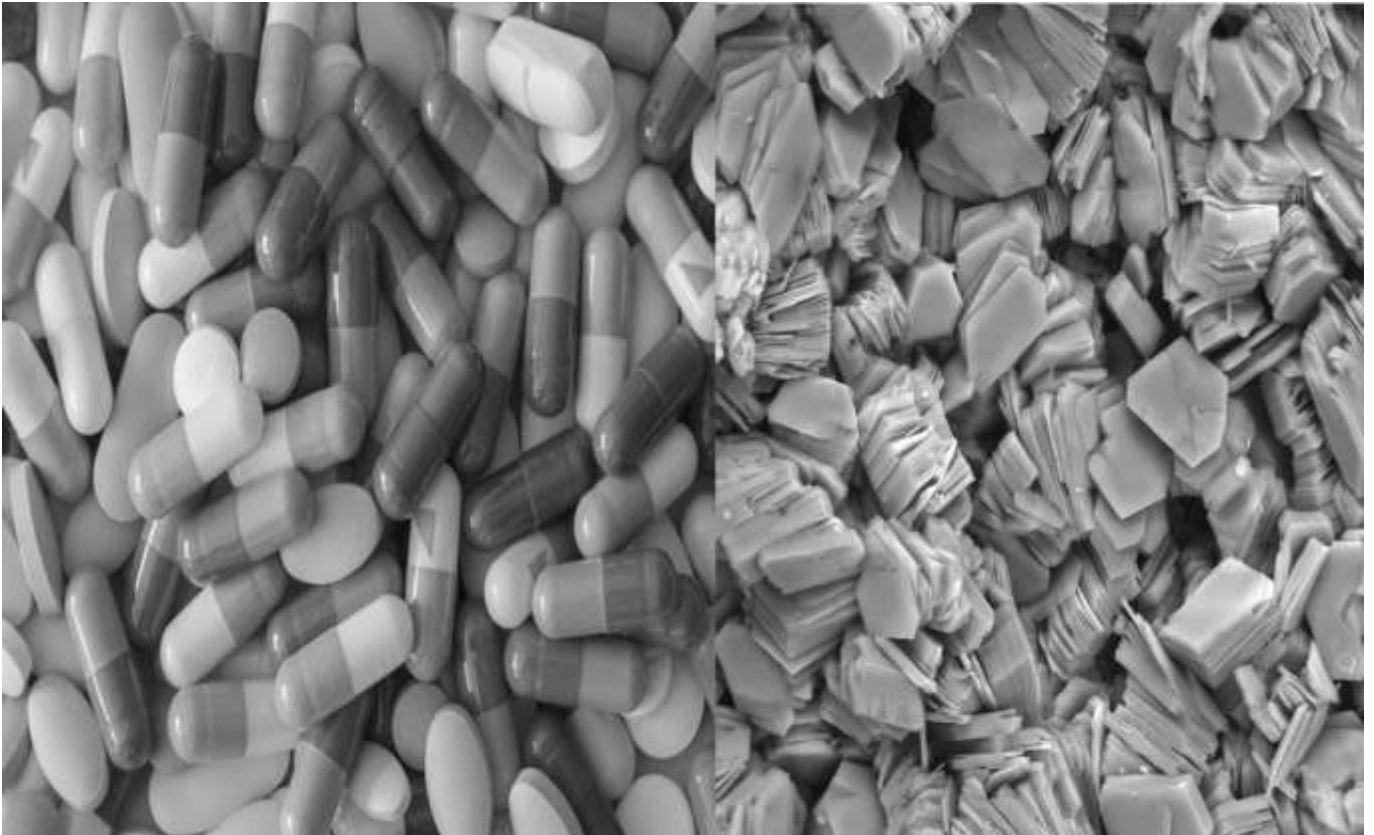


Доказательная геотехника



Интересная, но все же достаточно спорная статья инженера-геотехника, в которой утверждается, что критерии оценки качества экспериментальных исследований, которые были разработаны и апробированы в рамках развития доказательной медицины, могут распространяться и на прочие области научного знания, в том числе, на геотехнику. Например, по мнению автора, выполнение нескольких простых методических правил, таких как «самоослепление» и рандомизация, позволило бы сделать результаты исследований инженеров-геологов и геотехников менее предвзятыми и, зачастую, «менее положительными».

Статья публикуется в порядке дискуссии.

Мирный Анатолий Юрьевич

Старший научный сотрудник Геологического факультета МГУ им. М.В. Ломоносова, руководитель проекта «Независимая геотехника»

В 2019 году была опубликована книга Петра Талантова «0,05. Доказательная медицина от магии до поисков бессмертия», пользующаяся огромной популярностью у медицинских работников и исследователей в области биологии. Как следует из названия, книга посвящена истории развития доказательной медицины – подходе, при котором решение о применении того или иного метода лечения основывается на имеющихся доказательствах их эффективности и безопасности. При этом достоверность этих доказательств и проверок обеспечивается планированием исследований в соответствии с определенным набором правил.

Во-первых, в исследование включаются две группы – экспериментальная и контрольная, причем распределение по группам происходит случайным образом. Каждая из групп проверяется на репрезентативность, то есть на соответствие той группе пациентов, которую она представляет.

Во-вторых, сравнивается действие проверяемого препарата или метода с отсутствием лечения (плацебо), либо с наиболее распространенным на настоящий момент методом лечения. При этом сравнение результатов воздействия в двух группах выполняется с учетом статистической достоверности (например p -критерия), а не просто по абсолютным средним значениям оцениваемых параметров. В качестве пограничного значения принимается некоторый уровень значимости, например $p < 0,05$ (очень условно: только в 5% случаев между двумя группами не будет достоверной разницы), однако это не единственный способ.

В-третьих, применяется так называемое «ослепление», как пациентов, так и принимающих участие в эксперименте специалистов. Это означает, что ни непосредственный участник исследования, ни врач не знают, к какой группе относится пациент. Такой подход позволяет уменьшить или устранить источники экспериментальной предвзятости.

Очевидно, что такая сложная схема экспериментальных исследований требует значительных экономических затрат, что не всегда возможно для отдельных исследователей или научных организаций. В связи с этим существует проблема попыток прямого или косвенного влияния на результат исследования спонсирующих их коммерческих организаций – чаще всего, фармацевтических фирм. В качестве контрмеры разработаны методы проверки результатов, позволяющие выявить подтасовки и недобросовестные исследования. Кроме того, на основании мета-анализа результатов нескольких исследований могут быть сделаны общие выводы с учетом статистической мощности каждого из рассматриваемых опубликованных результатов.

Казалось бы, какое отношение это имеет к области инженерных изысканий и геотехнического строительства? Как ни странно, самое прямое: в рамках развития доказательной медицины были разработаны и – самое главное – апробированы критерии оценки качества экспериментальных исследований, которые могут распространяться и на прочие области научного знания.

Технические науки, и, в частности, геотехника в этом отношении находятся в более благоприятном положении относительно медицины по нескольким причинам:

- объект исследования – образец грунта или геотехническая конструкция – в отличие от пациентов врача, «не врет», то есть его реакция на воздействие определяется не восприятием лечения или отношением к экспериментатору, а исключительно условиями этого воздействия. При этом оценка параметров предельно объективна, так как для этого используются исключительно измерительные приборы.

- количество побочных факторов, влияющих на реакцию объекта, существенно меньше, чем в случае биологических систем. При этом каждый из этих факторов, в свою очередь, уже достаточно изучен и предсказуем на основании априорных представлений о физической природе процессов.

- точность измерения, как параметров воздействия, так и реакции объекта позволяет поддерживать низкий уровень погрешности для частных образцов/испытываемых конструкций.

- количество отдельных образцов в выборке может быть установлено исследователем на этапе планирования эксперимента. В медицинских исследованиях

количество пациентов-участников может изменяться в ходе эксперимента в силу различных причин.

Тем не менее, некоторые из решений доказательной медицины могли бы применяться, но не применяются при проведении исследований. Например, «ослепление» экспериментатора. Представим себе, что выполняется серия лабораторных опытов, определяется модуль деформации в определенных условиях. В идеальном случае непосредственно работать с образцом и испытательным оборудованием должен один специалист, а интерпретацию данных выполнять другой. Если образцы при этом имеют только идентификационный код, то можно исключить предвзятость. Но даже если опыт выполняется и интерпретируется одним человеком, можно реализовать кодирование образцов, используя сложные цифробуквенные коды, исключая случайное запоминание.

Большие вопросы вызывает объем исследуемой выборки – количество образцов в конкретном исследовании. Даже в случае лабораторных испытаний крайне редко группа включает более 30 образцов, а при полевых испытаниях или натуральных испытаниях конструкций еще меньше, вплоть до единичных испытаний. В результате на частных случаях строятся выводы, которые никак не могут носить общего характера. Подобные исследования следовало бы относить к так называемым «case stories» или, в медицинской терминологии, «клиническим случаям» – отдельным практическим результатам, представляющим интерес для исследователей.

Отдельная обширная тема – мета-анализ. Данный метод позволяет объединить результаты нескольких опубликованных исследований в общую выборку и проверить общую для них гипотезу с учетом статистической мощности каждого. Существуют поправки на влияние неопубликованных исследований – ведь обычно публикуются только положительные результаты, однако во многих исследованиях мог быть получен и отрицательный результат. Применительно к геотехнике проведение мета-анализа позволило бы перейти от обзора литературы и только качественной оценки чужих результатов к количественному использованию в совместной статистической обработке. В настоящее время использовать результаты опубликованных исследований крайне сложно – в работах они редко приводятся в достаточном объеме для повторной интерпретации, не говоря уже о больших расхождениях методики исследований.

Наконец, основная проблема – достоверность результатов. Можно привести простой пример: стоит задача оценки результатов закрепления грунтов основания. Выполнены испытания 30 образцов, отнесенных к одному ИГЭ, по результатам испытаний установлено среднее (оно же нормативное) значение модуля деформации равное 18 МПа. После этого выполнено закрепление, отобраны 12 контрольных образцов со средним значением модуля деформации 25 МПа. Казалось бы, все отлично, закрепление дало результаты. Однако если сопоставить доверительные интервалы, то может оказаться, что они пересекаются – проще говоря, нельзя обоснованно утверждать, что две эти выборки не могут быть объединены в одну. Вместо этого следовало бы воспользоваться критерием выделения РГЭ согласно ГОСТ 20522 и таким образом доказать, что элемент до закрепления и после закрепления являются двумя разными расчетными элементами.

Доказательной медициной вводится рейтинговая система оценки качества клинических рекомендаций, основанных на результатах исследований. На низшей ступени доказательности находятся рекомендации, выработанные группой экспертов. Классом выше – нерандомизированные исследования, то есть те, где распределение образцов между экспериментальной и контрольной группой осуществлено неслучайным образом (по

выбору экспериментатора), либо контрольная группа вовсе отсутствует. Далее следуют рандомизированные исследования, и замыкают рейтинг исследования с двойным ослеплением. Нетрудно заметить, что в области геотехнического строительства большинство исследований находится на двух нижних ступенях – примерно на уровне середины XVIII века, когда врач британского флота Джеймс Линд впервые клинически доказал эффективность цитрусовых против цинги.

Справедливости ради следует отметить, что аналогичные недостатки присущи и зарубежным геотехническим исследованиям – большинство авторов считают вполне уместным построение общих выводов на результатах испытаний 5–10 образцов без повторения опытов и тем более ослепления. Очевидно, что подобные публикации могут быть отнесены к категории «коротко об интересном» и не годятся для выработки практических рекомендаций. Между тем зачастую авторы занимаются поиском научных объяснений для статистических выбросов, и, самое главное, находят его!

Конечно же, напрашивается возражение: допустимо ли распространять методику, предназначенную для оценки эффективности лекарственных препаратов, на исследования свойств грунтов? Действительно, пример с закреплением вроде бы подходит хорошо – вот состояние «до воздействия» (контрольная группа), вот «после воздействия» (закрепленный грунт). А с чем сравнивать, если речь идет о других видах исследований, например, сопротивлении грунтов сдвигу или влиянии переуплотнения на объемную сжимаемость?

Все очень просто: в этом случае групп может оказаться несколько. Исследования с контрольной группой самые простые, в их ходе проверяется «нулевая гипотеза»: «0» – воздействие не вызывает реакции, «1» – воздействие вызывает достоверную реакцию. Но если между воздействием и реакцией подразумевается линейная связь, то следует сопоставлять между собой уже две группы, подверженные воздействию разной интенсивности (например, испытанные в трехосных приборах при двух различных давлениях в камере). При этом распределение образцов должно быть случайным. А если предполагается нелинейная связь, то таких групп должно быть не менее трех, чтобы установить характер нелинейности... Это уже относится к области факторного анализа.

Целью этой небольшой публикации вовсе не является снижение ценности выполняемых экспериментальных исследований: все практикующие специалисты знают, что обеспечение достаточного объема выборок в геотехнике связано с большими трудностями. Это и отбор нужного количества образцов, и их качество, и влияние пространственной изменчивости свойств. Но даже в этом случае выполнение нескольких простых методических правил, таких как «самоослепление» и рандомизация позволило бы сделать результаты наших исследований менее предвзятыми, и, зачастую, «менее положительными».

Нечего скрывать, и сам автор публиковал работы, результаты которых не выдержали бы критического анализа с точки зрения статистики. Все они скорее относятся к упомянутым выше case stories, и не претендуют на общее распространение выводов. Но что делать, когда результаты таких исследований оказываются частью нормативных технических документов? Когда частные результаты безо всяких оснований распространяются на общие совокупности? Об этом следует задуматься каждому неравнодушному специалисту и ответственному исследователю.