

Отсев и репрезентативность при разработке и валидации онлайн-сервисов проверки симптомов — тематическое исследование опросника Rheumatic?

Источник: Frontiers in AI — Medicine

Оригинал: <https://www.frontiersin.org/articles/10.3389/frai.2026.1815241>

валидация моделей

исследования

онлайн-диагностика

предвзятость данных

ревматология

Введение

Онлайн-сервисы для проверки симптомов часто разрабатываются и валидируются на данных, подверженных эффекту самоотбора и селективному отсеvu (selective attrition), что потенциально вносит систематические ошибки (biases) в прогностические модели.

Цели

Оценить закономерности набора, отбора и отсева участников в крупном голландском онлайн-сервисе проверки симптомов при скелетно-мышечных жалобах, а также оценить потенциальные систематические ошибки путем сравнения характеристик участников в зависимости от источников набора и с внешними целевыми популяциями.

Методы

Используя данные голландского онлайн-опросника **Rheumatic?** по скелетно-мышечным жалобам, мы сравнили исходные характеристики и ключевые самозаявленные симптомы у респондентов последующего (follow-up) опроса и лиц, не ответивших на него. Респонденты опроса дополнительно сравнивались в зависимости от источника набора в опросник, а именно: через клиники первичной медицинской помощи, клиники вторичной медицинской помощи или через различные онлайн-источники. Распределение по полу, возрасту и **ИМТ** (индекс массы тела) во всей исследуемой группе сравнивалось с внешними данными потенциальных целевых популяций пациентов первичного и вторичного звена медицинской помощи в Нидерландах.

Результаты

Общая исследуемая группа ответов на опросник составила 31 457 респондентов, из которых 50% (n = 15 591) ответили на последующий опрос. Участники исследования были преимущественно женщинами (76%), среднего возраста (одна треть — 50–60 лет), никогда не курившими (66%) и имели избыточный вес. В то время как участники, набранные через учреждения здравоохранения, напоминали целевые популяции, респонденты последующего опроса были старше, имели больше ревматических диагнозов (49% против 32%) и сообщали о большем количестве симптомов, чем лица, не ответившие на опрос. Характеристики участников варьировались в зависимости от источника набора: социальные сети привлекали более молодых женщин, в то время как каналы здравоохранения охватывали более разнообразные популяции с различными проявлениями симптомов.

Заключение

Закономерности набора и отсева привели к различиям в характеристиках участников. Набор через учреждения здравоохранения обеспечил участие лиц, напоминающих намеченные целевые популяции, а респонденты последующего опроса отличались от лиц, не ответивших на него, по некоторым пунктам. Понимание этих процессов отбора имеет важное значение при использовании данных о проверке симптомов из реальной клинической практики (real-world data) для разработки моделей.

Перевод выполнен: 15.05.2026 | ai4med.ru

Машинный перевод. Рекомендуем сверять с оригиналом при клиническом использовании.