Exploring Data Scraping on ClinicalTrials.gov to Identify Key Variables to Include in an EHR-based Recruitment Tool for Diabetes Clinical Trials

Sydney A. Lash, Emily R. Pfaff, and Klara R. Klein
University of North Carolina at Chapel Hill

UNC SCHOOL OF MEDICINE
North Carolina Translational and Clinical Sciences Institute

Introduction

- Recruitment challenges terminate ~20% of clinical trials and delay >85% of timelines.
- Limited access to clinical trials in rural areas leads to trials that do not represent disease populations or the practices where they receive care.
- Poor representation inhibits swift uptake of new therapies.[2] In diabetes, disease-modifying therapies reach <10% of eligible.
- Identifying methods to reduce recruitment inefficiencies and increase trial representativeness is key to translating therapies into routine care.
- We aim to build an electronic health record (EHR) based research tool to identify populations eligible for clinical trial participation in diabetes trials across North Carolina.

Methods

- We scraped ClinicalTrials.gov (CT.gov) for common diabetes clinical trial inclusion and exclusion (I/E) criteria and compared them with expert-identified criteria.
- We analyzed actively recruiting or recently active phase 2 and 3 adult trials (>18 years) in the U.S.
- We gathered input from 2 clinical research coordinators and 3 diabetes trial experts, including 2 drug/device trialists and 1 qualitative researcher.
- I/E were normalized for grouping terms into broader, clinically relevant diabetes-related categories by a clinical diabetes expert.

Results

- Among 203 reviewed trials on CT.gov, 115 unique terms were identified. 91 were deemed clinically relevant.
- After normalization, the CT.gov list contained 37 distinct terms.
- The final trialist list included 29 distinct terms aggregated from all investigators and CRCs.

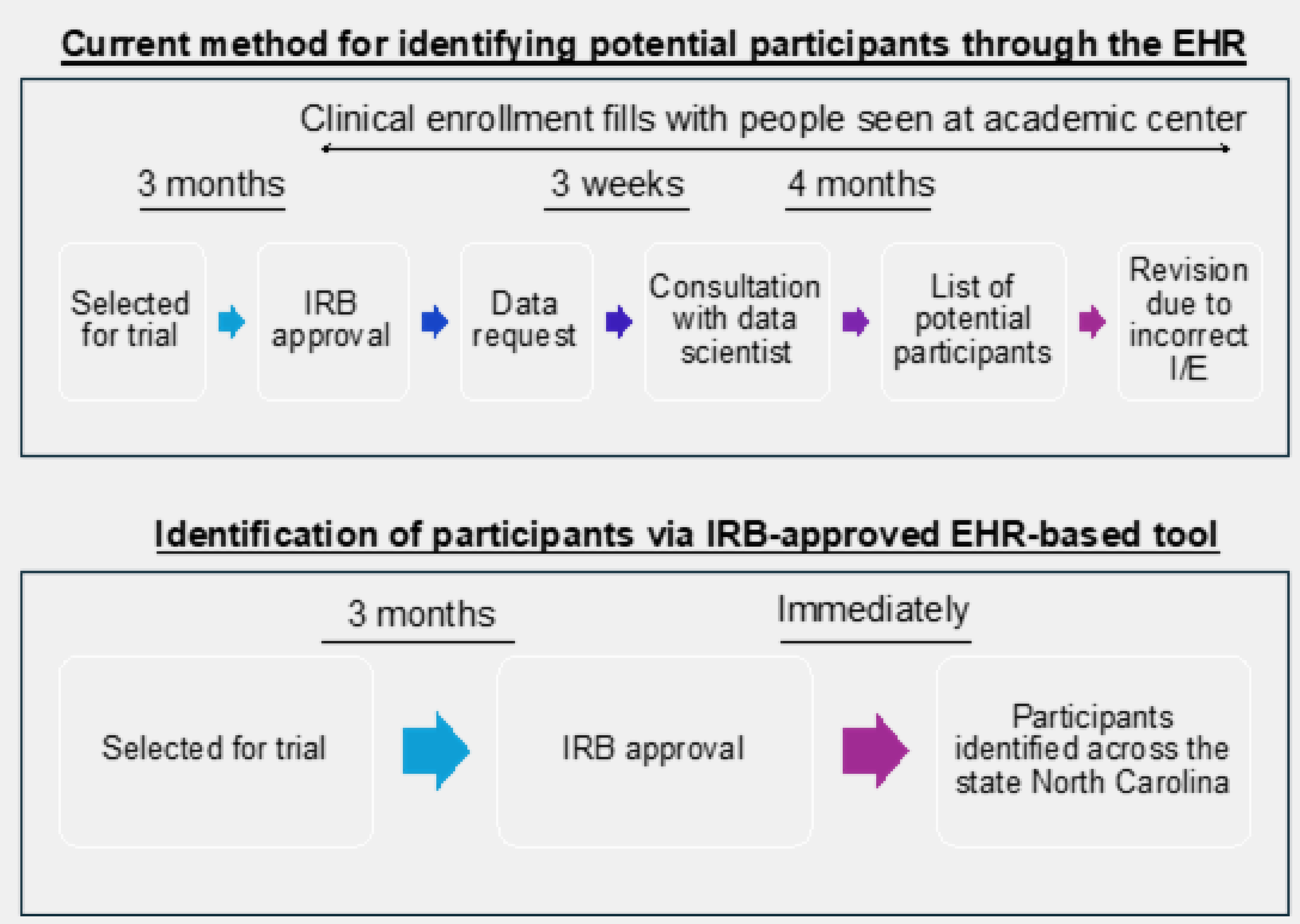

Figure 1. *Current and proposed methods for identification of clinical trial participants in North Carolina*

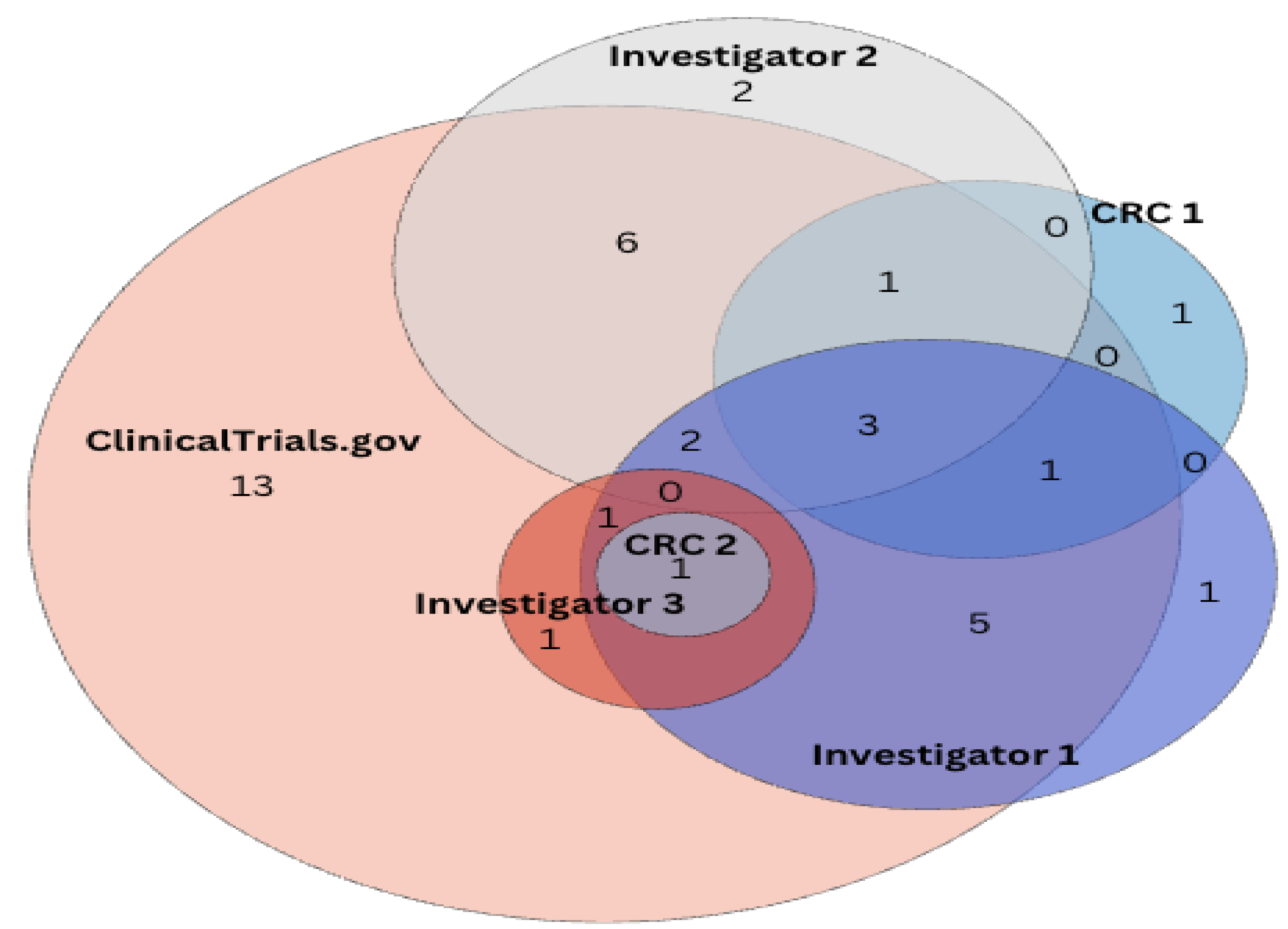

Figure 2. *CT.gov-derived key I/E criteria overlapped substantially with expert-derived criteria.*

- Nine CT.gov-derived normalized, clinically relevant terms were not identified by diabetes trial experts.
- Four expert-derived terms were not included on CT.gov.

Terms Common to CT.gov and Clinical Trial Experts

- Ability to communicate in English
- Age
- Autoimmune related
- BMI
- Cancer related
- CVD
- Diabetes duration
- Diabetes type
- Female reproductive system related
- Gender
- Incretin therapy related
- Hemoglobin A1c
- Immunosuppression or transplant
- Insulin use status
- Kidney related
- Liver disease related
- Mental health related
- Neurologic conditions
- Non-insulin glucose lowering medications
- Race/ethnicity
- Recent illness or surgery
- Retinopathy
- Steroid related
- Use of ACE inhibitors or ARBs
- Use of diabetes technology
- Weight related

Terms Unique to CT.gov

- Allergies
- Bone related
- C-peptide level
- Covid related
- Diabetes complications related
- DKA related
- Gastrointestinal disease related
- Hematology related
- Hypoglycemia related
- Pulmonary related
- Vaccination history

Terms Unique to Clinical Trial Experts

- Prior participation in trial
- Last endocrinology encounter
- Last office encounter

Figure 3. *Clinically relevant, normalized I/E criteria for diabetes trials as derived by CT.gov and trial experts*

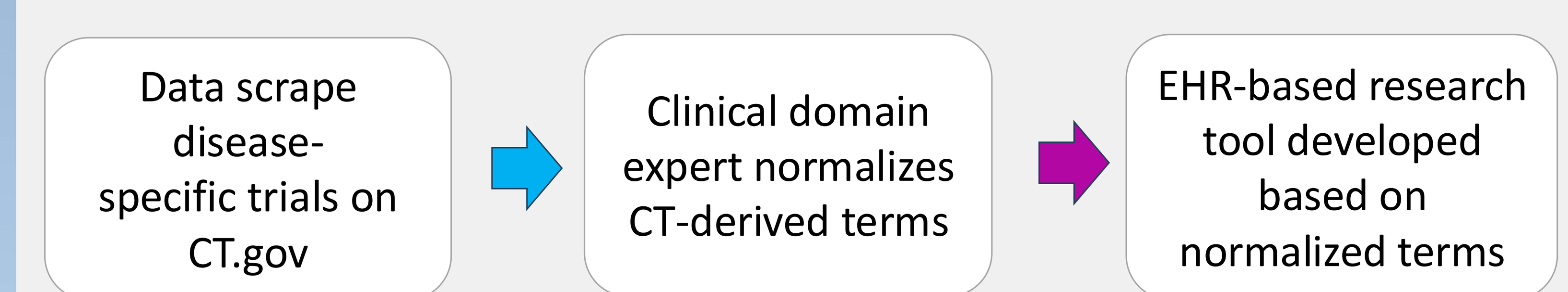

Figure 4. *Proposed method to identify variables for inclusion in a disease-based EHR recruitment tool*

Discussion

- Data scraping CT.gov generated a broad list of I/E for diabetes clinical trials.
- Clinical domain expertise is required to normalize terms to clinically relevant variables.
- Combining data-scraping of CT.gov with clinical domain expertise will identify a comprehensive, clinically relevant list of I/E criteria that can be leveraged for use in an EHR-based research tool.

The project described was supported by the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, through Grant Award Number K12TR004416. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

ABBREVIATIONS: CT.gov, clinicaltrials.gov; CRC, clinical research coordinator; DKA, diabetic ketoacidosis, I/E, inclusion and exclusion.