Да здравствует статистика, но не метафизика!

В номере, который сейчас держит в руках уважаемый читатель, опубликована очередная статья коллектива авторов из Федерального научно-клинического центра реаниматологии и реабилитологии, посвященная методам и правилам статистической обработки данных в медицинских исследованиях. Валерий Владимирович Лихванцев и его коллеги шаг за шагом воплощают в жизнь родившийся у нас еще летом 2020 г. план — сделать для нашего «Зеленого журнала» серию обзоров по всем узловым вопросам биомедицинской статистики.

Задача эта действительно важна, поскольку не только начинающие исследователи, но и маститые рецензенты топовых журналов, пусть и превосходно владея свой предметно-конкретной областью, подчас допускают грубые погрешности в трактовке и правилах представления результатов обработки данных. Чего стоят, например, такие замечания: *«Раз P>0,05, то авторы не имеют права писать о том, что средние величины в группах отличаются между собой...»* или *«В медицинских публикациях не принято писать P=0,0045 или P<10^{-6}, можно писать только P<0,05 или P<0,01...»!* Да почему же, оставаясь в пределах самой жесткой математической строгости, авторы не могут указать, что при нормальном распределении данных их средние значения в группах отличаются между собой (не утверждая при этом, что группы не относятся к одной генеральной совокупности!), почему они не могут указать точное значение доверительной вероятности нулевой гипотезы, если использованные программы позволяют им его рассчитать?! Это же описание вероятностных, а не детерминированных свойств объективной реальности! — так почему же не вести его по всем правилам математической теории вероятностей?

Да, в мировой медицинской науке торжествует культ статистики, и он исторически оправдан как путь выхода из безысходности многовековой эмпирической практики, многие позиции которой равно неопровержимы и недоказуемы. Но когда этот культ обретает свойства религиозного, отрываясь от вполне рациональной основы статистики и дрейфуя к аксиоматике заповедей и незыблемым вердиктам (построенным на *вероятностном анализе статистических гипотез!*), становится как-то не по себе...

Оказывается, ASA — это не только Американское общество анестезиологов, созданное в 1905 г., но и American Statistical Association, основанная в том же Бостоне, штат Массачусетс — правда, за целых семь лет до первого наркоза! В марте 2016 г. именно эта ASA пошла на экстраординарный шаг, опубликовав в своем главном журнале специальное обращение о статистической значимости величины доверительной вероятности нулевой гипотезы P. Шесть тезисов этого отрезвляющего документа заслуживают того, чтобы быть процитированными без купюр, но с комментариями, переведенными на язык медицины Elizabeth L. Whitlock и Catherine L. Chen — авторами завершающей 90-й главы «Интерпретация медицинской литературы» в 9-м издании «Анестезии Рональда Миллера» (2020).

1. Значение P может указывать на то, что данные несовместимы с определенной статистической моделью. *Если модель явно несовместима, значение P часто будет небольшим (например, потому, что <данные> не совпадают с допущениями модели). Малое значение P не подразумевает, что выбранная модель хорошо объясняет данные.*
2. Значение P не измеряет вероятность того, что исследуемая гипотеза верна, или вероятность того, что данные были получены только в результате случайности. *Лежащие в основе аналитические решения — которые могут быть очень сложными — вносят такой же вклад в релевантность анализа, как выбор модели и статистического теста.*
3. Научные выводы, деловые или политические решения не должны быть основаны только на том, что значение P вписывается в какой-либо порог. *Вывод не более или менее верен при P=0,04, чем при P=0,06; оба эти порога взяты произвольно!*
4. Получение правильного заключения невозможно без полной отчетности и прозрачности. *Значение P — это продукт аналитического планирования и самих данных; следовательно, должно быть представлено полное описание всех проверяемых гипотез (в том числе тех, где значение P считается «незначимым»), чтобы предотвратить выборочное или предвзятое суждение о статистически значимых результатах.*
5. Значение P, или статистическая значимость, не измеряет размер эффекта или важность результата. *Тонкие и клинически несущественные различия могут иметь крошечное значение P, особенно при высокоточных измерениях и/или большом размере выборки, а клинически важные различия могут быть «незначительными» если измерения неточные или размер выборки невелик. Статистическая значимость — не показатель клинической значимости!*
6. Само по себе значение P не обеспечивает хорошую меру доказательства относительно модели или гипотезы. *Значение P неотделимо от подлежащей методологии и, без контекста, дает неадекватную информацию о размере эффекта и настолько неясно, что не подлежит интерпретации.*

Давайте же не будем превращать наши сегодняшние оценки, основанные на достаточно условных договоренностях, в непреложный закон, вытекающий из не нуждающейся в доказательствах аксиомы. И еще раз поблагодарим проф. В.В. Лихванцева и его коллег, терпеливо объясняющих вполне земные основы обожествляемого нами жанра медицинской статистики.

Рискну в заключение привести здесь победный вывод одного клинического исследования из номера 3 нашего «Зеленого журнала» за 2044 г. — конечно, уже целиком и полностью электронного: *«При условии, что распределение изучаемых данных о пациентах в описанной нами клинической модели отлично от нормального, вероятность того, что пациенты основной и контрольной групп относятся к одной генеральной совокупности, оценивается U-критерием Wilcoxon—Mann—Whitney как меньшая, чем 5%...»*

Наука требует точности — и я уверен, что она ее в итоге получит.

К. Лебединский



Источники

1. Wasserstein RL, Lazar NA. The ASA's Statement on p-Values: Context, Process, and Purpose. *The American Statistician*. 2016;70:129-133. https://doi.org/10.1080/00031305.2016.1154108
2. Whitlock EL, Chen CL. Interpreting the Medical Literature. Chapter 90 in: Miller's Anesthesia. Ninth Edition. *Elsevier.* 2020;2:2813-2824.