**Веб-сервис распределения пациентов в группы при проведении клинических исследований**

И. В. Ворошин, В.Л. Лукинов

Большой частью клинических исследований являются медицинские исследования в параллельных группах объектов с целью выявления наилучшего способа вмешательства для лечения или профилактики. Для данного дизайна клинических исследований важной проблемой является контроль сопоставимости параллельных групп для чистоты выводов о результатах разных вмешательств в исследуемых группах.

Для обеспечения сопоставимости исследуемых групп в настоящее время применяются различные методы рандомизированного распределения объектов в группы и методы стратификации, а также комбинации этих методов.

Другой проблемой является «защита» схем распределения от недобросовестного изменения схем исследователями. Решением является «ослепление» исследователей, то есть сокрытие предварительной общей схемы.

Цель работы – разработка онлайн-сервиса для распределения пациентов в группы в клинических исследованиях по схеме двойного слепого исследования (иными словами, чтобы о разбиении на группы не знали ни участники групп, ни исследователи). Схема формирования групп будет скрыта до полного завершения исследования, принадлежность пациентов к группе будет выдаваться при его регистрации в исследование. Дополнительной особенностью будет ведение лога – записей всех событий, несколько уровней пользователей с разными правами доступа, возможность проведения многоцентровых исследований с ведением пациентов в разных центрах разными исследователями в рамках одного проекта, а также свободный доступ к схеме распределения после завершения проекта.

*Ключевые слова*: клиническое исследование, ослепление, рандомизация, стратификация, веб-сервис.

1. Введение

Рандомизированное клиническое исследование (РКИ) – это клиническое исследование, при котором пациенты случайным образом разбиваются по группам лечения и получают одно из двух или более объектов исследования. При этом группа, получившая новое лечение, называется экспериментальной, а группа, получившая стандартное лечение или плацебо, – контрольной. Такие группы являются параллельными по отношению друг к другу и будут получать выбранное лечение от начала до конца исследования.

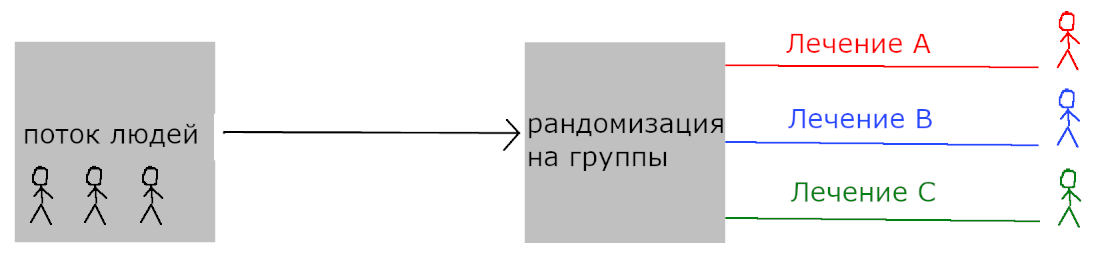


Рисунок 1. Схема проведения РКИ в параллельных группах

Важным моментом РКИ является максимальная схожесть групп для обеспечения уверенности в том, что результат исследования будет связан лишь с получаемым препаратом. Решить эту проблему помогают методы рандомизации и стратификации групп.

1. **Рандомизация групп в РКИ**
   1. **Рандомизация**

Рандомизацию можно определить, как процедуру случайного выбора элементов. Её разделяют на фиксированную (простая, блоковая, стратифицированная) и динамическую (метод «несимметричной монеты», адаптивная рандомизация). [5]

Методы динамической рандомизации заключаются в динамических изменениях вероятностей в зависимости от разницы количества набранных людей в ту или иную группу. Минус такого подхода заключаются в сложности проведения, а также необходимости частичного раскрытия данных. Рассмотрим поподробнее фиксированные методы.

Простая (обычная) рандомизация создаёт последовательность случайных значений. Наглядный пример – подбрасывание монетки, определяющей, в какую группу пойдет испытуемый. К этому типу также относятся использование таблиц случайных чисел и метод конвертов. Проблема этих способов заключается в формировании неравных групп по количеству участников, дисбаланса по различным признакам (возрасту, полу, степени заболевания и т.п.), возможности подтасовки и нарушении сокрытия распределения [2]. Поэтому к настоящему времени такой подход является устаревшим.

Метод блочной рандомизации справляется с проблемой сбалансированности групп по количеству участников, так как, пока не закончится потребность в увеличении количества испытуемых, будут формироваться и заполняться блоки небольших размеров, кратных сумме методов исследования. В этих блоках и проводится рандомизация – в каждом будет случайным образом формироваться последовательность из методов лечения, которые в сумме равны друг другу по количеству. Недостаток данного метода – то что исследователь, заранее зная последовательность, может «подтасовывать» пациентов по своему желанию или, зная информации о генераторе псевдослучайных чисел (ГПСЧ), спрогнозировать, в какую группу пойдет следующий испытуемый, что может серьезно повлиять на ход исследования и достоверность её результатов (например, если известен размер блока, предыдущие распределения внутри блоков и одна из групп уже укомплектована). Данная проблема может решаться случайным определением размера блоков с помощью ГПСЧ, либо сокрытием информации о величине блоков.

В стратифицированной рандомизации при формировании групп учитываются некоторые важные отличительные признаки, например, тяжесть состояния или возраст. По этим признакам пациенты разделяются на страты, в которых осуществляется рандомизированный отбор в одну из групп. Таким образом, сформированные выборки получатся репрезентативными по отношению ко всей группе в целом. Однако реализация данного способа является довольно сложной, особенно при необходимости учёта нескольких страт.

* 1. **Генерация случайных и псевдослучайных чисел**

На сегодняшний день актуальными методами рандомизации являются блочные методы, использующие генераторы случайных и псевдослучайных чисел (ГСЧ и ГПСЧ).

ГСПЧ использует в качестве точки опоры начальное число (seed, value), от которого программно по некоему алгоритму преобразований и генерируется последовательность чисел, по своим свойствам близкая к случайной. Эта последовательность сводится к формуле (линейный конгруэнтный генератор):



, где a – множитель, c – приращение, m – модуль.

Получается, проблема такой генерации заключается в невозможности создания полностью случайных последовательностей, так как каждое новое число зависит от предыдущего. То есть, наблюдая за полученными последовательностями, имеется возможность найти закономерность между числами, решив систему уравнений, и предсказать следующую генерацию [6, с. 161-176].

Главное отличие ГСЧ от ГСПЧ в том, что ГСЧ в качестве значений берет высококачественные числа, полученные аппаратным путём из естественных источников случайных чисел, физических шумов, таких как космическое излучение или ионизирующая радиация. Полученные таким путём числа также можно использовать в качестве задаваемых начальных значений (seed) для ГСПЧ. Недостатки таких генераторов включают в себя высокую стоимость, трудозатратность установки и настройки, более медленный процесс генерации, а также невозможность воспроизведения ранее сгенерированных последовательностей.

* 1. **Ослепление**

Следующей важной задачей является «ослепление» - т.е. сокрытие исследования от лиц, которые могут повлиять на процесс. метод и результат рандомизации принято скрывать от участников: пациентов – во избежание эффекта плацебо, и исследователей – в случае двойного слепого метода, чтобы нейтрализовать возможность недобросовестного вмешательства в распределение пациентов (такой принцип считается более надежным и эффективным для достоверности результатов). В случае исследования, в котором результат зависит от субъективной оценки исследователя, а также есть возможность сделать вывод по видимым признакам – результаты должны быть выведены независимым экспертом, которого тоже можно ослепить (ослепление третьих лиц).

В разрабатываемом сервисе функция ослепления будет задействована в сокрытии блоков распределения пациентов до завершения исследования.

1. **Основные особенности сервиса**

* 1. **Алгоритм рандомизации**

В отрытом источники уже существует несколько библиотек программирования для проведения рандомизации в клинических исследованиях (blockrand, randomizeR, pwr, experiment). Одна из них – пакет «blockrand» под язык программирования R [4]. В его функционал входит генерация блоковой рандомизации для РКИ, включая стратифицированный отбор и случайную последовательность сгенерированных блоков. При проведении стратифицированного клинического исследования необходимо запустить функцию по одному разу для каждой страты, а после объединить всё в одну таблицу стандартной функцией rbind.

Вот так выглядит вызов функции генерации списка рандомизации с не менее 60 назначениями и блоками случайных размеров по 3, 6 или 9 человек:

bfla <- blockrand(n = 60,

num.levels = 3, *# three treatments*

levels = c("CS", "CS/Tofa", "CS/Upa"), *# arm names*

stratum = "Bfail.LowAlb", *# stratum name*

id.prefix = "BfLA", *# stratum abbrev*

block.sizes = c(1,2,3), *# times arms = 3,6,9*

block.prefix = "BfLA") *# stratum abbrev*

bfla

n – минимальное количество назначений

num.levels задаёт сколько видов вмешательств будет использоваться

strateum – название страты

id.prefix задает приставку к идентификатору id

block.sizes – размерность блоков

block.prefix задаёт приставку к идентификатору блоков block.id



Рисунок 2. Часть получившейся таблицы распределения случайных размеров пакетом blockrand с тремя методами лечения

Создав таблицу распределения пациентов, при записи новых пациентов они будут заполнять блоки и соотноситься к одному из видов лечения. В теории исследователь, пользуясь этой схемой, должен распределять по группам людей из потока, но, этот момент нельзя проверить, и исследователь может воспользоваться этим слабым местом и распределять пациентов по своему усмотрению в целях фальсификации. Поэтому схемы распределения, как и вид лечения каждого участника будут скрыты до окончания проведения исследования – узнать схемы можно будет в конце при участии пользователя с должными высокими правами доступа.

* 1. **Шифрование**

Сохранить сгенерированные блоки распределения в виде файла-таблицы с его сокрытием от лиц представляется возможным с помощью алгоритмов шифрования данных на основе криптосистем с открытым ключом [1, с. 12-40]. Имея со стороны сервиса информацию, которую необходимо закодировать, а также используя секретный(е) ключ(и), получается зашифрованное сообщение. То есть благодаря таким алгоритмам, например, шифр Вернама, можно как шифровать, так и расшифровывать файлы любого типа.

* 1. **Логирование**

Логи (лог-файлы) — это файлы, которые содержат некоторую информацию работы сервера или ПК, куда заносятся определенные действия программы или пользователя. То есть под логированием понимается ведение таких файлов. Лог разрабатываемого сервиса будет содержать запись всех действий, совершаемых в проекте, в виде определенной структуры (таблицы). В каждой записи полагается сохранять:

1. время совершения операции, тип данных – дата и время;
2. вид совершения операции, тип данных – строка;
3. вызвавший операцию пользователь (actor), тип данных – идентификатор (id) пользователя;
4. результат операции (успех/неудача), тип данных – bool (1 или 0).
   1. **Уровни доступа пользователей**

1. Администратор сервиса. Доступные операции: удаление, блокировка проекта, управлением доступом других пользователей.
2. Администратор проекта – роль управления проектом. Доступные операции: создание проекта, предоставление доступа к проекту, назначение роли администратора другим пользователям.
3. Суперпользователь – пользователь с особыми правами. Получает доступ к логам операций в проекте.
4. Обычный пользователь. Имеет доступ к итоговой информации о проекте.

Литература

1. *Рябко Б. Я., Фионов А. Н.* Криптографические методы защиты информации: учебное пособие для вузов. М.: Горячая линия–Телеком, 2005. 229 с.
2. <https://trialsjournal.biomedcentral.com/articles/10.1186/1745-6215-13-90>
3. <https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-021-01303-z>
4. <https://bookdown.org/pdr_higgins/rmrwr/randomization-for-clinical-trials-with-r.html>
5. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3136079/>
6. *Niederreiter H.* Random Number Generation and Quasi-Monte Carlo Methods. – SIAM, 1992. – 247 p. <https://www.ricam.oeaw.ac.at/files/people/siambook_nied.pdf>

**Ворошин Илья Викторович**

Студент кафедры телекоммуникационных систем и вычислительных средств СибГУТИ (630102, Новосибирск, ул. Кирова, 86), e-mail: Iliya-Voroshin-2000@ya.ru.

**Лукинов Виталий Леонидович**

Доцент кафедры телекоммуникационных систем и вычислительных средств СибГУТИ (630102, Новосибирск, ул. Кирова, 86), e-mail: Vitaliy.lukinov@gmail.com.

**Web service for the distribution of patients into groups during clinical trials**

**I. Voroshin, V. Lukinov**

A large part of clinical research is medical research in parallel groups of objects in order to identify the best method of intervention for treatment or prevention. For this design of clinical trials, an important problem is to control the comparability of parallel groups for the purity of conclusions about the results of different interventions in the studied groups.

To ensure comparability of the studied groups, various methods of randomized distribution of objects into groups and stratification methods, as well as combinations of these methods, are currently used.

Another problem is the "protection" of distribution schemes from unscrupulous modification of schemes by researchers. The solution is to "blind" the researchers, that is, conceal the preliminary general scheme.

The aim of the work is to develop an online service for the distribution of patients into groups in clinical trials according to the double–blind study scheme (in other words, so that neither group participants nor researchers know about the division into groups). The scheme of group formation will be hidden until the study is fully completed, the patients' belonging to the group will be issued when it is registered in the study. An additional feature will be the logging of all events, several levels of users with different access rights, the possibility of conducting multicenter studies with patients in different centers by different researchers within the same project, as well as free access to the distribution scheme after the completion of the project.

*Keywords*: clinical trial, blinding, randomization, stratification, web service.