CONSIDERATIONS FOR IDENTIFYING AND CONDUCTING CLUSTER TRIALS

CONSIDERATIONS FOR IDENTIFYING AND CONDUCTING CLUSTER RANDOMIZED TRIALS

BY

AHMED AL-JAISHI, MSc

A THESIS

SUBMITTED TO THE DEPARTMENT OF

HEALTH RESEARCH METHODS, EVIDENCE, AND IMPACT

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Doctor of Philosophy (2021)

McMaster University
Hamilton, Ontario, Canada

Department of Health Research Methods,
Evidence, and Impact (HEI)

TITLE: Considerations for Identifying and Conducting Cluster Randomized Trials

AUTHOR: Ahmed A. Al-Jaishi, BSc (University of Guelph), MSc (Western University)

SUPERVISOR: Dr. Amit X. Garg

NUMBER OF PAGES: 175

LAY ABSTRACT

The cluster trial design randomly assigns groups of people to different treatment arms rather than individuals. Cluster trials are commonly used in research areas such as education, public health, and health service research. Examples of clusters can include villages/communities, worksites, schools, hospitals, hospital wards, and physicians. This dissertation aimed to (1) develop machine learning algorithms to identify cluster trials in bibliographic databases, (2) assess reporting of methodological and ethical elements in hemodialysis-related cluster trials, and (3) identified best practices for randomly assigning hemodialysis centers in cluster trials. We conducted three studies to address these aims. The results of this dissertation will help researchers quickly identify cluster trials in bibliographic databases (study 1) and inform the design and analyses of future Canadian trials conducted within the hemodialysis setting (study 2 & 3).

ABSTRACT

Background: The cluster randomized trial design randomly assigns groups of people to different treatment arms. This dissertation aimed to (1) develop machine learning algorithms to identify cluster trials in bibliographic databases, (2) assess reporting of methodological and ethical elements in hemodialysis-related cluster trials, and (3) assess how well two covariate-constrained randomization methods balanced baseline characteristics compared with simple randomization.

Methods: In study 1, we developed three machine learning algorithms that classify whether a bibliographic citation is a CRT report or not. We only used the information available in an article citation, including the title, abstract, keywords, and subject headings. In study 2, we conducted a systematic review of CRTs in the hemodialysis setting to review the reporting of key methodological and ethical issues. We reviewed CRTs published in English between 2000 and 2019 and indexed in MEDLINE or EMBASE. In study 3, we assessed how well two covariate-constrained randomization methods balanced baseline characteristics compared with simple randomization.

Results: In study 1, we successfully developed high-performance algorithms that identified whether a citation was a CRT. Our algorithms had greater than 97% sensitivity and 77% specificity in identifying CRTs. For study 2, we found suboptimal conduct and reporting of methodological issues of CRTs in the hemodialysis setting and incomplete reporting of key ethical issues. For study 3, where we randomized 72 clusters, constraining the randomization using historical information achieved a better balance on

baseline characteristics than simple randomization; however, the magnitude of benefit was modest.

Conclusions: This dissertation's results will help researchers quickly identify cluster trials in bibliographic databases (study 1) and inform the design and analyses of future Canadian trials conducted within the hemodialysis setting (study 2 & 3).

To my wife and two beautiful children

ACKNOWLEDGEMENTS

My doctoral experience at the ICES Kidney, Dialysis, Transplantation Program (ICES-KDT) and the Department of Health Research Methods, Evidence, and Impact has been nothing short of exceptional. Thanks to all staff, faculty, and friends for their guidance throughout my course work, academic life, and thesis writing process. Your support and encouragement did not go unnoticed.

I want to express my sincere gratitude to my mentor, Dr. Amit Garg, for all his encouragement, support, and direction. He always motivated me to excel, challenged me with insightful questions, and offered invaluable advice. This thesis is evidence showing a struggle to reflect a portion of his insight, mentorship, and experience. To him, I owe very much.

I want to thank members of my thesis committee, Drs. Lehana Thabane and PJ Devereaux. I was fortunate to have them as committee members. They mentored me in research, helped guide my career, and challenged me to think critically. Their vision, sincerity and motivation deeply inspire me.

A heartfelt thank you goes to my family for supporting me through all the highs and lows. To my wife and two beautiful children, the significance you give me each day is unparalleled. You teach me that while life might seem “random,” there is a message and lesson to everything. You are the people I want to make most proud. I adore you; I love you.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOTATION, DEFINITIONS, AND ABBREVIATIONS

Abbreviations

AUC curve: Area Under the receiver operating Characteristic curves

CBOW: Continuous Bag of Words Model

CIHI: Canadian Institute for Health Information

CCO: Cancer Care Ontario

CIHR: Canadian Institutes of Health Research

CONSORT: Consolidated Standards of Reporting Trials

CRT: Cluster Randomized Trials

CSV: Comma-Separated Values

HD: Hemodialysis

ICC: Intra-class Correlation Coefficient

ICES-KDT: ICES Kidney, Dialysis, Transplantation Program

iCT: innovative Clinical Trials

MOHLTC: Ministry of Health and Long-Term Care

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

REC: Research Ethics Committee

RCT= Randomized Controlled Trials

SVM: Support Vector Machines

TF-IDF: Term Frequency-Inverse Document Frequency

TPE: Tree of Parzen Estimators

DECLARATION OF ACADEMIC ACHIEVEMENT

I led each of the three studies in this thesis (2016 – 2021). The co-authors provided supervision or a supportive role as indicated below.

Study 1 (Chapter 2)

AAA conceived the overall project idea and led the study design. AAA developed the data extraction tool, and AAA and MDA extracted the study data for the training and validation datasets. AAA conducted all analyses. AAA wrote the initial draft of the manuscript with input from all authors. All authors contributed to the interpretation of results, made critical revisions, and approved the final manuscript. AA is the guarantor.

Study 2 (Chapter 3)

AXG and MT conceived the overall project idea and co-led the funding application with contributions from all authors. MT developed the data extraction tool and AAA, KC, and CEG extracted the study data. AAA, MT, and SND extracted details on trials considering the effect of clustering during sample size estimation and analysis. AAA, KC, CEG, and MT wrote the manuscript's initial draft with substantial input from all authors. All authors contributed to the interpretation of results, made critical revisions, and approved the final manuscript. AA is the guarantor.

Study 3 (Chapter 4)

AAA and AXG conceived and led the study design. SND, EM, PJD, and LT contributed to the study design. AA was responsible for data management and analysis. AA drafted the manuscript. All authors contributed to manuscript revision and approved the final manuscript. AA is the guarantor.

CHAPTER 1: INTRODUCTION

This chapter discusses a general overview of randomized trials and how cluster randomized trials (CRTs) differ from individually randomized trials. The second part of the chapter discusses considerations for conducting CRTs. The chapter ends with the relevance of CRTs in the hemodialysis setting. These sections set the stage for the three studies discussed in this thesis where we **(1)** used machine learning algorithms to identify CRTs in bibliographic databases, **(2)** assessed reporting of methodological and ethical elements in hemodialysis-related CRTs, and **(3)** identified best practices for using covariate constrained randomization in hemodialysis-related, registry-based CRTs.

Randomized trials

Randomized controlled trials (RCTs) play a central role in evidence-based medicine. The RCT study design is regarded as the "gold standard" for evaluating the efficacy and effectiveness of an intervention. There is broad acceptance that major public health interventions or clinical treatments should be based on large and rigorously conducted randomized trials.

In an RCT, participants are randomly assigned to different treatment arms (called randomization). If implemented correctly and with a sufficiently large sample size, randomization reduces the chances of bias and provides a rigorous tool to examine causal relationships between the intervention and outcome. This is because randomization enables researchers to assemble treatment groups comparable in every aspect other than the treatment condition. In addition to randomization, trialists also employ allocation

concealment, blinding, trial monitoring, strategies that minimize loss-to-follow-up, and intention-to-treat analyses to enhance RCTs' rigor. As such, we can confidently attribute any differences in group outcomes to the treatment condition. This is not possible with other study designs.

In contrast to individually randomized trials, CRTs randomized ***groups*** of participants to different treatment arms. In individual-level trials, participants are assumed to be independent of each other, and there is no interaction among trial participants after randomization. However, participants in a CRT are not independent and can interact with each other after randomization. Cluster trials are commonly used in research areas such as education, public health, and health service research. Examples of clusters can include villages/communities, worksites, schools, hospitals, hospital wards, and physicians.

In individual-level RCTs, the participant is the unit of randomization, and the intervention is applied to the participant, and analyses are conducted at the individual level. For CRTs, the cluster is the unit of randomization, and the intervention can be applied at the individual or cluster level. The analysis can also be conducted at the individual or cluster level.

Reasons for using a CRT design

The most obvious reason for using a CRT design is when the intervention is naturally applied to groups of people rather than individuals, such as water fluoridation to improve dental health in a community. The CRT design is also useful when there are situations where participants in the control arm may adopt the intervention, hence attenuating

potential treatment effects. There may also be logistical or administrative considerations that make it easier to implement the intervention at the group (or cluster) level.

Methodological and ethical issues in conducting CRTs

The CRT design has several disadvantages, and accordingly, its use must be carefully justified. These disadvantages include but are not limited to: **(1)** dependence (or clustering) among individuals within the same cluster – for example, outcomes of patients within a cluster are more alike than patients' outcomes in different clusters. This implies that larger sample sizes are required. **(2)** CRTs pose unique challenges to existing ethical, legal, and social frameworks for researchers, research ethics committees, regulators, health system managers, and funders; [1–4] **(3)** obtaining informed consent in CRTs can be more complicated than individual-level RCTs. When individual consent is required, authors should specify the nature of the consent because it can be obtained at one or more levels, e.g., individual- and cluster–level. In a CRT, participant consent can also be obtained differently for each treatment group and for different things (e.g., data collection, study interventions, and randomization.)

Randomized trials in nephrology

End-stage kidney disease and dialysis

When a person's kidney fails, they require renal replacement therapy (i.e., dialysis or a kidney transplant) to remove excess fluid and eliminate toxins from the blood. Hemodialysis (HD) provides a life-saving treatment option for 23,000 Canadians and over two million individuals worldwide living with kidney failure. Individuals being treated with hemodialysis have poor health outcomes and high mortality – 30 times

higher than the general population; where 20-40% of patients die within one year of starting hemodialysis.[5–7]

Hemodialysis-related randomized controlled trials

Kidney medicine, particularly for hemodialysis, conducts fewer clinical trials than any other internal medicine specialty.[8] Many trials suffer from low-quality reporting and study design.[8] Traditional explanatory and individual-level RCTs conducted in the field of hemodialysis also suffer from poor recruitment, inadequate sample sizes, and missing data.[8–13] As a result, many treatment strategies for patients on hemodialysis are based on expert opinion from physiology and clinical experience rather than clinical trial data. There is a need to design more trials with broader inclusion criteria to generate high-quality evidence supporting real-world practice.

A significant challenge to conducting traditional RCTs for several hemodialysis-based interventions is cross-group contamination. For example, one healthcare provider cares for many patients in a single hemodialysis center. If a healthcare provider observes some patients in the intervention arm have better outcomes than those in the control arm, they may apply it to patients in the control group. This action negates the purpose of the randomization and contaminates the control group. The CRT design protects against bias from cross-group contamination because intact groups (e.g., the entire center) are randomized to the same treatment.

Hemodialysis-related cluster randomized trials

The hemodialysis setting is well suited for conducting CRTs because (1) patients treated in-center have frequent and predictable encounters with the healthcare system, (2) there is

detailed and uniform data collection across centers, (3) electronic data systems are often used, and the data may be captured in local, provincial/state, or national registries and (4) delivery of care is administered by a small number of provider organizations.[14] Additionally, many changes implemented in the dialysis setting are often administered to all patients within a center as a policy determined by the hemodialysis director.

Thesis content and structure

This thesis is presented in a sandwich format. The projects were undertaken in the context of a larger Canadian Institutes of Health Research (CIHR) multi-year initiative that aims to develop recommendations for the design, conduct, and reporting of CRTs in the hemodialysis setting. Following this chapter, chapters 2 to 4 include manuscripts that are published or submitted for publication. The content of chapters 2 to 4 is stand-alone and does not need to be read sequentially.

Overview study 1: The manuscript for study 1 (chapter 2), titled "Machine learning algorithms to identify cluster randomized trials from MEDLINE and EMBASE," has been submitted for publication.

Cluster randomized trials are poorly indexed, making them difficult to retrieve from Medline, EMBASE, and other large bibliographic databases. Machine learning algorithms can improve retrieval and have proven highly accurate for identifying RCTs.[15–17] In this study, we developed machine learning algorithms to classify whether an article cited in bibliographic databases is a CRT or not. We based our classification on

the information available in an article bibliographic citation, including the title, abstract, keywords, and subject headings.

Overview study 2: The manuscript for study 2 (chapter 3), titled "Reporting of key methodological and ethical aspects of cluster trials in hemodialysis require improvement: a systematic review," is published in *Trials*. [18]

The hemodialysis setting is suitable for trials that use cluster randomization, where intact groups of individuals are randomized. However, CRTs are complicated in their design, analysis and reporting, and can pose ethical challenges. In this study, we aimed to systematically review and assess the reporting of key methodological and ethical issues for CRTs published in the hemodialysis setting.

Overview study 3: The manuscript for study 3 (chapter 4), titled " Simple compared to covariate-constrained randomization methods in balancing baseline characteristics: a case study of randomly allocating 72 hemodialysis centers in a cluster trial," has been submitted for publication.

Some parallel-group cluster randomized trials use covariate-constrained rather than simple randomization. This is done to increase the chance of balancing the groups on the cluster- and patient-level baseline characteristics. This study assessed how well two covariate-constrained randomization methods balanced baseline characteristics compared with simple randomization. We wanted to understand the best practices for randomizing hemodialysis centers into two parallel groups in Ontario, Canada.

Finally, chapter 5 includes a summary of the findings, overall implications from the three studies, future directions, and concluding remarks.

References

1. Weijer C, Grimshaw JM, Taljaard M, et al. Ethical issues posed by cluster randomized trials in health research. *Trials*. 2011;12(1):100. doi:10.1186/1745-6215-12-100

2. Califf RM, Sugarman J. Exploring the ethical and regulatory issues in pragmatic clinical trials. *Clin Trials*. 2015;12(5):436-441. doi:10.1177/1740774515598334

3. Sugarman J, Califf RM. Ethics and Regulatory Complexities for Pragmatic Clinical Trials. *JAMA*. 2014;311(23):2381. doi:10.1001/jama.2014.4164

4. Donner A, Klar N. *Design and Analysis of Cluster Randomization Trials in Health Research*. (Gooster L, Ueberberg A, eds.). Arnold; 2000.

5. Foley RN, Collins AJ. End-stage renal disease in the United States: an update from the United States Renal Data System. *J Am Soc Nephrol*. 2007;18(10):2644-2648. doi:10.1681/ASN.2007020220

6. Foote C, Ninomiya T, Gallagher M, et al. Survival of elderly dialysis patients is predicted by both patient and practice characteristics. *Nephrol Dial Transplant*. 2012;27(9):3581-3587. doi:10.1093/ndt/gfs096

7. Jassal SVS, Trpeski L, Zhu N, et al. Changes in survival among elderly patients initiating dialysis from 1990 to 1999. *CMAJ*. 2007;177(9):1033-1038. doi:10.1503/cmaj.061765

8. Strippoli GFM, Craig JC, Schena FP. The number, quality, and coverage of randomized controlled trials in nephrology. *J Am Soc Nephrol*. 2004;15(2):411-

419.

9. de Boer IH, Kovesdy CP, Navaneethan SD, et al. Pragmatic Clinical Trials in CKD: Opportunities and Challenges. *J Am Soc Nephrol*. 2016;27(10):2948-2954. doi:10.1681/ASN.2015111264

10. Covic A, Gusbeth-Tatomir P, Goldsmith D. Negative outcome studies in end-stage renal disease: how dark are the storm clouds? *Nephrol Dial Transplant*. 2007;23(1):56-61. doi:10.1093/ndt/gfm706

11. Palmer SC, Sciancalepore M, Strippoli GFM. Trial Quality in Nephrology: How Are We Measuring Up? *Am J Kidney Dis*. 2011;58(3):335-337. doi:10.1053/j.ajkd.2011.06.006

12. Deo A, Schmid CH, Earley A, Lau J, Uhlig K. Loss to Analysis in Randomized Controlled Trials in CKD. *Am J Kidney Dis*. 2011;58(3):349-355. doi:10.1053/j.ajkd.2011.04.023

13. Inrig JK, Califf RM, Tasneem A, et al. The landscape of clinical trials in nephrology: a systematic review of Clinicaltrials.gov. *Am J Kidney Dis*. 2014;63(5):771-780. doi:10.1053/j.ajkd.2013.10.043

14. Dember LM, Archdeacon P, Krishnan M, et al. Pragmatic Trials in Maintenance Dialysis: Perspectives from the Kidney Health Initiative. *J Am Soc Nephrol*. 2016;27(10):2955-2963. doi:10.1681/ASN.2016030340

15. Cohen AM, Smalheiser NR, McDonagh MS, et al. Automated confidence ranked

classification of randomized controlled trial articles: an aid to evidence-based medicine. *J Am Med Inf Assoc*. 2015;22(3):707-717. doi:doi:10.1093/jamia/ocu025

16. Wallace B, Noel-Storr A, Marshall I, Cohen A, Smalheiser N, Thomas J. Identifying reports of randomized controlled trials (RCTs) via a hybrid machine learning and crowdsourcing approach. *J Am Med Inf Assoc*. 2017;24(6):1165-1168. doi:doi:10.1093/jamia/ocx053

17. Marshall IJ, Noel-Storr A, Kuiper J, Thomas J, Wallace BC. Machine learning for identifying Randomized Controlled Trials: An evaluation and practitioner's guide. In: *Research Synthesis Methods*. Vol 9. John Wiley and Sons Ltd; 2018:602-614. doi:10.1002/jrsm.1287

18. Al-Jaishi AA, Carroll K, Goldstein CE, et al. Reporting of key methodological and ethical aspects of cluster trials in hemodialysis require improvement: A systematic review. *Trials*. 2020;21(1). doi:10.1186/s13063-020-04657-9

CHAPTER 2: STUDY 1

TITLE: Machine learning algorithms to identify cluster randomized trials from MEDLINE and EMBASE

AUTHORS: Ahmed A. Al-Jaishi, MSc [1,2]; Monica Taljaard, Ph.D. [3]; Melissa D. Al-Jaishi, BScN, MLIS [4]; Sheikh S. Abdullah, Ph.D. [5,6]; Lehana Thabane, Ph.D. [2]; PJ Devereaux, MD, Ph.D. [2]; Stephanie N. Dixon, Ph.D. [1,5,7,8]; Amit X. Garg, MD, Ph.D. [1,2,5,7]

Affiliations

[1] Lawson Health Research Institute, London, Ontario, Canada

[2] McMaster University, Department of Health Research Methods, Evidence, and Impact, Hamilton, ON, Canada

[3] Ottawa Hospital Research Institute, Clinical Epidemiology Program; and School of Epidemiology and Public Health, University of Ottawa, Ottawa, ON, Canada

[4] Western University, Arthur Labatt Family School of Nursing, London, Ontario, Canada

[5] ICES, Ontario, Canada

[6] Department of Computer Science, Western University, London, ON, Canada

[7] Department Medicine, Epidemiology and Biostatistics, Western University, London, ON, Canada

[8] Department of Mathematics and Statistics, University of Guelph, Guelph, ON, Canada

KEYWORDS: Cluster Randomized Controlled Trial; Machine Learning; Bibliographic Databases; Sensitivity and Specificity; Prediction

Citation: Al-Jaishi AA, Taljaard M, Al-Jaishi MD, et al. Machine learning algorithms to identify cluster randomized trials from MEDLINE and EMBASE. [submitted]

Abstract

Background: Cluster randomized trials (CRTs) are an increasingly important design. However, authors do not always adhere to requirements to explicitly identify the design as cluster randomized in titles and abstracts, making retrieval from bibliographic databases difficult. Machine learning algorithms may improve their identification and retrieval.

Aims: Develop machine learning algorithms that accurately determine whether a bibliographic citation is a CRT report. The information available in an article citation includes the title, abstract, keywords, and subject headings.

Methods: We trained, internally validated, and externally validated two convolutional neural networks and one support vector machines (SVM) algorithms to predict whether a citation is a CRT report or not. The algorithms' output was a probability from 0 to 1. We assessed algorithm performance using the area under the receiver operating characteristic (AUC) curves. Each algorithm's performance was assessed individually and together as an ensemble. We randomly selected 5000 from 87,633 citations to train and internally validate our algorithms. Of the 5000 selected citations, 589 (12%) were confirmed CRT reports. We then externally validated our algorithms on an independent set of 1916 randomized trial citations, with 665 (35%) being confirmed CRT reports.

Results: In internal validation, the ensemble algorithm discriminated best for identifying CRT reports with an AUC of 98.6% (95% confidence interval: 97.8%, 99.4%), sensitivity of 97.7% (94.3%, 100%), and specificity of 85.0% (81.8%, 88.1%). In external

validation, the ensemble algorithm had an AUC of 97.8 % (97.0%, 98.5%), sensitivity of 97.6% (96.4%, 98.6%), and specificity of 78.2% (75.9%, 80.4%)). All three individual algorithms performed well, but less so than the ensemble.

Conclusions: We successfully developed high-performance algorithms that identified whether a citation was a CRT report with high sensitivity and moderately high specificity. We provide open-source software to facilitate the use of our algorithms in practice.

Introduction

Randomized controlled trials (RCTs) are a robust study design to evaluate health interventions. Compared to individually randomized trials, which randomize individuals, cluster randomized trials (CRTs) allocate groups of people, such as medical practices, hospitals, nursing homes, schools, or even entire communities. Methodologists increasingly search MEDLINE, EMBASE, and other bibliographic databases for reports of CRTs. [1–6] Unfortunately, authors do not always adhere to the Consolidated Standards of Reporting Trials (CONSORT) Statement Extension for Cluster Randomized Trials requirements to explicitly identify the design as "cluster randomized" in the title or abstract of the report. As such, it is challenging to retrieve reports of CRTs from bibliographic databases. As of June 2020, we estimate that less than 0.1% of the 17.5 million citations in PubMed over the prior two decades were CRT reports; finding reports of CRTs in bibliographic databases is a problem akin to screening for rare diseases in the general population. [7]

A common practice for identifying CRT reports may involve using an established *database search filter*. [8] Search filters contain combinations of text strings and database tags developed by information specialists. An existing search filter captures over 90% of CRT-related articles. [8] However, this filter also captures many articles that are not CRT reports, and CRT reports represent only 10% to 15% of articles identified by the search filter. [8] Thus, a reviewer needs to screen ten records or more to identify one CRT report. This process is time-consuming, with a chance for human error.

Machine learning and text mining techniques can extract useful information from an article's citation (e.g., title and abstract) and have proved successful in classifying whether an article's citation is an RCT. [9–11] In this study, we developed and both internally and externally validated machine learning algorithms to accurately determine whether an article citation is a CRT report so it can be retrieved when searching bibliographic databases.

Methods

This section is organized into five subsections. First, we describe the data sets used to train, internally validate, and externally validate the machine learning algorithms. Second, we describe the machine learning algorithms. Third, we describe how we processed the data, trained our algorithms, and optimized each algorithm's hyperparameters. Fourth, we describe how we combined our models (referred to as an ensemble method) to boost the algorithm's predictive performance compared to a single model. Finally, we describe the evaluation metrics used to test our algorithms' overall performance.

Datasets

Training and internal validation

To identify article citations for our training and internal validation sets, we used a previously published CRT search filter in MEDLINE and EMBASE that yielded 87,633 citations published between January 1st, 2000, and December 31st, 2019 (see **Appendix 2-1** for additional details about the search). [8] We randomly selected 5000 citations from these records for training and internal validation. Two reviewers (AAA & MDA) independently classified whether each citation was a CRT report or not (over 97%

agreement between both reviewers). Inclusion criteria were primary or secondary reports of CRTs, CRT protocols, or pilot and feasibility of CRTs. Citations meeting those eligibility criteria were included regardless of the setting, clinical area, or cluster type. Exclusion criteria were trials reporting only baseline findings, quasi-randomized trials, studies reporting process evaluation or methods papers, individually randomized trials, observational studies, editorials, and mechanistic studies. The reviewers based their assessment primarily on the title and abstract, but the article's full text was reviewed in cases where the unit of randomization was unclear. We expected that 10% to 15% of the 5000 articles would be CRT reports. [8]

External validation dataset

We evaluated our algorithms' performance against an external dataset that included 1988 articles. These articles were confirmed primary reports of RCTs, of which 688 were CRT reports and the rest were individually randomized trials. This dataset has been described elsewhere. [12] Briefly, the authors identified pragmatic clinical trials using a sensitivity-maximizing pragmatic search filter; this search filter is independent of this study's CRT search filter. [13] The search filter for the external dataset was applied in MEDLINE on April 3rd, 2019, for the period between January 1st, 2014, and April 3rd, 2019.

From the 1988 articles, we removed 72 articles that were captured in the training or validation datasets. We applied this exclusion criterion to avoid data leakage that would artificially inflate the models' performance. See **Appendix 2-2** for additional details about the external dataset.

Machine learning algorithms

Convolutional neural networks

Although convolutional neural networks were initially developed and used for image classification, they have emerged as state-of-the-art models for text classification. [14–17] These models use low-dimensional (typically 50 to 300) continuous vectors to represent words (called word embeddings). The convolutional neural network algorithm takes an input text document, assigns learnable weights and biases to different words or phrases. The algorithm passes linear filters represented with corresponding weight vectors over word embeddings. The filters start at the beginning and move sequentially through the document. As such, each filter produces a vector that is proportional in size to the document length. Filter outputs are then combined by extracting the maximum value on each filter output vector (i.e., max-pooling). Finally, the algorithm concatenates these scalar values to form a vector representation of the entire document that becomes the input for the classification prediction layer (**eFigure 2-1**).

Support Vector Machines

Support vector machines identify the best hyperplane that separates classes (e.g., CRT vs. non-CRT report) in high-dimensional space. [18] This method uses kernel functions (i.e., a similarity function between a pair of records) that can be a linear, polynomial, sigmoid, or radial basis function. These kernel functions transform the data into the form necessary for prediction. We trained the models using the above-mentioned kernel functions and found the radial basis function performed best (see "Hyperparameter optimization"). **Appendix 2-3** and **eFigure 2-2** provide more details on the two parameters associated

with the radial basis function: gamma (kernel coefficient) and c (regularization parameter).

Data preprocessing, model features, and hyperparameter choices

Data preprocessing

We concatenated each citation's title, abstract, keywords, and subject headings. We conducted several data cleaning steps for each record, including putting text in lowercase, removing brackets/parentheses, punctuations, numbers, and words containing numbers. When a citation had a structured abstract, we removed the discussion and conclusion because these sections rarely contained relevant information about the study design. We then tokenized titles, abstracts, keywords, and subject headings. Finally, we removed stopwords such as "*and," "the," "we," and "was"* (*i.e.,* common words with low informational content).

Word embeddings

For the convolutional neural network models, we used word embeddings as feature parameters. A word embedding is a learned representation for text, where words that have a similar meaning (i.e., used in a similar context) will have similar representation in vector space (e.g., "mother," "father," "parent," "guardian" would have similar vector representation). An unsupervised neural network maps each word to one vector. We trained two different word embedding models: Word2Vec and FastText, with the skip-gram architecture and ten iterations (**Appendix 2-4**). [19,20] We trained the word embedding models using the 87,633 articles retrieved by our search strategy.

Term frequency-inverse document frequency

We used the Term Frequency-Inverse Document Frequency (TF-IDF) method to weigh the relative importance of unique words in our dataset for support vector machines. [18,21] The TF-IDF weights increase proportionally to the number of times a word appears in a document. These weights are offset by the number of records containing that word, which helps to adjust for expressions frequently appearing in the dataset (e.g., the word "random"). Information retrieval, text mining, and user modeling tasks commonly use the TF-IDF method as a weighting factor. [22]

Handling class imbalance

There are far fewer CRT than non-CRT reports, which posed a problem for standard learning algorithms that aim to maximize overall predictive accuracy. Given this class imbalance scenario, we observed high model accuracy by uniformly predicting the majority class (i.e., non-CRT reports). We handled class imbalance by (1) constructing a dataset that included all CRT reports but only a random subsample of non-CRT reports and (2) adjusting class weights where each CRT training example carried more weight than non-CRT reports. [23] **Table 2-1** shows the details of the search space and the chosen sampling ratio.

Hyperparameter optimization

It was not feasible to conduct a grid search over all specified hyperparameters for the convolutional neural network models. We used the *hyperopt* python library, which implements Bayesian hyperparameter optimization, to optimize the algorithm's hyperparameters and achieve the highest algorithm performance. [24] We implemented the Tree of Parzen Estimators (TPE) algorithm with 500 iterations.[25,26] We also optimized the

class weighting, the sampling ratio, and the L1 regularization strength. We examined the effect of different numbers and sizes of filters and differing dropout rates. Dropout rates influence the proportion of neural network connections randomly dropped during training, a strategy used to prevent overfitting. [27] Finally, we examined the effect of varying the vocabulary size where we retained the N most common words (e.g., 5000) from the training data. **Table 2-1** shows the full details of the search space and the chosen hyperparameters.

For support vector machines, we performed a grid search over all hyperparameters, including sampling ratio, class weights, kernel functions, kernel coefficient(s), regularization parameter, and word vectorization. We also compared the model's predictive ability when using unigrams, bigrams, and trigrams. **Table 2-1** provides the search spaces and chosen hyperparameters.

Ensembling

Ensemble learning helps improve results from machine learning models by combining two or more models to boost predictive performance compared to a single model. [28] We evaluated the convolutional neural networks and support vector machines individually and as an ensemble of all three algorithms. We estimated the final predicted probability that an article was a CRT report by calculating the average probability score across the three ensembled algorithms.

Evaluation methods

The algorithms output a probability score from 0 to 1 that an article citation was a CRT report. For each algorithm, we plotted the area under the receiver operating characteristic

curve (AUC) as the true positive rate by the false positive rate. The AUC values range between 0 to 1. The AUC value provides information about how well an algorithm can distinguish CRT from a non-CRT report; the closer the AUC value to 1, the better the algorithm predicts non-CRT articles as a non-CRT and CRT articles as a CRT. The bootstrap procedure in the pROC package in R 3.6.1 was used to estimate the AUCs confidence intervals. [29] As a secondary measure, we estimated the number needed to screen, defined as the average number of algorithm-positive articles that need to be manually read to retrieve one CRT report. The prevalence of CRT reports in the respective search strategy and domain area influences the number needed to screen and should be interpreted with caution.

To enable the classification of articles, we chose a threshold probability score to decide whether an article is a CRT report. An article with a probability score greater than the threshold was labeled as a CRT report. We aimed at choosing a probability threshold that would lead to the final algorithm's sensitivity score to be greater than 95% for the internal validation dataset without significantly harming the specificity.

The final algorithms were chosen using the single best-performing hyperparameters trained on the full training and internal validation datasets. We then assessed the best-performing algorithms on an external dataset. We conducted data processing and analyses using Python 3.7.7; **Appendix 2-5** describes the python libraries used for this project. [30,31,40,32–39]

Results

From the 5000 selected articles, 589 were confirmed to be CRT reports and the remaining articles were either not CRT reports or were otherwise ineligible. We classified the design for 850 (17%) of the 5000 articles based on the full text, while the remaining articles were classified based on the title and abstract alone; we reviewed the full text when the randomization unit was unclear. The 589 CRT citations had 111,492 words and the 4411 non-CRT citations had 816,167 words. **Figure 2-1** illustrates a scatter plot of terms associated with CRT and non-CRT reports. For example, from the interactive version of **Figure 2-1**, 67% of CRT-related articles compared to 1.7% for non-CRT articles contained the term "cluster" in their title, abstract, keyword, or as a subject heading.

Table 2-2 displays each algorithm's performance characteristics for the internal and external datasets. We evaluated the three machine learning algorithms separately and as an ensemble. The individual algorithms operated well. However, the ensemble discriminated best on the validation dataset with an AUC of 98.6% (95% confidence interval: 97.8%, 99.4%), sensitivity of 97.7% (94.3%, 100%), and specificity of 85.0% (81.8%, 88.1%); **Figure 2-2** shows the algorithm's performance. For the internal validation dataset, a person would need to screen 6.8 citations, on average, to identify one CRT report. That number dropped to 1.9 citations when using the ensemble algorithm.

For the external dataset (665 CRT reports of 1916 articles), the ensemble algorithm had an AUC of 97.8 % (97.0%, 98.5%), sensitivity of 97.6% (96.4%, 98.6%), and specificity

of 78.2% (75.9%, 80.4%)) [**eFigure 2-3**]. The number needed to read dropped from 2.9 to 1.4 citations after implementing the ensemble algorithm.

Discussion

We showed that CRT reports could be reliably classified using an ensemble of machine learning algorithms. We expect our algorithms' performance will improve overtime because (1) we will continue to fine-tune our algorithms as our repository of CRT reports increase, and (2) we expect better reporting of the CRT study design as both journals and the CONSORT Statement Extension for Cluster Trials mandate (or recommend) that authors publishing CRT reports include the study design explicitly in the title and abstract. [41]

For systematic reviews, our algorithms lead to a substantial reduction in the number of citations needed to screen with a low probability that CRT reports are excluded (sensitivity greater than 97%.) To facilitate the use of our algorithms in practice, we have made these algorithms available as an open-source software called *MLScreener* (MLScreener.com). The user conducts their database search with clinical terms using their desired search syntax combined with the existing CRT search filter (**Appendix 2-6**). The user must save the output of retrieved articles from their preferred bibliographic database in a .csv file. *MLScreener* takes the search result (in .csv format) as input and generates a sorted list that ranks articles from highest to lowest probability (as well as a binary classification of CRT or not). We recommend that researchers screen the sorted list of articles and make an informed decision to stop based on either a low likelihood of

seeing CRT reports in the remaining articles or when resources (time or money) have been spent.

As an illustration of our algorithms' application, we implemented our algorithms on a search strategy that identified articles for a systematic review of CRT reports aiming to capture ethical and methodological reporting issues in the dialysis setting. [42] The search strategy for this review identified 882 potentially relevant articles. Our ensemble algorithm correctly identified 33 of the 34 CRT-related articles that two independent screeners identified in their review; see **Appendix 2-7** for more details. The number of records that required screening was reduced by more than a half (note, this will differ depending on the prevalence of CRT reports in the relevant area).

We estimate that it would require an average of two minutes to screen and assess a single article's eligibility per reviewer for systematic reviews. In our application dataset (882 records), screening time required approximately 30 hours per reviewer, which would have been reduced to 15 hours per reviewer if we used our algorithms.

Limitations

First, in the absence of available databases of confirmed CRT reports, we created our training and internal validation datasets from a random sample of 5000 articles published between 2000 and 2019. We did not review the full text of all 5000 articles to judge whether they were a report of a CRT or not. Thus, we may have missed some CRT reports. However, it is unlikely we missed a significant proportion of articles given our algorithms' high discriminative ability on the external validation dataset, where the full-

text for each article was reviewed. Second, articles included in our external dataset were published between 2014 to 2019. Thus, we had no external validation for the period before 2014. We expect the reliability for classifying CRT reports to resemble the results reported here. Finally, we did not conduct any user analysis for the proposed *MLScreener* software. Thus, we are not aware of how users will engage and interact with our application.

Funding

Ahmed Al-Jaishi was supported by the Allied Health Doctoral Fellowship from the Kidney Foundation of Canada, CIHR Doctoral Award, and McMaster University Michael DeGroote Scholarship. Stephanie Dixon's research is supported through a SPOR Innovative Clinical Trial Multi-Year Grant (Grant number: MYG-151209) from the Canadian Institutes of Health Research (CIHR). Amit Garg was supported by the Dr. Adam Linton Chair in Kidney Health Analytics and a Clinician Investigator Award from the CIHR.

Author Contributions

AAA conceived the overall project idea and led the study design. AAA developed the data extraction tool, and AAA and MDA extracted the study data for the training and validation datasets. AAA conducted all analyses. AAA wrote the initial draft of the manuscript with substantial input from all authors. All authors contributed to the interpretation of results, made critical revisions, and approved the final manuscript. AA is the guarantor.

Acknowledgments

We are grateful to Dr. Monica Taljaard, Dr. Stuart G Nicholls, Ms. Kelly Carroll, Ms. Jennifer Zhang from the Ottawa Hospital Research Institute and Mr. Hayden Nix (Western University) for providing the datasets used for our algorithms' external validation.

Ethics approval and consent to participate

We did not apply for ethics approval as we conducted this study using published literature.

References

1. Hayes RJ, Moulton LH. *Cluster Randomised Trials*. CRC Press; 2009.

2. Eldridge S, Kerry SM. *A Practical Guide to Cluster Randomised Trials in Health Services Research*. John Wiley & Sons; 2012.

3. Awasthi S, Peto R, Read S, et al. Vitamin A supplementation every 6 months with retinol in 1 million pre-school children in north India: DEVTA, a cluster-randomised trial. *Lancet*. 2013;381(9876):1469-1477.

4. WHO | Vitamin A supplementation in infants and children 6–59 months of age. Accessed July 17, 2020. https://www.who.int/elena/titles/vitamina_children/en/

5. OSSU. Our Impact – Ontario SPOR SUPPORT Unit. Published 2020. Accessed July 17, 2020. https://ossu.ca/our-impact/

6. Waters E, Armstrong R, Swinburn B, et al. An exploratory cluster randomised controlled trial of knowledge translation strategies to support evidence-informed decision-making in local governments (The KT4LG study). *BMC Public Health*. 2011;11:34. doi:10.1186/1471-2458-11-34

7. Pubmed. Published 2020. https://pubmed.ncbi.nlm.nih.gov/?term=2000%3A2019%5Bdp%5D

8. Taljaard M, McGowan J, Grimshaw JM, et al. Electronic search strategies to identify reports of cluster randomized trials in MEDLINE: low precision will improve with adherence to reporting standards. *BMC Med Res Methodol*.

2010;10(1):15. doi:10.1186/1471-2288-10-15

9. Cohen AM, Smalheiser NR, McDonagh MS, et al. Automated confidence ranked classification of randomized controlled trial articles: an aid to evidence-based medicine. *J Am Med Inf Assoc*. 2015;22(3):707-717. doi:doi:10.1093/jamia/ocu025

10. Wallace B, Noel-Storr A, Marshall I, Cohen A, Smalheiser N, Thomas J. Identifying reports of randomized controlled trials (RCTs) via a hybrid machine learning and crowdsourcing approach. *J Am Med Inf Assoc*. 2017;24(6):1165-1168. doi:doi:10.1093/jamia/ocx053

11. Marshall IJ, Noel-Storr A, Kuiper J, Thomas J, Wallace BC. Machine learning for identifying Randomized Controlled Trials: An evaluation and practitioner's guide. In: *Research Synthesis Methods*. Vol 9. John Wiley and Sons Ltd; 2018:602-614. doi:10.1002/jrsm.1287

12. Nicholls SG, Carroll K, Hey SP, et al. A review of pragmatic trials found a high degree of diversity in design and scope, deficiencies in reporting and trial registry data, and poor indexing. *J Clin Epidemiol*. Published online March 2021. doi:10.1016/j.jclinepi.2021.03.021

13. Taljaard M, McDonald S, Nicholls SG, et al. A search filter to identify pragmatic trials in MEDLINE was highly specific but lacked sensitivity. *J Clin Epidemiol*. 2020;124:75-84. doi:10.1016/j.jclinepi.2020.05.003

14. Kim Y. Convolutional Neural Networks for Sentence Classification. *EMNLP 2014*

- *2014 Conf Empir Methods Nat Lang Process Proc Conf*. Published online August 25, 2014:1746-1751.

15. Zhang Y, Wallace B. A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification. Published online October 13, 2015.

16. Young T, Hazarika D, Poria S, Cambria E. Recent Trends in Deep Learning Based Natural Language Processing. Published online August 9, 2017.

17. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM*. 2017;60(6):84-90. doi:10.1145/3065386

18. Raschka S. *Python Machine Learning*. (Hussain A, Youe R, Rajani M, et al., eds.). Packt Publishing Ltd; 2015.

19. Bojanowski P, Grave E, Joulin A, Mikolov T. Enriching Word Vectors with Subword Information. *Trans Assoc Comput Linguist*. 2016;5:135-146.

20. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. In: *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*. International Conference on Learning Representations, ICLR; 2013.

21. Müller AC, Guido S. *Introduction to with Python Learning Machine*. (Schanafelt D, ed.). O'Reilly Media, Inc.; 2017.

22. Beel J, Gipp B, Langer S, Breitinger C. Research-paper recommender systems: a literature survey. *Int J Digit Libr*. 2016;17(4):305-338. doi:10.1007/s00799-015-0156-0

23. Buda M, Maki A, Mazurowski MA. A systematic study of the class imbalance problem in convolutional neural networks . *Neural Networks*. 2018;106:249-259.

24. Bergstra J, Komer B, Eliasmith C, Yamins D, Cox DD. Hyperopt: a Python library for model selection and hyperparameter optimization. *Comput Sci Discov*. 2015;8(1):14008. doi:10.1088/1749-4699/8/1/014008

25. Bergstra J, Bardenet R, Bengio Y, Kégl B. Algorithms for Hyper-Parameter Optimization. *NIPS*. 2011;24:2546-2554.

26. Bergstra J, Yamins D, Cox DD. Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. In: *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*. ICML'13. JMLR.org; 2013:I–115–I–123.

27. Srivastava N, Hinton G, Krizhevsky A, Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J Mach Learn Res*. 2014;15(56):1929-1958.

28. Dietterich TG. Ensemble Methods in Machine Learning. In: *Lecture Notes in Computer Science*. Vol 1857. Springer; 2000:1-15. doi:10.1007/3-540-45014-9_1

29. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77.

30. Kessler JS. Scattertext: a Browser-Based Tool for Visualizing how Corpora Differ. *CoRR*. 2017;abs/1703.0.

31. Rehurek R, Sojka P. Software Framework for Topic Modelling with Large Corpora. In: *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. ELRA; 2010:45-50.

32. Abadi M, Agarwal A, Barham P, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. Published online 2016.

33. Oliphant TE. *Guide to NumPy*. 2nd ed. CreateSpace Independent Publishing Platform; 2015.

34. Hunter JD. Matplotlib: A 2D graphics environment. *Comput Sci Eng*. 2007;9(3):90-95. doi:10.1109/MCSE.2007.55

35. Buitinck L, Louppe G, Blondel M, et al. API design for machine learning software: experiences from the scikit-learn project. In: *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*. ; 2013:108-122.

36. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in {P}ython. *J Mach Learn Res*. 2011;12:2825-2830.

37. McKinney W. Data Structures for Statistical Computing in Python. In: *Proceedings of the 9th Python in Science Conference*. ; 2010:51-56.

38. The pandas development team. pandas-dev/pandas: Pandas. Published online February 2020. doi:10.5281/zenodo.3509134

39. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Published online 2019.

40. Chollet F. Keras. Published 2015. https://github.com/fchollet/keras

41. Campbell MK, Piaggio G, Elbourne DR, Altman DG, CONSORT Group. Consort 2010 statement: extension to cluster randomised trials. *BMJ*. 2012;345:e5661.

42. Al-Jaishi AA, Carroll K, Goldstein CE, et al. Reporting of key methodological and ethical aspects of cluster trials in hemodialysis require improvement: A systematic review. *Trials*. 2020;21(1). doi:10.1186/s13063-020-04657-9

Tables and Figures

Table 2-1: Hyperparameter search space for support vector machines and convolutional neural networks.

Hyperparameter	Values checked	Chosen value
For all models		
Sampling ratio (Non-CRT:CRT)	(1411:589), (2411:589), (3411:589), (4411:589)	3411 : 589
Class weights (Non-CRT:CRT)	(1:1), (1:5), (0.59:3.4), (1:17), (1:20)	0.59 : 3.4
Convolutional neural network – Word2Vec		
Max length of each abstract	100, 150, 200, 250, 300, 350	300
Batch size (distribution)	Uniform distribution (10, 30)	11
Learning rate (distribution)	Uniform distribution (0.0005, 0.005)	0.0047
Dropout rate (distribution)	Uniform distribution (0.1, 0.5)	0.29
Number of filters (distribution)	Uniform distribution (64, 1526)	923
Kernel size (distribution)	Uniform distribution (3, 12)	8
Number of epochs (distribution)	Uniform distribution (3, 20)	7
Constraint applied to the kernel matrix (distribution)	1, 1.5, 2, 2.5, 3	2
Optimizer (distribution)	Adadelta, Adam	Adam
Embedding	Skip-gram; CBOW	Skip-gram
Embedding dimensions	50, 100, 200, 300	100
Number of embedding iterations	5, 10, 15, 20	10
Convolutional neural network – FastText		
Max length of each abstract	100, 150, 200, 250, 300, 350	300
Batch size (distribution)	Uniform distribution (10, 30)	16
Learning rate (distribution)	Uniform distribution (0.0005, 0.005)	0.0026
Dropout rate (distribution)	Uniform distribution (0.1, 0.5)	0.47
Number of filters (distribution)	Uniform distribution (64, 1526)	532
Kernel size (distribution)	Uniform distribution (3, 12)	11

Number of epochs (distribution)	Uniform distribution (3, 20)	14
Constraint applied to the kernel matrix (distribution)	1, 1.5, 2, 2.5, 3	2
Optimizer (distribution)	Adadelta, Adam	Adam
Embedding	Skip-gram; CBOW	Skip-gram
Embedding dimensions	50, 100, 200, 300	100
Number of embedding iterations	5, 10, 15, 20	10
Support vector machines		
Kernel	linear, polynomial, sigmoid, or radial basis function	radial basis function
Kernel coefficient	1, 0.1, 0.01, 0.001, 0.0001	0.001
Regularization parameter	1, 10, 100, 1000	100
Ngrams	1, 1 to 2, 1 to 3, 1 to 4	1-gram and bi-gram (1 to 2)
Word Vectorization	Bag of Words, TF-IDF	TF-IDF

CRT: Cluster randomized trial; Ngrams: a sequence of n words from a text document; TF-IDF: Term Frequency-Inverse Document Frequency; CBOW: Continuous Bag of Words Model

Table 2-2: Model metrics for the internal and external validation datasets.

Dataset	**AUC, % (95% CI)**	**True Positive Rate sensitivity, % (95% CI)**	**False Positive Rate 1-specificity, % (95% CI)**	**Number needed to screen (95% CI)**
Internal validation This dataset had 600 articles, with ~15% being CRTs Number needed to read: 6.8 **				
Convolutional neural network – Word2Vec	98.2 (96.9, 99.5)	96.6 (92.0, 100)	13.9 (10.7, 17.0)	1.8 (1.6, 2.1)
Convolutional neural network – FastText	98.4 (97.3, 99.5)	89.8 (83.0, 96.6)	3.5 (2.0, 5.1)	1.2 (1.1, 1.3)
Support vector machines	97.2 (95.7, 98.8)	97.7 (94.3, 100)	19.9 (16.4, 23.2)	2.2 (1.9, 2.6)
Ensemble	98.6 (97.8, 99.4)	97.7 (94.3, 100)	15.0 (11.9, 18.2)	1.9 (1.7, 2.2)
External validation This dataset had 1916 articles, with ~35% being CRTs Number needed to read: 2.9 **				
Convolutional neural network – Word2Vec	97.9 (97.2, 98.6)	97.0 (95.6, 98.2)	20.8 (18.5, 23.0)	1.4 (1.3, 1.5)
Convolutional neural network – FastText	97.7 (97.0, 98.4)	91.7 (89.8, 93.8)	4.8 (3.7, 6.0)	1.1 (1.1, 1.1)
Support vector machines	96.8 (96.0, 97.6)	97.3 (96.1, 98.5)	32.2 (29.7, 34.9)	1.6 (1.6, 1.7)
Ensemble	97.8 (97.0, 98.5)	97.6 (96.4, 98.6)	21.8 (19.6, 24.1)	1.4 (1.4, 1.5)

** The number needed to read was calculated as one divided by the % of articles that are CRTs. AUC curve = Area under the receiver operating characteristic curve. CI = confidence interval.

Figure 2-1: Scatter text visualization of words and phrases used in our dataset. Points are colored blue or red based on related terms with cluster-randomized trials (CRT) or non-CRT citations. The dataset consisted of 589 CRT (111,492 words) and 4411 non-CRTcitations (816,167 words). The terms associated with each category are under "Top CRT" and "Top Non-CRT" headings. Interactive version of the figure: **(Interactive Figure 1)** (*Note: The file size for the interactive figure is large and can take several minutes to load in a browser*)

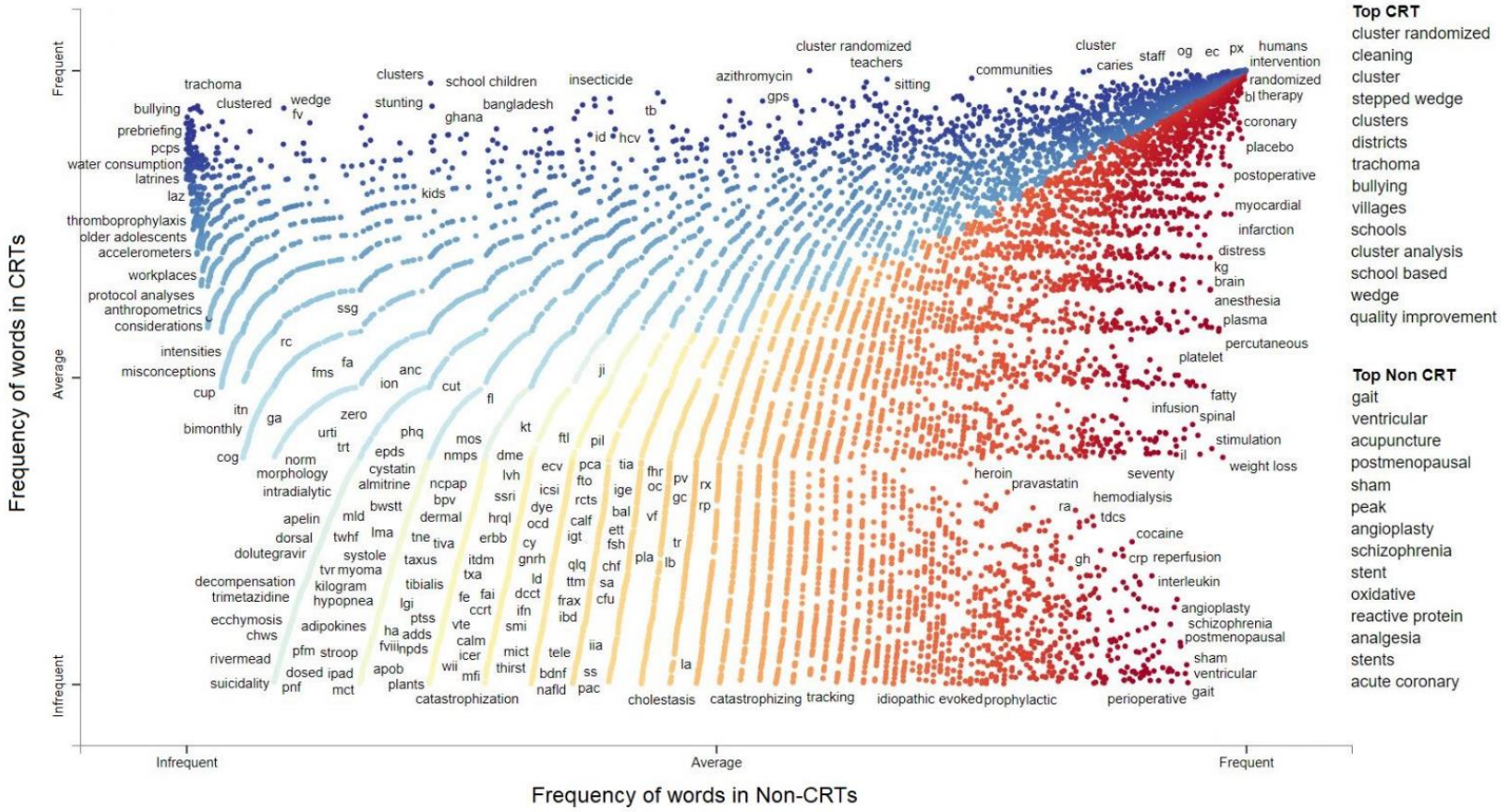
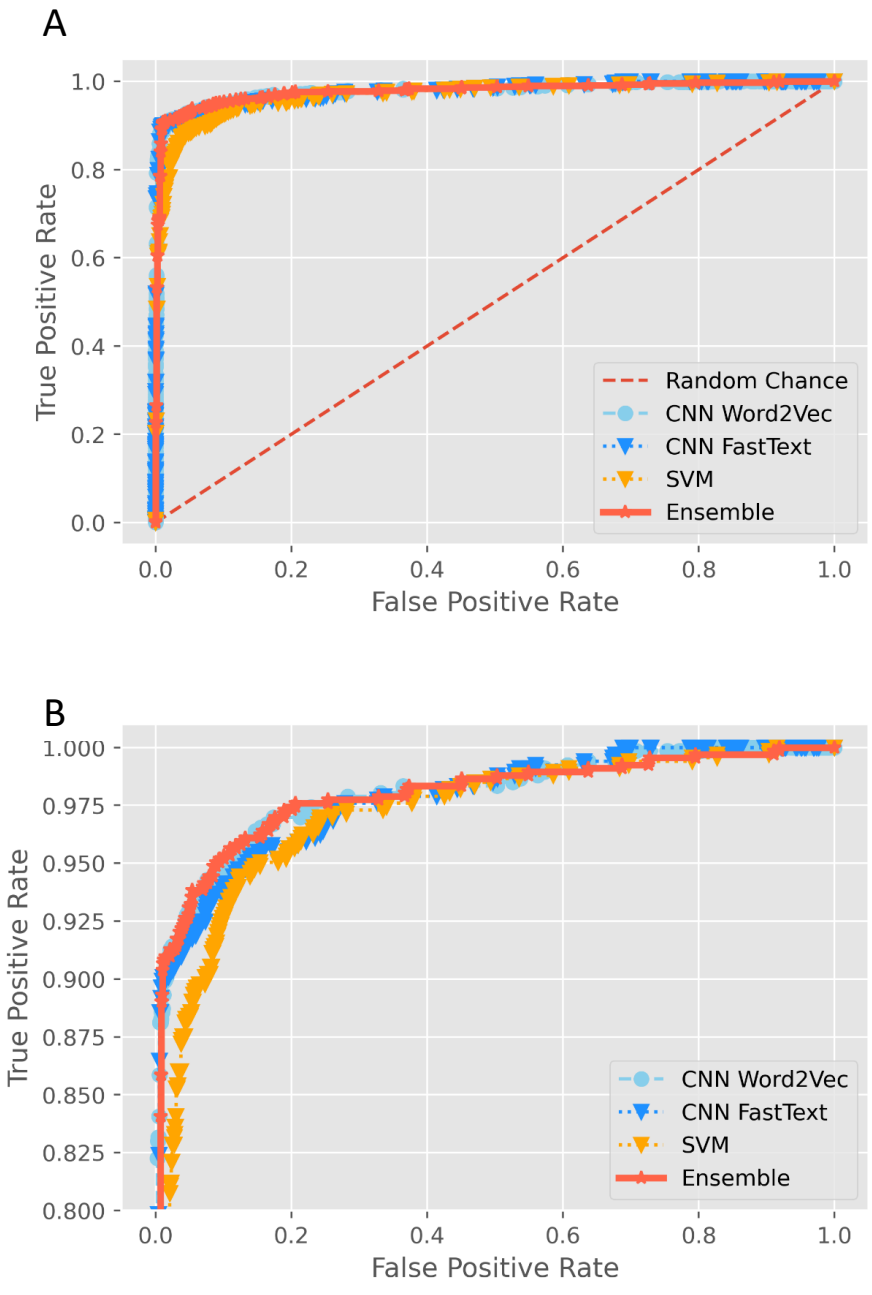


Figure 2-2: Receiver operating characteristic curves of (1) convolutional neural network using Word2Vec word embedding, (2) convolutional neural network using FastText word embedding, (3) support vector machine (SVM), and (4) ensemble model. Figure **A** depicts a zoomed-out version, and Figure **B** zoomed to accentuate variability in the models' receiver operating characteristic curves.



Appendix

Appendix 2-1: Justification for using a CRT search filter.

We created a dataset of CRT and non-CRT articles to train and internally validate our machine learning algorithms. Given that the prevalence of CRTs is less than 0.1% in bibliographic databases, we would have needed to screen at least 500,000 articles to identify 500 CRT articles, an inefficient and cumbersome process. Therefore, we used a CRT search filter to increase the prevalence of relevant articles retrieved in our bibliographic search, where we expected about 10% to 15% of articles would be CRTs. [1]

Search syntax to identify articles in Medline between Jan 1st, 2000 and Dec 31st, 2019 in EMBASE Classic+ EMBASE, Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R).

Search No.	Search	Records
1	randomized controlled trial.pt.	525702
2	animals/	8037797
3	humans/	32495688
4	2 not (2 and 3)	5739131
5	1 not 4	514156
6	(cluster$ adj2 randomi$).tw.	30283
7	((communit$ adj2 intervention$) or (communit$ adj2 randomi$)).tw.	20137
8	group$ randomi$.tw.	8732
9	6 or 7 or 8	57622
10	intervention?.tw.	2425049
11	cluster analysis/	122853
12	health promotion/	176542
13	program evaluation/	80902
14	health education/	159941
15	10 or 11 or 12 or 13 or 14	2837625
16	9 or 15	2848506
17	16 and 5	107591
18	limit 17 to yr="2000 - 2019"	87633

Appendix 2-2: Additional details for the external dataset

We evaluated our algorithms' performance against an external dataset that included 1988 articles. These articles were confirmed primary reports of RCTs, of which 688 were CRT reports and the rest were individually randomized trials. This dataset has been described elsewhere. [2] Briefly, the authors identified pragmatic clinical trials using a sensitivity-maximizing pragmatic search filter; this search filter is independent of this study's CRT search filter. [3] The search filter for the external dataset was applied in MEDLINE on April 3rd, 2019, for the period between January 1st, 2014, and April 3rd, 2019. The 2014 date was chosen because this was the first date the National Library of Medicine began indexing pragmatic clinical trials as a publication type. The authors identified 4337 pragmatic articles from the search strategy, of which we used 1988 trials that were registered in ClinicalTrials.gov. From these 1988 articles, we removed 72 articles that were captured in the training or validation datasets. We applied this exclusion criterion to avoid data leakage that would artificially inflate the models' performance.

Appendix 2-3: Description of the gamma and c parameters

The gamma parameter is a cut-off parameter for the Gaussian sphere; increasing gamma increases the training samples' reach and a softer decision boundary (**eFigure 2-2**).[4] The c parameter controls the penalty for misclassification. Large values of c correspond to a more substantial error penalty imposed on the model for misclassifying a record (**eFigure 2-2**).[4]

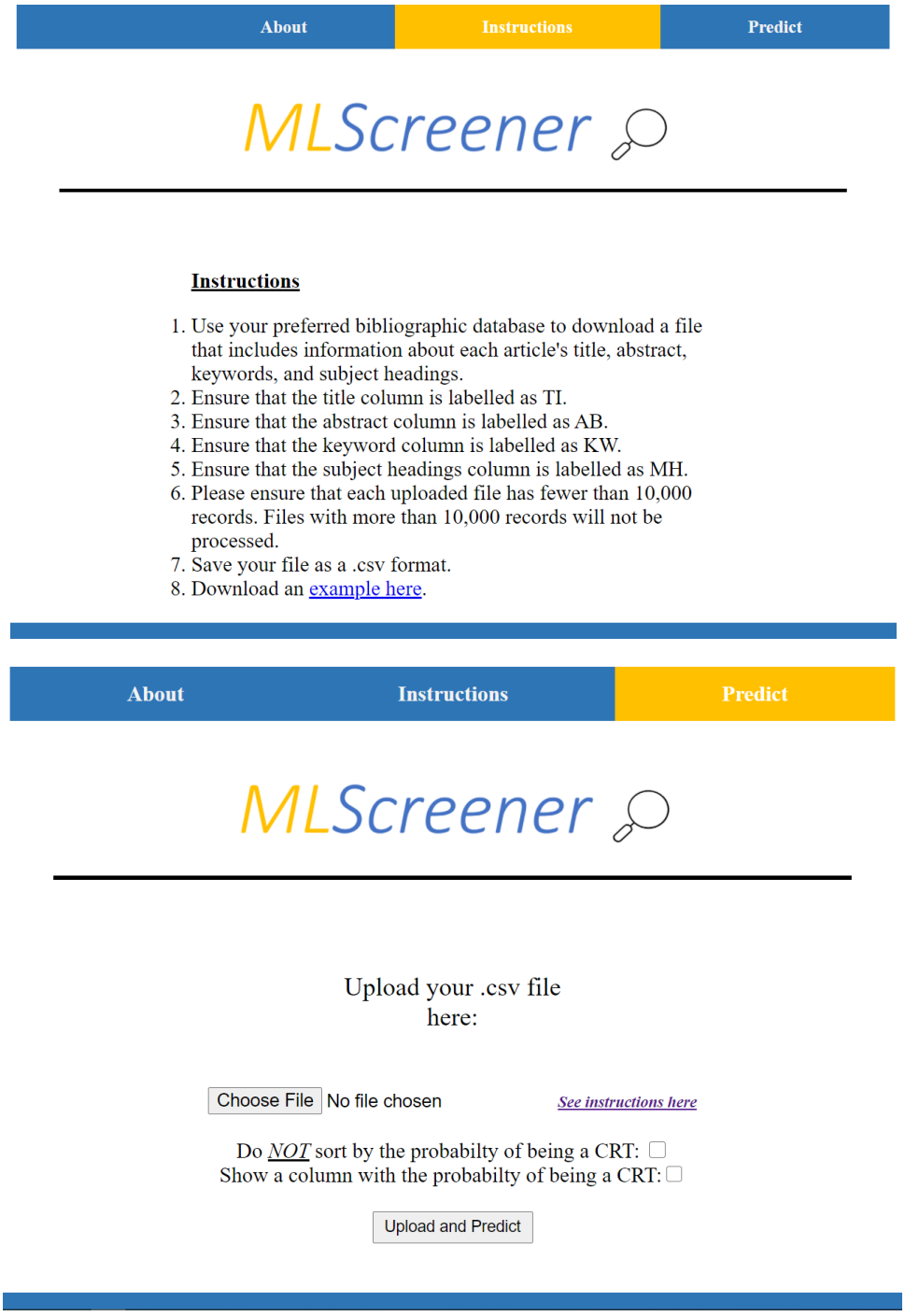
Appendix 2-4: Continuous skip-gram architecture

In the continuous skip-gram architecture, the model uses the current word to predict the surrounding window of context words; we set the window to 5 words in our setting. The skip-gram architecture weighs nearby context words more heavily than more distant context words.

Appendix 2-5: Python libraries used for this project.

1. *pandas* (version 0.25.1) is a package that is intended for working with relational data tables.
2. *NumPy* (version 1.18.1) is a library for scientific computing, which can contain n-dimensional array objects.
3. *sklearn* (version 0.22.1) is a machine learning library for Python. It features various machine learning algorithms like the support vector machine used in this project.
4. *re* is a regular expression library.
5. *string* is a library with a collection of string constants.
6. *nltk* (version 3.4.5) is a Natural Language Toolkit (NLTK) library used for natural language processing.
7. *pickle* library helps create a portable serialized representation of Python objects (e.g., a data frame).
8. *gensim* (version 3.4.0) is a library that is often used for natural language processing and information retrieval.
9. *keras* (version 2.2.4-tf) is a deep learning framework for developing and evaluating deep learning models.
10. *math* is a module that provides access to mathematical functions.
11. *tensorflow* (version 2.0.0) is a library used for fast numerical computing.
12. *Scattertext* (0.0.2.28) is a package that lets you interactively visualize how two categories of text are different from each other.
13. *matplotlib* (version 3.2.1) is an object-oriented plotting library.
14. *hyperopt* (version 0.2.3) is a Bayesian optimization library that allows for the automatic search of data preparation methods, machine learning algorithms, and model hyperparameters for classification and regression tasks.

Appendix 2-6: Screenshots of the front-end tool used for our machine learning algorithm.



Appendix 2-7: Details and results for applying our model on a dialysis-related dataset.

This dataset included 882 articles from a search strategy that combined two search filters to identify CRT reports and dialysis studies. [1,5] Two reviewers screened titles and abstracts to identify CRT reports in the hemodialysis setting. We excluded CRT reports unrelated to in-center hemodialysis. We applied our algorithms to this dataset to evaluate whether it reduced the number of articles needed to screen without excluding hemodialysis-related CRT reports.

The **eTable** below showed the results when our model was applied in practice to classify CRT reports in the dialysis setting. The models correctly classified 33 of the 34 dialysis-related CRT reports. The number of articles needed to screen to capture a single CRT was reduced from 26 to 11.4.

eTable: Number of relevant articles retrieved with and without machine learning algorithm using the demonstration dataset (n=882 records). The research objective was to review CRTs in the hemodialysis setting to report key methodological and ethical issues.

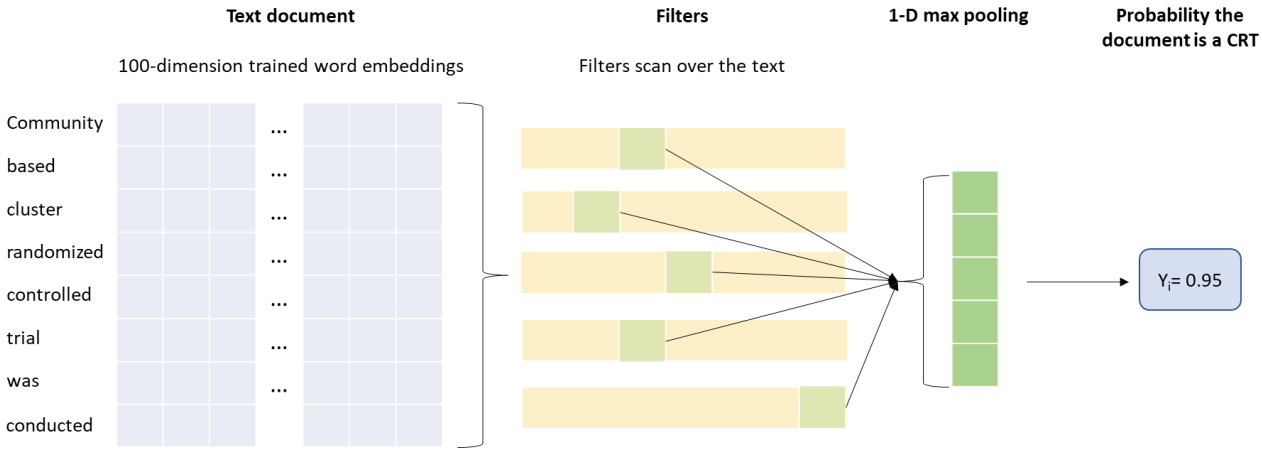
	Number of relevant CRT articles identified	Number of non-relevant articles captured
Manual screening	34*	848 [++]
Ensemble	33**	343 [++]

* There was a total of 36 cluster randomized trials (CRTs) conducted in the hemodialysis setting. Two of the 36 CRT articles were identified in the included articles' reference list and are not included above. The title and abstract of these two had no indication they utilized a CRT study design and these articles were not picked up by the CRT search filter. [6,7]

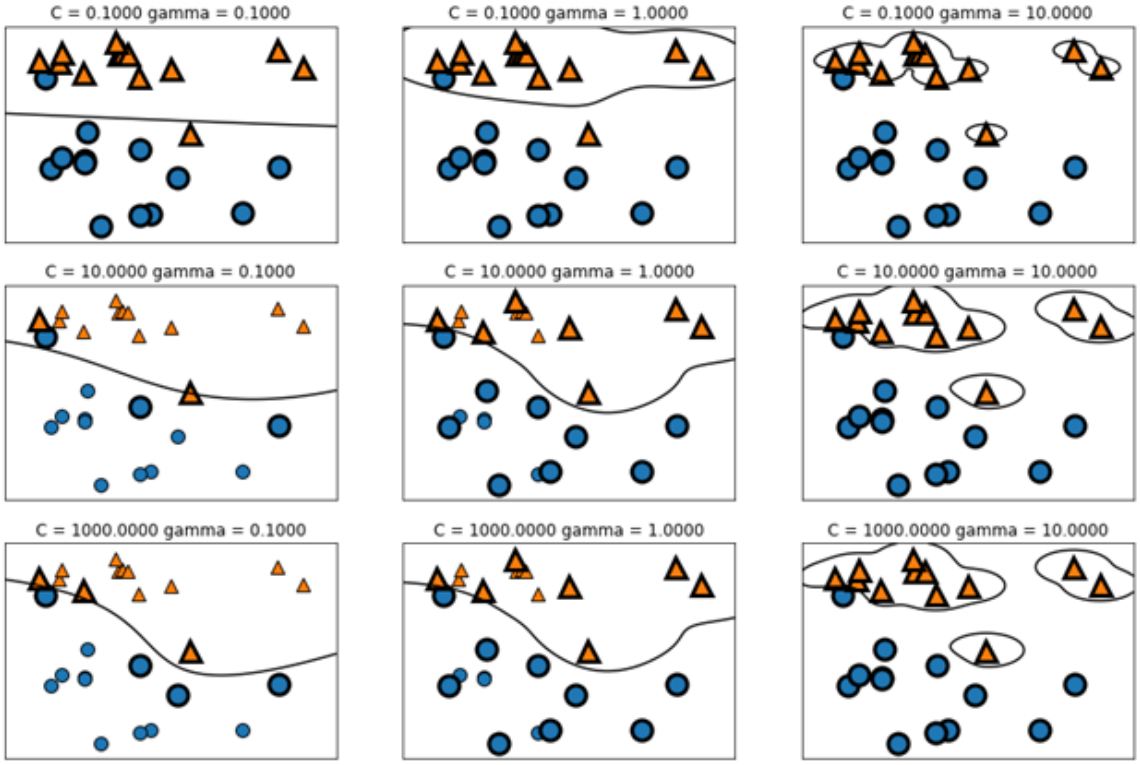
** The missed article stated it was a "group randomized trial" but later stated, "Patients at participating dialysis centers were randomized… [8]

++ A proportion of these articles were CRTs unrelated to the hemodialysis setting and were not relevant to the respective review's research objective.

eFigure 2-1: A general architecture for a convolutional neural work used for text classification.

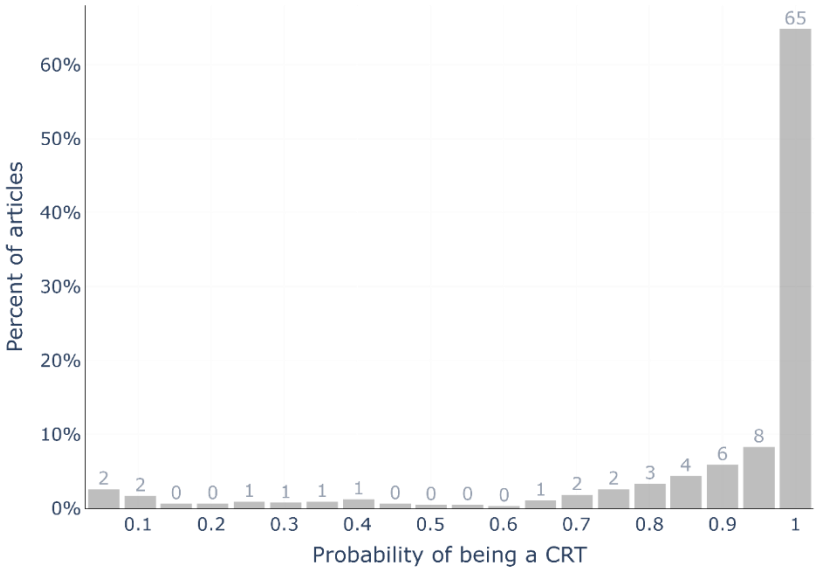


eFigure 2-2: This figure shows the decision boundaries and support vectors for different settings of the parameters regularization parameter (C) and Kernel coefficient (gamma). We used the mglearn package to create this figure. [9]

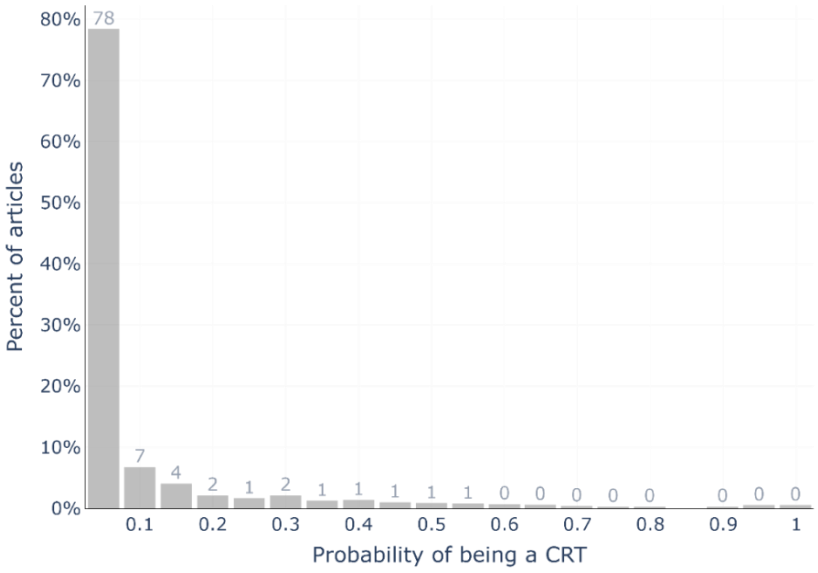


eFigure 2-3: Probability plot for the CRTs in the first external data classified as a CRT (Figure A, 665 CRTs) and non-CRTs classified as a CRT (Figure B, 1251 non-CRTs). The x-axis depicts the stacked ensemble model's prediction of the article being classified as a CRT. The y-axis depicts the proportion of all documents that had the corresponding probability.

A



B



References for chapter 2 appendix

1. Taljaard M, McGowan J, Grimshaw JM, et al. Electronic search strategies to identify reports of cluster randomized trials in MEDLINE: low precision will improve with adherence to reporting standards. *BMC Med Res Methodol*. 2010;10(1):15. doi:10.1186/1471-2288-10-15

2. Nicholls SG, Carroll K, Hey SP, et al. A review of pragmatic trials found a high degree of diversity in design and scope, deficiencies in reporting and trial registry data, and poor indexing. *J Clin Epidemiol*. 2020;Submitted.

3. Taljaard M, McDonald S, Nicholls SG, et al. A search filter to identify pragmatic trials in MEDLINE was highly specific but lacked sensitivity. *J Clin Epidemiol*. 2020;124:75-84. doi:10.1016/j.jclinepi.2020.05.003

4. Raschka S. *Python Machine Learning*. (Hussain A, Youe R, Rajani M, et al., eds.). Packt Publishing Ltd; 2015.

5. Iansavichus A V, Haynes RB, Lee CWC, et al. Dialysis search filters for PubMed, Ovid MEDLINE, and Embase databases. *Clin J Am Soc Nephrol*. 2012;7(10):1624-3161. doi:10.2215/CJN.02360312

6. Kauric-Klein Z. Improving blood pressure control in end stage renal disease through a supportive educative nursing intervention. *Nephrol Nurs J*. 2012;39(3):217-228.

7. Howren MB, Kellerman QD, Hillis SL, Cvengros J, Lawton W, Christensen AJ.

Effect of a Behavioral Self-Regulation Intervention on Patient Adherence to Fluid-Intake Restrictions in Hemodialysis: a Randomized Controlled Trial. *Ann Behav Med*. 2016;50(2):167-176. doi:10.1007/s12160-015-9741-0

8. Waterman AD, Peipert JD. An Explore Transplant Group Randomized Controlled Education Trial to Increase Dialysis Patients' Decision-Making and Pursuit of Transplantation. *Prog Transplant*. 2018;28(2):174-183. doi:10.1177/1526924818765815

9. Müller AC, Guido S. *Introduction to with Python Learning Machine*. (Schanafelt D, ed.). O'Reilly Media, Inc.; 2017.

CHAPTER 3: STUDY 2

TITLE: Reporting of key methodological and ethical aspects of cluster trials in hemodialysis require improvement: a systematic review

AUTHORS: Ahmed A. Al-Jaishi, MSc [1,2,3]; Kelly Carroll, MA [4]; Cory E. Goldstein, MA [5]; Stephanie N. Dixon, Ph.D. [1,3,6,7]; Amit X. Garg, MD, Ph.D. [1,2,3,6]; Stuart G. Nicholls, Ph.D. [4]; Jeremy M. Grimshaw, MBChB, Ph.D. [4,9]; Charles Weijer, MD, Ph.D. [5,6]; Jamie Brehaut, Ph.D. [4,8]; Lehana Thabane, Ph.D. [2]; PJ Devereaux, MD, Ph.D. [2]; Monica Taljaard, Ph.D. [4,8]

AFFILIATIONS:

[9] Lawson Health Research Institute, London, Ontario, Canada

[10] McMaster University, Department of Health Research Methods, Evidence, and Impact, Hamilton, ON, Canada

[11] ICES, London, Ontario, Canada

[12] Ottawa Hospital Research Institute, Clinical Epidemiology Program, Ottawa, ON, Canada

[13] Western University, Rotman Institute of Philosophy, London, Ontario, Canada

[14] Department Medicine, Epidemiology and Biostatistics, Western University, London, ON, Canada

[15] Department of Mathematics and Statistics, University of Guelph, Guelph, ON, Canada

[16] University of Ottawa, School of Epidemiology Public Health, Ottawa, ON, Canada

[17] University of Ottawa, Department of Medicine, Ottawa, ON, Canada

KEYWORDS: Cluster Randomized Controlled Trial; Systematic Review; Ethics; Informed Consent; Hemodialysis

CITATION: Al-Jaishi AA, Carroll K, Goldstein CE, et al. Reporting of key methodological and ethical aspects of cluster trials in hemodialysis require improvement: A systematic review. *Trials*. 2020;21(1). doi:10.1186/s13063-020-04657-9

Abstract

Background: The hemodialysis setting is suitable for trials that use cluster randomization, where intact groups of individuals are randomized. However, cluster randomized trials (CRTs) are complicated in their design, analysis, and reporting and pose ethical challenges. We reviewed CRTs in the hemodialysis setting for reporting of key methodological and ethical issues.

Methods: We conducted a systematic review of CRTs in the hemodialysis setting, published in English between 2000 and 2019 and indexed in MEDLINE or EMBASE. Two reviewers extracted data and study results were summarized using descriptive statistics.

Results: We identified 26 completed CRTs and five study protocols of CRTs. These studies randomized hemodialysis centers (n=17, 55%), hemodialysis shifts (n=12, 39%), healthcare providers (n=1, 3%), and nephrology units (n=1, 3%). Trials included a median of 28 clusters with a median cluster size of 20 patients. Justification for using a clustered design was provided by 15 trials (48%). Methods that accounted for clustering were used during sample size calculation in 14 (45%), during analyses in 22 (71%), and during both sample size calculation and analyses in 13 trials (42%). Among all CRTs, 26 (84%) reported receiving research ethics committee approval; patient consent was reported in 22 trials: 10 (32%) reported the method of consent for trial participation and 12 (39%) reported no details about how consent was obtained or its purpose. Four trials (13%) reported receiving waivers of consent and the remaining 5 (16%) provided no or unclear information about the consent process.

Conclusion: There is an opportunity to improve the conduct and reporting of essential methodological and ethical issues in future CRTs in hemodialysis.

Registration: We conducted this systematic review using a pre-specified protocol that was not registered.

Introduction

Patients on hemodialysis (HD) are often excluded from clinical trials and many trials in the hemodialysis setting suffer from poor recruitment, inadequate sample sizes, and poor adherence to allocated treatment and treatment contamination (1–4). Cluster randomized trials (CRTs) randomize intact groups of individuals (rather than independent individuals) to different arms. This design can offer a logistically convenient method to produce high-quality evidence, effectively avoid treatment contamination, and be better received by participants and healthcare staff when delivered to a group of individuals rather than select patients. The CRT is an attractive design in the HD setting, where interventions are often delivered at the center level and staff follows the same protocol for patients under their care.

Cluster randomization, however, introduces methodological issues that need to be addressed during the design and analysis stages (5,6). First, it may not be possible to identify and recruit participants until the cluster has been randomized. This increases the risk of selection bias because knowledge of the allocated arm can influence both the identification of potential participants and their decisions to participate. Second, because outcomes are usually correlated within clusters, CRTs are statistically less efficient than individual-level randomized trials. As such, the CONSORT Statement for Cluster Randomized Trials requires that studies report how clustering was considered in both sample size calculation and analysis. Failing to account for clustering in the sample size calculation implies that the study may not have adequate power to detect meaningful differences between the groups. While failing to account for clustering in the analysis

implies that standard errors of treatment effects will be under-estimated, increasing the risk of spurious statistical significance (5–8).

The CRT design also raises complex ethical issues. *The Ottawa Statement on the Ethical Design and Conduct of Cluster Randomized Trials* offers ethical guidance, providing 15 recommendations for those who design, conduct, and review CRTs (9–13). For example, ethical issues that may challenge researchers include: When is a study considered research? Who is the research subject? And from whom, how, and when must informed consent be obtained? A summary of the Ottawa Statement recommendations and applicability of these recommendations for CRTs conducted in the hemodialysis setting is provided in **Appendix 3-1**.

While CRTs offer a promising approach to conducting clinical trials within the hemodialysis setting, this design may have unique methodological and ethical requirements for conduct and reporting in this setting. In the present study, we conducted a descriptive analysis of how CRTs in hemodialysis report key methodological (accounting for clustering effects and reporting of the intra-class correlation coefficient) and ethical issues (regarding the elements highlighted in the Ottawa Statement). This review will serve as a foundational step in a multi-year initiative that seeks to develop recommendations for the ethical design, conduct, and reporting of CRTs in the hemodialysis setting.

Materials and Methods

Protocol and registration

We conducted this systematic review using a pre-specified protocol and reported our results according to published guidelines (PRISMA Checklist: **Appendix 3-2**) (14).

Studies eligible for review

We did not set any limits on the country of study and included published primary reports of CRTs or study protocols of CRTs with an unpublished primary report. We included English-language reports published between January 2000 and November 2019 that involved: **1)** patients on in-center hemodialysis; or **2)** patients on in-center hemodialysis as a subgroup in a larger study of non-in-center hemodialysis patients. When we found a CRT study protocol with an identified completed trial, we used the protocol to supplement any missing information from the final published report. Other reports such as secondary analyses, conference abstracts, and pilot or feasibility CRTs were excluded. We excluded feasibility and pilot trials because they have different methodological (15) and ethical considerations than full-scale CRTs.

Information sources

We implemented a search syntax on November 30th, 2019, to identify published reports in MEDLINE and EMBASE.

Search

Our search strategy combined two published search filters designed to identify publications related to CRT (16) and dialysis (17) studies (**Appendix 3-3**). Two reviewers (AAA and KC) screened titles and abstracts of articles. AAA manually searched for additional articles in bibliographies of all included articles, lists of articles that cited the included studies in Google Scholar, and "Similar articles" in PubMed. The

complete list of included studies was also reviewed by an expert in the field (AXG) to capture additional studies that may have been missed.

Study selection

We retrieved the full text of any article considered potentially relevant by any reviewer. Full-text articles were assessed for study eligibility by two reviewers (AAA and KC), with disagreements resolved through discussion. Agreement between the two reviewers was evaluated using the Kappa statistic (18).

Data collection process

We utilized a data abstraction form that was pilot tested on three studies by three reviewers (AAA, KC, and CEG). After that, two reviewers (AAA and either KC or CEG) independently extracted data from each manuscript. AAA and either MT or SND extracted details on whether trials accounted for clustering during sample size estimation and analysis. After each set of three studies, data extractions were compared within the pair and disagreements were resolved by consensus. Details of extracted data are highlighted in **Appendix 3-4.** We extracted data on study characteristics, methodological characteristics, data collection method, the justification for using a CRT design, type of intervention, information regarding research ethics committee (REC) review, gatekeepers (i.e., an individual or body that represents the interests of cluster members, clusters, or organizations (19)), informed consent procedures, and any information about harm-benefit assessment or protection of vulnerable populations.

Analysis

We summarised results using frequencies for categorical variables and medians with interquartile ranges for continuous variables. Given the small number of included studies,

we did not test changes in reporting over time or the association between reporting ethical elements and study characteristics. For all our analyses, we used R (Version 3.6.2) (20).

Results

Characteristics of included studies

The study flow diagram is presented in **Figure 3-1**. We screened 777 citations and retrieved 29 full-text articles to assess eligibility. We identified another seven articles by reviewing citation links (n total=36). We had an almost perfect between-reviewer agreement on which studies met the criteria for review (kappa statistic 0.96, 95% confidence interval: 0.91 to 1.00). Five articles were excluded after the full-text review (21–25). Thus, 31 articles were included in this review: 26 completed studies and 5 study protocols (26–56).

Study characteristics for the included trials are presented in **Table 3-1**. The 31 trials were published in 19 journals. Nineteen trials (61%) recruited patients from the United States, three (10%) were from the United Kingdom, three (10%) from Australia / New Zealand, and seven (23%) from other countries (some trials were multi-national and these categories are not mutually exclusive).

Reporting of methodological characteristics

Table 3-2 provides a description of the reporting of study characteristics. Thirty trials (97%) utilized a parallel arm design and one trial (3%) used a stepped-wedge design. All trials were designed as superiority trials. The types of randomized clusters were hemodialysis centers (n=17; 55%), hemodialysis shifts or sessions (n=12; 39%), providers or professionals (n=1; 3%), and nephrology units (n=1; 3%; it was not clear how a “nephrology unit” was defined). Clusters were randomly allocated to the treatment

arm using unrestricted (n=8; 26%), pair-matched (n=4; 13%), stratified (n=4; 13%), split-cluster (n=11, 35% [i.e., day shifts within centers]), covariate-constrained randomization (n=1, 3%), or an unreported method of allocation (n=3, 10%).

The median (25^{th}, 75^{th} percentile) number of clusters included per trial was 28 (12, 43), and all trials used 1:1 randomization. One trial (3%) had one cluster per arm, and six trials (19%) had fewer than the minimum recommendation of four clusters per arm (6,7). The median number of participants per trial was 228 (120, 1723). All trials included patients (as opposed to providers alone) as the research participants with a median number of 20 (8, 32) patients per cluster.

One study (3%) reported the intra-class correlation coefficient (ICC) for their primary outcome. **Table 3-3** describes whether and how clustering was accounted for during sample size estimation and analysis. Fourteen trials (45%) accounted for clustering during sample size estimation for the primary outcome, three (10%) did not account for clustering, two (6%) accounted for clustering but using a different outcome measure than the primary outcome, one (3%) it was unclear, and 11 (35%) did not report a sample size or power estimate. At the analysis stage, 22 trials (71%) accounted for clustering using either an individual-level analysis adjusting for clustering or using a cluster-level summary method. The remaining nine trials (29%) either did not account for clustering in their primary analysis, or it was unclear if clustering was accounted for in the analysis. A total of 13 trials (42%) accounted for clustering in sample size calculation and analysis.

Reporting of justification for cluster randomization

Of all 31 trials, 15 trials (48%) reported a justification for using a cluster randomized design (**Table 3-3**). Thirteen trials (42%) reported using a CRT design to avoid contamination, and two trials (6%) reported using a CRT design to avoid contamination and for logistical or administrative convenience.

Reporting of intervention type and target population

Table 3-4 lists the types of intervention used in each arm of included trials. The most common type of study intervention was health promotion or an educational intervention (n=22 trials; 71%), for which patients were the intended recipients. Six trials (19%) examined a direct patient therapeutic intervention – for example, intradialytic resistance training or antimicrobial barrier caps for central venous catheters. Among all trials, the intervention was necessarily administered at the cluster level (e.g., education of providers) for 18 trials (58%). In the control arm, 23 trials (74%) utilized "usual care," four (13%) used some form of augmented care (usual care plus some minimal elements of active intervention), three (10%) used an active control, and one (3%) used an attention-placebo. Four trials (13%) utilized interventions that included an educational or quality improvement component targeting health professionals (e.g., transplant education and engagement activities). Both prevalent and incident patients on hemodialysis were included in 22 trials (71%), eight trials (26%) included only prevalent patients, and one trial (3%) included only incident patients on hemodialysis.

Data collection procedures

Data collection procedures in the intervention and control arm were similar for most trials (**Table 3-4**). In the intervention arm, 30 (97%) trials used local routinely collected data

(e.g., medical charts or electronic medical records) as the primary source for data collection. Eleven trials (35%) used clinical registry data and 24 (77%) supplemented routinely collected data with additional sources: self-administered questionnaires (n=18; 58%), interviewer-administered questionnaires (n=9; 29%), specimen collection or physical examination not required for usual patient care (n=4; 13%), as well as active data collection (n=5; 16%); for example, using case report forms.

Gatekeepers

Five trials (16%) reported that a gatekeeper provided permission for clusters to participate in the study (**Table 3-5**). For the remaining trials (84%), no information about gatekeepers was provided.

Reporting of research ethics review

We found that 26 trials (84%) reported REC approval, one (3%) reported that the study was exempt from REC review, and four (13%) did not report whether the study was reviewed by a REC (**Table 3-5**).

Reporting of consent procedures

One trial (3%) reported they received an exemption from ethics review, three (10%) received a waiver of consent from the REC (see **Appendix 3-5**), 22 trials (71%) reported obtaining consent from patients, and five (16%) trials either did not discuss the consent process or it was unclear if patients provided informed consent. For the 22 trials (71%) that reported obtaining consent from patients, written or verbal informed consent was reported in 10 trials (45%) for the study intervention and eight (36%) for data collection; the remaining trials provided no details about the method of consent for study intervention or data collection (**Table 3-4**).

Among the remaining 27 trials that did not receive an exemption from ethics review or had a waiver of consent, the timing of consent took place before randomization for seven trials (26%), after randomization for ten (37%), and was unclear for the remaining ten trials (37%).

The ability for participants to opt-out of the data collection was reported in seven trials (23%); three trials (10%) reported patients could not opt-out of data collection, and the ability to opt-out was unclear for the remaining 21 trials (68%) (**Table 3-5**).

Assessment of benefit-harm and protections for vulnerable groups

Kidney disease disproportionally affects individuals traditionally considered vulnerable (i.e., patients with comorbidities, dementia, lower education levels, lower health literacy, and who reside in rural or remote areas). Although vulnerable subgroups may have been included in our reviewed trials, none reported additional protections.

Discussion

The hemodialysis population is suitable for the CRT design, especially for interventions implemented at the center level; in our review, approximately 60% of trials utilized an intervention that was necessarily administered at the cluster level. This review presents a descriptive analysis of the reporting of key methodological and ethical characteristics of CRTs involving hemodialysis patients. Guidance on CRTs' reporting is provided in the CONSORT extension for CRTs, while the Ottawa Statement is currently the only guidance document specific to the ethical design and conduct of CRTs in health research (7,13). While several studies were published before the dissemination of the CONSORT,

the Ottawa Statement, or both, the interpretation of our results would not change had we presented our results based on the period pre- and post-publication of these statements.

We found that cluster randomized trials in hemodialysis have low methodological quality and sub-optimally report ethical considerations around this design. While many of the identified issues are not unique to the hemodialysis setting, we consider three issues that require special attention: (1) taking clustering into account at the sample size estimation and analysis stages; (2) methodological and contamination issues around designs that randomize shifts within hemodialysis centers; and (3) reporting on how the rights of vulnerable participants are protected.

First, patients on hemodialysis within the same center have similar characteristics compared to patients from other centers. For example, small satellite hemodialysis centers might have medically stable patients compared to large academic centers that might treat sicker patients requiring close medical monitoring. It is concerning that more than half of included trials did not report a method that appropriately accounts for within-cluster correlation when estimating sample size, and more than a quarter of trials did not account for clustering in the analysis, putting the study results at an increased risk of spurious statistical significance (6–8,13). Adjusting for clustering is especially important in this setting because there is generally high practice variation between hemodialysis centers and low variation within centers (57–59), which increases the ICC (60).

There is limited information in the literature to inform estimates of the ICC for patients' outcomes on hemodialysis; thus, researchers in hemodialysis must rely on estimates from

other disciplines or historical data. As such, completed trials need to report the observed ICC or design effect estimates for their outcomes so that the community can begin to build a repository that might help in the design of future trials. In our review, only one trial reported an ICC (45).

Second, a common experimental design was to randomize shifts within hemodialysis centers (e.g., Mon, Wed, Fri versus Tues, Thu, Sat). This type of randomization requires additional considerations in design and conduct. For example, the same healthcare staff will care for patients dialyzing in a single center in both trial arms. Contamination of the trial's two arms can still occur if staff observe better patient outcomes in one arm and then begin to implement the treatment in clusters (i.e., shifts) in the other arm. This design type also requires additional considerations in the analysis because clustering can occur at two levels, i.e., center and shift. It was beyond this review to assess whether authors reported the appropriate analyses accounting for this type of experimental design.

Third, authors should report how vulnerable participants' rights are protected, especially those with limited health literacy who may not be capable of providing informed consent. When including these subgroups, it raises ethical concerns about the extent to which these participants are truly informed. There are no clear standards for "how much" understanding is adequate (61). Additionally, lower education levels, lower health literacy, and a participant's primary language are all associated with poor comprehension of the informed consent process (62). These characteristics are significant in the hemodialysis setting, where vulnerable participants are overrepresented: patients often

have multiple comorbidities, are members of a socially marginalized group, live in rural or remote locations, or have cognitive impairments (63–66).

In general, trials in this setting were small, with both a limited number of clusters and patients within clusters. One trial randomized only one cluster to each arm and a fifth of reviewed trials had four or fewer clusters (41,54). Randomizing two clusters effectively precludes any inferences about the intervention because it is impossible to disentangle natural variation between clusters from the effect of the intervention (67). While some have suggested that parallel arm CRTs should have at least four clusters per study arm (6,7), with such a small number of clusters, the study may be severely underpowered, parametric statistical tests (e.g., t-tests) may not meet the assumption of normality, and there is a high risk of baseline imbalances between trial arms that might complicate the interpretation of the trial results (68).

There is room for improved reporting of consent procedures. When consent is required, study authors ought to report adequate details to assess what consent was for (e.g., enrollment, receiving the interventions, data collection), as well as from whom (e.g., patients, providers, etc.), when (before or after randomization), and how (e.g., written, verbal) consent was obtained (7,13). The timing of informed consent was either not reported or took place after randomization for 20 trials. Post-randomization consent, especially when the study is unblinded, is a key risk of bias that can introduce selection bias through differential recruitment (7). When applicable, researchers must justify how

their study meets the criteria for a waiver or alteration of informed consent as outlined by national regulations or international guidelines (69–71).

Our study has several limitations. We were unable to examine changes in quality of reporting over time or factors associated with better reporting due to the small number of trials. When a study protocol was published for one of the completed trials, we used both references to complete study extraction; however, we did not have access to the original research ethics submissions or non-peer-reviewed study protocols, did not follow-up with study authors, and did not conduct a search of any trial registries or Green Open Access options (e.g., ResearchGate). Thus, our results are based exclusively on what was reported in peer-reviewed published articles; for example, we are aware of other ongoing CRTs not included here because no study protocol or a primary report was available at the time of our search (72,73).

Our study also has several strengths. We utilized an abstraction tool developed and refined over several studies (74–77). It is unlikely that a substantial number of relevant primary trials were missed, as we combined two validated search strategies supplemented with an extensive manual search of reference resources (16,17). To reduce the risk of misclassification of trial characteristics and reporting practices, we used consensus between two reviewers who independently extracted information from published articles.

Conclusion

There is suboptimal conduct and reporting of methodological issues of CRTs in the hemodialysis setting and incomplete reporting of key ethical issues. The *Ottawa Statement on the Ethical Design and Conduct of Cluster Randomized Trials* provides

specific recommendations for CRTs but did not consider the hemodialysis setting's unique characteristics (13). This systematic review was conducted as a first step to describe key study design characteristics and document reporting of ethical practices in CRTs in the hemodialysis setting. Our future work builds on this review's information to explore the views/perceptions of researchers and patients regarding the ethical issues for CRTs in the hemodialysis setting.

Ethics approval and consent to participate

We did not apply for ethics approval as we conducted a systematic review of published literature.

Disclosures

CW receives consulting income from Eli Lilly & Company Canada. Other authors have nothing to disclose.

We declare that the results presented in this paper have not been published previously in whole or part.

Consent for Publication

Consent for publication was obtained from all authors.

Availability of data and material

Study data can be made available upon request.

Acknowledgments

We are grateful to Risa Shorr for creating and executing the search strategy for this review.

Ahmed Al-Jaishi was supported by the Allied Health Doctoral Fellowship from the Kidney Foundation of Canada, CIHR Doctoral Award, and McMaster University Michael DeGroote Scholarship. Cory E. Goldstein and Stephanie Dixon's research is supported

through a SPOR Innovative Clinical Trial Multi-Year Grant (Grant number: MYG-151209) from the Canadian Institutes of Health Research (CIHR). Amit Garg was supported by the Dr. Adam Linton Chair in Kidney Health Analytics and a Clinician Investigator Award from the CIHR. Jeremy Grimshaw holds a Canada Research Chair in Health Knowledge Transfer and Uptake.

Funding

We received funding for this review from partnering organizations, including Lawson Health Research Institute, Ontario Renal Network, Dialysis Clinic Inc., Heart and Stroke Foundation of Canada, and Canadian Institutes of Health Research (CIHR) Innovative Clinical Trials (iCT) Initiative. Funding is also provided by the Ontario Strategy for Patient-Oriented Research SUPPORT Unit, which is supported by the Canadian Institutes of Health Research and the Province of Ontario. Funding bodies had no role in the study's design, data collection, analysis, and interpretation of data and writing the manuscript.

Authors' contributions

AXG and MT conceived the overall project idea and co-led the funding application with substantial contributions from all authors. MT developed the data extraction tool and AAA, KC, and CEG extracted the study data. AAA, MT, and SND extracted details on trials considering the effect of clustering during sample size estimation and analysis. AAA, KC, CEG and MT wrote the initial draft of the manuscript with substantial input from all authors. All authors contributed to the interpretation of results, made critical revisions and approved the final manuscript.

References

1. Strippoli GFM, Craig JC, Schena FP. The number, quality, and coverage of randomized controlled trials in nephrology. Journal of the American Society of Nephrology : JASN 2004; 15: 411–419.

2. Palmer SC, Sciancalepore M, Strippoli GFM. Trial Quality in Nephrology: How Are We Measuring Up? American Journal of Kidney Diseases 2011; 58: 335–337.

3. Deo A, Schmid CH, Earley A, Lau J, Uhlig K. Loss to Analysis in Randomized Controlled Trials in CKD. American Journal of Kidney Diseases 2011; 58: 349–355.

4. Inrig JK, Califf RM, Tasneem A et al. The landscape of clinical trials in nephrology: a systematic review of Clinicaltrials.gov. American journal of kidney diseases : the official journal of the National Kidney Foundation 2014; 63: 771–780.

5. Donner A, Klar N. Design and Analysis of Cluster Randomization Trials in Health Research. Arnold, London: 2000.

6. Hayes RJ, Moulton LH. Cluster randomised trials. CRC Press, 2009.

7. Campbell MK, Piaggio G, Elbourne DR, Altman DG, CONSORT Group. Consort 2010 statement: extension to cluster randomised trials. BMJ (Clinical research ed.) 2012; 345: e5661.

8. Campbell MJ, Walters SJ. How to design, analyse and report cluster randomised trials in medicine and health related research. John Wiley & Sons, 2014.

9. Klar N, Donner A. Ethical Challenges Posed by Cluster Randomization. In: Wiley Encyclopedia of Clinical Trials. John Wiley & Sons, Inc., Hoboken, NJ, USA: 2007;

10. Goldstein CE, Weijer C, Taljaard M et al. Ethical Issues in Pragmatic Cluster-Randomized Trials in Dialysis Facilities. American Journal of Kidney Diseases 2019; 74: 659–666.

11. Eldridge SM, Ashby D, Feder GS. Informed patient consent to participation in cluster randomized trials: an empirical exploration of trials in primary care. CliniTrials 2005; 2: 91–98.

12. Edwards SJL, Braunholtz DA, Lilford RJ, Stevens AJ. Ethical issues in the design and conduct of cluster randomised controlled trials. BMJ 1999; 318: 1407–1409.

13. Weijer C, Grimshaw JM, Eccles MP et al. The Ottawa Statement on the Ethical Design and Conduct of Cluster Randomized Trials. PLoS medicine 2012; 9: e1001346.

14. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred Reporting

Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. J Clin Epidemiol 2009; 62: 1006–1012.

15. Eldridge SM, Chan CL, Campbell MJ et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. Pilot and feasibility studies 2016; 2: 64.

16. Taljaard M, McGowan J, Grimshaw JM et al. Electronic search strategies to identify reports of cluster randomized trials in MEDLINE: low precision will improve with adherence to reporting standards. BMC medical research methodology 2010; 10: 15.

17. Iansavichus A V, Haynes RB, Lee CWC et al. Dialysis search filters for PubMed, Ovid MEDLINE, and Embase databases. Clin J Am Soc Nephrol 2012; 7: 1624–3161.

18. Cohen J. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 1960; 20: 37–46.

19. Gallo A, Weijer C, White A et al. What is the role and authority of gatekeepers in cluster randomized trials in health research? Trials 2012; 13: 116.

20. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2019;

21. Huml AM, Sullivan CM, Pencak JA, Sehgal AR. Accuracy of dialysis medical records in determining patients' interest in and suitability for transplantation. Clinical transplantation 2013; 27: 541–545.

22. Leonardis D, Mallamaci F, Enia G et al. The MAURO study: baseline characteristics and compliance with guidelines targets. Journal of nephrology 2012; 25: 1081–1090.

23. Dai C-Y, Lin S-C, Peng H-L et al. Effectiveness of Vestibular Rehabilitation in Hemodialysis Patients With Dizziness. Rehabilitation nursing : the official journal of the Association of Rehabilitation Nurses 2017; 42: 125–130.

24. Carswell C, Reid J, Walsh I, McAneney H, Noble H. Implementing an arts-based intervention for patients with end-stage kidney disease whilst receiving haemodialysis: a feasibility study protocol. Pilot and feasibility studies 2019; 5: 1.

25. Griva K, Rajeswari M, Nandakumar M et al. The combined diabetes and renal control trial (C-DIRECT) - a feasibility randomised controlled trial to evaluate outcomes in multi-morbid patients with diabetes and on dialysis using a mixed methods approach. BMC nephrology 2019; 20: 2.

26. Sehgal AR, Leon JB, Siminoff LA, Singer ME, Bunosky LM, Cebul RD. Improving the quality of hemodialysis treatment: a community-based randomized controlled trial to overcome patient-specific barriers. JAMA : the journal of the American Medical Association 2002; 287: 1961–1967.

27. McClellan WM, Hodgin E, Pastan S, McAdams L, Soucie M. A randomized evaluation of two health care quality improvement program (HCQIP) interventions to improve the adequacy of hemodialysis care of ESRD patients: feedback alone versus intensive intervention. Journal of the American Society of Nephrology : JASN 2004; 15: 754–760.

28. Wileman V, Farrington K, Chilcot J et al. Evidence that self-affirmation improves phosphate control in hemodialysis patients: a pilot cluster randomized controlled trial. Annals of behavioral medicine : a publication of the Society of Behavioral Medicine 2014; 48: 275–281.

29. Karavetian M, de Vries N, Elzein H, Rizk R, Bechwaty F. Effect of behavioral stage-based nutrition education on management of osteodystrophy among hemodialysis patients, Lebanon. Patient Education and Counseling 2015; 98: 1116–1122.

30. Bennett PN, Fraser S, Barnard R et al. Effects of an intradialytic resistance training programme on physical function: A prospective stepped-wedge randomized controlled trial. Nephrology Dialysis Transplantation 2016; 31: 1302–1309.

31. Graham-Brown MPM, March DS, Churchward DR et al. Design and methods of CYCLE-HD: improving cardiovascular health in patients with end stage renal disease using a structured programme of exercise: a randomised control trial. BMC nephrology 2016; 17: 69.

32. Howren MB, Kellerman QD, Hillis SL, Cvengros J, Lawton W, Christensen AJ. Effect of a Behavioral Self-Regulation Intervention on Patient Adherence to Fluid-Intake Restrictions in Hemodialysis: a Randomized Controlled Trial. Annals of behavioral medicine : a publication of the Society of Behavioral Medicine 2016; 50: 167–176.

33. Wileman V, Chilcot J, Armitage CJ et al. Evidence of improved fluid management in patients receiving haemodialysis following a self-affirmation theory-based intervention: A randomised controlled trial. Psychology and Health 2016; 31: 100–114.

34. Hymes JL, Mooney A, Van Zandt C, Lynch L, Ziebol R, Killion D. Dialysis Catheter–Related Bloodstream Infections: A Cluster-Randomized Trial of the ClearGuard HD Antimicrobial Barrier Cap. American Journal of Kidney Diseases 2017; 69: 220–227.

35. Patzer RE, Paul S, Plantinga L et al. A Randomized Trial to Reduce Disparities in Referral for Transplant Evaluation. Journal of the American Society of Nephrology : JASN 2017; 28: 935–942.

36. Patzer RE, Smith K, Basu M et al. The ASCENT (Allocation System Changes for Equity in Kidney Transplantation) Study: a Randomized Effectiveness-Implementation Study to Improve Kidney Transplant Waitlisting and Reduce

Racial Disparity. Kidney international reports 2017; 2: 433–441.

37. Brunelli SM, Van Wyck DB, Njord L, Ziebol RJ, Lynch LE, Killion DP. Cluster-Randomized Trial of Devices to Prevent Catheter-Related Bloodstream Infection. Journal of the American Society of Nephrology : JASN 2018; 29: 1336–1343.

38. Leon JB, Albert JM, Gilchrist G et al. Improving Albumin Levels Among Hemodialysis Patients: A Community-Based Randomized Controlled Trial. American Journal of Kidney Diseases 2006; 48: 28–36.

39. Delmas P, O'Reilly L, Cara C et al. Effects on nurses' quality of working life and on patients' quality of life of an educational intervention to strengthen humanistic practice among hemodialysis nurses in Switzerland: a protocol for a mixed-methods cluster randomized controlled trial. BMC nursing 2018; 17: 47.

40. Griva K, Nandakumar M, Ng JH, Lam KFY, McBain H, Newman SP. Hemodialysis Self-management Intervention Randomized Trial (HED-SMART): A Practical Low-Intensity Intervention to Improve Adherence and Clinical Markers in Patients Receiving Hemodialysis. American Journal of Kidney Diseases 2018; 71: 371–381.

41. Huang B, Li Z, Wang Y et al. Effectiveness of self-management support in maintenance haemodialysis patients with hypertension: A pilot cluster randomized controlled trial. Nephrology 2018; 23: 755–763.

42. Milazi M, Douglas C, Bonner A. A bundled phosphate control intervention (4Ds) for adults with end-stage kidney disease receiving haemodialysis: A cluster randomized controlled trial protocol. Journal of advanced nursing 2018;

43. Sullivan CM, Barnswell K V, Greenway K et al. Impact of Navigators on First Visit to a Transplant Center, Waitlisting, and Kidney Transplantation: A Randomized, Controlled Trial. Clinical journal of the American Society of Nephrology : CJASN 2018; 13: 1550–1555.

44. Waterman AD, Peipert JD. An Explore Transplant Group Randomized Controlled Education Trial to Increase Dialysis Patients' Decision-Making and Pursuit of Transplantation. Progress in transplantation (Aliso Viejo, Calif.) 2018; 28: 174–183.

45. Dember LM, Lacson E, Brunelli SM et al. The TiME Trial: A Fully Embedded, Cluster-Randomized, Pragmatic Trial of Hemodialysis Session Duration. Journal of the American Society of Nephrology 2019; 30: 890–903.

46. Sullivan C, Sayre SS, Leon JB et al. Effect of Food Additives on Hyperphosphatemia Among Patients With End-stage Renal Disease. JAMA 2009; 301: 629.

47. Bond TC, Patel PR, Krisher J et al. A group-randomized evaluation of a quality

improvement intervention to improve influenza vaccination rates in dialysis centers. American Journal of Kidney Diseases 2011; 57: 283–290.

48. Sullivan C, Leon JB, Sayre SS et al. Impact of navigators on completion of steps in the kidney transplant process: A randomized, controlled trial. Clinical Journal of the American Society of Nephrology 2012; 7: 1639–1645.

49. Pradel FG, Suwannaprom P, Mullins CD, Sadler J, Bartlett ST. Short-term impact of an educational program promoting live donor kidney transplantation in dialysis centers. Progress in transplantation (Aliso Viejo, Calif.) 2008; 18: 263–272.

50. Song M-K, Unruh ML, Manatunga A et al. SPIRIT trial: A phase III pragmatic trial of an advance care planning intervention in ESRD. Contemporary clinical trials 2018; 64: 188–194.

51. Locatelli F, Covic A, MacDougall IC, Scherhag A, Wiecek A. Effect of computer-assisted European Best Practice Guideline implementation on adherence and target attainment: ORAMA results. Journal of Nephrology 2009; 22: 662–674.

52. Kauric-Klein Z. Improving blood pressure control in end stage renal disease through a supportive educative nursing intervention. Nephrology nursing journal : journal of the American Nephrology Nurses' Association 2012; 39: 217–228.

53. Bennett PN, Miller MD, Woodman RJ, Hill K, Murray S, Gleadle JM. Nutrition screening by nurses in dialysis. Journal of Clinical Nursing 2013; 22: 723–732.

54. Karavetian M, Ghaddar S. Nutritional education for the management of osteodystrophy (NEMO) in patients on haemodialysis: A randomised controlled trial. Journal of Renal Care 2013; 39: 19–30.

55. Weisbord SD, Mor MK, Green JA et al. Comparison of symptom management strategies for pain, erectile dysfunction, and depression in patients receiving chronic hemodialysis: A cluster randomized effectiveness trial. Clinical Journal of the American Society of Nephrology 2013; 8: 90–99.

56. Rosenblum A, Wang W, Ball LK, Latham C, Maddux FW, Lacson E. Hemodialysis catheter care strategies: A cluster-randomized quality improvement initiative. American Journal of Kidney Diseases 2014; 63: 259–267.

57. Fuller DS, Robinson BM. Facility Practice Variation to Help Understand the Effects of Public Policy: Insights from the Dialysis Outcomes and Practice Patterns Study (DOPPS). Clinical Journal of the American Society of Nephrology 2017; 12: 190–199.

58. Fuller DS, Zepel L, Bieber BA, Robinson BM, Pisoni RL. Hemodialysis Facility Variation in Hospitalization and Transfusions Using Medicare Claims: The DOPPS Practice Monitor for US Dialysis Care. American Journal of Kidney Diseases 2016; 67: 337–340.

59. Al Salmi I, Larkina M, Wang M et al. Missed Hemodialysis Treatments: International Variation, Predictors, and Outcomes in the Dialysis Outcomes and Practice Patterns Study (DOPPS). American journal of kidney diseases : the official journal of the National Kidney Foundation 2018; 72: 634–643.

60. Campbell MK, Grimshaw JM, Elbourne DR. Intracluster correlation coefficients in cluster randomized trials: empirical insights into how should they be reported. BMC medical research methodology 2004; 4: 9.

61. Falagas ME, Korbila IP, Giannopoulou KP, Kondilis BK, Peppas G. Informed consent: how much and what do patients understand? The American Journal of Surgery 2009; 198: 420–435.

62. Sherlock A, Brownie S. Patients' recollection and understanding of informed consent: a literature review. ANZ Journal of Surgery 2014; 84: 207–210.

63. United States Renal Data System. 2018 USRDS annual data report: Epidemiology of kidney disease in the United States. Bethesda, MD: 2018.

64. Moist LM, Fenton S, Kim JS et al. Canadian Organ Replacement Register (CORR): reflecting the past and embracing the future. Canadian journal of kidney health and disease 2014; 1: 26.

65. Liyanage T, Ninomiya T, Jha V et al. Worldwide access to treatment for end-stage kidney disease: a systematic review. The Lancet 2015; 385: 1975–1982.

66. Green JA, Mor MK, Shields AM et al. Associations of Health Literacy With Dialysis Adherence and Health Resource Utilization in Patients Receiving Maintenance Hemodialysis. American Journal of Kidney Diseases 2013; 62: 73–80.

67. Varnell SP, Murray DM, Baker WL. An evaluation of analysis options for the one-group-per-condition design. Can any of the alternatives overcome the problems inherent in this design? Evaluation review 2001; 25: 440–453.

68. Taljaard M, Teerenstra S, Ivers NM, Fergusson DA. Substantial risks associated with few clusters in cluster randomized and stepped wedge designs. Clinical trials (London, England) 2016; 13: 459–463.

69. Canadian Institutes for Health Research, Natural Sciences and Engineering Research Council of Canada, Social Sciences and Humanities Research Council of Canada, Canadian Institutes for Health Research the Natural Sciences, Natural Sciences and Engineering Research Council of Canada, Social Sciences and Humanities Research Council of Canada. Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans. 2014. Ottawa: 2014.

70. Council for International Organizations of Medical Sciences, World Health Organization. International ethical guidelines for health-related research involving

humans.

71. Guideline For Good Clinical Practice E6(R1). 1996.

72. HiLo | A Pragmatic Clinical Trial [Internet]. [cited 2020 Jan 20] Available from: https://hilostudy.org/

73. Outcomes of a Higher Versus Lower Hemodialysate Magnesium Concentration [Internet]. [cited 2020 January 20th] Available from: https://clinicaltrials.gov/ct2/show/NCT04079582

74. Taljaard M, Goldstein CE, Giraudeau B et al. Cluster over individual randomization: are study design choices appropriately justified? Review of a random sample of trials. Clinical Trials 2019; : Under Review.

75. Ivers NM, Taljaard M, Dixon S et al. Impact of CONSORT extension for cluster randomised trials on quality of reporting and study methodology: review of random sample of 300 trials, 2000-8. BMJ 2011; 343: d5886–d5886.

76. Taljaard M, Hemming K, Shah L, Giraudeau B, Grimshaw JM, Weijer C. Inadequacy of ethical conduct and reporting of stepped wedge cluster randomized trials: Results from a systematic review. Clinical trials (London, England) 2017; 14: 333–341.

77. Taljaard M, McRae AD, Weijer C et al. Inadequate reporting of research ethics review and informed consent in cluster randomised trials: review of random sample of published trials. BMJ (Clinical research ed.) 2011; 342: d2496.

Tables and Figures

Table 3-1: Included studies and their characteristics.

First Author	Year	Country	Intervention Arm (Number)		Control Arm (Number)		Type of Cluster	Type of Patients ₳	Type of Intervention ₱	Primary Outcome
			Clusters	Patients	Clusters	Patients				
Sehgal(26)	2002	USA	21	85	23	84	Individual providers	Prevalent only	2 and 3	Change in Kt/V and achievement of facility Kt/V goal
McClellan(27)	2004	USA	21	2237	20	2044	HD units	Prevalent and incident	1, 2, and 3	Proportion of patients whose urea reduction ratio was ≥65%
Leon(38)	2006	USA	21	86	23	94	HD units	Prevalent only	3	Serum albumin level
Pradel(49)	2008	USA	14	107	14	107	Shifts in HD unit	Prevalent and incident	3	See ¥
Locatelli(51)	2009	EU**	NR*	321	NR	278	Nephrology Unit	Prevalent and incident	2	Proportions of patients with hemoglobin >11g/dL, serum ferritin >100 µg/L, hypochromic red cell count <10%, or transferrin saturation >20%
Sullivan(46)	2009	USA	14	145	14	134	Shifts in HD unit	Prevalent only	3	Serum phosphorus level

Bond(47)	2011	USA	38	3157	39	3135	HD units	Prevalent and incident	2 and 3	Change in influenza vaccination rates
Kauric-Klein(52)	2012	USA	NR	59	NR	59	HD Units	Prevalent only	3	Changes in systolic blood pressure over time (primary outcome not explicitly stated)
Sullivan(48)	2012	USA	11	92	12	75	HD units	Prevalent and incident	3	Number of transplant process steps completed
Bennett(53)	2013	AUS / NZ	2	38	2	41	HD units	Prevalent and incident	2	Rate of referral to dietetic services for nutrition support
Karavetian(54)	2013	Lebanon	1	37	1	24	Shifts in HD unit	Prevalent and incident	3	Patient knowledge score £
Weisbord(55)	2013	USA	9	100	9	120	Shifts in HD unit	Prevalent and incident	2	Changes in scores on pain, erectile dysfunction, and depression surveys
Rosenblum(56)	2014	USA	216	4609	216	4551	HD units	Prevalent and incident	2 and 4	Positive blood culture rate
Wileman(28)	2014	UK	6	45	6	45	Shifts in HD unit	Prevalent and incident	3	Serum phosphate level
Karavetian(29)	2015	Lebanon	6	88	6	96	Shifts in HD unit	Prevalent and incident	3	Serum phosphorus level
Bennett(30)	2016	AUS / NZ	15	171	15	171	HD units	Prevalent and incident	3 and 4	30-second sit-to-stand test

Graham-Brown(31)	2016	UK	3	NA***	3	NA***	Shifts in HD unit	Prevalent only	4	Left ventricular mass
Howren(32)	2016	USA	11	61	11	58	Shifts in HD unit	Prevalent and incident	3	Unclear: Mean interdialytic weight gain across for periods **or** Fluid nonadherent as defined by an interdialytic weight gain >2.5 kg over four weeks
Wileman(33)	2016	UK	6	49	6	40	Shifts in HD unit	Prevalent and incident	3	Interdialytic weight gain
Hymes(34)	2017	USA	20	1245	20	1225	HD units	Prevalent and incident	2 and 4	Positive blood culture rate
Patzer(35)	2017	USA	67	4203	20	1225	HD units	Prevalent and incident	1, 2, and 3	Facility level transplant referral rate
Patzer(36)	2017	USA	NA***	NA***	NA***	NA***	HD units	Prevalent and incident	1, 2, and 3	Co-primary outcomes of (i) change in the proportion of patients waitlisted, and (ii) disparity reduction in the proportion of patients waitlisted in a dialysis facility after one year

Brunelli(37)	2018	USA	20	826	20	845	HD units	Prevalent and incident	4	Positive blood culture rate
Delmas(39)	2018	Switzerland	NR	NR	NR	NR	HD Units	Prevalent only	1	Nurse quality of working life
Griva(40)	2018	Singapore	14	101	14	134	Shifts in HD unit	Prevalent only	3	Serum potassium/phosphate levels and interdialytic weight gains
Huang(41)	2018	China	1	46	1	44	Shifts in HD unit	Prevalent and incident	3	Blood pressure monitored before each hemodialysis
Milazi(42)	2018	AUS / NZ	3	60	3	60	Shifts in HD unit	Prevalent and incident	3	Serum phosphate level
Song(50)	2018	USA	NA***	NA***	NA***	NA***	HD units	Prevalent only	3 and 5	Patient and surrogate self-reported preparedness for end-of-life decision making
Sullivan(43)	2018	USA	20	1041	20	836	HD units	Prevalent and incident	3	Placement on kidney transplant waiting list
Waterman(44)	2018	USA	10	133	10	120	HD units	Prevalent and incident	3	Patients' readiness to allow someone to be a living donor
Dember(45)	2019	USA	133	1938	133	2532	HD units	Incident only	4	Death

* Locatelli et al. did not report the number of clusters randomized to each arm; however, the authors reported a total of 53 nephrology units participated in the trial.

** Included countries from Bulgaria, Croatia, Poland, Romania, and Serbia and Montenegro

*** This was a study protocol of an ongoing trial, and thus the final sample size used (or to be used) in the analysis was not available.

₳ We defined patients as "prevalent" if they were on hemodialysis for at least six months and "incident" if they are newly starting or started hemodialysis less than six months before baseline.

¥ Study assessed three distinct behaviors to explore patients' readiness to pursue living donor kidney transplant: (1) considering a living donor kidney transplant, (2) talking with family or friends about living donor kidney transplant, and (3) asking someone to be a kidney transplant donor.

£ Patient knowledge questionnaire was utilized to assess patients' knowledge of kidney disease, renal diet, phosphate binders, and vitamin D therapy.

₱ 1 = Educational/ quality improvement interventions targeted at health professionals (e.g., transplant education and engagement activities targeting health professionals, etc.); 2 = Quality improvement interventions targeted at the organization of health care or health services delivery (e.g., nutrition screening, change in catheter exit-site care, etc.); 3 = Patient health promotion or educational intervention (e.g., education about benefits of resistance exercise program, dietary counseling, education on avoiding foods with phosphorus additives, etc.); 4= Direct patient therapeutic intervention (e.g., intradialytic resistance training, antimicrobial barrier caps for catheters, etc.) and; 5 = Other

NR=Not reported, USA=United States of America, EU=European Union, UK=United Kingdom, AUS / NZ = Australia / New Zealand, NA=Not applicable, g/dL=grams per deciliter, µg/L= micrograms per liter, Kt/V = fractional urea clearance represented by K=dialyzer clearance of urea, t=dialysis time, V=distribution volume of urea

Table 3-2: Reporting of study characteristics.

Component	**Number of Studies (%) (N total=31)**
Trial Design	
Parallel arm	30 (97%)
Stepped-wedge design	1 (3%)
Types of Randomized Clusters	
Hemodialysis centers	17 (55%)
Hemodialysis shifts or sessions	12 (39%)
Providers or professionals	1 (3%)
Nephrology units ₱	1 (3%)
Method of Random Allocation	
Completely randomized design (unrestricted randomization)	8 (26%)
Stratified design	4 (13%)
Pair-matched design	4 (13%)
Split-cluster (i.e. shifts within a hemodialysis center)	11 (35%)
Covariate-constrained	1 (3%)
Not reported	3 (10%)
Number of Clusters per trial [median (25^{th}, 75^{th} percentile)]¥	28 (12, 43)
Number of Patients per Trial [median (25^{th}, 75^{th} percentile)]₳	228 (120, 1723)
Number of Patients per Cluster [median (25^{th}, 75^{th} percentile)]€, ɤ	20 (8, 32)

₱ It is not clear how a "nephrology unit" was defined.

Estimate is based on ¥ 32, ₳ 29, and € 28 trials. Missing data may have resulted from not reporting or the study being a protocol with no final information on the number of clusters/patients being available.

ɤ For each study, we estimated the average cluster size by dividing the total number of patients recruited by the number of clusters (e.g., 200 patients recruited in a trial / 10 clusters = 20 patients per cluster). We then took the median of the calculated average of patients per cluster from each trial.

Table 3-3: Reporting of a) how clustering was considered during sample size estimation and analysis; and b) the justification for using a cluster randomized design.

Did sample size/power calculations account for the cluster design?	**N=31 trials (%)**
Not presented Ⱥ	11 (35%)
Yes, used patient-level data and accounted for clustering (e.g., random-effects model)	11 (35%)
Yes, used cluster-level summaries	3 (10%)
No, used patient-level data without accounting for clustering	3 (10%)
Unclear	1 (3%)
Other ¥	2 (6%)
Did the analysis for the primary outcome account for clustering?	
Yes, used patient-level data and accounted for clustering	17 (55%)
Yes, used cluster-level summaries	5 (16%)
No, used patient-level data without accounting for clustering ₱	7 (23%)
Unclear / Other ¥	2 (6%)
Justification for utilizing a cluster randomized design (categories were not mutually exclusive)	
None provided	16 (52%)
Avoid contamination	15 (48%)
Logistical or administrative convenience	2 (6%)

Ⱥ One study presented power calculation, but it was a post-hoc power analysis

¥ This may have included using an inappropriate method for the proposed primary outcome, or the study accounted for clustering but not based on the primary outcome measure (e.g., they assumed a continuous outcome, but the primary endpoint was a proportion).

₱ One study accounted for repeated events within patients but did not report accounting for within-cluster correlation; another study reported using a generalized linear mixed model but did not specify whether they accounted for the effect of the cluster as a random effect.

Table 3-4: Summary of results for type(s) of interventions, data collection procedures, reporting of participant consent procedures for study interventions and data collection, the timing of any participant consent, whether participants can opt-out of the intervention or data collection.

Component	**Intervention arm n (%)**	**Control arm n (%)**
Type(s) of interventions (i.e., all components of intervention)¥	**N total=31**	**N total= 8****
Educational/ quality improvement interventions targeted at health professionals (e.g., transplant education and engagement activities targeting health professionals, etc.)	4 (13%)	0 (0%)
Quality improvement interventions targeted at organization of health care or health services delivery (e.g., nutrition screening, change in catheter exit-site care, etc.)	10 (32%)	2 (25%)
Patient health promotion or educational intervention (e.g., education about benefits of resistance exercise program, dietary counseling, education on avoiding foods with phosphorus additives, etc.)	22 (71%)	4 (50%)
Direct patient therapeutic intervention (e.g., intradialytic resistance training, antimicrobial barrier caps for catheters, etc.)	6 (19%)	1 (12%)
Other €	1 (3%)	1 (12%)
Types of Data collection procedures ¥	**N total=31**	**N total=31**
Routinely collected outcomes extracted locally from existing patient medical records (physical charts or electronic records)	30 (97%)	30 (97%)
Data query from clinical data registry or other central sources of routinely collected data (e.g., administrative data)	11 (35%)	11 (35%)
Specimen collection or physical examination that was not required for usual patient care	4 (13%)	4 (13%)
Interviewer-administered patient questionnaires done face-to-face or by telephone that was not required for usual patient care	9 (29%)	9 (29%)
Self-administered patient questionnaires (done by mail, e-mail, or internet) that were not required for usual patient care	18 (58%)	16 (52%)
Other ₳	5 (16%)	2 (6%)
Reporting of participant consent procedures for *study interventions*	**N total=31**	**N total=31**
Reported written informed consent	9 (29%)	10 (32%)
Reported verbal informed consent	1 (3%)	0 (0%)
Reported informed consent but no details about the method or what consent was for	12 (39%)	11 (35%)
Reported the study was exempt from REC review, received a waiver of consent, or explicitly stated no consent	4 (13%)	4 (13%)
Unclear if participants consented	1 (3%)	2 (6%)

Component	**Intervention arm n (%)**	**Control arm n (%)**
Not mentioned	4 (13%)	4 (13%)
Reporting of participant consent procedures for *data collection*	**N total=31**	**N total=31**
Reported written informed consent	7 (22%)	6 (19%)
Reported verbal informed consent	1 (3%)	1 (3%)
Reported informed consent but no details about the method or what consent was for	14 (45%)	14 (45%)
Reported the study was exempt from REC review, received a waiver of consent, or explicitly stated no consent	4 (13%)	4 (13%)
Unclear if participants consented	1 (3%)	2 (6%)
Not mentioned	4 (13%)	4 (13%)

¥ The responses to these questions were not mutually exclusive.

₳ Active data collection, including using case report form.

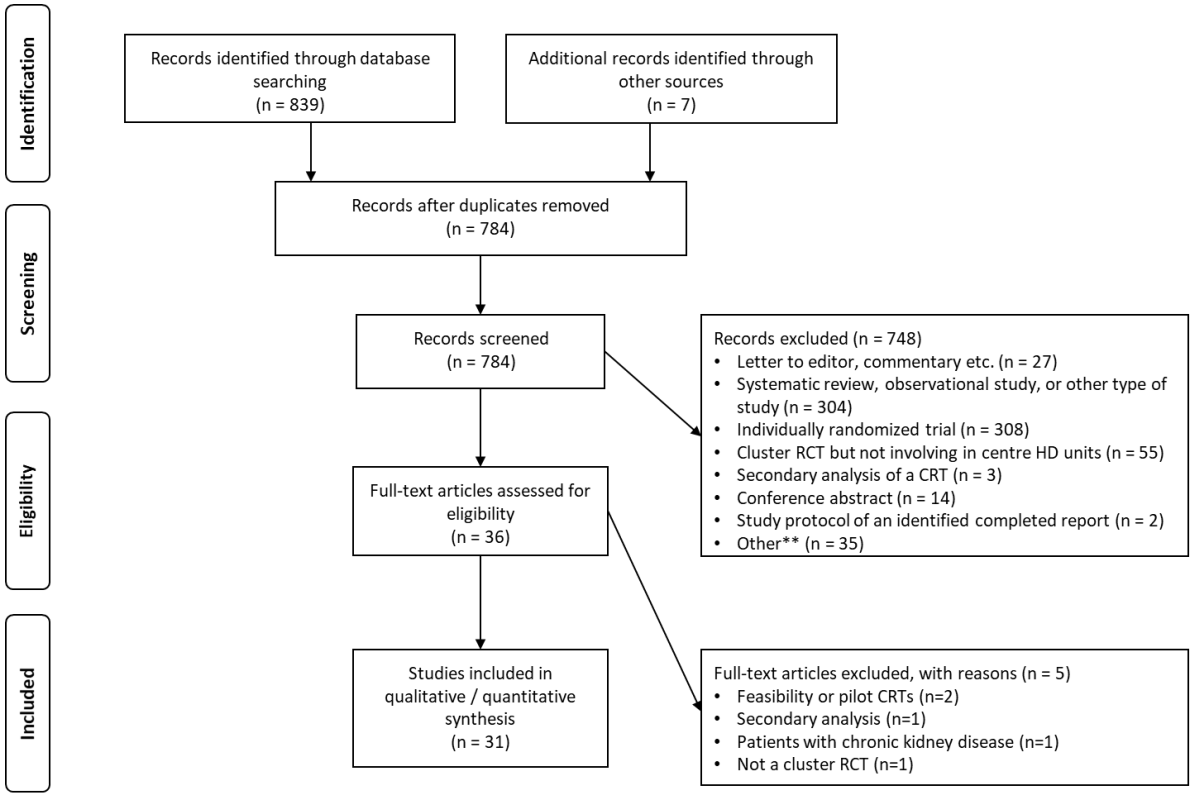
€ Surrogate decision-maker educational intervention in the intervention arm; Audit feedback from the previous year in the control arm.

These questions were not applicable when the comparator arm was ***usual care*.

Table 3-5: Summary of results for reported information about gatekeepers, research ethics committee review, the timing of any participant consent, and whether participants can opt-out of the intervention or data collection.

Component	**Number of Trials** **N total = 31 (%)**
Whether a gatekeeper was identified that allowed access to each cluster	
Yes – a clearly identified individual or body	3 (10%)
Yes – but the gatekeeper not clearly identified	2 (6%)
No gatekeeper information provided	26 (84%)
Reporting of research ethics review	
Stated REC approval	26 (84%)
Stated REC exempt (specify reason)	1 (3%)
Not reported	4 (13%)
Timing of any participant consent	
Not applicable	4 (13%)
Any consent was ***before*** randomization of clusters	7 (23%)
Any consent was ***after*** randomization of clusters	10 (32%)
The timing of consent was unclear and could not be deduced from the report	10 (32%)
Whether participants can opt-out of the data collection	
Yes - it is clearly reported that participants could opt out of data collection	7 (23%)
No - participants could not opt-out of data collection	3 (10%)
Not reported or Unclear if participants could opt-out	21 (68%)

Figure 3-1: Flow diagram of study selection.



**Other: One manuscript described the statistical plan for a main publication not related to cluster randomized trials, two described a program of research not related to the target population, and two were duplicate records not previously removed.

Abbreviation: RCT, randomized controlled trial; CRT, cluster randomized trial.

Appendix

Appendix 3-1: Recommendations from the Ottawa Statement.

Ethical Issue	**Recommendation Number**	**Recommendation**	**How trials in our review reported on these recommendations**
Justification for using cluster randomization	1	Researchers should provide a clear rationale for the use of the cluster randomized design and adopt statistical methods appropriate for this design.	Cluster randomized trials in the hemodialysis setting often fail to provide a clear rationale for the choice of cluster randomization. Providing a clear rationale is especially important in the case of individual-level interventions, in which a patient-randomized trial would have been possible. Researchers should explain why cluster randomization benefits outweigh the disadvantages of increased sample size and risks of bias in such circumstances. The effects of clustering must be considered during sample size calculation and analysis to avoid an underpowered study and spurious statistical significance, respectively. Additionally, when there are multiple levels of clustering (e.g., multiple hemodialysis shifts or providers within centres), all levels of clustering may need to be taken into consideration in the design and analysis.

Research ethics committee review	2	Researchers must submit a cluster randomized trial involving human research participants for approval by a research ethics committee before commencing.	Some cluster randomized trials fail to report receiving research ethics approval. However, all research involving human participants must seek and obtain research ethics committee approval; this includes trials that evaluate minimal-risk quality improvement interventions (e.g., altering the hemodialysis central venous catheter care procedure compared to usual care).
Identifying research participants	3	Researchers should clearly identify the research participants in cluster randomized trials. A research participant can be identified as an individual whose interests may be affected as a result of study interventions or data collection procedures, that is, an individual: (1) who is the intended recipient of an experimental (or control) intervention; or (2) who is the direct target of an experimental (or control) manipulation of his/her environment; or (3) with whom an investigator interacts for the purpose of collecting data about that individual; or (4) about whom an investigator obtains identifiable private information for the purpose of collecting data about that individual.	Most trials in the hemodialysis setting include patients as research participants, while a few include health professionals as participants. Health professionals are commonly overlooked as research participants; researchers and research ethics committees should pay particular attention when testing interventions that specifically target these individuals.

		Unless one or more of these criteria is met, an individual is not a research participant.	
Obtaining informed consent	4	Researchers must obtain informed consent from human research participants in a cluster randomized trial, unless a waiver of consent is granted by a research ethics committee under specific circumstances.	When informed consent is sought from participants, trials ought to report adequate details to assess the purpose of the consent (e.g., enrollment, receiving the intervention, data collection), from whom (e.g., patients, provider), when consent (before or after randomization), and how consent is obtained (e.g., written, oral).
	5	When participants' informed consent is required, but recruitment of participants is not possible *before* randomization of clusters, researchers must seek participants' consent for trial enrollment as soon as possible after cluster randomization—that is, as soon as the potential participant has been identified, but before the participant has undergone any study interventions or data collection procedures.	Trials in hemodialysis settings often do not report the timing of informed consent or whether the consent procedures occur after randomization of clusters. Obtaining consent post-randomization compromises randomization and increases the risk of recruitment bias. However, in hemodialysis trials, patients may need to be prospectively recruited after randomization (e.g., new patients starting hemodialysis treatment). Some authors put in place protections to reduce the risks of bias, including: (1) ensuring that participants are recruited by individuals who are blinded to the cluster's allocation;

			(2) putting in place a standardized participant identification mechanism in all trial arms; and (3) developing mechanisms whereby participants cannot become unblinded before consent or entry into the study.
	6	A research ethics committee may approve a waiver or alteration of consent requirements when (1) the research is not feasible without a waiver or alteration of consent, and (2) the study interventions and data collection procedures pose no more than minimal risk. Canada's Tri-Council Policy Statement also require that: (3) the waiver or alteration is unlikely to adversely affect the welfare or the rights of participants, and (4) there is a plan to inform participants of the trial and the intervention, and participants can refuse the intervention	Hemodialysis trials utilizing waivers of consent commonly fail to adequately report how their study meets the waiver criteria or alteration of informed consent.
	7	Researchers must obtain informed consent from professionals or other service providers who are research participants unless conditions for a waiver or alteration of consent are met.	While a few trials involve health care professionals as research participants, these individuals are commonly overlooked as research participants. Examples of health care providers who may be research participants in the hemodialysis setting include dialysis nurses, nurse educators, nurse practitioners, nephrologists,

			physiotherapists, kinesiologists, and pharmacists targeted by knowledge translation interventions.
Gatekeepers	8	Gatekeepers should not provide proxy consent on behalf of individuals in their cluster.	Gatekeepers in the hemodialysis setting may include but are not limited to medical directors, administrators, clinical care providers, patient and family advisory boards, members of the dialysis provider organization leadership, payers, or representatives of governmental organizations. Gatekeepers can provide permission for the center to participate in the trial but cannot consent to study participation on behalf of patients who are research participants.
	9	When a cluster randomized trial may substantially affect cluster or organizational interests, and a gatekeeper possesses the legitimate authority to make decisions on its behalf, the researcher should obtain the gatekeeper's permission to enrol the cluster or organization in the trial. Such permission does not replace the need for the informed consent of research participants.	Trials in the hemodialysis setting rarely report the role of gatekeepers. Gatekeepers play an important role in hemodialysis trials as they usually provide permission for their cluster to be recruited or randomized to different arms of the trial.
	10	When cluster randomized trial interventions may substantially affect cluster interests, researchers should seek	Hemodialysis trials rarely report whether any gatekeeper consultations take place. Researchers may consult with gatekeepers

		to protect cluster interests through cluster consultation to inform study design, conduct, and reporting. Where relevant, gatekeepers can often facilitate such a consultation.	to facilitate implementation of the intervention, ensure high adherence to the assigned treatment protocol, and minimize or reduce protocol violations.
Assessing benefits and harms	11	The researcher must ensure that the study intervention is adequately justified. The benefits and harms of the study intervention must be consistent with competent practice in the field of study relevant to the cluster randomized trial.	Many aspects of clinical care in the hemodialysis setting are guided by clinical opinion and physiologic studies. There is a high degree of practice variations between hemodialysis centres and health care providers, and this complicates the assessment of benefits and harms, which is often not reported in trials in the hemodialysis setting.
	12	Researchers must adequately justify the choice of the control condition. When the control arm is usual practice or no treatment, individuals in the control arm must not be deprived of effective care or programmes to which they would have access, were there no trial.	Hemodialysis cluster randomized trials typically compare the effectiveness of existing, widely used interventions in the setting of routine clinical practice (i.e., usual care). Most trials utilize “usual care” as the control arm. However, there are instances where trials might conduct a head-to-head comparison between two interventions.
	13	Researchers must ensure that data collection procedures are adequately justified. The risks of data collection procedures must (1) be minimised consistent with sound design and (2) stand	In the hemodialysis setting, trials commonly use routinely collected data (e.g., medical charts or electronic medical records) as a primary source for data collection and often supplement this information with other data sources (e.g.,

		in reasonable relation to the knowledge to be gained.	questionnaires, specimen collection, physical examination, administrative data, etc.).
Protecting vulnerable participants	14	Clusters may contain some vulnerable participants. In these circumstances, researchers and research ethics committees must consider whether additional protections are needed.	Kidney disease disproportionally affects individuals with multiple comorbidities, live in rural or remote locations, have dementia, lower education levels, and lower health literacy. These characteristics are also associated with poor comprehension of the informed consent processes. Although vulnerable subgroups may have been included in the trials, none report putting protections for vulnerable populations.
	15	When individual informed consent is required, and there are individuals who may be less able to choose participation freely because of their position in a cluster or organizational hierarchy, research ethics committees should pay special attention to recruitment, privacy, and consent procedures for those participants.	Researchers and research ethics committees must ensure the rights are protected for vulnerable participants (e.g., trainees or nurses) in an organizational setting, where a superior (e.g., medical director) might influence the participation in the study. However, without access to the original research ethics committee submission, it may not be possible to determine to what degree research ethics committees are attentive to recruitment, privacy, and consent procedures for such vulnerable participants.

Appendix 3-2: PRISMA Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3 / 4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	4
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	4
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	4 / 5

Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	5 / Appendix 3-3
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	5
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	5 / 6
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	5-6 / Appendix 3-4
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	NA – 1
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	6
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	NA – 2
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	NA – 1
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	NA – 2
RESULTS			

Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	6 / Figure 3-1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	6 / Table 3-1
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	NA – 1
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	7 to 10
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	NA – 2
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	NA – 1
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	NA – 2
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	10 to 13
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	13
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	14

FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	15

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit **www.prisma-statement.org**.

NA – 1: This research aimed not to assess a specific health outcome within selected studies. However, instead, we aimed to capture key ethical and methodological elements within selected hemodialysis cluster randomized trials.

NA – 2: While we do report some summary statistics (e.g., medians, range, etc.), the aim of this paper was not to capture measures of association for a specific outcome. Also, as described in the manuscript, we could not conduct any meaningful subgroup analyses given the small number of studies eligible for inclusion.

Appendix 3-3: Search syntax to identify relevant articles in Medline between January 1st, 2000 and July 20th, 2018 in EMBASE Classic+ EMBASE, Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R).

Database: Ovid MEDLINE(R) ALL <1946 to November 30, 2019>

---

1. (dialy* or hemodi*).mp. or haemodi*.tw. or end-stage renal.tw. or endstage renal.tw. or end-stage kidney.tw. or esrd.tw. or renal replacement.mp. or uremia.mp. or uraemia.mp. or exp "Uremia"/ or capd.tw. or hemofilt*.mp. or haemofilt*.mp. or hyperphosphataemia.tw. or hyperphosphatemia.tw. or uremic patient*.tw. or uraemic patient*.tw. or secondary hyperparathyroidism.tw. or renal osteodystrophy.mp. or intradialy*.tw. or hyperoxaluria.mp. or tenckhoff*.tw. or autosomal dominant polycystic kidney.ti. or ccpd.tw. (241390)
2. exp Renal Insufficiency, Chronic/ (103332)
3. exp *kidney failure/ (113586)
4. (indwelling catheter/ or central venous catheterization/) and heparin.mp. (970)
5. ("Severity of Illness Index"/ or vascular.ti. or *"Anemia"/ or anemi*.ti. or anaemi*.ti. or nephrogenic.tw. or amyloid*.mp. or rhabdomyolysis.mp.) and *"Kidney Disease"/ (3849)
6. chronic.mp. and *"Kidney Disease"/ (11683)
7. ((kidney transplant* or renal transplant*) and (candidates or wait* list*)).tw. (2726)
8. encapsulating.tw. and scleros*.mp. (623)
9. or/1-8 (316044)
10. (random* and cluster*).mp. (31050)
11. (cluster* adj3 rct*).tw. (524)
12. (cluster* adj3 trial*).tw. (7928)
13. (communit* adj2 intervention*).tw. (6015)
14. (random* adj2 (group* or communit*)).tw. (20526)
15. or/10-14 (56430)
16. animals/ not humans/ (4443017)
17. 15 not 16 (51170)
18. 9 and 17 (448)
19. limit 18 to yr="2000 -Current" (383)
20. ("2017 06 23*" or "2017 06 24*" or "2017 06 25*" or "2017 06 26*" or "2017 06 27*" or "2017 06 28*" or "2017 06 29*" or "2017 06 3*" or 2017 07* or 2017 08* or 2017 09* or 2017 1* or 2018*).dt. (1375531)
21. 19 and 20 (32)

Database: EMBASE ***Classic+*** EMBASE ***<1947 to November 30, 2019>***

---

1. (dialy* or hemodi*).mp. or haemodi*.tw. or end-stage renal.tw. or endstage renal.tw. or end-stage kidney.tw. or esrd.tw. or renal replacement.mp. or uremia.mp. or uraemia.mp. or exp "Uremia"/ or capd.tw. or hemofilt*.mp. or haemofilt*.mp. or hyperphosphataemia.tw. or hyperphosphatemia.tw. or uremic patient*.tw. or uraemic patient*.tw. or secondary hyperparathyroidism.tw. or renal osteodystrophy.mp. or intradialy*.tw. or hyperoxaluria.mp. or tenckhoff*.tw. or autosomal dominant polycystic kidney.ti. or ccpd.tw. (357289)
2. *chronic kidney failure/ (39936)
3. *kidney failure/ or *end stage renal disease/ (45353)
4. (indwelling catheter/ or central venous catheterization/) and heparin.mp. (880)
5. ("Severity of Illness Index"/ or vascular.ti. or *"Anemia"/ or anemi*.ti. or anaemi*.ti. or nephrogenic.tw. or amyloid*.mp. or rhabdomyolysis.mp.) and (exp *kidney failure/ or *"Kidney Disease"/) (10012)
6. chronic.mp. and *"Kidney Disease"/ (9516)
7. ((kidney transplant* or renal transplant*) and (candidates or wait* list*)).tw. (5136)
8. encapsulating.tw. and scleros*.mp. (800)
9. or/1-8 (402695)
10. (cluster* and random*).tw. (34468)
11. (cluster* adj3 rct*).tw. (739)
12. (cluster* adj3 trial*).tw. (9762)
13. (communit* adj2 intervention*).tw. (7650)
14. (random* adj2 (group* or communit*)).tw. (30199)
15. controlled clinical trial/ and cluster analysis/ (952)
16. randomized controlled trial/ and cluster analysis/ (1870)
17. or/10-16 (71363)
18. (exp animal/ or nonhuman/) not exp human/ (6666585)
19. 17 not 18 (62060)
20. 9 and 19 (574)
21. limit 20 to yr="2000 -Current" (518)
22. ("20170623*" or "20170624*" or "20170625*" or "20170626*" or "20170627*" or "20170628*" or "20170629*" or "2017063*" or 2017 07* or 2017 08* or 2017 09* or 2017 1* or 2018*).dc. (1045427)
23. 21 and 22 (43)

Appendix 3-4: Extracted data

We extracted data on the following: **(1)** study characteristics, including the year of publication, country of study recruitment, country's level of development (for identification of emerging and developing economies, we used World Economic Outlook database); **(2)** methodological characteristics, including study design, method of random allocation, unit of randomization, number of clusters and patients analyzed (or sample size estimated in the published protocol), data collection method, sample size estimation and whether clustering was taken into account, whether the analysis considered the effect of clustering; **(3)** justification for using a CRT design; **(4)** type of intervention and whom the intervention was targeting; **(5)** information regarding REC review, including which committee (e.g., local, central REC, etc.) reviewed the ethics application; **(6)** who provided access to the cluster and the role they played (i.e., any gatekeeper information). We defined a "Gatekeeper" as an individual or body that represents the interests of cluster members, clusters, or organizations.[1] Gatekeepers may give permission to enroll the cluster in the trial, and when appropriate, give researchers the permission to approach eligible participants to enroll in the study**; (7)** information about informed consent procedures, including how (if at all) consent to the intervention and data collection was obtained, the timing of participant consent, and what information (if any) was disclosed to participants during the consent procedure; and **(8)** any information about harm-benefit assessment or protection of vulnerable populations.

Appendix 3-5: Reported information about waiver of consent for the four studies that reported a waiver of informed patient consent or research ethic committee exemption.

Intervention and comparison groups	**Outcome**	**Reported information about waiver of consent**
Compared hemodialysis centres that used Clear Guard HD Antimicrobial Barrier Caps versus hemodialysis centres that use standard CVC caps	Positive blood culture rate as an indicator of bloodstream infection rate	*"The informed consent waiver resulted in broad inclusion and ease of conducting the study."*[2]
Compared hemodialysis centres that used Clear Guard antimicrobial barrier caps with hemodialysis centres that used Tego hemodialysis connectors plus Curos disinfecting caps.	Blood culture positivity rate	*"The informed consent waiver was important for conducting the study in a pragmatic manner, adherence to the prescribed intervention, and broad inclusion"*[3]
Compared hemodialysis that used 2% chlorhexidine with 70% alcohol swab sticks for exit-site care and 70% alcohol pads to perform "scrub the hubs" in dialysis-related central venous catheter care procedures with hemodialysis centres that used usual care.	Positive blood cultures for estimating bloodstream infection rates.	*"…during its inception, because no investigational products were used and care processes in the intervention were all within standard clinical practice, this minimal-risk QI [Quality Improvement] initiative was not submitted for institutional review board review.";*[4]
Dialysis facilities randomized to the intervention adopted a default session duration of ≥4.25 hours (255 minutes) for patients initiating maintenance hemodialysis. If the treating nephrologist felt that the ≥4.25-hour duration was not appropriate for an individual patient, shorter treatments could be prescribed to achieve session durations as close to 4.25 hours as possible. Dialysis	Death	*"The trial was conducted under a waiver of the requirement for informed consent on the basis of criteria specified in the Common Rule [45 CFR Part 46.116(c)]. Patients in both intervention and usual care facilities were given written information about the trial that included the facility's randomized assignment. Patients were provided with telephone access to the research teams at the dialysis provider organizations to obtain additional information and/or to opt out of having their clinical data included in the trial dataset. Patients meeting the eligibility criteria were enrolled in the*

facilities randomized to usual care had no trial-driven approach to session duration.		*trial unless they opted out of data sharing."*[5]

References for chapter 3 appendix

1. Gallo A, Weijer C, White A, et al. What is the role and authority of gatekeepers in cluster randomized trials in health research? *Trials*. 2012;13(1):116. doi:10.1186/1745-6215-13-116

2. Hymes JL, Mooney A, Van Zandt C, Lynch L, Ziebol R, Killion D. Dialysis Catheter–Related Bloodstream Infections: A Cluster-Randomized Trial of the ClearGuard HD Antimicrobial Barrier Cap. *Am J Kidney Dis*. 2017;69(2):220-227. doi:10.1053/j.ajkd.2016.09.014

3. Brunelli SM, Van Wyck DB, Njord L, Ziebol RJ, Lynch LE, Killion DP. Cluster-Randomized Trial of Devices to Prevent Catheter-Related Bloodstream Infection. *J Am Soc Nephrol*. 2018;29(4):1336-1343. doi:10.1681/ASN.2017080870

4. Rosenblum A, Wang W, Ball LK, Latham C, Maddux FW, Lacson E. Hemodialysis catheter care strategies: A cluster-randomized quality improvement initiative. *Am J Kidney Dis*. 2014;63(2):259-267. doi:10.1053/j.ajkd.2013.08.019

5. Dember LM, Lacson E, Brunelli SM, et al. The TiME Trial: A Fully Embedded, Cluster-Randomized, Pragmatic Trial of Hemodialysis Session Duration. *J Am Soc Nephrol*. 2019;30(5):890-903. doi:10.1681/ASN.2018090945

CHAPTER 4: STUDY 3

TITLE: Simple compared to covariate-constrained randomization methods in balancing baseline characteristics: a case study of randomly allocating 72 hemodialysis centers in a cluster trial

AUTHORS: Ahmed A. Al-Jaishi, MSc [1,2,3]; Stephanie N. Dixon, PhD [1,3,4,5]; Eric McArthur, MSc [1,3]; PJ Devereaux, MD, PhD [2]; Lehana Thabane, PhD [2]; Amit X. Garg, MD, PhD [1,2,3,4]

AFFILIATIONS:

[18] Lawson Health Research Institute, London, Ontario, Canada

[19] McMaster University, Department of Health Research Methods, Evidence, and Impact, Hamilton, ON, Canada

[20] ICES, London, Ontario, Canada

[21] Department Medicine, Epidemiology and Biostatistics, Western University, London, ON, Canada

[22] Department of Mathematics and Statistics, University of Guelph, Guelph, ON, Canada

KEYWORDS: cluster randomized trial; covariate-constrained; randomization; balanced allocation, restricted randomization

CITATION: Al-Jaishi AA, Dixon SN, McArthur, E, et al. Simple compared to covariate-constrained randomization methods in balancing baseline characteristics: a case study of randomly allocating 72 hemodialysis centers in a cluster trial. [submitted]

Abstract

Background and aim: Some parallel-group cluster randomized trials use covariate-constrained rather than simple randomization. This is done to increase the chance of balancing the groups on cluster- and patient-level baseline characteristics. This study assessed how well two covariate-constrained randomization methods balanced baseline characteristics compared with simple randomization.

Methods: We conducted a mock three-year cluster randomized trial, with no active intervention, that started April 1st, 2014, and ended March 31st, 2017. We included a total of 11,832 patients from 72 hemodialysis centers (clusters) in Ontario, Canada. We randomly allocated the 72 clusters into two groups in a 1:1 ratio on a single date using individual- and cluster-level data available up to April 1st, 2013. Initially, we generated 1,000 allocation schemes using simple randomization. Then, as an alternative, we performed covariate-constrained randomization based on historical data from these centers. In one analysis, we restricted on a set of 11 individual-level prognostic variables; in the other, we restricted on principal components generated using 29 baseline historical variables.

We created 300,000 different allocations for the covariate-constrained randomizations, and we restricted our analysis to the 30,000 best allocations. We then randomly sampled 1,000 schemes from the 30,000 best allocations. We summarized our results with each randomization approach as the median (25th, 75th percentile) number of balanced

baseline characteristics. There were 156 baseline characteristics, and a variable was balanced when the between-group standardized difference was ≤ 10%.

Results: The three randomization techniques had at least 125 of 156 balanced baseline characteristics in 90% of sampled allocations. The median number of balanced baseline characteristics using simple randomization was 147 (142, 150). The corresponding value for covariate-constrained randomization using 11 prognostic characteristics was 149 (146, 151), while for principal components, the value was 150 (147, 151). The median number of balanced baseline characteristics using the two covariate-constrained randomizations were statistically different from simple randomization (p-value < 0.0001).

Conclusion: In this setting with 72 clusters, constraining the randomization using historical information achieved better balance on baseline characteristics compared with simple randomization; however, the magnitude of benefit was modest.

Introduction

The cluster randomized trial (CRT) study design is useful when the interventions are naturally implemented on groups of individuals.[1,2] In contrast to individually randomized trials, CRTs randomly allocate groups rather than independent individuals. Simple randomization is the most basic and straightforward type of random allocation. Each "randomized unit" is assigned purely by chance. However, suppose the total number of randomized units is small (e.g., fewer than 20 units). In that case, simple randomization may result in a moderate to a high probability of imbalance in baseline characteristics between the trial arms.[3] In two-group, parallel-arm, individual-level trials, some have suggested that including at least 1,000 participants per group is required to provide sufficient protection against the imbalance of baseline characteristics.[4] In the CRT setting, it is often impossible to have such a large number of randomized units. In a systematic review of 300 CRTs, 50% of trials randomly allocated fewer than 21 clusters, and 75% allocated fewer than 52 clusters.[5]

Observing between-group differences in a trial's baseline characteristics complicates the interpretation of observed treatment effects and threatens the trial's internal validity.[6–8] Other randomization techniques may help minimize the risk of imbalance on baseline measured characteristics when using parallel arm CRT designs.[8] These techniques are described as "restricted" or "constrained" and include stratification, matching, minimization, and covariate-constrained randomization. All restricted methods require *a priori* knowledge about participating clusters and the baseline measures used for the restriction process.

Covariate-constrained randomization can provide a better baseline balance than other allocation methods (e.g., simple random allocation, stratification, and minimization).[3,8–10] This manuscript focuses on covariate-constrained randomization, where we constrained the randomization process using two sets of baseline characteristics (either constraining on a set of prognostic variables or principal components.) Principal components are a small set of artificial variables that explain most of the variance about a larger group of variables.

Covariate-constrained randomization limits the potential schemes available for selection among all possible allocations (called the randomization space). This method simultaneously balances several measured cluster- or individual-level characteristics to ensure that the two treatment arms are similar at baseline.[8,9] Briefly, the covariate-constrained randomization process includes **(i)** *a priori* identifying and specifying a limited number of key prognostic cluster- or individual-level variables associated with the outcome that will be used to constrain the randomization process (or a function of baseline characteristics, for example, principal components); **(ii)** when there are 20 or more clusters[7], either enumerating all or generating at least 100,000 allocation schemes; **(iii)** for each allocation scheme, estimating balance on the selected baseline characteristics according to some predefined balance metric (e.g., absolute differences, standardized differences, or another measure[11]); **(iv)** choosing a constrained randomization space containing a subset of allocations that are balanced on the constrained baseline characteristics (e.g., 10% of the best allocations[11–13]); and **(v)**

randomly selecting one allocation scheme from the constrained randomization space that will be used for the trial.

There is a trade-off between the potential for a better balance achieved on the constrained baseline characteristics and the potential concerns with highly restricted randomization.[9,12] These trade-offs can include **(i)** jeopardizing the appearance of impartiality, for example, if pairs of clusters always (or never) appear in the same arm [9,12]; **(ii)** a departure from the nominal Type I error when clusters with correlated outcomes have a very high or very low probability of being included in the same trial arm [9,12]; and **(iii)** a loss in statistical power when variables used in the constrained randomization do not associate with the trial outcome.[9,12] Also, covariate-constrained randomization uses historical data on recruited clusters to capture baseline information on demographics, patients' medical histories, and historical rates of the outcomes.[14–16] However, historical data are usually several months to years old at randomization. In an "open cohort" setting, information available at the randomization date also cannot account for new participants entering the cohort during the trial period. Thus, the balance achieved at the time of randomization with historical information does not guarantee a balance of the baseline characteristics during the trial period. It is important to note that the randomization design (i.e., constrained variables) needs to be considered at the analysis stage.[17–19]

We conducted this study to understand the best practices for randomizing hemodialysis centers into two parallel groups in Ontario, Canada. The lessons learned from this study will help our group make informed decisions about randomization processes for several CRTs that we plan to advance.

Motivating example

In the hemodialysis setting, the CRT is an attractive design when implementing interventions at the dialysis center level.[15,20,21] The CRT design offers logistical and administrative advantages such as simplifying the trial organization when evaluating policy- or cluster-level intervention.[1,22]

Suppose that we wish to undertake a CRT with hemodialysis centers in Ontario, Canada. In this example, we used historical data from administrative data sources to conduct covariate-constrained randomization. The trial period was three years, from April 1st, 2014, to March 31st, 2017. The primary outcome was a composite of time-to-first event for cardiovascular-related death or non-fatal major cardiovascular-related hospitalization (hospital admission for myocardial infarction, stroke, or congestive heart failure).

Objectives

This paper compared randomization methods for a two-arm, parallel-group CRT, with the intent that all individuals within a given randomized center will receive the same intervention. We randomized a moderate number of clusters (i.e., hemodialysis centers) using either simple randomization or covariate-constrained randomization with pre-trial historical records (called the *Population for Randomization*). We performed the randomization on a single date and allowed patients to enter the cohort throughout the study period. We compared simple randomization to covariate-constrained randomization on balance achieved on a set of baseline characteristics during a three-year trial period (called the *Trial Population*). We constrained either on prognostic variables or principal components.

Our secondary aim was to assess whether, in the absence of any intervention, the allocation schemes selected through the constrained randomization process preserved **(i)** a null treatment effect on the primary outcome and **(ii)** a 5% nominal Type I error rate.

Methods

Design and setting

We used a CRT design of outpatient hemodialysis centers in Ontario, Canada, that cared for a minimum of 15 patients. In 2013, Ontario had approximately 13.5 million residents with universal access to healthcare and physician services.[23] In the same period, Ontario had 26 regional dialysis programs that oversaw over 100 hemodialysis centers caring for about 8,000 in-center patients in the outpatient setting.[24]

Data sources

We ascertained center- and patient-level characteristics using records from linked healthcare databases in Ontario, Canada (**Appendix 4-1**).[25–38] These datasets were linked using unique encoded identifiers and analyzed at ICES.[39]

Patients

We included two populations of patients, the *Population for Randomization* and the *Trial Population*. The *Population for Randomization* included patients who were actively receiving in-center hemodialysis on April 1st, 2013. The *Trial Population* included an open cohort of patients who received in-center hemodialysis on April 1st, 2014 or began receiving in-center hemodialysis during the trial period.

Baseline characteristics

We identified two cluster- and 86 individual-level (total 88) baseline characteristics to describe each cohort (**Appendix 4-2**); the cluster-level characteristics included the center size and historical rate for the primary outcome. There were 23 continuous, 58 binary,

and 14 categorical baseline characteristics. Nine continuous baseline characteristics were also featured as categorical variables. We created a new binary (or "dummy") variable to indicate each level of a category's presence or absence. In total, we evaluated 156 continuous or binary candidate baseline characteristics.

Randomization process

Sequence generation:

We randomly allocated the 72 hemodialysis centers into two groups in a 1:1 ratio on a single date. Initially, we generated 1,000 random allocation schemes using simple (unconstrained) randomization that required no information on baseline characteristics. This number of random allocations produced an estimate within 0.5% accuracy of the true hazard ratio of 1.00 with a 5% significance level and a standard deviation of 0.08.[40] Then, as an alternative, we performed the covariate-constrained randomization using pre-trial historical records, which ended April 1st, 2013 (see next section). Using PROC PLAN in SAS version 9.4 (SAS Institute Inc., NC Cary), we generated 300,000 unique allocation schemes of the 72 centers (**Appendix 4-3**). Greene (2017) suggested performing at least 100,000 allocations when there are at least 20 clusters; with our computational capacity, we enumerated 300,000 allocations.

Covariate-constrained randomization:

We performed the covariate-constrained randomization in the following series of steps using baseline characteristics of the *Population for Randomization*.[6,8,9,41]

Step 1: Randomly selected 300,000 allocation schemes from the 4.43×10^{20} possible allocation schemes.

Step 2: For each of the 300,000 allocation schemes, we restricted the randomization space using one of two constraining criteria.[8]

i. We constrained the allocation on a set of 11 baseline characteristics deemed prognostic for the outcome, based on prior literature or clinical experience (**Appendix 4-4a**).

ii. We constrained the allocation on principal components. A principal component analysis is a dimensionality reduction technique whereby a dataset with many variables is transformed into a smaller set of artificial variables (called principal components). These principal components ideally retain some or most of the meaningful properties of the original set of variables. We used the principal components to account for some of the variation in the observed data and as criterion variables in our constrained randomization process (**Appendix 4-4b**).

We compared baseline differences between the two arms using standardized differences,[42,43] which describes the differences between group means or proportions relative to the pooled standard deviation.

Step 3: For each allocation scheme from the *Population for Randomization*, we counted the number of constrained variables with a standardized difference greater than 10% and calculated the sum of the constrained variables' standardized differences.[42,44] We added a penalty of ten units to the sum of standardized differences for each imbalanced constrained variable. We imposed this penalty to favor allocation schemes that had the least number of imbalanced constrained baseline characteristics. For example, if the sum

of standardized differences was two and three constrained variables were imbalanced, the penalized sum of standardized differences would be 32.

From the 300,000 randomization schemes, we constrained the randomization space to the 30,000 best allocation schemes, based on the smallest sum of the penalized standardized differences.[11–13] From the 30,000 best allocations, we randomly sampled 1,000 allocations to reduce the computational time for analysis.[11,12]

Statistical Analysis

For the 1,000 sampled schemes, we **(i)** estimated the percentage of times each center was allocated to each arm, **(ii)** estimated the percentage of times each combination of center pairs appeared in the same group,[41] and **(iii)** calculated the standardized difference of all 156 baseline characteristics for the *Trial Population*. We then estimated the percentage of time each of the 156 baseline characteristics was balanced among the 1,000 sampled randomization schemes, **(iv)** calculated the median (25^{th}, 75^{th} percentile) number of baseline characteristics balanced for the *Trial Population*, **(v)** we used the Kruskal-Wallis H Test to evaluate the null hypothesis that the population medians from the three randomization techniques are equal,[45] and finally **(vi)** estimated the unadjusted and adjusted hazard ratio between the randomized arms, for the time-to-first event of the composite outcome of cardiovascular-related death or a non-fatal cardiovascular-related hospitalization (see definition of outcome in **Appendix 4-5;** this is a primary outcome for future trials that is highly relevant to patients and their providers). We estimated the hazard ratio using a generalized-estimating-equations extension for the Cox proportional hazard model, with an exchangeable covariance matrix to account for within-center clustering.[22,46] For each of the 1,000 sampled randomization schemes, the models were

fitted to patient-level data from the *Trial Population*. We conducted unadjusted and another analysis adjusting for the randomization design (i.e., adjusted analyses using the constrained baseline characteristics by adding these variables into the model).

We stopped following patients on March 31st, 2017, or earlier if they died. We summarized the hazard ratios as the mean with the 2.5th and 97.5th percentiles, which corresponded to the hazard ratio estimate with a 95% confidence interval.[47] We expected to observe no between-group differences in the event rate of our primary outcome approximately 95% of the time (i.e., a nominal Type I error of 5%). The use of 1,000 randomizations allowed us to detect a Type I error between 3.6% and 6.4% as not significantly different than 5%; we used a standard test based on the normal approximation to the binomial distribution as described by Rosner (1995).[48]

Results

Characteristics of cohorts

The *Population for Randomization* (n=5,812) included all patients receiving in-center hemodialysis on April 1st, 2013. The *Trial Population* (n=11,832) included patients receiving hemodialysis on April 1st, 2014 (n=5,410) and patients who started in-center hemodialysis during the three-year trial period (n=6,412). The *Trial Population* included 4,415 patients (37%) that were also in the *Population for Randomization*. The median (25th, 75th percentile) number of patients in each center for the *Population for Randomization* was 61 (28, 105) and for the *Trial Population* was 131 (55, 227).

The *Population for Randomization* and the *Trial Population* differed on several baseline characteristics (**Table 4-1** and **Appendix 4-2**). However, the differences were mostly attributed to the inherent differences between prevalent and new patients starting

hemodialysis (e.g., length of time on dialysis, number of dialysis sessions in the prior year, healthcare services utilization, and general practitioner visits the preceding year.)

Results from the principal component analysis

We subjected 29 of the 156 baseline characteristics to principal component analysis (**Appendix 4-4b**). We retained ten principal components that accounted for 61% of the 29 baseline characteristics variance. **Appendix 4-6** and **Appendix 4-7** show results from the principal component analysis.

Randomization of hemodialysis centers

Each of the 72 participating centers had an approximately 50% chance of being randomized to either trial arm (see **Appendix 4-8** for the process and hardware specification). We observed that some pairs of centers were allocated to different trial arms at a different probability than we might have expected if we had used simple randomization (**Figure 4-1 A** to **C**). These pairs of centers tended to be large and generally had over 225 patients.

Balance of baseline characteristics

Table 4-2 shows the balance for a select set of baseline characteristics by the method of constraining. In the *Trial Population*, both sets of constrained variables were generally well balanced between the two arms, regardless of the randomization method. The constrained randomizations generally provided a slightly better balance. **Appendix 4-9** shows the percentage of times each of the 156 baseline characteristics (from the *Trial Population*) were balanced across the 1,000 randomization schemes for the three allocation methods. **Table 4-3** shows a summary of the number of baseline characteristics balanced across randomization schemes. The *Trial Population* had at least 125 of 156

(80%) balanced baseline characteristics in 90% of simple randomization schemes. By comparison, the constrained methods always had slightly more balanced baseline characteristics (at least 85% of the 156 baseline characteristics were balanced in 90% of sampled allocations). **Table 4-3** also shows the median (25^{th}, 75^{th} percentile) number of balanced baseline characteristics across the 1,000 sampled randomization schemes by allocation method. The distributions for the number of balanced baseline characteristics were statistically different for the three allocation methods (p-value < 0.0001).

Cardiovascular-related death or major cardiovascular-related hospitalization

We followed patients for an average of 1.7 years, and there were 2,260 events over the three-year follow-up. The event rate of the primary outcome was 11 per 100 person-years. **Table 4-4** shows the unadjusted and adjusted analyses for simple and covariate-constrained randomization methods. Across the 1,000 simple randomization schemes for the *Trial Population*, the mean unadjusted hazard ratio (2.5^{th}, 97.5^{th} percentile) was 1.01 (0.87, 1.16), and 5.9% of allocation schemes produced a confidence interval for the hazard ratio that did not contain the null value of 1.00. Compared to simple randomizations, constrained randomizations had similar unadjusted hazard ratios, with slightly narrower 95% confidence intervals. The Type I error tended to be somewhat lower than the nominal level for some constrained methods than the unconstrained approach.

Adjusted analyses for the constrained methods produced narrower confidence intervals than the unadjusted analyses. However, the Type I error was within the acceptable range only when models adjusted for the ten principal components; the Type I error was outside the expected range for all other adjusted analyses. We also explored the results when

adjusting for aggregate-level baseline characteristics as used in the randomization, which aligned with the results when we adjusted for individual-level variables (results not shown).

Discussion

This empirical study presented an example of using historical data to conduct covariate-constrained randomization that balances baseline characteristics for a parallel, two-group, cluster randomized trial. Compared to simple randomization, we showed that constraining the random allocation using a historical cohort (i.e., a *Population for Randomization*) provides a better balance on baseline characteristics. However, we randomized a moderate number of clusters, and the magnitude of benefit was modest. Our results also suggested that model-based adjustment for the constrained variables produced treatment effects with the nominal Type I error that is narrower than those produced with simple randomization. However, researchers should constrain prognostic variables and adjust for the constrained variables at the analysis stage; otherwise, the Type I error might deviate from the nominal level described in previous reports.[1,9,11,12,17,18]

In a review of 300 CRTs published between 2000 and 2008, Wright et al.[49] found significant discrepancies between the restricted randomization used at the design stage and covariate adjustments at the analysis stage. Wright et al.[49] identified 174 CRTs that used design-based restricted randomization. However, only 30 (17.2%) of these studies reported an adjusted analysis for all the constrained variables.

From an analysis perspective, the analysis should account for the design that uses covariate-constrained randomization.[1,9,11,12] Otherwise, the Type I error may deviate from the nominal level because clusters with highly correlated outcomes get separated into different treatment arms (as observed in **Figure 4-1 B** and **C**).[9] Splitting correlated clusters into different treatment arms tends to **(i)** lower the Type I error below the nominal level (in the unadjusted analyses), and **(ii)** decrease power slightly, although we might still expect substantial gains in power due to the assurance of balance on prognostic baseline characteristics.[9,50] Several analytical techniques can test for treatment effects and take into account the study design. These methods include mixed-effects models, bias-corrected generalized estimating equations, and randomization-based permutation tests.

In our motivating example, we used an analysis for the time-to-first event. In contrast, previous studies have focused their investigations primarily on continuous or binary outcomes.[1,9,11,12] Our results add to this literature showing a generalized estimating equation-based approach can yield results that maintain the nominal Type I error after adjusting for the covariate-constrained design. However,

This study has some limitations. First, the *Trial Population* included a large percentage of patients (37%) included in the *Population for Randomization*. Thus, our results may not apply to other designs, for example, CRTs where the *Population for Randomization* and the *Trial Population* are the same or settings where cluster- and patient-level profiles change rapidly over time. Second, some historical data may lag by more than one year; thus, these results may not be applicable for *Populations at Randomization* less than or

more than a year old. Third, our example cohort randomized a moderately large number of clusters; a previous review reported that 75% of published CRTs randomized fewer than 52 clusters. Covariate-constrained randomization may provide a better baseline balance compared to simple randomization when there are fewer clusters. Finally, our secondary objective does not constitute a formal test of the Type I error. Computer simulations with more control over the generated data would be better suited. As such, the reader should interpret these results cautiously.

Conclusions and guidance for future trials

Although covariate-constrained randomization approaches used in this setting had modest improvement for balance, there may be substantial improvements in statistical power.[12] We propose the following recommendations (**Box 4-1**) for CRTs based on the empirical comparisons presented in this paper and other published literature. It is worth noting that these recommendations are based on a single setting, and while we anticipate similar findings in different contexts, a more formal statistical comparison would be beneficial.

1. Identify prognostic variables *a priori* using background literature, historical data, or previous trials. Previous work for individual-level randomized controlled trials showed increases in statistical power when analyses prespecified covariates strongly associated with the outcome. The adjusted covariates had a more considerable impact on statistical power when the prevalence was moderate to high (between 10% and 50%).[19,51–53]
2. Researchers should consider generating all (or at least 1,000) simple randomizations to identify baseline characteristics that are always or almost always balanced

(e.g.,>95% of the time) between treatment arms. There would be no need to include these baseline characteristics in the constraining process; however, researchers can have these variables in the model-based adjustment to improve the estimates' precision. Importantly, all prognostic variables should be specified *a priori*.[53]

3. Carefully consider the number of baseline characteristics used during the constraining process. Evidence from our study (and previous simulation studies) showed that over-constraining could result in clusters with highly correlated outcomes having a lower probability of being included in the same trial arm. Thus, over-constraining can lead to a Type I error below the nominal level and slightly decrease power.[9,50]
4. Researchers can use a dimensionality-reduction method (e.g., principal component analysis) to reduce many dimensions of the prognostic variables to several criterion variables used in the constrained randomization process.[54] As above, all analyses should account for the dimensionality-reduction criterion at the analytic stage.
5. While the constraining process utilizes aggregate patient-level and cluster-level data, investigators should consider missingness when constraining the randomization on these variables. When appropriate, variables with missing data should be imputed before aggregating the variable at the cluster level.[55]
6. Researchers should consider constraining the randomization space to the 10% best allocations. Furthermore, researchers should enumerate all possible randomization schemes when fewer than 20 clusters or at least 100,000 randomization schemes.[12]

Declarations

Ethics approval and consent to participate

We had the authorization to use data in this project under section 45 of Ontario's Personal Health Information Protection Act, which does not require a research ethics board review. The dataset from this study is held securely in coded form at ICES. ICES is an independent, non-profit research institute whose legal status under Ontario's health information privacy law allows it to collect and analyze healthcare and demographic data, without consent, for health system evaluation and improvement.

Consent for publication

Not applicable.

Availability of data and materials

While data sharing agreements prohibit ICES from making our study dataset publicly available, access may be granted to those who meet prespecified criteria for confidential access, available at www.ices.on.ca/das. The full dataset creation plan and underlying analytic code can be requested from the authors on the understanding that the computer programs may rely upon coding templates or macros that are unique to ICES and are therefore either inaccessible or may require modification.

Competing interests

Authors have nothing to disclose.

Funding

Ahmed Al-Jaishi was supported by the Allied Health Doctoral Fellowship from the Kidney Foundation of Canada, CIHR Doctoral Award, and McMaster University Michael DeGroote Scholarship. Stephanie Dixon’s research is supported through a SPOR Innovative Clinical Trial Multi-Year Grant (Grant number: MYG-151209) from CIHR.

Amit Garg was supported by the Dr. Adam Linton Chair in Kidney Health Analytics and a Clinician Investigator Award from the CIHR.

We received funding for this study from partnering organizations, including Lawson Health Research Institute, Ontario Renal Network, Dialysis Clinic Inc., Heart and Stroke Foundation of Canada, and Canadian Institutes of Health Research (CIHR) Innovative Clinical Trials Initiative (Grant number: MYG-151209). Funding was also provided by the Ontario Strategy for Patient-Oriented Research SUPPORT Unit, which is supported by the Canadian Institutes of Health Research and the Province of Ontario. Funding bodies had no role in the study's design, analysis, interpretation of data, or writing the manuscript.

Authors' contributions

AAA and AXG conceived and led the study design. SND, EM, PJD, and LT contributed to the study design. AA was responsible for data management and analysis. AA drafted the manuscript. All authors contributed to manuscript revision and approved the final manuscript. AA is the guarantor.

Acknowledgments

This study was supported by ICES, which is funded by an annual grant from the Ontario Ministry of Health and Long-Term Care (MOHLTC). Parts of this material are based on data and information compiled and provided by CIHI, Cancer Care Ontario (CCO), MOHLTC, and Ontario Service Reports. We thank IMS Brogan Inc. for use of their Drug Information Database. The analyses, conclusions, opinions, and statements expressed herein are solely those of the authors and do not reflect those of the data sources; no endorsement is intended or should be inferred.

Tables and Figures

Box 4-1: Guidance for conducting covariate-constrained randomization.

1. Identify prognostic baseline characteristics *a priori* using background literature, historical data, or previous trials.
2. Generate all (or at least 1,000) simple randomizations to identify baseline characteristics that are always balanced between treatment arms (e.g.,≥95% of the time).
3. Carefully consider the number of variables added to the constraining process or consider using a dimensionality-reduction method for many variables (e.g., principal component analysis).
4. Consider the amount of missingness of constrained baseline characteristics before randomization.
5. Enumerate all possible allocation schemes when there are fewer than 20 clusters or at least 100,000 allocations otherwise.

Table 4-1: Select baseline characteristics. The Population for Randomization included patients on hemodialysis as of April 1, 2013. The Trial Population included an open cohort of patients receiving in-center hemodialysis on April 1, 2014, or began receiving in-center hemodialysis during the trial period between April 1, 2014, and March 31, 2017.

Baseline characteristic	**VALUE**	**Population for Randomization**	**Trial Population**
Centers	Number of centers (n Patients)	72 (n=5,812)	72 (n=11,832)
Center Size [1]	Mean ± Standard deviation	81 (69)	164 (137)
The composite outcome of CV-related death or major CV-related hospitalization [2]	Historic rate per 100 person-years (cluster standard deviation)	10 (3.7)	11 (3.3)
Age (years)	Mean ± Standard deviation	67 (15)	66 (15)
Sex	Male	3373 (58%)	7069 (60%)
Living in a rural area	Yes	359 (6%)	809 (7%)
Etiology for End-stage kidney disease	Diabetes	2194 (38%)	4472 (38%)
	Glomerulonephritis/autoimmune diseases	882 (15%)	1575 (13%)
	Drug-induced nephropathy	83 (1%)	159 (1%)
	Polycystic kidney disease	229 (4%)	426 (4%)
	Renal vascular disease	1115 (19%)	1852 (16%)
	Other	738 (13%)	2221 (19%)
	Unknown	571 (10%)	1127 (10%)
Race	Asian	475 (8%)	854 (7%)
	Black	562 (10%)	1022 (9%)
	White	3698 (64%)	7598 (64%)
	Other	1038 (18%)	2173 (18%)
	Unknown	39 (1%)	185 (2%)
First dialysis modality	Home hemodialysis	33 (1%)	84 (1%)
	In-center hemodialysis	5215 (90%)	10529 (89%)
	Peritoneal dialysis	564 (10%)	1219 (10%)
First type of vascular access	Arteriovenous graft	103 (2%)	198 (2%)
	Arteriovenous fistula	1044 (18%)	1999 (17%)
	Central venous catheter	3927 (68%)	8157 (69%)
	Peritoneal catheter	456 (8%)	1092 (9%)
	Unknown	282 (5%)	386 (3%)
	Arteriovenous graft	223 (4%)	376 (3%)

Last vascular access used before the index date	Arteriovenous fistula	2159 (37%)	3461 (29%)
	Central venous catheter	3376 (58%)	7544 (64%)
Patient in the Ontario Drug Benefit in the six months before the index date	Yes [4]	4494 (86%)	10196 (86%)
Number of unique hypertensive prescriptions in the six months before the index date	Mean ± Standard deviation	2 (2)	2 (2)
Prescribed hypertensive drugs [3]	Angiotensin-converting enzyme (ACE) inhibitors	1157 (26%)	2199 (22%)
	Angiotensin II Receptor Blocker	1258 (28%)	2413 (24%)
	Beta-Blockers	2649 (59%)	5604 (55%)
	Calcium Channel Blocker	2392 (53%)	5453 (53%)
	Diuretics	1612 (36%)	4242 (42%)
Prior CABG/PCI	Yes	1234 (21%)	2612 (22%)
Coronary Artery Disease (with angina)	Yes	3541 (61%)	6861 (58%)
Heart failure	Yes	2862 (49%)	6177 (52%)
Diabetes mellitus	Yes	3402 (59%)	7244 (61%)
Depression	Yes	1528 (26%)	3161 (27%)
Ischemic Stroke	Yes	261 (4%)	551 (5%)
Lower extremity amputation	Yes	298 (5%)	554 (5%)
Lung disease (COPD)	Yes	2276 (39%)	4666 (39%)
Myocardial infarction	Yes	1454 (25%)	2909 (25%)
Major Cancer	Yes	841 (14%)	1803 (15%)
Peripheral vascular disease	Yes	1699 (29%)	3055 (26%)
Modified Charlson comorbidity score [5]	Mean ± Standard deviation	4 (2)	4 (2)
	2	1931 (33%)	3631 (31%)
	3	519 (9%)	1028 (9%)
	4	1382 (24%)	2883 (24%)
	5+	1980 (34%)	4290 (36%)
Having a kidney transplant before the index date	Yes	71 (1%)	148 (1%)
Number of days spent in long-term	Mean ± Standard deviation	37 (164)	25 (133)

care in the year before the index date			
Number of days spent in the hospital in the year before the index date	Mean ± Standard deviation	11 (28)	17 (31)
Number of hospital admissions in the year before the index date	Mean ± Standard deviation	1 (1)	1 (1)
Time since starting dialysis (days)	Mean ± Standard deviation	1847 (1836)	1327 (1782)

[1] Population for Randomization included patients that were on hemodialysis as of April 1, 2013 index date. The Trial Population included patients on hemodialysis as of April 1, 2014, and any patient who started in-center hemodialysis at one of the 72 participating centers during the three-year trial period. Follow-up ended March 31, 2017. The index date was the first date patients entered the respective cohort.

[2] The composite outcome of cardiovascular-related death or hospitalization for myocardial infarction, ischemic stroke, congestive heart failure.

[3] Percentages presented only for patients eligible to receive the Ontario Drug Benefit (ODB) plan in the six months before the index date (i.e., 4494 for Randomization Cohort and 10196 for the Trial cohort). CABG/PCI = Coronary artery bypass grafting / percutaneous coronary intervention.

[4] Ontario residents are eligible for the Ontario Drug Benefit program include people 65 years or older, on social assistance, residing in homes for special care and long-term care homes, people receiving professional home care services, and registrants in the Trillium Drug Program.

[5] This is an adapted version of the Charlson comorbidity index explicitly created for use in patients with ESRD. This version has a modified weighting scheme specific to dialysis patients.

Table 4-2: The percentage of times each of the baseline characteristics was balanced across each of the 1000 randomizations schemes in the Trial Population.

Baseline characteristic	Value	Constrained randomization method		
		Unrestricted / Simple	Prognostic baseline characteristics	Principal components
Center Size	Mean ± Standard deviation	32.9%	41.8%	38.7%
Composite outcome of CV-related death and major CV-related hospitalization	Rate (per 100 person-year)	32.5%	36.2%	33.5%
Age (years)	Mean ± Standard deviation	95.3%	99.8%	99.2%
	< 65	97.8%	99.7%	99.9%
	65 to 74	100.0%	100.0%	100.0%
	75 to 84	100.0%	100.0%	100.0%
	85 to 105	99.5%	100.0%	99.9%
Sex	Male	100.0%	100.0%	100.0%
Living in a rural area	Yes	63.0%	84.2%	65.8%
Etiology for End-stage kidney disease	Diabetes	93.0%	94.5%	95.0%
	Glomerulonephritis/auto immune diseases	96.3%	100.0%	99.5%
	Drug-induced nephropathy	100.0%	99.9%	100.0%
	Polycystic kidney disease	100.0%	100.0%	100.0%
	Renal vascular disease	97.5%	97.6%	96.7%
	Other	88.3%	91.9%	91.6%
Race	Asian	75.0%	81.3%	88.1%
	Black	73.4%	95.9%	91.9%
	White	45.6%	64.0%	90.2%
	Other	56.6%	65.7%	77.5%
	Unknown	93.2%	93.7%	93.6%
First dialysis modality	Home hemodialysis	100.0%	99.8%	99.9%
	In-center hemodialysis	97.8%	98.6%	99.9%
	Peritoneal dialysis	97.4%	98.7%	99.8%
First vascular access used at dialysis start	AV Graft	99.9%	100.0%	100.0%
	Fistula	98.9%	99.1%	99.4%
	Catheter	93.5%	96.2%	99.4%
	PD Catheter	98.8%	99.0%	100.0%
	Unknown	92.4%	93.8%	94.3%
Most recent vascular access before the index date	AV Graft	98.7%	99.8%	98.9%
	Fistula	91.9%	94.8%	97.7%

	Catheter	89.9%	94.0%	97.4%
Patients 65+ years in ODB in the 6 months before the index date	Yes	97.5%	99.3%	99.4%
Unique hypertensive drugs six months before the index date	Mean ± Standard deviation	97.1%	99.9%	99.5%
Prescribed hypertensive drugs	Angiotensin-converting enzyme (ACE) inhibitors	99.4%	99.3%	99.5%
	Angiotensin II Receptor Blocker	90.7%	96.1%	96.9%
	Beta-Blockers	99.7%	100.0%	99.9%
	Calcium Channel Blocker	98.1%	100.0%	99.6%
	Diuretics	91.9%	97.0%	95.6%
CABG/PCI	Yes	99.4%	99.5%	100.0%
Heart failure	Yes	96.8%	100.0%	99.8%
Diabetes mellitus	Yes	99.0%	100.0%	100.0%
Ischemic Stroke	Yes	100.0%	100.0%	100.0%
Lower extremity amputation	Yes	99.9%	100.0%	100.0%
Lung disease (COPD)	Yes	99.0%	99.6%	100.0%
Myocardial infarction	Yes	99.2%	100.0%	100.0%
Major Cancer	Yes	100.0%	100.0%	100.0%
Peripheral vascular disease	Yes	90.7%	97.2%	91.4%
Modified Charlson comorbidity Score	Mean ± Standard deviation	96.8%	99.9%	100.0%
	2	97.7%	100.0%	100.0%
	3	100.0%	100.0%	100.0%
	4	100.0%	100.0%	100.0%
	5+	98.9%	100.0%	100.0%
Having a kidney transplant before the index date.	Yes	100.0%	100.0%	100.0%
Number of hospital admissions in the year before the index date	Mean ± Standard deviation	93.9%	98.4%	98.4%
	0	78.4%	76.4%	81.1%
	1 to 3	99.5%	99.6%	99.9%
	4 to 6	99.4%	99.6%	99.5%
	7 to 9	100.0%	99.9%	100.0%
	10+	92.1%	92.0%	94.6%
Long term care facility utilization in the year before the index date	Yes	81.3%	86.6%	86.1%
Time since the first date on dialysis (days)	Mean ± Standard deviation	88.1%	94.0%	94.4%

Table 4-3: Summary of the balanced baseline characteristics for the Trial Population.

Criteria	Constrained randomization method			
	Unconstrained / Simple	**Prognostic baseline characteristics**	**Principal components**	**P-value****
11 prognostic characteristics ++				
Number of constrained baseline characteristics that were balanced in all 1000 (100%) sampled allocations	0 of 11 (0%) ¥	2 of 11 (18%)	2 of 11 (18%)	<0.0001
Number of constrained baseline characteristics that were balanced in at least 950 (95%) sampled allocations	6 of 11 (55%)	10 of 11 (91%)	7 of 11 (64%)	
Number of constrained baseline characteristics that were balanced in at least 900 (90%) sampled allocations	8 of 11 (73%)	10 of 11 (91%)	9 of 11 (82%)	
Median (25^{th}, 75^{th} percentile) number of baseline characteristics that were balanced across the 1000 selected randomization schemes	10 (9, 11) *	11 (10, 11)	10 (10, 11)	
29 baseline characteristics used in the principal component analysis**				
Number of constrained baseline characteristics that were balanced in all 1000 (100%) sampled allocations	8 of 29 (28%)	12 of 29 (41%)	12 of 29 (41%)	<0.0001
Number of constrained baseline characteristics that were balanced in at least 950 (95%) sampled allocations	19 of 29 (66%)	23 of 29 (79%)	25 of 29 (86%)	
Number of constrained baseline characteristics that were balanced in at least 900 (90%) sampled allocations	24 of 29 (83%)	25 of 29 (86%)	26 of 29 (90%)	

Median (25th, 75th percentile) number of baseline characteristics that were balanced across the 1000 selected randomization schemes	27 (26, 28)	28 (27, 28)	28 (28, 29)	
All 156 available baseline characteristics				
Number of constrained baseline characteristics that were balanced in all 1000 (100%) sampled allocations	41 of 156 (26%)	46 of 156 (28%)	55 of 156 (35%)	<0.0001
Number of constrained baseline characteristics that were balanced in at least 950 (95%) sampled allocations	104 of 156 (67%)	115 of 156 (74%)	118 of 156 (76%)	
Number of constrained baseline characteristics that were balanced in at least 900 (90%) sampled allocations	125 of 156 (80%)	132 of 156 (85%)	134 of 156 (86%)	
Median (25th, 75th percentile) number of baseline characteristics that were balanced across the 1000 selected randomization schemes	147 (142, 150)	149 (146, 151)	150 (147, 151)	

The Trial Population included patients on hemodialysis as of April 1, 2014, and new patients who started in-center hemodialysis during the three-year follow-up. We conducted simple randomization without any restrictions.

¥ For example, for simple randomization, 2 of the 11 chosen prognostic baseline characteristics were always balanced across 1000 randomly sampled allocation schemes.

*For example, for simple randomization, 500 of 1000 allocation schemes had at least ten balanced baseline characteristics out of the 11 prognostic baseline characteristics. As such, there is a 50% probability that a randomly selected allocation will have at least 10 of the 11 prognostic baseline characteristics balanced and a 75% probability that at least 9 of the 11 prognostic baseline characteristics will be balanced.

** We used the Kruskal-Wallis H Test to determine whether the three randomization techniques' median number of balanced characteristics was the same.

++ Prognostic baseline characteristics: Constraining on a set of baseline characteristics that thought to be important *a priori* and included the following patient-level information: age at index date, living in a rural area, Black race, Modified Charlson comorbidity index, number of hospital visits in the previous 12 months, number of unique drugs the patient was prescribed in the six months before the index date, as well as history in the last five years of diagnosis for peripheral vascular disease, congestive heart failure, coronary artery disease, myocardial infarction, number of nephrology consults in the previous 12 months before the index date.

** Results are shown for the 29 baseline characteristics included in the principal component analysis. We did not include any cluster-level baseline characteristics in the constraining process.

Table 4-4: Mean hazard ratio (2.5^{th}, 97.5^{th} percentile) for the composite outcome during a 3-year follow-up of patients on in-center hemodialysis.

Baseline characteristics adjusted in the analysis	**Mean HR (2.5^{th}, 97.5^{th} percentile)**	**Width of CI Ɪ**	**Type 1 error***
Unadjusted analyses			
Simple (i.e., unconstrained) randomization	1.01 (0.87, 1.16)	0.280	5.9% Ⱥ
Constrained on a minimal set of baseline characteristics Ƴ	1.00 (0.89, 1.12)	0.233	3.2%
Constrained on a minimal set of baseline characteristics Ƴ and historical rate of the primary outcome	1.00 (0.88, 1.13)	0.250	4.4% Ⱥ
Constrained on a minimal set of baseline characteristics Ƴ and cluster size at time of randomization	1.00 (0.88, 1.14)	0.260	5.2% Ⱥ
Constrained on a minimal set of baseline characteristics Ƴ, historical rate of the primary outcome, and cluster size at time of randomization	1.00 (0.88, 1.13)	0.247	4.5% Ⱥ
Constrained on 10 principal components	1.01 (0.89, 1.12)	0.234	3.3%
Constrained on 10 principal components and historic rate of primary outcome	1.00 (0.88, 1.14)	0.261	5.2% Ⱥ
Constrained on 10 principal components and cluster size at time of randomization	1.00 (0.87, 1.14)	0.264	4.1% Ⱥ
Constrained on ten principal components, the historical rate of the primary outcome, and cluster size at time of randomization	1.00 (0.89, 1.13)	0.239	3.1%
Adjusted for constrained baseline characteristics**			
Constrained on a minimal set of baseline characteristics Ƴ	1.00 (0.89, 1.12)	0.232	8.6%
Constrained on a minimal set of baseline characteristics Ƴ and historical rate of the primary outcome	1.00 (0.89, 1.12)	0.223	8.3%
Constrained on a minimal set of baseline characteristics Ƴ and cluster size at time of randomization	1.00 (0.89, 1.11)	0.221	9.8%
Constrained on a minimal set of baseline characteristics Ƴ, historical rate of the primary outcome, and cluster size at time of randomization	1.00 (0.90, 1.11)	0.216	9.6%
Constrained on 10 principal components	1.00 (0.90, 1.11)	0.203	5.2% Ⱥ
Constrained on 10 principal components and historic rate of primary outcome	1.00 (0.90, 1.11)	0.201	6.0% Ⱥ

Constrained on 10 principal components and cluster size at time of randomization	1.00 (0.90, 1.11)	0.203	6.3% ~~A~~
Constrained on ten principal components, the historical rate of the primary outcome, and cluster size at time of randomization	1.00 (0.91, 1.11)	0.201	6.4% ~~A~~

All randomization methods had 1000 randomization schemes. The cohort included patients on dialysis as of April 1, 2014, and any patient who started in-center hemodialysis at one of the 72 participating centers during the three-year follow-up.

HR= Hazard ratio; Width of CI = width of confidence interval (i.e., upper minus lower confidence limit).

X The confidence interval's width may not be equal to the difference between the lower and upper confidence limits because of rounding.

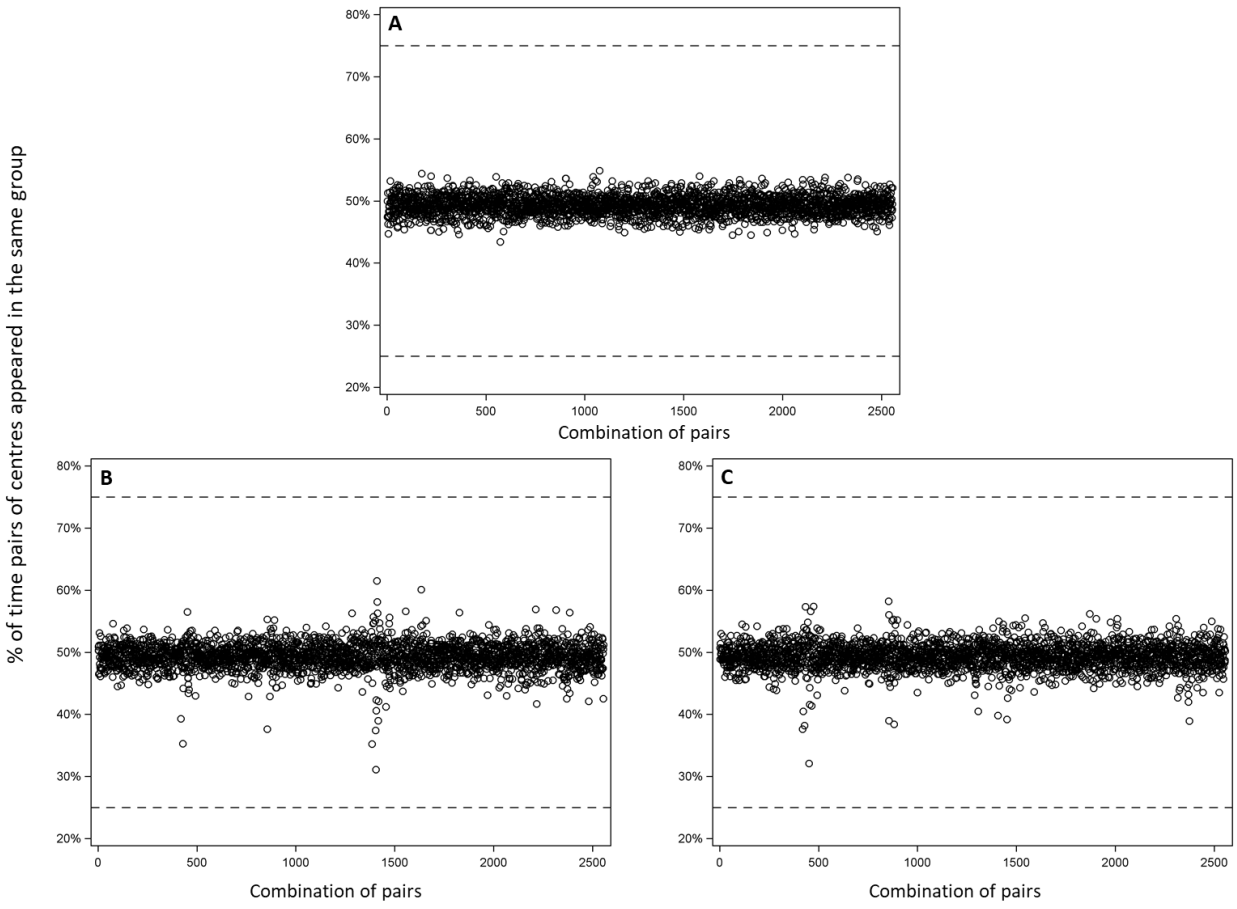
ϒ Included patient-level information: age, living in a rural area, Black race, Modified Charlson comorbidity index, number of hospital visits in the previous 12 months, number of unique drugs the patient was prescribed in the six months before the index date, as well as history in the last five years of diagnosis for peripheral vascular disease, congestive heart failure, coronary artery disease, myocardial infarction, number of nephrology consults in the previous 12 months before the index date.

* Type 1 error in the various constrained scenarios. Note: The nominal Type 1 error is 5%. The observed Type 1 error was within an "acceptable range" if it fell within the 95% confidence interval of the nominal value (i.e., between 3.6% and 6.4%).

**** Adjusted analyses included baseline characteristics used in the constraining process.

~~A~~ An acceptable Type 1 error was observed for this method (i.e., between 3.6% and 6.4%).

Figure 4-1: Percentage of time each pair of centers were randomly allocated to the same group (i.e., Center 1 with Center 2, Center 1 with Center 3, Center 1 with Center 4, …, Center 71 with Center 72). There were a total of 2556 unique center pairs. **(A)** Centers randomly allocated without any constraints (i.e., simple randomization) would appear in the same arm approximately 50% of the time. **(B)** Constraining on a subset of 11 prognostic baseline characteristics; **(C)** Constraining on ten principal components from a Principal Component Analysis. The horizontal dashed lines show center pairs (if any) allocated to the same arm 25% or 75% of the time.[41]



References

1. Hayes RJ, Moulton LH. *Cluster Randomised Trials*. CRC Press; 2009.
2. Eldridge S, Kerry SM. *A Practical Guide to Cluster Randomised Trials in Health Services Research*. John Wiley & Sons; 2012.
3. Perry M, Faes M, Reelick MF, Olde Rikkert MGM, Borm GF. Studywise minimization: A treatment allocation method that improves balance among treatment groups and makes allocation unpredictable. *J Clin Epidemiol*. 2010;63(10):1118-1122. doi:10.1016/j.jclinepi.2009.11.014
4. Chu R, Walter SD, Guyatt G, et al. Assessment and Implication of Prognostic Imbalance in Randomized Controlled Trials with a Binary Outcome – A Simulation Study. Gong Y, ed. *PLoS One*. 2012;7(5):e36677. doi:10.1371/journal.pone.0036677
5. Ivers NM, Taljaard M, Dixon S, et al. Impact of CONSORT extension for cluster randomised trials on quality of reporting and study methodology: review of random sample of 300 trials, 2000-8. *BMJ*. 2011;343(sep26 1):d5886-d5886. doi:10.1136/bmj.d5886
6. Raab GM, Butcher I. Balance in cluster randomized trials. *Stat Med*. 2001;20(3):351-365. doi:10.1002/1097-0258(20010215)20:3<351::AID-SIM797>3.0.CO;2-C
7. Carter BR, Hood K, Fisher R, et al. Balance algorithm for cluster randomized trials. *BMC Med Res Methodol*. 2008;8(1):65. doi:10.1186/1471-2288-8-65
8. Ivers NM, Halperin IJ, Barnsley J, et al. Allocation techniques for balance at baseline in cluster randomized trials: a methodological review. *Trials*. 2012;13(1):120. doi:10.1186/1745-6215-13-120
9. Moulton LH. Covariate-based constrained randomization of group-randomized trials. *Clin Trials*. 2004;1(3):297-305. doi:10.1191/1740774504cn024oa
10. Xiao L, Lavori PW, Wilson SR, Ma J. Comparison of dynamic block randomization and minimization in randomized trials: A simulation study. *Clin Trials*. 2011;8(1):59-69. doi:10.1177/1740774510391683
11. Li F, Lokhnygina Y, Murray DM, Heagerty PJ, DeLong ER. An evaluation of constrained randomization for the design and analysis of group-randomized trials. *Stat Med*. 2016;35(10):1565-1579. doi:10.1002/sim.6813
12. Li F, Turner EL, Heagerty PJ, Murray DM, Vollmer WM, DeLong ER. An evaluation of constrained randomization for the design and analysis of group-randomized trials with binary outcomes. *Stat Med*. 2017;36(24):3791-3806. doi:10.1002/sim.7410

13. Yu H, Li F, Gallis JA, Turner EL. cvcrand: A Package for Covariate-constrained Randomization and the Clustered Permutation Test for Cluster Randomized Trials. *R J*. 2019;11(2):1-14.

14. Dickinson LM, Beaty B, Fox C, et al. Pragmatic Cluster Randomized Trials Using Covariate Constrained Randomization: A Method for Practice-based Research Networks (PBRNs). *J Am Board Fam Med*. 2015;28(5):663-672. doi:10.3122/jabfm.2015.05.150001

15. Al-Jaishi AA, McIntyre CW, Sontrop JM, et al. Major Outcomes With Personalized Dialysate TEMPerature (MyTEMP): Rationale and Design of a Pragmatic, Registry-Based, Cluster Randomized Controlled Trial. *Can J Kidney Heal Dis*. 2020;7:1-18. doi:10.1177/2054358119887988

16. Dempsey AF, Pyrznawoski J, Lockhart S, et al. Effect of a health care professional communication training intervention on adolescent human papillomavirus vaccination a cluster randomized clinical trial. *JAMA Pediatr*. 2018;172(5). doi:10.1001/jamapediatrics.2018.0016

17. Ford I, Norrie J, Ahmadi S. Model inconsistency, illustrated by the cox proportional hazards model. *Stat Med*. 1995;14(8):735-746. doi:10.1002/sim.4780140804

18. Ford I, Norrie J. The role of covariates in estimating treatment effects and risk in long-term clinical trials. *Stat Med*. 2002;21(19):2899-2908. doi:10.1002/sim.1294

19. Kahan BC, Jairath V, Doré CJ, Morris TP. The risks and rewards of covariate adjustment in randomized trials: An assessment of 12 outcomes from 8 studies. *Trials*. 2014;15(1):139. doi:10.1186/1745-6215-15-139

20. Outcomes of a Higher Versus Lower Hemodialysate Magnesium Concentration. Accessed January 20, 2020. https://clinicaltrials.gov/ct2/show/NCT04079582

21. HiLo | A Pragmatic Clinical Trial. Accessed January 20, 2020. https://hilostudy.org/

22. Donner A, Klar N. *Design and Analysis of Cluster Randomization Trials in Health Research*. (Gooster L, Ueberberg A, eds.). Arnold; 2000.

23. Statistics Canada. Population estimates, quarterly. Published August 12, 2020. Accessed August 12, 2020. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710000901

24. Webster G, Wu J, Williams B, Ivis F, de Sa E, Hall N. *Canadian Organ Replacement Register Annual Report: Treatment of End-Stage Organ Failure in Canada 2003 - 2012*.; 2014.

25. Moist LM, Trpeski L, Na Y, Lok CE. Increased Hemodialysis Catheter Use in Canada and Associated Mortality Risk: Data from the Canadian Organ

Replacement Registry 2001-2004. *Clin J Am Soc Nephrol*. 2008;3(6):1726-1732. doi:10.2215/CJN.01240308

26. Ellwood AD, Jassal SV, Suri RS, Clark WF, Na Y, Moist LM. Early Dialysis Initiation and Rates and Timing of Withdrawal From Dialysis in Canada. *Clin J Am Soc Nephrol*. 2012;8:1-6. doi:10.2215/CJN.01000112

27. Saczynski JS, Andrade SE, Harrold LR, et al. A systematic review of validated methods for identifying heart failure using administrative data. *Pharmacoepidemiol Drug Saf*. 2012;21 Suppl 1:129-140. doi:10.1002/pds.2313

28. Pladevall M, Goff DC, Nichaman MZ, et al. An assessment of the validity of ICD Code 410 to identify hospital admissions for myocardial infarction: The Corpus Christi Heart Project. *Int J Epidemiol*. 1996;25(5):948-952.

29. Tamariz L, Harkins T, Nair V. A systematic review of validated methods for identifying ventricular arrhythmias using administrative and claims data. *Pharmacoepidemiol Drug Saf*. 2012;21 Suppl 1:148-153. doi:10.1002/pds.2340

30. Moist LM, Richards HA, Miskulin D, et al. A Validatiaon Study of the Canadian Organ Replacement Register. *Clin J Am Soc Nephrol*. 2011;6(4):813-818. doi:10.2215/CJN.06680810

31. Oliver MJ, Quinn RR, Garg AX, Kim SJ, Wald R, Paterson JM. Likelihood of Starting Dialysis after Incident Fistula Creation. *Clin J Am Soc Nephrol*. 2012;7(3):466-471. doi:10.2215/CJN.08920811

32. Perl J, Wald R, McFarlane P, et al. Hemodialysis Vascular Access Modifies the Association between Dialysis Modality and Survival. *J Am Soc Nephrol*. 2011;22(6):1113-1121. doi:10.1681/ASN.2010111155

33. Quinn RR, Laupacis A, Austin PPC, et al. Using administrative datasets to study outcomes in dialysis patients: a validation study. *Med Care*. 2010;48(8):745-750. doi:10.1097/MLR.0b013e3181e419fd

34. Al-Jaishi AA, Moist LM, Oliver MJ, et al. Validity of administrative database code algorithms to identify vascular access placement, surgical revisions, and secondary patency. *J Vasc Access*. Published online March 12, 2018:112972981876200. doi:10.1177/1129729818762008

35. Schultz SE, Rothwell DM, Chen ; Z, Tu ; K, Chen Z, Tu K. Identifying cases of congestive heart failure from administrative data: a validation study using primary care patient records. *Chronic Dis Inj Can*. 2013;33(3):160-166.

36. Hennessy S, Leonard CE, Freeman CP, et al. Validation of diagnostic codes for outpatient-originating sudden cardiac death and ventricular arrhythmia in Medicaid and Medicare claims data. *Pharmacoepidemiol Drug Saf*. 2010;19(6):555-562. doi:10.1002/pds.1869

37. Hussain MA, Mamdani M, Saposnik G, et al. Validation of Carotid Artery Revascularization Coding in Ontario Health Administrative Databases. *Clin Investig Med Médecine Clin Exp*. 2016;39(2):E73-8.

38. Longenecker JC, Coresh J, Klag MJ, et al. Validation of comorbid conditions on the end-stage renal disease medical evidence report: the CHOICE study. Choices for Healthy Outcomes in Caring for ESRD. *J Am Soc Nephrol*. 2000;11(3):520-529.

39. ICES. Privacy at ICES. Accessed November 25, 2019. https://www.ices.on.ca/Data-and-Privacy/Privacy-at-ICES

40. Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. *Stat Med*. 2006;25(24):4279-4292.

41. Greene EJ. A SAS Macro for Covariate-Constrained Randomization of General Cluster-Randomized and Unstratified Designs. *J Stat Softw*. 2017;77. doi:10.18637/jss.v077.c01

42. Austin PC. Using the Standardized Difference to Compare the Prevalence of a Binary Variable Between Two Groups in Observational Research. *Commun Stat Simul Comput*. 2009;38(6):1228-1234. doi:10.1080/03610910902859574

43. Mamdani M, Sykora K, Li P, et al. Reader ' s guide to critical appraisal of cohort studies : 2 . Assessing potential for confounding. *BMJ*. 2005;330:960-962.

44. Yang D, Dalton JE. A unified approach to measuring the effect size between two groups using SAS ®. *Pap 335-2012 Present 2012 SAS Glob Forum*. Published online 2012:1-6.

45. Kruskal WH, Wallis A. Use of Ranks in One-Criterion Variance Analysis. *J Am Stat Assoc*. 1952;47(260):583-621.

46. Lin DY. Cox regression analysis of multivariate failure time data: The marginal approach. *Stat Med*. 1994;13(21):2233-2247. doi:10.1002/sim.4780132105

47. Wicklin R. *Simulating Data with SAS ®*. SAS Institute Inc.; 2013.

48. Rosner B. *Fundamentals of Biostatistics*. Duxbury Press; 1995.

49. Wright N, Ivers N, Eldridge S, Taljaard M, Bremner S. A review of the use of covariates in cluster randomized trials uncovers marked discrepancies between guidance and practice. *J Clin Epidemiol*. 2015;68(6):603-609. doi:10.1016/j.jclinepi.2014.12.006

50. Freedman LS, Green SB, Byar DP. Assessing the gain in efficiency due to matching in a community intervention study. *Stat Med*. 1990;9(8):943-952. doi:10.1002/sim.4780090810

51. Hernández A V., Steyerberg EW, Habbema JDF. Covariate adjustment in

randomized controlled trials with dichotomous outcomes increases statistical power and reduces sample size requirements. *J Clin Epidemiol*. 2004;57(5):454-460. doi:10.1016/j.jclinepi.2003.09.014

52. Hernández A V., Eijkemans MJC, Steyerberg EW. Randomized controlled trials with time-to-event outcomes: How much does prespecified covariate adjustment increase power? *Ann Epidemiol*. 2006;16(1):41-48. doi:10.1016/j.annepidem.2005.09.007

53. Raab GM, Day S, Sales J. How to Select Covariates to Include in the Analysis of a Clinical Trial. *Control Clin Trials*. 2000;21(4):330-342. doi:10.1016/S0197-2456(00)00061-1

54. Silipo R, Widmann M. 3 New Techniques for Data-Dimensionality Reduction in Machine Learning – The New Stack. Published August 9, 2019. Accessed August 26, 2020. https://thenewstack.io/3-new-techniques-for-data-dimensionality-reduction-in-machine-learning/

55. Fiero MH, Huang S, Oren E, Bell ML. Statistical analysis and handling of missing data in cluster randomized trials: a systematic review. *Trials*. 2016;17(1):72. doi:10.1186/s13063-016-1201-z

Appendix

Appendix 4-1: Common Data Sources used for Population-Based Studies

Database (Source)	Description	Key Data Variables
Health Services		
Discharge Abstract Database (CIHI)	Hospital discharge abstracts for acute, chronic, and rehabilitative care (1988 onward)	Diagnoses; Procedures; Comorbidities; Length of Stay
National Ambulatory Care Reporting System (CIHI)	ED visits, same-day surgery, outpatient clinics (e.g., dialysis, cancer clinics) (2002 onward)	Reason for the visit; Triage level; Interventions; Mode of arrival
Ontario Drug Benefit Database (MOHLTC)	Claims for prescribed drugs covered by the Ontario Drug Formulary for adults aged 65+ and those receiving social assistance (1990 onward)	Drug ID number; Drug quantity; Cost
Ontario Health Insurance Plan (MOHLTC)	Reimbursement claims made by fee-for-service physicians and community-based labs (1991 onward)	Service provided; Diagnosis codes; Fee paid; Physician specialty
Registry		
Canadian Organ Replacement Register (CIHI)	Collects and records the incidence, prevalence, treatment changes, and outcomes of all chronic dialysis and solid organ transplant patients in Canada. Data is collected by voluntary completion of survey forms for each patient at dialysis initiation and at yearly follow-up (2001 onward)	Hemodialysis start; vascular access use; nephrology referral; comorbid and baseline conditions
Ontario Renal Reporting System	Collects and records the incidence, prevalence, treatment changes, and outcomes of all chronic dialysis and solid organ transplant patients in Canada. Data collected is mandated by the Ontario Renal Network for each patient at dialysis initiation and yearly follow-up (2010 onward)	Hemodialysis start; vascular access use; nephrology referral; comorbid and baseline conditions
Population and Demographics		
Registered Persons Database (MOHLTC)	Demographic information about all Ontarians that ever had an Ontario Health Card Number. (1990 onward)	Date of birth; Date of death; Sex; Geographic information
Office of the Registrar General- Deaths (ORGD)	ORGD is an annual dataset containing information on all deaths registered in Ontario starting on January 1, 1990.	**Note**: Information on cause of death lags other variables by ~2 years.

Care Providers		
ICES Physicians Database	This data set contains yearly information about all physicians in Ontario (1992 onward)	Annual demographics; Specialization; Workload

MOHTC: Ministry of Health and Long-term Care, CIHI – Canadian Institutes for Health Information

Appendix 4-2: Complete list of 156 Baseline characteristics for the randomization and trial population cohorts

Baseline characteristic	**VALUE**	**Population for Randomization**	**Trial Population**
Center Size [1]	Mean ± Standard deviation	81 (69)	164 (137)
The composite outcome of CV-related death and major CV-related hospitalization [2]	Historic rate (per 100 person-years)	10	10
Age (years)	Mean ± Standard deviation	67 (15)	66 (15)
	< 65	2348 (40%)	4913 (42%)
	65 to 74	1450 (25%)	3060 (26%)
	75 to 84	1464 (25%)	2853 (24%)
	85 to 105	550 (9%)	1006 (9%)
Sex	Male	3373 (58%)	7069 (60%)
Living in a rural area	Yes	359 (6%)	809 (7%)
Neighbourhood Income Quintile	1	1682 (29%)	3748 (32%)
	2	1317 (23%)	2665 (23%)
	3	1070 (18%)	2132 (18%)
	4	960 (17%)	1799 (15%)
	5	740 (13%)	1467 (12%)
	Missing	43 (1%)	21 (0%)
Etiology for ESKD	Diabetes	2194 (38%)	4472 (38%)
	Glomerulonephritis/autoimmune diseases	882 (15%)	1575 (13%)
	Drug-induced nephropathy	83 (1%)	159 (1%)
	Polycystic kidney disease	229 (4%)	426 (4%)
	Renal vascular disease	1115 (19%)	1852 (16%)
	Other	738 (13%)	2221 (19%)
	Unknown	571 (10%)	1127 (10%)
Race	Asian	475 (8%)	854 (7%)
	Black	562 (10%)	1022 (9%)
	White	3698 (64%)	7598 (64%)
	Other	1038 (18%)	2173 (18%)
	Unknown	39 (1%)	185 (2%)
First dialysis modality	Home hemodialysis	33 (1%)	84 (1%)
	In-center hemodialysis	5215 (90%)	10529 (89%)
	Peritoneal dialysis	564 (10%)	1219 (10%)
Initial vascular access used at dialysis start	Arteriovenous graft	103 (2%)	198 (2%)
	Arteriovenous fistula	1044 (18%)	1999 (17%)

	Catheter	3927 (68%)	8157 (69%)
	Peritoneal catheter	456 (8%)	1092 (9%)
	Unknown	282 (5%)	386 (3%)
Most recent vascular access before the index date	Arteriovenous graft	223 (4%)	376 (3%)
	Arteriovenous fistula	2159 (37%)	3461 (29%)
	Catheter	3376 (58%)	7544 (64%)
Patients < 65 years in Ontario Drug benefit in the six months before the index date	Yes	1596 (27%)	3431 (29%)
Patients 65+ years in Ontario Drug Benefit in the six months before the index date	Yes	3398 (58%)	6765 (57%)
Angiotensin-converting enzyme (ACE) inhibitors	Yes	1157 (26%)	2199 (22%)
Angiotensin II Receptor Blocker	Yes	1258 (28%)	2413 (24%)
Alpha-Blockers	Yes	408 (9%)	1073 (11%)
Benzodiazepine	Yes	1060 (24%)	1884 (18%)
Beta-Blockers	Yes	2649 (59%)	5604 (55%)
Centrally Acting Antiadrenergic	Yes	143 (3%)	338 (3%)
Calcium Channel Blocker	Yes	2392 (53%)	5453 (53%)
Diuretics	Yes	1612 (36%)	4242 (42%)
Midodrine	Yes	250 (6%)	517 (5%)
Vasodilators	Yes	226 (5%)	778 (8%)
Anti-psychotics	Yes	309 (7%)	657 (6%)
Anti-depressants	Yes	1270 (28%)	2531 (25%)
Number of unique hypertensive prescriptions in the six months before the index date	Mean ± Standard deviation	2 (2)	2 (2)
Number of hypertensive subclasses prescribed in the six months before the index date	Mean ± Standard deviation	2 (1)	2 (2)
Abdominal aortic aneurysm repair/aortic bypass	Yes	34 (1%)	78 (1%)

Atrial Fibrillation/Flutter	Yes	914 (16%)	2045 (17%)
Acute Kidney Injury in the six months before the index date	Yes	1431 (25%)	3934 (33%)
Alcoholism	Yes	97 (2%)	256 (2%)
Lower extremity amputation	Yes	298 (5%)	554 (5%)
Arrhythmia	Yes	1431 (25%)	3119 (26%)
CABG/PCI	Yes	1234 (21%)	2612 (22%)
Coronary Artery Disease (with angina)	Yes	3541 (61%)	6861 (58%)
Heart failure	Yes	2862 (49%)	6177 (52%)
Diabetes mellitus	Yes	3402 (59%)	7244 (61%)
Dementia	Yes	892 (15%)	1830 (15%)
Depression	Yes	1528 (26%)	3161 (27%)
Having any type of Fracture	Yes	604 (10%)	1092 (9%)
Fracture of the Humerus	Yes	26 (0%)	48 (0%)
Fracture of the Pelvis	Yes	80 (1%)	130 (1%)
Fracture of the Femur	Yes	136 (2%)	267 (2%)
Fracture of the Hip	Yes	302 (5%)	526 (4%)
Fracture of the wrist	Yes	237 (4%)	445 (4%)
Hypertension	Yes	5629 (97%)	11453 (97%)
Hypotension	Yes	669 (12%)	1382 (12%)
Ischemic Stroke	Yes	261 (4%)	551 (5%)
Subarachnoid Hemorrhage	Yes	6 (0%)	20 (0%)
Liver disease	Yes	697 (12%)	1521 (13%)
Lung disease (COPD)	Yes	2276 (39%)	4666 (39%)
Myocardial infarction	Yes	1454 (25%)	2909 (25%)
Malignancy (excluding skin cancer)	Yes	2339 (40%)	5001 (42%)
Major Cancer	Yes	841 (14%)	1803 (15%)
Other Serious Illness that could shorten life expectancy to less than five years	Yes	845 (15%)	1949 (16%)
Peripheral vascular disease	Yes	1699 (29%)	3055 (26%)
Having a kidney transplant before the index date	Yes	71 (1%)	148 (1%)

Smoker	Yes	733 (13%)	1564 (13%)
Syncope	Yes	248 (4%)	494 (4%)
Venous thromboembolism	Yes	344 (6%)	779 (7%)
Stroke/Transient ischemic attack (TIA)	Yes	1018 (18%)	2029 (17%)
Body mass index	Mean ± Standard deviation	28 (8)	28 (8)
	Underweight	280 (5%)	414 (3%)
	Normal	1816 (31%)	3526 (30%)
	Overweight	1599 (28%)	3376 (29%)
	Obese I	913 (16%)	2040 (17%)
	Obese II	433 (7%)	962 (8%)
	Obese III	348 (6%)	799 (7%)
	Missing	423 (7%)	715 (6%)
Modified Charlson comorbidity Score	Mean ± Standard deviation	4 (2)	4 (2)
	2	1931 (33%)	3631 (31%)
	3	519 (9%)	1028 (9%)
	4	1382 (24%)	2883 (24%)
	5+	1980 (34%)	4290 (36%)
Abdominal/Renal ultrasound	Yes	5064 (87%)	10611 (90%)
Chest x-ray	Yes	5690 (98%)	11567 (98%)
Coronary angiogram	Yes	1267 (22%)	2540 (21%)
Coronary revascularization	Yes	681 (12%)	1418 (12%)
Echocardiography	Yes	5168 (89%)	10440 (88%)
Holter monitoring	Yes	1583 (27%)	3401 (29%)
Stress test	Yes	3442 (59%)	6969 (59%)
Carotid endarterectomy	Yes	18 (0%)	44 (0%)
Number of cardiology visits in the year before the index date	Mean ± Standard deviation	3 (5)	5 (7)
	0	1547 (27%)	2283 (19%)
	1 to 3	2465 (42%)	4666 (39%)
	4 to 6	977 (17%)	2196 (19%)
	7 to 9	352 (6%)	1046 (9%)
	10+	471 (8%)	1641 (14%)
Number of general practitioner visits in the year before the index date	Mean ± Standard deviation	9 (13)	13 (16)
	0	1547 (27%)	2283 (19%)
	1 to 3	2465 (42%)	4666 (39%)
	4 to 6	977 (17%)	2196 (19%)
	7 to 9	352 (6%)	1046 (9%)
	10+	471 (8%)	1641 (14%)

Number of nephrology consults in the year before the index date	Mean ± Standard deviation	8 (14)	12 (17)
	0	1547 (27%)	2283 (19%)
	1 to 3	2465 (42%)	4666 (39%)
	4 to 6	977 (17%)	2196 (19%)
	7 to 9	352 (6%)	1046 (9%)
	10+	471 (8%)	1641 (14%)
Number of days spent in the hospital in the year before the index date	Mean ± Standard deviation	11 (28)	17 (31)
	0	1547 (27%)	2283 (19%)
	1 to 3	2465 (42%)	4666 (39%)
	4 to 6	977 (17%)	2196 (19%)
	7 to 9	352 (6%)	1046 (9%)
	10+	471 (8%)	1641 (14%)
Number of hospitalization visits in the year before the index date	Mean ± Standard deviation	1 (1)	1 (1)
	0	1547 (27%)	2283 (19%)
	1 to 3	2465 (42%)	4666 (39%)
	4 to 6	977 (17%)	2196 (19%)
	7 to 9	352 (6%)	1046 (9%)
	10+	471 (8%)	1641 (14%)
Number of emergency department visits in the year before the index date	Mean ± Standard deviation	2 (3)	3 (3)
	0	1547 (27%)	2283 (19%)
	1 to 3	2465 (42%)	4666 (39%)
	4 to 6	977 (17%)	2196 (19%)
	7 to 9	352 (6%)	1046 (9%)
	10+	471 (8%)	1641 (14%)
Number of days spent in long-term care in the year before the index date	Mean ± Standard deviation	37 (164)	25 (133)
Long term care facility utilization in the year before the index date	Yes	568 (10%)	1082 (9%)
Number of dialysis sessions in the year before the index date	Mean ± Standard deviation	140 (36)	107 (48)
Time since the first date on dialysis (days)	Mean ± Standard deviation	1847 (1836)	1327 (1782)
Height (cm) before starting dialysis	Mean ± Standard deviation	175 (63)	170 (42)
Weight (kg) before starting dialysis	Mean ± Standard deviation	87 (71)	83 (50)
Urea test result before starting dialysis	Mean ± Standard deviation	33 (21)	32 (22)

Hemoglobin test results before starting dialysis	Mean ± Standard deviation	97 (23)	97 (35)
Creatinine test result before starting dialysis	Mean ± Standard deviation	632 (342)	634 (331)
eGFR using CKD EPI	Mean ± Standard deviation	9 (6)	9 (6)
Serum albumin test result before starting dialysis	Mean ± Standard deviation	32 (7)	32 (7)

[1] Population for Randomization included patients that were on hemodialysis as of April 1, 2013, and Trial Population included patients that were on hemodialysis as of April 1, 2014, and any patient that started in-center hemodialysis at one of the 72 participating centers during the three-year follow-up. [2] Composite outcome of cardiovascular-related death or hospitalization for myocardial infarction, ischemic stroke, congestive heart failure. [3] Percentage presented only for patients eligible to receive the Ontario Drug Benefit (ODB) plan in the six months before the index date (i.e., 4494 for Randomization Cohort and 10196 for the Trial cohort). CABG/PCI = Coronary artery bypass grafting (CABG) / percutaneous coronary intervention; GN = Glomerulonephritis; ESKD=End-stage kidney disease.

Appendix 4-3: Randomization of the 72 clusters using PROC PLAN in SAS.

```
**********************************;
**Creating randomizations Schemes**;
**********************************;

proc plan seed=14424;
        factors set=300000 group=72 / noprint;
        output out=a;
run;

proc sort data=a;
        by set;
run;

data b;
        label Set="The allocation number"
                Center_Number="Unique Cluster Number"
                Arm="Treatment arm"
        ;
        set a;
        by set;
        retain Center_Number;
        Arm=(group LE 36);

        if first.set then
                Center_Number=0;
        Center_Number=Center_Number+1;
        drop group;
run;
```

Appendix 4-4a: Prognostic baseline characteristics that were thought to be relevant a priori or correlated with the outcome from previous literature.

Prognostic factors included the following patient-level information: age at index date, living in a rural area, Black race, Modified Charlson comorbidity index, number of hospital visits in the previous 12 months, number of unique drugs the patient was prescribed in the six months before the index date, as well as history in the last five years of diagnosis for peripheral vascular disease, congestive heart failure, coronary artery disease, myocardial infarction, number of nephrology consults in the previous 12 months before the index date. We also included cluster-level baseline characteristics such as center size and the historical rate of the primary composite outcome of cardiovascular-related death or hospitalization for myocardial infarction, ischemic stroke, or heart failure.

We excluded several prognostic factors from above (e.g., diabetes) because these baseline characteristics were almost always balanced in the non-constrained setting (See **Appendix 4-9**). As such, constraining on these baseline characteristics would not have influenced the results for the constrained randomization.

Appendix 4-4b: The 156 baseline characteristics from the Population for Randomization were subjected to a principal component analysis using one as prior communality estimates; communalities refers to the estimate of the variances for the principal components.[1] We dropped 127 baseline characteristics that loaded on more than one component because these baseline characteristics are not pure measures of any single construct. Thus, 29 baseline characteristics were included in the analysis. We used the principal axis method to extract the components, followed by a varimax (orthogonal) rotation.[2] We retained principal components for rotation when the eigenvalues were greater than one. In interpreting the rotated factor pattern, an item loaded on a given component if the factor loading was equal to or greater than 40% for that component and less than 40% for the others.

Appendix 4-5: Algorithm for capturing primary composite outcome.

Outcome	Algorithm	Position of code	Performance
Cardiovascular-related death ~~A~~, ¥	ORGD: Leading Cause of Death LCD_33 = Chronic rheumatic heart disease LCD_34 = Hypertensive disease LCD_35 = Ischemic heart disease LCD_36 = Pulmonary heart disease and related LCD_37 = Nonrheumatic valve disorders LCD_38 = Cardiomyopathy LCD_39 = Cardiac arrest LCD_40 = Cardiac arrhythmias LCD_41 = Heart failure and complications, ill-defined heart disease LCD_42 = Cerebrovascular diseases LCD_43 = Atherosclerosis LCD_44 = Aortic aneurysm and dissection	N/A	Not available
Cardiovascular-related death	**ICD-10:** I00 - I78 AND Discharge disposition of '07' or death in the Registered Persons Database during the hospital stay	Primary Diagnosis	RPDB has an accuracy of 99% for capturing death [3]
Hospital admission with ischemic stroke	***ICD-10:*** I63 (excl. I63.6), I64, H341	Primary Diagnosis	PPV= 85% [4,5]
Hospital admission with myocardial infarction	***ICD-10:*** I21, I22	Primary Diagnosis	Sn= 89%, PPV= 87% [6]
Hospital admission with heart failure	**ICD-10:** I50	Primary Diagnosis	Sn=61% , Sp=98%, PPV=66%[7]

Abbreviations: ICD = International Classification of Disease; OHIP = Ontario Health Insurance Plan; Dischdisp=Discharge disposition; Sn=Sensitivity; PPV= Positive Predictive Value; LCD=Leading Cause of Death; ORGD=Office of Registrar General - Deaths.

~~A~~ Due to the time lag in data capture, deaths from ORGD will only capture events for the follow-up period between April 3, 2017, and December 31, 2020. These events capture both in- and out-of-hospital cardiovascular-related deaths. For the remaining study period, we will only be able to capture in-hospital deaths using ICD-10 codes.

¥ Personal communication with Dr. Jack Tu, who was part of a working group conducting validation of this outcome using existing Ontario clinical trial data as the reference standard.

Appendix 4-6: Results from Principal component analysis (PCA).

Eigenvalues of the Correlation Matrix			
Principal component	**Eigenvalue****	**The proportion of variance explained**	**Cumulative variance explained**
1	3.67	0.13	0.13
2	2.19	0.08	0.20
3	2.04	0.07	0.27
4	1.72	0.06	0.33
5	1.60	0.06	0.39
6	1.53	0.05	0.44
7	1.33	0.05	0.49
8	1.30	0.05	0.53
9	1.13	0.04	0.57
10	1.03	0.04	0.61
11	0.99	0.03	0.64
12	0.89	0.03	0.67
13	0.83	0.03	0.70
14	0.82	0.03	0.73
15	0.81	0.03	0.75
16	0.75	0.03	0.78
17	0.70	0.02	0.80
18	0.67	0.02	0.83
19	0.66	0.02	0.85
20	0.58	0.02	0.87
21	0.55	0.02	0.89
22	0.51	0.02	0.91
23	0.50	0.02	0.92
24	0.49	0.02	0.94
25	0.45	0.02	0.96
26	0.36	0.01	0.97
27	0.32	0.01	0.98
28	0.29	0.01	0.99
29	0.28	0.01	1.00

Note: There were 29 baseline characteristics used in the PCA. These included: Age at index date, Male, White Race, Modified Charlson comorbidity score, history of coronary artery disease (with angina), history of heart failure, history of ischemic stroke, history of stroke or transient ischemic attack, history of myocardial infarction, history of diabetes mellitus, history of a major

malignancy, history of malignancy (excluding skin cancer), history of coronary revascularization, history of depression, history of fractures, history of femur fracture, history of hip fracture, had a late referral to a nephrologist before starting renal replacement therapy, first modality used was in-center hemodialysis, first vascular access was a central venous catheter, last vascular access used before the index date was a central venous catheter, younger than 65 and at least one prescription in ODB in the 6 months before the index date, had at least one prescription of calcium channel blocker in the six months before the index date, had at least one diuretics prescription in the 6 months before the index date, number of days spent in the hospital in the 12 months before the index date, number of hospitalization visits in the 12 months before the index date, number of general practitioner visits in the 12 months before the index date, number of cardiology visits in the 12 months before the index date, number of emergency department visits in the 12 months before the index date.

** See definition of Eigenvalue at:[8] http://mathworld.wolfram.com/Eigenvalue.html

PC= Principal component; We used the first ten principal components in the analysis, which explained 61% of the variation in the baseline data.

Appendix 4-7: We used the principal axis method to extract the principal components. A varimax (orthogonal) rotation followed the principal axis method. Only the first ten components displayed eigenvalues greater than 1 (see **Appendix 4-6**), and the results of a scree test also suggested that only the first ten components were meaningful. Therefore, we retained the first ten components for rotation.

Baseline characteristic	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9	Factor 10
History of coronary artery disease (with angina)	7	69*	1	-4	-7	0	0	-1	-6	11
Had at least one prescription of calcium channel blocker in the six months before the index date	0	-8	0	5	-1	1	0	72*	5	-10
History of congestive heart failure	16	53*	4	-12	12	3	3	16	12	-5
Malignancy (excluding skin cancer)	4	1	0	-12	-2	1	84*	-2	-1	12
Number of cardiology visits in the 12 months before the index date	66*	31	-3	6	-2	-9	-2	-5	-6	-2
Modified Charlson comorbidity score	28	58*	7	-1	9	29	22	22	13	-23
History of coronary revascularization	1	66*	-1	7	-4	-6	-2	-10	-12	1
History of diabetes mellitus	9	39	3	-3	2	20	3	39	7	-41*
Late referral to a nephrologist	-2	-2	-2	14	57*	-3	4	-20	14	-4
Depression	14	10	11	16	-6	18	4	26	24	49*

Use of diuretics in the 6 months before the index date	-1	10	-2	-10	1	-5	-3	69*	-4	12
Number of emergency department visits in the 12 months before the index date	69*	5	1	19	3	4	3	13	4	14
First modality used was in-center hemodialysis	3	-2	4	-13	70*	1	-5	18	-32	6
History of fractures	7	1	84*	-4	1	2	3	2	4	6
Number of general practitioner visits in the 12 months before the index date	54*	0	7	-18	4	13	4	0	7	3
Number of days spent in the hospital in the 12 months before the index date	71*	-1	7	-12	0	2	-3	-10	4	0
Number of hospitalization visits in the 12 months before the index date	83*	13	2	11	2	3	6	6	5	5
History of ischemic stroke	7	1	-2	1	-2	83*	-3	-3	1	-1
History of myocardial infarction	4	73*	1	-12	5	5	-6	-3	-3	9
History of a major malignancy	3	0	0	-2	4	-3	87*	-1	-1	-3
Male	0	16	-7	6	9	1	5	-7	-66*	0
White Race	2	7	4	-15	6	1	8	-6	-5	78*
Patients was younger than 65 and at least one prescription in ODB in	0	-5	-2	87*	1	-2	-4	9	-3	1

the 6 months before the index date										
Age at index date	-2	14	7	-84*	-2	6	11	18	9	10
History of femur fracture	3	3	78*	1	1	0	-3	-2	2	-2
History of hip fracture	3	4	89*	-5	1	1	1	-1	3	5
First vascular access was a central venous catheter	6	8	2	-2	82*	3	2	6	16	1
The last access used before the index date was a central venous catheter	13	11	1	-7	28	2	3	-5	70*	3
History of stroke or transient ischemic attack	4	8	3	-8	3	83*	0	1	-1	7

Note: Values above were multiplied by 100 and rounded to the nearest integer. Values greater than the absolute value of 40 were flagged by an '*.'

Appendix 4-8: Hardware specification and optimization for running the constrained randomization process.

It took approximately 1 second to evaluate each randomization scheme's balance or a total of 83 hours (of CPU time) to assess all 300,000 allocation schemes (see **below** for hardware specification and optimization). From the 300,000 allocations, we constrained the randomization space to the 30,000 best allocations (i.e., 10% of the randomization space) and randomly selected 1000 allocations. All 1000 sampled allocations schemes were balanced on all constrained baseline characteristics in the Population for Randomization, regardless of the constraining method.

Hardware: We used a Windows 10 Intel(R) Core(TM) i7-7500U CPU @ 2.70GHz, 2904 Mhz with 2 Cores and 4 Logical Processors. This hardware had 12GB RAM.

Optimization: Rather than running the 300,000 allocation schemes sequentially, we parallelized the process by utilizing the "RSUBMIT" statement in SAS. Parallel processes allowed us to execute three statements in a remote SAS session with three logical processors. This method reduced our computation time by approximately a third.

Appendix 4-9: The percentage of times each of the 156 baseline characteristics was balanced across 1000 randomization schemes for the three techniques.

Baseline characteristic	**Value**	**Constrained randomization method**		
		Unrestricted / Simple	**Prognostic baseline characteristics**	**Principal components**
Center Size	Mean ± Standard deviation	32.9%	41.8%	38.7%
Composite outcome of CV-related death and major CV-related hospitalization ++	Rate (per person-year)	32.5%	36.2%	33.5%
Age (years)	Mean ± Standard deviation	95.3%	99.8%	99.2%
	< 65	97.8%	99.7%	99.9%
	65 to 74	100.0%	100.0%	100.0%
	75 to 84	100.0%	100.0%	100.0%
	85 to 105	99.5%	100.0%	99.9%
Sex	Male	100.0%	100.0%	100.0%
Living in a rural area	Yes	63.0%	84.2%	65.8%
Neighbourhood Income Quintile	1	77.5%	77.0%	75.3%
	2	98.8%	98.5%	98.8%
	3	98.7%	98.7%	99.0%
	4	91.9%	95.6%	93.6%
	5	89.0%	91.9%	92.8%
	Missing	100.0%	100.0%	100.0%
Etiology for ESKD	Diabetes	93.0%	94.5%	95.0%
	Glomerulonephritis/autoimmune diseases	96.3%	100.0%	99.5%

	Drug-induced nephropathy	100.0%	99.9%	100.0%
	Polycystic kidney disease	100.0%	100.0%	100.0%
	Renal vascular disease	97.5%	97.6%	96.7%
	Other	88.3%	91.9%	91.6%
Race	Asian	75.0%	81.3%	88.1%
	Black	73.4%	95.9%	91.9%
	White	45.6%	64.0%	90.2%
	Other	56.6%	65.7%	77.5%
	Unknown	93.2%	93.7%	93.6%
First dialysis modality	Home hemodialysis	100.0%	99.8%	99.9%
	In-center hemodialysis	97.8%	98.6%	99.9%
	Peritoneal dialysis	97.4%	98.7%	99.8%
Initial vascular access used at dialysis start	AV Graft	99.9%	100.0%	100.0%
	Fistula	98.9%	99.1%	99.4%
	Catheter	93.5%	96.2%	99.4%
	PD Catheter	98.8%	99.0%	100.0%
	Unknown	92.4%	93.8%	94.3%
Most recent vascular access before the index date	AV Graft	98.7%	99.8%	98.9%
	Fistula	91.9%	94.8%	97.7%

	Catheter	89.9%	94.0%	97.4%
Patients < 65 years in ODB in the 6 months before the index date	Yes	99.8%	99.7%	99.9%
Patients 65+ years in ODB in the 6 months before the index date	Yes	97.5%	99.3%	99.4%
Angiotensin-converting enzyme (ACE) inhibitors	Yes	99.4%	99.3%	99.5%
Angiotensin II Receptor Blocker	Yes	90.7%	96.1%	96.9%
Alpha-Blockers	Yes	88.6%	92.0%	92.9%
Benzodiazepine	Yes	98.3%	98.6%	99.6%
Beta-Blockers	Yes	99.7%	100.0%	99.9%
Centrally Acting Antiadrenergic	Yes	96.0%	94.9%	95.1%
Calcium Channel Blocker	Yes	98.1%	100.0%	99.6%
Diuretics	Yes	91.9%	97.0%	95.6%
Midodrine	Yes	71.6%	73.9%	71.6%
Vasodilators	Yes	96.6%	98.1%	97.4%
Anti-psychotics	Yes	100.0%	100.0%	100.0%
Anti-depressants	Yes	96.0%	98.7%	99.4%
Number of unique hypertensive prescriptions in the six months before the index date	Mean ± Standard deviation	97.1%	99.9%	99.5%
Number of hypertensive subclasses prescribed in the six months before the index date	Mean ± Standard deviation	93.3%	98.9%	97.7%

Abdominal aortic aneurysm repair/aortic bypass	Yes	100.0%	100.0%	100.0%
Atrial Fibrillation/Flutter	Yes	99.7%	100.0%	100.0%
Acute Kidney Injury in the six months before the index date	Yes	97.0%	98.8%	99.1%
Alcoholism	Yes	100.0%	100.0%	100.0%
Lower extremity amputation	Yes	99.9%	100.0%	100.0%
Arrhythmia	Yes	99.9%	100.0%	100.0%
CABG/PCI	Yes	99.4%	99.5%	100.0%
Coronary Artery Disease (with angina)	Yes	96.3%	98.7%	96.1%
Heart failure	Yes	96.8%	100.0%	99.8%
Diabetes mellitus	Yes	99.0%	100.0%	100.0%
Dementia	Yes	98.9%	99.7%	99.7%
Depression	Yes	97.4%	99.1%	100.0%
Having any type of Fracture	Yes	100.0%	100.0%	100.0%
Fracture of the Humerus	Yes	100.0%	100.0%	100.0%
Fracture of the Pelvis	Yes	100.0%	100.0%	100.0%
Fracture of the Femur	Yes	100.0%	100.0%	100.0%
Fracture of the Hip	Yes	100.0%	100.0%	100.0%
Fracture of the wrist	Yes	100.0%	100.0%	100.0%
Hypertension	Yes	100.0%	100.0%	100.0%
Hypotension	Yes	100.0%	100.0%	99.9%
Ischemic Stroke	Yes	100.0%	100.0%	100.0%
Subarachnoid Hemorrhage	Yes	100.0%	100.0%	100.0%
Liver disease	Yes	99.8%	99.6%	98.9%

Lung disease (COPD)	Yes	99.0%	99.6%	100.0%
Myocardial infarction	Yes	99.2%	100.0%	100.0%
Malignancy (excluding skin cancer)	Yes	95.1%	94.5%	98.5%
Major Cancer	Yes	100.0%	100.0%	100.0%
Other Serious Illness that could shorten life expectancy to less than five years	Yes	55.2%	56.7%	59.5%
Peripheral vascular disease	Yes	90.7%	97.2%	91.4%
Having a kidney transplant before the index date	Yes	100.0%	100.0%	100.0%
Smoker	Yes	96.8%	99.5%	100.0%
Syncope	Yes	100.0%	100.0%	100.0%
Venous thromboembolism	Yes	99.9%	99.8%	99.9%
Stroke/Transient ischemic attack (TIA)	Yes	100.0%	100.0%	100.0%
Body mass index	Mean ± Standard deviation	93.4%	98.5%	99.6%
	Underweight	100.0%	100.0%	100.0%
	Normal	99.2%	99.5%	99.9%
	Overweight	100.0%	100.0%	100.0%
	Obese I	100.0%	100.0%	100.0%
	Obese II	100.0%	100.0%	100.0%
	Obese III	99.4%	99.9%	100.0%
	Missing	96.2%	98.7%	96.3%
Modified Charlson comorbidity Score	Mean ± Standard deviation	96.8%	99.9%	100.0%

	2	97.7%	100.0%	100.0%
	3	100.0%	100.0%	100.0%
	4	100.0%	100.0%	100.0%
	5+	98.9%	100.0%	100.0%
Abdominal/Renal ultrasound	Yes	100.0%	100.0%	100.0%
Chest x-ray	Yes	100.0%	99.9%	100.0%
Coronary angiogram	Yes	96.7%	97.1%	96.6%
Coronary revascularization	Yes	100.0%	100.0%	100.0%
Echocardiography	Yes	98.5%	99.6%	99.7%
Holter monitoring	Yes	99.2%	99.9%	99.5%
Stress test	Yes	94.9%	96.4%	95.5%
Carotid endarterectomy	Yes	100.0%	100.0%	100.0%
Number of cardiology visits in the year before the index date	Mean ± Standard deviation	81.1%	81.4%	86.9%
	0	78.4%	76.4%	81.1%
	1 to 3	99.5%	99.6%	99.9%
	4 to 6	99.4%	99.6%	99.5%
	7 to 9	100.0%	99.9%	100.0%
	10+	92.1%	92.0%	94.6%
Number of general practitioner visits in the year before the index date	Mean ± Standard deviation	59.9%	55.6%	57.4%

	0	78.4%	76.4%	81.1%
	1 to 3	99.5%	99.6%	99.9%
	4 to 6	99.4%	99.6%	99.5%
	7 to 9	100.0%	99.9%	100.0%
	10+	92.1%	92.0%	94.6%
Number of nephrology consults in the year before the index date	Mean ± Standard deviation	77.4%	95.5%	86.9%
	0	78.4%	76.4%	81.1%
	1 to 3	99.5%	99.6%	99.9%
	4 to 6	99.4%	99.6%	99.5%

	7 to 9	100.0%	99.9%	100.0%
	10+	92.1%	92.0%	94.6%
Number of days spent in the hospital in the year before the index date	Mean ± Standard deviation	92.2%	98.1%	98.7%
	0	78.4%	76.4%	81.1%
	1 to 3	99.5%	99.6%	99.9%
	4 to 6	99.4%	99.6%	99.5%
	7 to 9	100.0%	99.9%	100.0%
	10+	92.1%	92.0%	94.6%
Number of hospitalization visits in the year before the index date	Mean ± Standard deviation	93.9%	98.4%	98.4%

	0	78.4%	76.4%	81.1%
	1 to 3	99.5%	99.6%	99.9%
	4 to 6	99.4%	99.6%	99.5%
	7 to 9	100.0%	99.9%	100.0%
	10+	92.1%	92.0%	94.6%
Number of emergency department visits in the year before the index date	Mean ± Standard deviation	90.8%	97.1%	97.7%
	0	78.4%	76.4%	81.1%
	1 to 3	99.5%	99.6%	99.9%
	4 to 6	99.4%	99.6%	99.5%

	7 to 9	100.0%	99.9%	100.0%
	10+	92.1%	92.0%	94.6%
Number of days spent in long-term care in the year before the index date	Mean ± Standard deviation	99.3%	100.0%	99.9%
Long term care facility utilization in the year before the index date	Yes	81.3%	86.6%	86.1%
Number of dialysis sessions in the year before the index date	Mean ± Standard deviation	84.2%	83.4%	83.9%
Time since the first date on dialysis (days)	Mean ± Standard deviation	88.1%	94.0%	94.4%
Height (cm) before starting dialysis	Mean ± Standard deviation	100.0%	100.0%	100.0%
Weight (kg) before starting dialysis	Mean ± Standard deviation	99.1%	99.9%	99.9%
Urea test result before starting dialysis	Mean ± Standard deviation	99.5%	99.8%	100.0%
Hemoglobin test results before starting dialysis	Mean ± Standard deviation	85.1%	87.7%	88.6%
Creatinine test result before starting dialysis	Mean ± Standard deviation	77.8%	88.5%	97.1%

eGFR using CKD EPI	Mean ± Standard deviation	83.9%	89.1%	96.6%
Serum albumin test result before starting dialysis	Mean ± Standard deviation	67.5%	69.3%	72.8%

++ The covariate constrained randomization also included two cluster-level baseline characteristics: cluster size at the time of randomization and the historical rate of cardiovascular-related death and hospitalization for myocardial infarction, ischemic stroke, and heart failure.

References for chapter 4 appendix

1. UCLA: Statistical Consulting Group. Factor analysis: SAS annotated output. Accessed February 21, 2020. https://stats.idre.ucla.edu/sas/output/factor-analysis/

2. Hatcher L, O'Rourke N. *A Step-by-Step Approach to Using SAS® for Factor Analysis and Structural Equation Modeling*. SAS Institute; 2013.

3. Quinn RR, Laupacis A, Austin PPC, et al. Using administrative datasets to study outcomes in dialysis patients: a validation study. *Med Care*. 2010;48(8):745-750. doi:10.1097/MLR.0b013e3181e419fd

4. Kokotailo R a, Hill MD. Coding of stroke and stroke risk factors using international classification of diseases, revisions 9 and 10. *Stroke*. 2005;36(8):1776-1781. doi:10.1161/01.STR.0000174293.17959.a1

5. Andrade SE, Harrold LR, Tjia J, et al. A systematic review of validated methods for identifying cerebrovascular accident or transient ischemic attack using administrative data. *Pharmacoepidemiol Drug Saf*. 2012;21 Suppl 1:100-128. doi:10.1002/pds.2312

6. Juurlink D, Preyra C, Croxford R, et al. *Canadian Institute for Health Information Discharge Abstract Database : A Validation Study*.; 2006.

7. Schultz SE, Rothwell DM, Chen ; Z, Tu ; K, Chen Z, Tu K. Identifying cases of congestive heart failure from administrative data: a validation study using primary care patient records. *Chronic Dis Inj Can*. 2013;33(3):160-166.

8. Weisstein EW. Eigenvalue. Accessed November 20, 2019. http://mathworld.wolfram.com/Eigenvalue.html

CHAPTER 5: CONCLUSION

6.1 Summary of findings

This dissertation aimed to **(1)** develop machine learning algorithms to identify CRTs in bibliographic databases, **(2)** assess reporting of methodological and ethical elements in hemodialysis-related CRTs, and **(3)** identify best practices for using covariate constrained randomization in hemodialysis-related, registry-based CRTs. We conducted three studies to address these research objectives.

Study 1 developed machine learning algorithms that have high sensitivity (>97%) and moderate to good specificity (>77%) in identifying CRTs from bibliographic databases. We created an open-source application to enable the use of our algorithms in practice.

Study 2 found that CRTs in hemodialysis have low methodological quality and sub-optimally report important ethical elements. There is an opportunity to improve the conduct and reporting of essential methodological and ethical issues in future CRTs in hemodialysis.

Study 3 assessed how well two covariate-constrained randomization methods balanced baseline characteristics compared with simple randomization. In a setting with 72 clusters, we found that constraining the randomization using historical information achieved a better balance on baseline characteristics than simple randomization. We concluded the study by proposing several recommendations for best practices for conducting covariate constrained randomization.

6.2 Study implications and future directions

In study 1, we created an open-source application for identifying CRTs using the information in the title, abstract, keywords and subject headings. Although we developed and tested our algorithms using EMBASE and MEDLINE, our application can be used for citations from any bibliographic database. The availability of our tool can dramatically decrease the time systematic review researchers spend screening CRT articles. If we had used this tool for our systematic review (study 2), we would have cut our screening time by at least one-half. We aim to continue training and fine-tuning our machine learning algorithms to improve the specificity of the algorithms. Beyond CRTs, our approach to using machine learning for article retrieval can be applied to identify citations for other designs, such as pragmatic trials.

In study 2, we found that CRTs in the HD setting often had poor reporting with regards to participant consent, where studies often did not have adequate details to assess why consent was sought (e.g., for enrollment, receiving the interventions, or data collection), from whom (e.g., from patients or providers), when (before or after randomization), and how (e.g., written or oral). Future research should expand to include all pragmatic randomized controlled trials conducted in the HD setting (i.e., individual and cluster-level randomized trials). Practices around participant consent may be related to trial design and considerations of risk. Ultimately, the goal is to standardize reporting practices for participant informed consent and guide trials in the HD setting to improve the reporting of consent procedures supported by well-grounded moral reasoning.

In study 3, we carefully thought through the covariate constrained allocation procedure for CRTs in the hemodialysis setting. The best practices identified in study 3 will inform choices we make in future trials, but it is not easy to generalize our results to other settings. Future work in this area should use formal statistical simulations.

6.2 Lessons learned

Each project had its unique set of challenges. In study 1, the biggest challenge was the time it took to train our algorithms. At the early stages of the project, I used a laptop with limited computing power resulting in slow progress. However, purchasing more powerful hardware was fruitful and sped the algorithm training processes by nearly 12 times. Another challenge was identifying credible resources to learn, develop and deploy the machine learning algorithms. While traditional learning (e.g., courses and textbooks) is valuable, I found resources on YouTube and Machine Learning blogs extremely helpful.

In study 2, I found the peer-review process to be time-consuming. As a result, we were asked by reviewers to update our systematic review twice. I quickly learned not to be discouraged by the added work because it is part of the scientific process.

In study 3, I learned to take my time to create highly efficient computer programming and creating detailed “read me” notes. My first computer program for this project was highly inefficient. Running the program took several days and, at times, crashed because of low computing memory. Taking the time to plan and execute the computer program resulted in more efficient code that ran several thousand times faster. Additionally, I found it extremely useful to write detailed programming notes. These notes helped when I

returned to work on the project several months later to address co-author and peer-review comments.

6.3 Concluding remarks

Regarding treatment effects, randomized trials provide high-quality evidence, and areas in medicine (e.g., cardiology, oncology) with the highest number of randomized controlled trials have experienced the most transformative improvements in outcomes and patient care. This dissertation addressed specific CRT-related knowledge gaps. Our results will directly inform the review, design and analysis of future hemodialysis CRTs.

References

1. Li F, Turner EL, Heagerty PJ, Murray DM, Vollmer WM, DeLong ER. An evaluation of constrained randomization for the design and analysis of group-randomized trials with binary outcomes. *Stat Med*. 2017;36(24):3791-3806. doi:10.1002/sim.7410