing evidence based on a suite of survey paper and systematic

review; b) when evidence is scarce, we will seek expert consensus. Our informatics and clinical team will meet and iteratively review the selection results and provide feedback on their agreement to each selected risk / protective factors using 5-point Likert Scale. We will closely engage other stakeholders, especially our patient advisory council in periodic review of our findings gleaned by the clinical team. We will actively collect feedback and incorporate their suggestions into the study's design, conduct, data interpretation. The quantitative component will be described later under each Aim section.

C.8.5. Evaluation Plan. The quantitative validation portion, model performance is evaluated by measuring the degree of agreement between $F_t(X)$ and y_t on all types of testing sets (internal, temporal, external) with metrics measuring discrimination and calibration. Specifically, discrimination will be measured by area-under-receiver-operating-curve (AUROC), sensitivity/recall, specificity and positive predicted value (PPV)/precision as well as area-under-precision-recall curve (AUPRC), while calibration will be measured both globally using goodness-of-fit statistics (e.g., Hosmer-Lemeshow test) and locally using observed-to-expected-ratios within each risk stratum. Comparisons among models will be further assessed by net reclassification improvement (NDI) and integrated discrimination improvement (IDI).

C.8.6. Subgroup Analyses. We will first examine *clinically important subgroups* for IS decision making individually including recipients that are: AA or other minority groups; living donor source, children, the elderly, women of childbearing age, preemptive transplant, retransplant, highly sensitized, survivors of previous malignancies, and others. We will not report results if cell sample size is insufficient. We will then exploit data-driven approaches such as clustering algorithms (k-means,[95] hierarchical,[96] density-based methods[97]) to discover *phenotypic subgroups* who may benefit the most or the least from different treatment regimes.

C.9. Sensitivity Analysis. The sensitivity of the results to possible variation in input values will be assessed with several methods. First, in a one-way analysis, the effect on the results will be assessed by varying the input values to the 5th and 95th percentiles of their expected distributions based on standard errors estimated from the data. Second, a 10,000 iterations simulation will be run with input values drawn from their expected distributions based again on standard errors estimated from the data. Third, utility values and discount rates, which are drawn from the literature, will be varied along plausible ranges if estimates of their distributional characteristics are not available in the data.

C10. Decision Support Tool. This project will result in novel clinical support tool building on the current CISTEM complication calculator to provide insight into the estimated risk of graft failure, kidney function (eGFR) and rate of patient prioritized complications (**C.4.1**) given a patient's history and current allograft function. The presentation of these results will be refined to provide patient accessible presentation based on focus group feedback (**C.4.2**).

C.11. Potential problems and proposed solutions.

C.11.1. Despite the team's deep experience in data linkage, it is possible the hash linkages cannot be generated to link the PCORI CDM with all data sources (USRDS, SRTR, and Medicare advantage). In this case, the analysis can be performed using only SRTR/CDM or USRDS/CDM. In the former, clinical outcomes occurring outside of PCORI participating institutions may not be captured, however complete data on graft failure and death will be available to determine long term outcomes using captured events. Similarly, if only USRDS/CDM linkage is possible, the analyses will be restricted to Medicare patients with full data within the USRDS. However, as Medicare is the payer for > 60% of transplants, we believe a robust analysis is possible.

C.11.2 Development of ML algorithms requires robust data sources. While the GPC/SRTR/USRDS linkage represents the largest collection of multicenter linked clinical and claims data, it may still be insufficient. In this case, additional cases will be identified in the STAR and OneFlorida collaborative groups and used for model development. In this case alternative methods will be used for validation including bootstrapping, resampling, and data partitioning.

C.11.3 All real-world analyses are subject to the risk of missing or inaccurate data. We propose to address this by using robust computed phenotypes that include both clinical elements (laboratory values) and administrative data. Computed interpolation will be used for missing data as appropriate. Subjects will be censored at last documented follow-up at the transplant center to reduce type II errors caused by loss of follow-up.

C.11.4. The format of predicted outcomes reporting will be designed based on the long experience of the research team in the practice of transplantation. However, the formats which are most useful to a wider audience of practitioners and, importantly, to patients, may vary. If the models are predictive, we will propose ancillary studies to optimize data presentation based on patient feedback. In year 5, we will conduct patient focus groups to improve the data presentation and display.

C.12. Timeline.

Milestones	Pre	Year 1				Year 2				Year 3				Year 4				Year 5			
Pre-grant submission approvals from PCORNet collaboratives																					
IRB approval, data use agreements																					
Integration of SRTR and USRDS data with GPC data, application of AI techniques																					
Study analyses and interpretations																					
Integration of SRTR and USRDS data with STAR and OneFlorida, application of AI techniques																					
Validation of GPC integrations in STAR and OneFlorida																					
Conference presentations, Manuscript preparation, data sharing preparation																					

Bibliography and References Cited

1. Axelrod DA, Naik AS, Schnitzler MA, Segev DL, Dharnidharka VR, Brennan DC, Bae S, Chen J, Massie A, Lentine KL. National Variation in Use of Immunosuppression for Kidney Transplantation: A Call for Evidence-Based Regimen Selection. Am J Transplant. 2016;16(8):2453-62. doi: 10.1111/ajt.13758. PubMed PMID: 26901466.

2. Dharnidharka VR, Schnitzler MA, Chen J, Brennan DC, Axelrod D, Segev DL, Schechtman KB, Zheng J, Lentine KL. Differential risks for adverse outcomes 3 years after kidney transplantation based on initial immunosuppression regimen: a national study. Transpl Int. 2016. doi: 10.1111/tri.12850. PubMed PMID: 27564782.

3. Obayemi J, Keating B, Callans L, Lentine KL, Schnitzler MA, Caliskan Y, Xiao H, Dharnidharka VR, Mannon RB, Axelrod DA. Impact of CYP3A5 Status on the Clinical and Financial Outcomes Among African American Kidney Transplant Recipients. Transplant Direct. 2022;8(10):e1379. Epub 20220915. doi: 10.1097/TXD.0000000000001379. PubMed PMID: 36204191; PMCID: PMC9529042.

4. Axelrod DA, Chang SH, Lentine KL, Schnitzler MA, Norman D, Olyaei A, Malinoski D, Dharnidharka V, Segev D, Istre GR, Lockridge JB. The Clinical and Economic Benefit of CMV Matching in Kidney Transplant: A Decision Analysis. Transplantation. 2022;106(6):1227-32. Epub 20220706. doi: 10.1097/TP.0000000000003887. PubMed PMID: 34310099.

5. Axelrod DA, Caliskan Y, Schnitzler MA, Xiao H, Dharnidharka VR, Segev DL, McAdams-DeMarco M, Brennan DC, Randall H, Alhamad T, Kasiske BL, Hess G, Lentine KL. Economic impacts of alternative kidney transplant immunosuppression: A national cohort study. Clin Transplant. 2020;34(4):e13813. Epub 20200311. doi: 10.1111/ctr.13813. PubMed PMID: 32027049.

6. Lam NN, Schnitzler MA, Axelrod DA, Xiao H, McAdams-DeMarco M, Segev DL, Massie AB, Dharnidharka VR, Naik AS, Muzaale AD, Hess GP, Kasiske BL, Lentine KL. Outcome implications of benzodiazepine and opioid co-prescription in kidney transplant recipients. Clin Transplant. 2020;34(9):e14005. Epub 20200803. doi: 10.1111/ctr.14005. PubMed PMID: 32510628; PMCID: PMC7722087.

7. Dharnidharka VR, Naik AS, Axelrod DA, Schnitzler MA, Zhang Z, Bae S, Segev DL, Brennan DC, Alhamad T, Ouseph R, Lam NN, Nazzal M, Randall H, Kasiske BL, McAdams-Demarco M, Lentine KL. Center practice drives variation in choice of US kidney transplant induction therapy: a retrospective analysis of contemporary practice. Transpl Int. 2018;31(2):198-211. doi: 10.1111/tri.13079. PubMed PMID: 28987015.

8. Nazzal M, Lentine KL, Naik AS, Ouseph R, Schnitzler MA, Zhang Z, Randall H, Dharnidharka VR, Segev DL, Kasiske BL, Hess GP, Alhamad T, McAdams-Demarco M, Axelrod DA. Center-driven and Clinically Driven Variation in US Liver Transplant Maintenance Immunosuppression Therapy: A National Practice Patterns Analysis. Transplant Direct. 2018;4(7):e364. Epub 20180613. doi: 10.1097/TXD.0000000000000800. PubMed PMID: 30046654; PMCID: PMC6056277.

9. Axelrod DA, Schnitzler MA, Xiao H, Naik AS, Segev DL, Dharnidharka VR, Brennan DC, Lentine KL. The Changing Financial Landscape of Renal Transplant Practice: A National Cohort Analysis. Am J Transplant. 2017;17(2):377-89. doi: 10.1111/ajt.14018. PubMed PMID: 27565133.

10. Naik AS, Dharnidharka VR, Schnitzler MA, Brennan DC, Segev DL, Axelrod D, Xiao H, Kucirka L, Chen J, Lentine KL. Clinical and economic consequences of first-year urinary tract infections, sepsis, and pneumonia in contemporary kidney transplantation practice. Transpl Int. 2016;29(2):241-52. doi: 10.1111/tri.12711. PubMed PMID: 26563524; PMCID: PMC4805426.

11. Axelrod D, Segev DL, Xiao H, Schnitzler MA, Brennan DC, Dharnidharka VR, Orandi BJ, Naik AS, Randall H, Tuttle-Newhall JE, Lentine KL. Economic Impacts of ABO-Incompatible Live Donor Kidney Transplantation: A National Study of Medicare-Insured Recipients. Am J Transplant. 2016;16(5):1465-73. doi: 10.1111/ajt.13616. PubMed PMID: 26603690; PMCID: PMC4844838.

12. Lentine KL, Naik AS, Schnitzler M, Axelrod D, Chen J, Brennan DC, Segev DL, Kasiske BL, Randall H, Dharnidharka VR. Variation in Comedication Use According to Kidney Transplant Immunosuppressive Regimens: Application of Integrated Registry and Pharmacy Claims Data. Transplant Proc. 2016;48(1):55-8. doi: 10.1016/j.transproceed.2015.12.024. PubMed PMID: 26915843; PMCID: PMC4950501.

13. Lentine KL, Anyaegbu E, Gleisner A, Schnitzler MA, Axelrod D, Brennan DC, Dharnidharka VR, Abraham E, Tuttle-Newhall JE. Understanding medical care of transplant recipients through integrated registry and pharmacy claims data. Am J Nephrol. 2013;38(5):420-9. Epub 2013/11/13. doi: 10.1159/000356092. PubMed PMID: 24216747.

14. Dharnidharka VR, Naik AS, Axelrod D, Schnitzler MA, Xiao H, Brennan DC, Segev DL, Randall H, Chen J, Kasiske B, Lentine KL. Clinical and Economic Consequences of Early Cancer After Kidney Transplantation in Contemporary Practice. Transplantation. 2017;101(4):858-66. doi: 10.1097/TP.0000000000001385. PubMed PMID: 27490413; PMCID: PMC5346345.

15. Unterrainer C, Opelz G, Dohler B, Susal C, Collaborative Transplant S. Pretransplant cancer in kidney recipients in relation to recurrent and de novo cancer incidence posttransplantation and implications for graft and patient survival. Transplantation. 2018. Epub 2018/11/13. doi: 10.1097/TP.0000000000002459. PubMed PMID: 30418430.

16. Schnitzler MA, Skeans MA, Axelrod DA, Lentine KL, Randall HB, Snyder JJ, Israni AK, Kasiske BL. OPTN/SRTR 2016 Annual Data Report: Economics. Am J Transplant. 2018;18 Suppl 1:464-503. doi: 10.1111/ajt.14564. PubMed PMID: 29292607.

17. Hippen BE, Reed AI, Ketchersid T, Maddux FW. Implications of the Advancing American Kidney Health Initiative for kidney transplant centers. Am J Transplant. 2019. Epub 2019/09/29. doi: 10.1111/ajt.15619. PubMed PMID: 31561276.

18. Wang Y, Hemmelder MH, Bos WJW, Snoep JD, de Vries APJ, Dekker FW, Meuleman Y. Mapping health-related quality of life after kidney transplantation by group comparisons: a systematic review. Nephrol Dial Transplant. 2021;36(12):2327-39. doi: 10.1093/ndt/gfab232. PubMed PMID: 34338799; PMCID: PMC8643597.

19. Poggio ED, Augustine JJ, Arrigain S, Brennan DC, Schold JD. Long-term kidney transplant graft survival-Making progress when most needed. Am J Transplant. 2021;21(8):2824-32. Epub 20210208. doi: 10.1111/ajt.16463. PubMed PMID: 33346917.

20. Palepu S, Prasad GV. New-onset diabetes mellitus after kidney transplantation: Current status and future directions. World J Diabetes. 2015;6(3):445-55. Epub 2015/04/22. doi: 10.4239/wjd.v6.i3.445. PubMed PMID: 25897355; PMCID: PMC4398901.

21. Cen C, Fang HX, Yu SF, Liu JM, Liu YX, Zhou L, Yu J, Zheng SS. Association between ADIPOQ gene polymorphisms and the risk of new-onset diabetes mellitus after liver transplantation. Hepatobiliary Pancreat Dis Int. 2017;16(6):602-9. Epub 2018/01/03. doi: 10.1016/S1499-3872(17)60069-9. PubMed PMID: 29291779.

22. Bhat M, Mara K, Dierkhising R, Watt KDS. Immunosuppression, Race, and Donor-Related Risk Factors Affect De novo Cancer Incidence Across Solid Organ Transplant Recipients. Mayo Clin Proc. 2018;93(9):1236-46. Epub 2018/08/02. doi: 10.1016/j.mayocp.2018.04.025. PubMed PMID: 30064826.

23. Engels EA, Pfeiffer RM, Fraumeni JF, Jr., Kasiske BL, Israni AK, Snyder JJ, Wolfe RA, Goodrich NP, Bayakly AR, Clarke CA, Copeland G, Finch JL, Fleissner ML, Goodman MT, Kahn A, Koch L, Lynch CF, Madeleine MM, Pawlish K, Rao C, Williams MA, Castenson D, Curry M, Parsons R, Fant G, Lin M. Spectrum of cancer risk among US solid organ transplant recipients. JAMA. 2011;306(17):1891-901. Epub 2011/11/03. doi: 10.1001/jama.2011.1592. PubMed PMID: 22045767; PMCID: 3310893.

24. Kutinova A, Woodward RS, Ricci JF, Brennan DC. The incidence and costs of sepsis and pneumonia before and after renal transplantation in the United States. Am J Transplant. 2006;6(1):129-39. Epub 2006/01/26. doi: 10.1111/j.1600-6143.2005.01156.x. PubMed PMID: 16433767.

25. Ekberg H, Grinyo J, Nashan B, Vanrenterghem Y, Vincenti F, Voulgari A, Truman M, Nasmyth-Miller C, Rashford M. Cyclosporine sparing with mycophenolate mofetil, daclizumab and corticosteroids in renal allograft recipients: the CAESAR Study. Am J Transplant. 2007;7(3):560-70. Epub 2007/01/19. doi: 10.1111/j.1600-6143.2006.01645.x. PubMed PMID: 17229079.

26. Flechner SM, Glyda M, Cockfield S, Grinyo J, Legendre C, Russ G, Steinberg S, Wissing KM, Tai SS. The ORION study: comparison of two sirolimus-based regimens versus tacrolimus and mycophenolate mofetil in renal allograft recipients. Am J Transplant. 2011;11(8):1633-44. Epub 2011/06/15. doi: 10.1111/j.1600-6143.2011.03573.x. PubMed PMID: 21668635.

27. Holdaas H, Rostaing L, Seron D, Cole E, Chapman J, Fellstrom B, Strom EH, Jardine A, Midtvedt K, Machein U, Ulbricht B, Karpov A, O'Connell PJ, Investigators A. Conversion of long-term kidney transplant recipients from calcineurin inhibitor therapy to everolimus: a randomized, multicenter, 24-month study. Transplantation. 2011;92(4):410-8. Epub 2011/06/24. doi: 10.1097/TP.0b013e318224c12d. PubMed PMID: 21697773.

28. Lebranchu Y, Thierry A, Toupance O, Westeel PF, Etienne I, Thervet E, Moulin B, Frouget T, Le Meur Y, Glotz D, Heng AE, Onno C, Buchler M, Girardot-Seguin S, Hurault de Ligny B. Efficacy on renal function of early conversion from cyclosporine to sirolimus 3 months after renal transplantation: concept study. Am J Transplant. 2009;9(5):1115-23. Epub 2009/05/09. doi: 10.1111/j.1600-6143.2009.02615.x. PubMed PMID: 19422337.

29. Vincenti F, Charpentier B, Vanrenterghem Y, Rostaing L, Bresnahan B, Darji P, Massari P, Mondragon-Ramirez GA, Agarwal M, Di Russo G, Lin CS, Garg P, Larsen CP. A phase III study of belatacept-based immunosuppression regimens versus cyclosporine in renal transplant recipients (BENEFIT study). Am J Transplant. 2011;10(3):535-46. PubMed PMID: 20415897.

30. Vincenti F, Schena FP, Paraskevas S, Hauser IA, Walker RG, Grinyo J. A randomized, multicenter study of steroid avoidance, early steroid withdrawal or standard steroid therapy in kidney transplant recipients. Am J Transplant. 2008;8(2):307-16. PubMed PMID: 18211506.

31. Buchler M, Caillard S, Barbier S, Thervet E, Toupance O, Mazouz H, Hurault de Ligny B, Le Meur Y, Thierry A, Villemain F, Heng AE, Moulin B, Morin MP, Noel C, Lebranchu Y, Group S. Sirolimus versus cyclosporine in kidney recipients receiving thymoglobulin, mycophenolate mofetil and a 6-month course of steroids. Am J Transplant. 2007;7(11):2522-31. Epub 2007/09/18. doi: 10.1111/j.1600-6143.2007.01976.x. PubMed PMID: 17868057.

32. Ekberg H, Tedesco-Silva H, Demirbas A, Vitko S, Nashan B, Gurkan A, Margreiter R, Hugo C, Grinyo JM, Frei U, Vanrenterghem Y, Daloze P, Halloran PF. Reduced exposure to calcineurin inhibitors in renal transplantation. N Engl J Med. 2007;357(25):2562-75. Epub 2007/12/21. doi: 357/25/2562 [pii]

10.1056/NEJMoa067411. PubMed PMID: 18094377.

33. Guba M, Pratschke J, Hugo C, Kramer BK, Nohr-Westphal C, Brockmann J, Andrassy J, Reinke P, Pressmar K, Hakenberg O, Fischereder M, Pascher A, Illner WD, Banas B, Jauch KW, Group SM-S. Renal function, efficacy, and safety of sirolimus and mycophenolate mofetil after short-term calcineurin inhibitor-based quadruple therapy in de novo renal transplant patients: one-year analysis of a randomized multicenter trial. Transplantation. 2010;90(2):175-83. Epub 2010/05/14. doi: 10.1097/TP.0b013e3181e11798. PubMed PMID: 20463641.

34. Zhang H, Zang C, Xu Z, Zhang Y, Xu J, Bian J, Morozyuk D, Khullar D, Zhang Y, Nordvig AS, Schenck EJ, Shenkman EA, Rothman RL, Block JP, Lyman K, Weiner M, Carton TW, Wang F, Kaushal R. Machine Learning for Identifying Data-Driven Subphenotypes of Incident Post-Acute SARS-CoV-2 Infection Conditions with Large Scale Electronic Health Records: Findings from the RECOVER Initiative. medRxiv. 2022. Epub 20220608. doi: 10.1101/2022.05.21.22275412. PubMed PMID: 35665007; PMCID: PMC9164516.

35. Song X, Yu ASL, Kellum JA, Waitman LR, Matheny ME, Simpson SQ, Hu Y, Liu M. Cross-site transportability of an explainable artificial intelligence model for acute kidney injury prediction. Nat Commun. 2020;11(1):5668. Epub 20201109. doi: 10.1038/s41467-020-19551-w. PubMed PMID: 33168827; PMCID: PMC7653032.

36. Ravindhran B, Chandak P, Schafer N, Kundalia K, Hwang W, Antoniadis S, Haroon U, Zakri RH. Machine learning models in predicting graft survival in kidney transplantation: meta-analysis. BJS Open. 2023;7(2). doi: 10.1093/bjsopen/zrad011. PubMed PMID: 36987687; PMCID: PMC10050937.

37. Peloso A, Moeckli B, Delaune V, Oldani G, Andres A, Compagnon P. Artificial Intelligence: Present and Future Potential for Solid Organ Transplantation. Transpl Int.

2022;35:10640. Epub 20220704. doi: 10.3389/ti.2022.10640. PubMed PMID: 35859667; PMCID: PMC9290190.

38. Wesselman H, Ford CG, Leyva Y, Li X, Chang CH, Dew MA, Kendall K, Croswell E, Pleis JR, Ng YH, Unruh ML, Shapiro R, Myaskovsky L. Social Determinants of Health and Race Disparities in Kidney Transplant. Clin J Am Soc Nephrol. 2021;16(2):262-74. Epub 20210128. doi: 10.2215/CJN.04860420. PubMed PMID: 33509963; PMCID: PMC7863655.

39. Johnston A, Jones WS, Hernandez AF. The ADAPTABLE Trial and Aspirin Dosing in Secondary Prevention for Patients with Coronary Artery Disease. Curr Cardiol Rep. 2016;18(8):81. doi: 10.1007/s11886-016-0749-2. PubMed PMID: 27423939.

40. Song X, Afshar A, Statland J, Waitman L, editors. Neurologist Care for Amyotrophic Lateral Sclerosis-A utilization and outcome study. MUSCLE & NERVE; 2022: WILEY 111 RIVER ST, HOBOKEN 07030-5774, NJ USA.

41. Marquis-Gravel G, Roe MT, Robertson HR, Harrington RA, Pencina MJ, Berdan LG, Hammill BG, Faulkner M, Muñoz D, Fonarow GC, Nallamothu BK, Fintel DJ, Ford DE, Zhou L, Daugherty SE, Nauman E, Kraschnewski J, Ahmad FS, Benziger CP, Haynes K, Merritt JG, Metkus T, Kripalani S, Gupta K, Shah RC, McClay JC, Re RN, Geary C, Lampert BC, Bradley SM, Jain SK, Seifein H, Whittle J, Roger VL, Effron MB, Alvarado G, Goldberg YH, VanWormer JL, Girotra S, Farrehi P, McTigue KM, Rothman R, Hernandez AF, Jones WS. Rationale and Design of the Aspirin Dosing-A Patient-Centric Trial Assessing Benefits and Long-term Effectiveness (ADAPTABLE) Trial. JAMA Cardiol. 2020;5(5):598-607. doi: 10.1001/jamacardio.2020.0116. PubMed PMID: 32186653.

42. Ahmad FS, Ricket IM, Hammill BG, Eskenazi L, Robertson HR, Curtis LH, Dobi CD, Girotra S, Haynes K, Kizer JR, Kripalani S, Roe MT, Roumie CL, Waitman R, Jones WS, Weiner MG. Computable Phenotype Implementation for a National, Multicenter Pragmatic Clinical Trial: Lessons Learned From ADAPTABLE. Circ Cardiovasc Qual Outcomes. 2020;13(6):e006292. Epub 20200529. doi: 10.1161/circoutcomes.119.006292. PubMed PMID: 32466729; PMCID: PMC7321832.

43. Nichols GA, Desai J, Elston Lafata J, Lawrence JM, O'Connor PJ, Pathak RD, Raebel MA, Reid RJ, Selby JV, Silverman BG, Steiner JF, Stewart WF, Vupputuri S, Waitzfelder B. Construction of a multisite DataLink using electronic health records for the identification, surveillance, prevention, and management of diabetes mellitus: the SUPREME-DM project. Prev Chronic Dis. 2012;9:E110. Epub 20120607. doi: 10.5888/pcd9.110311. PubMed PMID: 22677160; PMCID: PMC3457753.

44. Furmanchuk A, Liu M, Song X, Waitman LR, Meurer JR, Osinski K, Stoddard A, Chrischilles E, McClay JC, Cowell LG, Tachinardi U, Embi PJ, Mosa ASM, Mandhadi V, Shah RC, Garcia D, Angulo F, Patino A, Trick WE, Markossian TW, Rasmussen-Torvik LJ, Kho AN, Black BS. Effect of the Affordable Care Act on diabetes care at major health centers: newly detected diabetes and diabetes medication management. BMJ Open Diabetes Res Care. 2021;9(1). doi: 10.1136/bmjdrc-2021-002205. PubMed PMID: 34187842; PMCID: PMC8245434.

45. Graham J, Iverson A, Monteiro J, Weiner K, Southall K, Schiller K, Gupta M, Simard EP. Applying computable phenotypes within a common data model to identify heart failure patients for an implantable cardiac device registry. Int J Cardiol Heart Vasc.

2022;39:100974. Epub 20220219. doi: 10.1016/j.ijcha.2022.100974. PubMed PMID: 35242997; PMCID: PMC8861122.

46. Song X, Waitman LR, Yu AS, Robbins DC, Hu Y, Liu M. Longitudinal Risk Prediction of Chronic Kidney Disease in Diabetic Patients using Temporal-Enhanced Gradient Boosting Machine: Retrospective Cohort Study. JMIR Medical Informatics. 2020;8(1):e15510. doi: 10.2196/15510.

47. Raynaud M, Aubert O, Divard G, Reese PP, Kamar N, Yoo D, Chin C-S, Bailly É, Buchler M, Ladrière M, Le Quintrec M, Delahousse M, Juric I, Basic-Jukic N, Crespo M, Silva HT, Linhares K, Ribeiro De Castro MC, Soler Pujol G, Empana J-P, Ulloa C, Akalin E, Böhmig G, Huang E, Stegall MD, Bentall AJ, Montgomery RA, Jordan SC, Oberbauer R, Segev DL, Friedewald JJ, Jouven X, Legendre C, Lefaucheur C, Loupy A. Dynamic prediction of renal survival among deeply phenotyped kidney transplant recipients using artificial intelligence: an observational, international, multicohort study. The Lancet Digital Health. 2021;3(12):e795-e805. doi: 10.1016/s2589-7500(21)00209-0.

48. Li K, Luo S. Dynamic predictions in Bayesian functional joint models for longitudinal and time-to-event data: An application to Alzheimer's disease. Statistical Methods in Medical Research. 2019;28(2):327-42. doi: 10.1177/0962280217722177.

49. Dong JJ, Cao J, Gill J, Miles C, Plumb T. Functional joint models for chronic kidney disease in kidney transplant recipients. Stat Methods Med Res. 2021;30(8):1932-43. Epub 20210510. doi: 10.1177/09622802211009265. PubMed PMID: 33970050.

50. Rizopoulos D. The R Package JMbayes for Fitting Joint Models for Longitudinal and Time-to-Event Data Using MCMC. Journal of Statistical Software. 2016;72(7):1 - 46. doi: 10.18637/jss.v072.i07.

51. Hickey GL, Philipson P, Jorgensen A, Kolamunnage-Dona R. A comparison of joint models for longitudinal and competing risks data, with application to an epilepsy drug randomized controlled trial. Journal of the Royal Statistical Society Series A (Statistics in Society). 2018;181(4):1105-23.

52. Lunceford JK, Davidian M, Tsiatis AA. Estimation of survival distributions of treatment policies in two-stage randomization designs in clinical trials. Biometrics. 2002;58(1):48-57. doi: 10.1111/j.0006-341x.2002.00048.x. PubMed PMID: 11890326.

53. Almirall D, Chronis-Tuscano A. Adaptive Interventions in Child and Adolescent Mental Health. J Clin Child Adolesc Psychol. 2016;45(4):383-95. Epub 20160616. doi: 10.1080/15374416.2016.1152555. PubMed PMID: 27310565; PMCID: PMC4930370.

54. Lavori PW, Dawson R. Adaptive treatment strategies in chronic disease. Annu Rev Med. 2008;59:443-53. doi: 10.1146/annurev.med.59.062606.122232. PubMed PMID: 17914924; PMCID: PMC2739674.

55. Chakraborty B, Murphy SA. Dynamic Treatment Regimes. Annual Review of Statistics and Its Application. 2014;1(1):447-64. doi: 10.1146/annurev-statistics-022513-115553.

56. Anzabi Zadeh S, Street WN, Thomas BW. Optimizing warfarin dosing using deep reinforcement learning. J Biomed Inform. 2023;137:104267. Epub 20221207. doi: 10.1016/j.jbi.2022.104267. PubMed PMID: 36494060.

57. Chao Yu JL, Fellow, IEEE, and Shamim Nemati. Reinforcement Learning in Healthcare: A Survey. arXiv pre-print server. 2020. doi: 10.48550/arxiv.1908.08796.

58. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000;11(5):550-60. doi: 10.1097/00001648-200009000-00011. PubMed PMID: 10955408.

59. Mahar RK, McGuinness MB, Chakraborty B, Carlin JB, Ijzerman MJ, Simpson JA. A scoping review of studies using observational data to optimise dynamic treatment regimens. BMC Medical Research Methodology. 2021;21(1). doi: 10.1186/s12874-021-01211-2.

60. Dickerman BA, Giovannucci E, Pernar CH, Mucci LA, Hernán MA. Guideline-Based Physical Activity and Survival Among US Men With Nonmetastatic Prostate Cancer (Supp). American Journal of Epidemiology. 2019;188(3):579-86. doi: 10.1093/aje/kwy261.

61. Zhao Y, Zeng D, Socinski MA, Kosorok MR. Reinforcement Learning Strategies for Clinical Trials in Nonsmall Cell Lung Cancer. Biometrics. 2011;67(4):1422-33. doi: 10.1111/j.1541-0420.2011.01572.x.

62. Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. Nature Medicine. 2018;24(11):1716-20. doi: 10.1038/s41591-018-0213-5.

63. Dickinson DM, Bryant PC, Williams MC, Levine GN, Li S, Welch JC, Keck BM, Webb RL. Transplant data: sources, collection, and caveats. Am J Transplant. 2004;4 Suppl 9:13-26. Epub 2004/04/29. doi: 10.1111/j.1600-6135.2004.00395.x. PubMed PMID: 15113352.

64. Lentine KL, Vijayan A, Xiao H, Schnitzler MA, Davis CL, Garg AX, Axelrod D, Abbott KC, Brennan DC. Cancer diagnoses after living kidney donation: linking U.S. Registry data and administrative claims. Transplantation. 2012;94(2):139-44. Epub 2012/07/25. doi: 10.1097/TP.0b013e318254757d. PubMed PMID: 22825543.

65. Raths D. Healthcare Innovation. 2021. [cited 2023]. Available from: https://www.hcinnovationgroup.com/clinical-it/learning-health-systems-research/news/21240260/pcori-funds-8-research-networks-for-pcornets-phase-3.

66. Center PDRNO. 2020. [cited 2023]. Available from: https://github.com/PCORnet-DRN-OC/PCORnet-Data-Curation.

67. Patient-Centered Outcomes Research Institute. Common Data Model (CDM) Specification, Version 6.0 [cited 2022 January 6]. Available from: https://pcornet.org/wp-content/uploads/2022/01/PCORnet-Common-Data-Model-v60-2020_10_221.pdf.

68. Patient-Centered Outcomes Research Institute. 2020. [cited 2023]. Available from: https://pcornet.org/wp-content/uploads/2022/01/PCORnet-Common-Data-Model-v60-2020_10_221.pdf.

69. Waitman LR, Aaronson LS, Nadkarni PM, Connolly DW, Campbell JR. The Greater Plains Collaborative: a PCORnet Clinical Research Data Network. J Am Med Inform Assoc. 2014;21(4):637-41. Epub 20140428. doi: 10.1136/amiajnl-2014-002756. PubMed PMID: 24778202; PMCID: PMC4078294.

70. Bernstam EV, Applegate RJ, Yu A, Chaudhari D, Liu T, Coda A, Leshin J. Real-World Matching Performance of Deidentified Record-Linking Tokens. Appl Clin Inform. 2022;13(4):865-73. Epub 20220727. doi: 10.1055/a-1910-4154. PubMed PMID: 35896508; PMCID: PMC9474266.

71. Kiernan D, Carton T, Toh S, Phua J, Zirkle M, Louzao D, Haynes K, Weiner M, Angulo F, Bailey C, Bian J, Fort D, Grannis S, Krishnamurthy AK, Nair V, Rivera P, Silverstein J, Marsolo K. Establishing a framework for privacy-preserving record linkage among electronic health record and administrative claims databases within PCORnet(®), the National Patient-Centered Clinical Research Network. BMC Res Notes. 2022;15(1):337-. doi: 10.1186/s13104-022-06243-5. PubMed PMID: 36316778.

72. Canterberry M, Kaul AF, Goel S, Lin PD, Block JP, Nair VP, Ma Q, Carton TW. The Patient-Centered Outcomes Research Network Antibiotics and Childhood Growth Study: Implementing Patient Data Linkage. Popul Health Manag. 2020;23(6):438-44. Epub 20191217. doi: 10.1089/pop.2019.0089. PubMed PMID: 31855123; PMCID: PMC7397429.

73. Marsolo K, Kiernan D, Toh S, Phua J, Louzao D, Haynes K, Weiner M, Angulo F, Bailey C, Bian J, Fort D, Grannis S, Krishnamurthy AK, Nair V, Rivera P, Silverstein J, Zirkle M, Carton T. Assessing the impact of privacy-preserving record linkage on record overlap and patient demographic and clinical characteristics in PCORnet®, the National Patient-Centered Clinical Research Network. Journal of the American Medical Informatics Association. 2022;30(3):447-55. doi: 10.1093/jamia/ocac229.

74. Kho AN, Cashy JP, Jackson KL, Pah AR, Goel S, Boehnke J, Humphries JE, Kominers SD, Hota BN, Sims SA, Malin BA, French DD, Walunas TL, Meltzer DO, Kaleba EO, Jones RC, Galanter WL. Design and implementation of a privacy preserving electronic health record linkage tool in Chicago. Journal of the American Medical Informatics Association. 2015;22(5):1072-80. doi: 10.1093/jamia/ocv038.

75. Ahmad FS, Ricket IM, Hammill BG, Eskenazi L, Robertson HR, Curtis LH, Dobi CD, Girotra S, Haynes K, Kizer JR, Kripalani S, Roe MT, Roumie CL, Waitman R, Jones WS, Weiner MG. Computable Phenotype Implementation for a National, Multicenter Pragmatic Clinical Trial. Circulation: Cardiovascular Quality and Outcomes. 2020;13(6). doi: 10.1161/circoutcomes.119.006292.

76. Lentine KL, Smith JM, Hart A, Miller J, Skeans MA, Larkin L, Robinson A, Gauntt K, Israni AK, Hirose R, Snyder JJ. OPTN/SRTR 2020 Annual Data Report: Kidney. Am J Transplant. 2022;22 Suppl 2:21-136. doi: 10.1111/ajt.16982. PubMed PMID: 35266618.

77. Song X. Embedded Ensemble Feature Selection 2020 [04/12/2022]. Available from: https://github.com/sxinger/EEFS.

78. Lundberg SM, Lee S-I, editors. A Unified Approach to Interpreting Model Predictions. NIPS; 2017.

79. Song X, Liu M, Waitman LR, Patel A, Simpson SQ. Clinical factors associated with rapid treatment of sepsis. PLoS One. 2021;16(5):e0250923. Epub 20210506. doi: 10.1371/journal.pone.0250923. PubMed PMID: 33956846; PMCID: PMC8101717.

80. Choi E, Schuetz A, Stewart WF, Sun J. Using recurrent neural network models for early detection of heart failure onset. J Am Med Inform Assoc. 2017;24(2):361-70. doi: 10.1093/jamia/ocw112. PubMed PMID: 27521897; PMCID: PMC5391725.

81. Zeng Z, Tang X, Liu Y, He Z, Gong X. Interpretable recurrent neural network models for dynamic prediction of the extubation failure risk in patients with invasive mechanical ventilation in the intensive care unit. BioData Mining. 2022;15(1). doi: 10.1186/s13040-022-00309-7.

82. Liu X, Liu C, Huang R, Zhu H, Liu Q, Mitra S, Wang Y. Long short-term memory recurrent neural network for pharmacokinetic-pharmacodynamic modeling. Int J Clin Pharmacol Ther. 2021;59(2):138-46. doi: 10.5414/cp203800. PubMed PMID: 33210994.

83. Shinozaki T, Suzuki E. Understanding Marginal Structural Models for Time-Varying Exposures: Pitfalls and Tips. Journal of Epidemiology. 2020;30(9):377-89. doi: 10.2188/jea.je20200226.

84. Yang W, Joffe MM. Subtle issues in model specification and estimation of marginal structural models. Pharmacoepidemiol Drug Saf. 2012;21(3):241-5. Epub 20120116. doi: 10.1002/pds.2306. PubMed PMID: 22509500.

85. Sato T, Matsuyama Y. Marginal structural models as a tool for standardization. Epidemiology. 2003;14(6):680-6. doi: 10.1097/01.EDE.0000081989.82616.7d. PubMed PMID: 14569183.

86. Talbot D, Atherton J, Rossi AM, Bacon SL, Lefebvre G. A cautionary note concerning the use of stabilized weights in marginal structural models. Stat Med. 2015;34(5):812-23. Epub 20141119. doi: 10.1002/sim.6378. PubMed PMID: 25410264.

87. Naimi AI, Cole SR, Westreich DJ, Richardson DB. A comparison of methods to estimate the hazard ratio under conditions of time-varying confounding and nonpositivity. Epidemiology. 2011;22(5):718-23. doi: 10.1097/EDE.0b013e31822549e8. PubMed PMID: 21747286; PMCID: PMC3155387.

88. Daniel RM, Cousens SN, De Stavola BL, Kenward MG, Sterne JA. Methods for dealing with time-dependent confounding. Stat Med. 2013;32(9):1584-618. Epub 20121203. doi: 10.1002/sim.5686. PubMed PMID: 23208861.

89. Breskin A, Cole SR, Westreich D. Exploring the Subtleties of Inverse Probability Weighting and Marginal Structural Models. Epidemiology. 2018;29(3):352-5. doi: 10.1097/ede.0000000000000813. PubMed PMID: 29384789; PMCID: PMC5882514.

90. Li R, Hu S, Lu M, Utsumi Y, Chakraborty P, Sow DM, Madan P, Li J, Ghalwash M, Shahn Z, Lehman L-w. G-Net: a Recurrent Network Approach to G-Computation for Counterfactual Prediction Under a Dynamic Treatment Regime. In: Subhrajit R, Stephen P, Emma R, Girmaw Abebe T, Luis O, Fabian F, Yuyin Z, Liyue S, Ghada Z, Purity M, Ayah Z, Matthew BAM, Emily A, editors. Proceedings of Machine Learning for Health; Proceedings of Machine Learning Research: PMLR; 2021. p. 282--99.

91. Anderssen N, Malterud K. Oversampling as a methodological strategy for the study of self-reported health among lesbian, gay and bisexual populations. Scand J Public Health. 2017;45(6):637-46. Epub 20170704. doi: 10.1177/1403494817717407. PubMed PMID: 28675963.

92. James T, Mukadam N, Sommerlad A, Pour HR, Knowles M, Azocar I, Livingston G. Protection against discrimination in national dementia guideline recommendations: A systematic review. PLOS Medicine. 2022;19(1):e1003860. doi: 10.1371/journal.pmed.1003860.

93. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring Fairness in Machine Learning to Advance Health Equity. Ann Intern Med. 2018;169(12):866-72. Epub 20181204. doi: 10.7326/m18-1990. PubMed PMID: 30508424; PMCID: PMC6594166.

94. Fletcher RR, Nakeshimana A, Olubeko O. Addressing Fairness, Bias, and Appropriate Use of Artificial Intelligence and Machine Learning in Global Health. Frontiers in Artificial Intelligence. 2021;3. doi: 10.3389/frai.2020.561802.

95. Reese JT, Blau H, Casiraghi E, Bergquist T, Loomba JJ, Callahan TJ, Laraway B, Antonescu C, Coleman B, Gargano M, Wilkins KJ, Cappelletti L, Fontana T, Ammar N, Antony B, Murali TM, Caufield JH, Karlebach G, McMurry JA, Williams A, Moffitt R, Banerjee J, Solomonides AE, Davis H, Kostka K, Valentini G, Sahner D, Chute CG, Madlock-Brown C, Haendel MA, Robinson PN, Spratt H, Visweswaran S, Flack JE, Yoo YJ, Gabriel D, Alexander GC, Mehta HB, Liu F, Miller RT, Wong R, Hill EL, Thorpe LE, Divers J. Generalisable long COVID subtypes: findings from the NIH N3C and RECOVER programmes. eBioMedicine. 2023;87:104413. doi: 10.1016/j.ebiom.2022.104413.

96. Sadeghi B, Cheung RCY, Hanbury M. Using hierarchical clustering analysis to evaluate COVID-19 pandemic preparedness and performance in 180 countries in 2020. BMJ Open. 2021;11(11):e049844. Epub 20211109. doi: 10.1136/bmjopen-2021-049844. PubMed PMID: 34753756; PMCID: PMC8578186.

97. Arauz-Garofalo G, Jodar M, Vilanova M, de la Iglesia Rodriguez A, Castillo J, Soler-Ventura A, Oliva R, Vilaseca M, Gay M. Protamine Characterization by Top-Down Proteomics: Boosting Proteoform Identification with DBSCAN. Proteomes. 2021;9(2). Epub 20210430. doi: 10.3390/proteomes9020021. PubMed PMID: 33946530; PMCID: PMC8162566.

EXHIBIT B

Data Flow Documentation

This Data Flow Documentation ("**Documentation**") describes and illustrates how data flows from the Site Participants and CISTEM2 Data Coordinating Center ("**DCC"**) at University of Missouri ("**MU"**). Both Scientific Registry of Transplant Recipient (SRTR) registry and the approved study sites within PCORnet will submit requested data to DCC, where data linkage and integration will be performed. Only the physical aspects of the movement of data through the networks are addressed is this Documentation, but not the legal, administrative or regulatory requirements for data transfer or use of the data.

Please note that unless otherwise specified, all capitalized terms used in this Documentation have the same meaning assigned to them as in the PCORnet and GPC Data Sharing.

I. DESCRIPTION.

A. GPC-Hosted Data Infrastructure. The CISTEM2 database will be managed leveraging the GPC-hosted data infrastructure, which is a HIPAA-compliant cloud-based data lake and can only be accessed from approved virtual machines in compliance with security and privacy regulations required by hosting sensitive data such as Medicare claims and registry data. This environment is annually certified by center of Medicare and Medicaid Data Privacy and Safeguard Program (CMS DPSP).

B. PCORnet Common Data Model. In the PCORnet Common Data Model (CDM), which is based on the FDA Sentinel Initiative CDM (www.sentinelsystem.org), each partner network securely collects and stores data behind its own firewall, and maps it to the same consistent format (i.e., with the same variable name, attributes, and other metadata). It leverages standard terminologies and coding systems for healthcare (including ICD, SNOMED, CPT, HCPSC, and LOINC) to enable interoperability with, and responsiveness to, evolving data standards. The PCORnet CDM (www.pcornet.org/pcornet-common-data-model) is maintained and managed by the DCC.

C. Utilizing Datavant Software. The Datavant software solution enables Privacy-Preserving-Record Linkage (PPRL) through the use of de-identified, encrypted tokens that can be linked across data sources. CISTEM2 sites with existing SLAs have already installed the Datavant De-ID software to generate multiple, encrypted identifiers (hash tokens) based on different permutations of personally identifiable information (PII) held within source systems. SRTR has also installed this software. The underlying PII is processed by the Datavant De-ID software, where it is irreversibly hashed (i.e., cannot regenerate the PII from the hash value) into a series of Master Tokens using the Datavant Master Seed. The key-based hashing process means that the same PII processed at one site will result in different hashes from those produced from another site. In order to match tokens from different sites, the site-specific Master Tokens are then transformed into Transit tokens that are specifically directed at a third-party site using a site-specific key. This means that if the transit tokens are accidentally sent to the wrong location, the transit tokens cannot be processed by the incorrect site. The receiving site can then transform the transit tokens from different sites into tokens that can be used to match the same patients across different sites. The same PII always generates the same set of Master Tokens, but it is

never present in any output or log stream from the De-ID software. Only the site-specific encryption tokens are written to the output file.

D. CISTEM2 Cohort Identification. CISTEM2 cohort includes all kidney transplant (Kx) patients who are included in the SRTR registry with an established linkage between SRTR registry and USRDS database. DCC will first leverage Datavant linkage to identify the crosswalk population between GPC EHR cohort and SRTR registry to generate site-specific finder files. The finder files will then be disseminated to participating sites for clinical data extraction (i.e., CDM-CISTEM2) on the crosswalk Kx patients.

E. Data Submission and Integration. DCC will utilize agreed upon security transfer methods for data submission. Participating sites can also choose to adopt additional layer of data encryption following AES-256 with password protection. Upon receiving data from participating sites, SRTR, USRDS and CMS. DCC will perform data linkage and integration using the mapping among four source-specific, masked identifiers: PATID, HASH_ID, OPTN_HID, and BENE_ID. HASH_ID is the study-specific transit token generated by Datavant software.

II. ILLUSTRATION. The diagram below illustrates the data flow, linkage and transfer process.