Приложение № 6

к Методическим рекомендациям по подготовке дизайна

аудита соответствия и аудита эффективности

# Потенциальные проблемы валидности

**1. Точность и процессы измерения (операционная валидность)**

1.1. При соблюдении **операционной валидности** собираемые данные позволяют измерить то, что планировалось измерить (для количественных данных) или отражают понятия и явления, которые планировалось изучить (для качественных данных).

1.2. Возможные проблемы операционной валидности:

1)проблеманекорректной операционализациивозникает в ситуации недостаточности знаний об интересующей характеристике предмета, когда данная характеристика не поддается измерению, или это измерение затруднено (например, ввиду того, что издержки измерения высоки). В этих случаях могут использовать так называемые прокси-показатели, которые служат показателем-заменителем, наилучшей альтернативой для измерения. Например, показатель ВВП на душу населения является прокси-показателем экономического благосостояния и качества жизни;

|  |
| --- |
| **Пример нарушения операционной валидности** |
| При операционализации понятия «бедность» показатель «доход ниже уровня прожиточного минимума» представляет собой некорректный прокси-показатель, если интересует аспект деприваций – если исследователя интересуют возможные ограничения доступа населения к различным благам: к доступному жилью, качественному здравоохранению, качественному школьному образованию, чистой воде и т. п. (такие ограничения имеются далеко не у одних только малоимущих, а у малоимущих не всегда есть такие ограничения) |

2) проблема преднамеренного введения в заблуждение возникает, когда представитель источника информации (например, автор ответа на запрос Счетной палаты или респондент) намеренно искажает факты, чтобы скрыть проблему или защитить себя;

3) проблема непреднамеренного введения в заблуждениевозникает в результате ошибки воспоминаний (например, при прошествии длительного периода времени после того, как произошло исследуемое событие) или несвоевременного обновления данных;

|  |
| --- |
| **Пример нарушения операционной валидности** |
| Исполнитель или участник программы непреднамеренно дает ложную информацию из-за ошибочных воспоминаний о фактах или событиях |

4)проблема стремления к социальной приемлемости и дискомфорта при оценкепри передаче данных о себе в случае, когда это является некомфортным или даже неприемлемым для определенных групп;

|  |
| --- |
| **Пример нарушения операционной валидности** |
| Респондент говорит то, что, по его мнению, интервьюер хочет услышать для получения одобрения |

5)проблема отложенного эффекта возникает в ситуации, когда эффекты реализуются позднее момента их измерения (то есть тогда, когда измерение проводится преждевременно);

|  |
| --- |
| **Пример нарушения операционной валидности** |
| Оценка воздействия программ продвижения здорового образа жизни на уровень смертности от болезней сердечно-сосудистой системы может иметь отложенный эффект |

6)изменение определений и методик расчета.Переопределение понятий или способов расчета показателей делает данные, собранные в различные периоды времени или на различных территориях, несопоставимыми;

|  |
| --- |
| **Примеры нарушения операционной валидности** |
| До 2021 года уровень бедности определялся как доля населения с доходами ниже величины прожиточного минимума, которая определялась как стоимость потребительской корзины для основных социально-демографических групп населения. Начиная с 2021 года величина прожиточного минимума стала определяться как доля в 44,2 % медианного дохода за предыдущий год. Такое радикальное изменение методики расчета прожиточного минимума ведет к тому, что уровень бедности, вычисленный как доля населения с доходами ниже величины прожиточного минимума, определенной по новой методике, окажется несопоставимым с уровнями бедности за предыдущие годы.  В методологию расчета показателей федерального проекта «Промышленный экспорт» и национального проекта «Международная кооперация и экспорт» начиная с 2021 года (2020 год – базовый) многократно вносились изменения. Так, в январе 2021 года в методику расчета показателей внесены изменения, которые, в том числе предусматривают изменение перечня кодов единой Товарной номенклатуры внешнеэкономической деятельности Евразийского экономического союза, что привело к несопоставимости оценок несырьевого неэнергетического экспорта в различные годы реализации федерального проекта |

*7*) искажение, связанное с использованием одного метода *–* использование только одного метода для сбора данных о явлении (например, самооценка);

|  |
| --- |
| **Примеры нарушения операционной валидности** |
| Использование только результатов анкетирования получателей для оценки эффективности мер поддержки в форме субсидий.    Использование только результатов интервьюирования получателей пособий для оценки эффективности мер поддержки в форме этих пособий |

8) отсутствие понимания культурного контекста *–* невозможность корректной интерпретации понятий и результатов измерений без учета точек зрения участников программы;

9) проблемавозможности неоднозначного восприятиявозникает, когда вопросы слишком сильно зависят от субъективных оценок значения вариантов ответа. Вопросы, в которых респондентам предлагается провести различие между прилагательными, которые могут интерпретироваться по-разному, приводят к несопоставимым ответам;

|  |
| --- |
| **Пример нарушения операционной валидности** |
| Интерпретация уровней оценок влияния рисков («низкое», «среднее», «высокое») может быть различной у различных экспертов |

10) ресурсное ограничение возможности сбора и ввода данных.Высокая рабочая нагрузка и (или) отсутствие технических возможностей могут привести к разрозненности практик сбора и ввода данных у разных сотрудников (например, сотрудников объекта аудита (контроля), в разных местах;

|  |
| --- |
| **Пример нарушения операционной валидности** |
| Работники органов социальной защиты могут не иметь времени для своевременного ввода данных в информационные системы; технические возможности ввода данных могут быть ограничены (например, недостаточное количество и качество компьютеров) |

11) недостаточная профессиональная подготовка сборщиков данных. Недостаточная подготовка сборщиков данных, интервьюеров, наблюдателей и (или) кодировщиков может сделать сбор и ввод данных недостаточно надежными.

|  |
| --- |
| **Пример нарушения операционной валидности** |
| Чрезмерно сжатые сроки и экономия за счет сокращения времени обучения могут привести к тому, что сотрудники (например, сотрудники объекта аудита (контроля) не будут готовы обеспечить корректное проведение сбора и ввода данных |

**2. Утверждения о причинно-следственных связях (внутренняя валидность)**

2.1. **Внутренняя валидность** отражает то, в какой степени наблюдаемые результаты обусловлены рассматриваемыми причинами.

2.2. Возможные проблемы внутренней валидности:

1. вмешательство иных событий*.* Наблюдаемые изменения связаны не только с мерой, а с каким-либо иным фактором или произошедшим событием. При этом мера может не быть ключевым фактором, повлиявшим на изменения;

|  |
| --- |
| **Пример нарушения внутренней валидности** |
| Увеличение объемов экспорта может отражать более благоприятные внешнеторговые условия (снижение тарифов или импортных пошлин, изменение курса валюты), а не являться лишь результатом субсидирования экспортеров |

1. отсутствие учета социокультурных особенностей;

|  |
| --- |
| **Пример нарушения внутренней валидности** |
| При анализе результативности реализации программ трудоустройства женщин может быть целесообразным изучение и учет социально-культурных представлений о семье и разделении обязанностей между супругами. Обусловленные традициями данной местности представления о роли женщины как исключительно жены и матери могут выступать в качестве социокультурного барьера, ограничивающего участие женщин в оплачиваемой занятости в целом и в программах трудоустройства в частности |

1. изменения во времени*.* Наблюдаемые изменения обусловлены не мерой, а тем, что участники меры (граждане, организации) меняются со временем (взрослеют, развиваются и т. д.);

|  |
| --- |
| **Пример нарушения внутренней валидности** |
| С возрастом проблемы со здоровьем у людей зачастую становятся более выраженными, что может приводить к недооценке фактического (успешного) достижения результатов мероприятий проекта «Активное долголетие» |

1. мера реализована не полностью.Оценка меры может быть преждевременной, если какие-либо факторы (например, нехватка ресурсов, ошибки процессов) привели к проблемам реализации или неполной реализации меры;

|  |
| --- |
| **Пример нарушения внутренней валидности** |
| Аудит реализации социального контракта в пилотных регионах показал, что центры службы социальной защиты, которые непосредственно работали с получателями мер государственной поддержки, были недоукомплектованы. Работники органов социальной защиты высказывались о нехватке методологических и обучающих материалов. В итоге социальный контракт в большинстве случаев реализовывался как прямой необусловленный трансферт населению, не позволяющий малоимущим семьям выйти из бедности в долгосрочной перспективе, хотя цель этой меры как раз заключается в предоставлении малоимущим семьям инструментов для улучшения своего благосостояния на долгий срок |

1. регрессия к среднему.Наблюдаемый результат меры может быть связан с тем, что она применялась к получателям, ранее продемонстрировавшим некоторые экстремальные значения показателя (например, очень низкие финансовые показатели для организаций), которые получились такими из-за сложившихся внешних обстоятельств. При этом при отсутствии меры получатели могли бы продемонстрировать улучшение показателя исключительно в силу того, что прошлое наблюдение оказалось «случайно плохим», и в реальности значение показателя лежит намного ближе к среднему по всей генеральной совокупности. Поскольку деятельность экономических агентов подвержена воздействию большого количества случайных внешних факторов, такое поведение характерно для многих экономических и финансовых показателей, и приписывать в таком случае изменение только воздействию реализованной меры некорректно;

|  |
| --- |
| **Пример нарушения внутренней валидности** |
| Оценивается программа, направленная на улучшение успеваемости школьников. До и после программы проводилось тестирование. Распределение оценок генеральной совокупности школьников не изменилось. В то же время в выборке из 10 % школьников, имевших до программы худшие результаты, средний результат после программы стал выше – не все школьники из этой выборки снова вошли в худшие 10 % после проведения программы. Аналогично, среднее по выборке показавших 10 % лучших результатов сместилось вниз. Это можно объяснить, например, тем, что худшие 10 % школьников в день проведения первого тестирования испытывали недомогание, стресс или волнение, но во время второго тестирования чувствовали себя хорошо. В свою очередь, аналогичные причины могут иметься и у школьников, которые написали успешно первое тестирование, но провалили второе |

1. смещение отбора.Наблюдаемые изменения обусловлены существовавшими ранее различиями между участниками в экспериментальной и контрольной группах, а не мерой;

|  |
| --- |
| **Пример нарушения внутренней валидности** |
| При проведении оценки программы по трудоустройству людей с инвалидностью в составе контрольной группы оказываются участники московской организации общества глухих, а в составе экспериментальной (подвергшейся воздействию программы) – люди с различными типами особенностей здоровья. Поскольку общество инвалидов профильное, туда попадают люди с конкретными особенностями здоровья: потребности в работе для глухих будут отличаться от таковых для слабовидящих людей или людей, которые пользуются инвалидной коляской для передвижения. Выборка будет смещена как по составу, так и по региональной структуре. Это не позволит обобщить полученные выводы на всех людей с ограниченными возможностями здоровья и для разных групп инвалидности |

1. некорректное установление направления причинно-следственного влияния. Отсутствие ясности в отношении того, какой показатель изменился первым, может привести к ошибке определения того, какой показатель является причиной, а какая следствием.Кроме того, возможна ситуация, когда оба показателя влияют друг на друга (показатель Х влияет на показатель Y, а показатель Y – на показатель Х);

|  |
| --- |
| **Пример нарушения внутренней валидности** |
| Направление взаимосвязи бедности и здоровья населения не является очевидным. С одной стороны, низкие доходы населения приводят к снижению доступности качественного питания и другим факторам риска для здоровья. С другой стороны, плохое здоровье может оказывать влияние на возможность найти высокооплачиваемую работу и на уровень производительности труда. Таким образом, оба показателя – уровень бедности и здоровье населения – могут оказывать влияние друг на друга |

1. эффект смешивания последствий нескольких мервозникает в условиях реализации нескольких мер одновременно, что делает сложным обособленную оценку каждого из эффектов.

**3. Возможность обобщения (внешняя валидность)**

3.1. **Внешняя валидность** определяет, в какой степени можно распространить полученные результаты на объекты (группы участников, контексты или обстоятельства), выходящие за рамки анализируемых.

Возможные проблемы внешней валидности:

1. последствия отбора.Результаты оценки меры могут быть применимы только к анализируемым участникам или контексту.

Смещение при отборе по превосходству или по тяжести проблем.

Смещение при отборе по превосходству возникает при анализе успешных практик реализации меры, так как характеристики этих практик могут повышать шансы на получение более высокого целевого результата меры. Смещение при отборе по тяжести проблем возникает при анализе проблемных аспектов реализации меры, когда отбор проводится на основе факторов тяжести проявления проблемы.

|  |
| --- |
| **Пример нарушения внешней валидности** |
| Проблема отбора по превосходству может возникнуть в ситуации, когда при оценке результатов цифровизации обучения школьников в выборку попадают только школы, которые специализируются на изучении компьютерных технологий, гимназии и лицеи, где есть доступный интернет, компьютеры, а также где педагоги регулярно повышают квалификацию в сфере информационных технологий, активно в образовательном процессе пользуются специальным программным обеспечением. При этом другие школы могут выстраивать процесс цифрового образования с нуля, иначе видеть и приоритизировать задачи цифрового образования. Для таких школ программы цифровизации обучения школьников могут показывать другие результаты, их проблемы и риски реализации могут быть иными.    Проблема отбора по тяжести проблемы может возникнуть в ситуации оценки экологического мониторинга в Российской Федерации при выборе регионов для анализа на основе степени распространенности и тяжести проблем при осуществлении мониторинга |

Смещение при отборе по доступностивозникает при анализе практик реализации меры на основе доступности для анализа (например, территориальной близости, готовности респондентов к взаимодействию и т. д.).

|  |
| --- |
| **Пример нарушения внешней валидности** |
| Использование опросов, проведенных через Российский союз промышленников и предпринимателей (РСПП), приводит к обращению к смещенной выборке предприятий с точки зрения представительства различных федеральных округов Российской Федерации |

Отбор по объемам финансирования, по охвату выгодоприобретателейтакже может приводить к смещенным результатам, так как из фокуса внимания выпадают объекты с отсутствием финансирования в данной области (низким/высоким охватом получателей), что может оказывать влияние на результаты анализа;

1. временные эффекты*.* Смещение может возникать при использовании данных, которые относятся к периоду, нерелевантному исследуемой проблеме;

|  |
| --- |
| **Пример нарушения внешней валидности** |
| Оценки результативности программы снижения ДТП, полученные с использованием данных за летние месяцы, могут в значительной степени отличаться от оценок за зимние месяцы |

1. географические эффекты*.* Смещение может возникать при проведении оценки в определенном регионе или в определенном контексте, из-за чего результаты оценки не могут быть распространены на другие условия.

|  |
| --- |
| **Пример нарушения внешней валидности** |
| Результаты оценки эффективности программы субсидирования малого бизнеса в городской местности одного региона могут быть непригодны для прогнозирования результатов аналогичной программы в сельской местности другого региона |

**4. Валидность статистических выводов**

Возможные проблемы валидности статистических выводов:

1)нерепрезентативность выборки. В случае отсутствия репрезентативности результаты, полученные с использованием выборки (например, выборки объектов аудита (контроля), регионов, муниципалитетов, проектов, отдельных документов и т. д.), не могут быть использованы для общих выводов для генеральной совокупности;

|  |
| --- |
| **Пример нарушения статистической валидности** |
| Выводы о причинах бедности, полученные на основе анализа бедности в одном регионе, не могут быть распространены на остальные регионы, поскольку между регионами имеются значительные различия в структуре экономики, населения, занятости и т. п. Определенные меры государственной политики, которые были результативны в одном регионе, могут не дать аналогичного результата в других регионах |

1. слишком малый объем выборки.Следствием слишком малого объема выборки может быть статистическая незначимость эффекта, несмотря на его наличие. Кроме того, для малых объемов выборки добавление или исключение даже одного наблюдения может приводить к значительным изменениям в количественных оценках;

|  |
| --- |
| **Пример нарушения статистической валидности** |
| Использование для регрессионного анализа 5–7 наблюдений может приводить к сформулированным выше проблемам. Например, оценка взаимосвязи уровня бедности и экономического роста за период 1996–2019 гг. будет отличаться от оценки за период 2001–2007 гг. или 2012–2019 гг. Для анализа подобных взаимосвязей, когда выборка содержит только несколько наблюдений, необходимо использовать иные методы (например, качественные методы) или изменить дизайн, чтобы рассматривать другие прокси-показатели, для которых есть достаточное количество наблюдений |

1. применение статистических методов, непригодных для имеющихся данных (в частности, для имеющегося типа данных).Применение методов, не пригодных для определенного типа данных, может приводить к некорректным результатам;

|  |
| --- |
| **Пример нарушения статистической валидности** |
| Использование метода линейной регрессии для анализа успеваемости, если оценки выражены по 5-балльной шкале (даже по 4-балльной – как правило, они меняются от 2 до 5), также может дать предсказанные значения оценок либо меньше 1 (в том числе отрицательные), либо больше 5, что не имеет смысла.    Применение метода линейной регрессии с использованием выборки показателей по регионам за один год для анализа влияния меры государственной политики некорректно, поскольку не происходит сравнения показателей в одном и том же регионе до и после реализации данной меры, а вместо этого происходит сравнение показателей в разных регионах. В то же время регионы обладают значительными различиями в структуре экономики, населения, занятости и т.п., обусловленными различной предысторией развития этих регионов, а результаты реализации мер государственной политики могут зависеть от этой предыстории. Кроме того, меры госполитики, давшие какой-то результат в одном регионе, могут не дать результата в других регионах.    Метод линейной регрессии в общем случае, равно как вычисление корреляций между показателями, не является способом выявления причинно-следственных взаимосвязей между этими показателями. Другими словами, коэффициенты регрессионной модели в общем случае не могут быть интерпретированы как выводы о причинно-следственной связи.  Для количественной оценки величины причинно-следственной связи требуется специальный дизайн (например, метод разность разностей или другие квазиэкспериментальные методы) |

1. нарушение условий применения статистического метода.Применение статистического метода с нарушением его предположений может быть причиной неверных или незначимых результатов;

|  |
| --- |
| **Пример нарушения статистической валидности** |
| Простая линейная регрессия неприменима при явных нелинейных эффектах. Например, если требуется оценить взаимосвязи между инвестициями в очистные сооружения и снижением выбросов вредных веществ, следует иметь в виду, что количество выбросов при применении все более современных технологий не может быть уменьшено до нуля, так что будет наблюдаться «эффект насыщения», когда произойдет приближение к технологическому пределу снижения выбросов, и эффект от дальнейших инвестиций будет все меньше и меньше, а дальнейшее снижение выбросов будет все дороже и дороже. В такой ситуации требуется применение других методов (например, нелинейного метода наименьших квадратов или иных методов, учитывающих нелинейность эффектов).    Использование метода линейной регрессии для нестационарных временных рядов может дать некорректные оценки для взаимосвязи между этими рядами, поскольку значимая корреляция между ними может быть обусловлена исключительно их нестационарностью (так называемый эффект «мнимой регрессии» или «мнимой корреляции»). В этом случае следует использовать специальные методы анализа временных рядов (проверка стационарности, проверка коинтеграции, преобразование временных рядов к стационарному виду, выделение тренда и сезонности и т. д.).    Использование метода линейной регрессии может дать некорректные результаты в том случае, когда имеется взаимное влияние между зависимым и объясняющим показателями, например, между ВВП и доходами населения (это частный случай **эндогенности** в регрессионных моделях). В этом случае для оценки взаимосвязи между зависимым и объясняющим показателями требуется использовать специальные методы (например, метод инструментальных переменных) |

1. проблемы измерений.Проблемы измерений могут включать в себя ошибки при измерениях, ненадежность прокси-показателей, ограниченность пределов измерения;

|  |
| --- |
| **Примеры нарушения статистической валидности** |
| Если возраст – важный фактор модели, но в выборке присутствуют только люди в возрасте от 21 года до 28 лет, то результаты могут оказаться смещенными и не могут быть распространены на другие возрастные категории.    Если шкала ответов представлена вариантами «хорошо», «прекрасно», «замечательно», то у разных респондентов эти слова могут означать разную степень удовлетворенности |

1. ошибкиспецификации.Невключение важных показателей в модель, либо включение неважных.

|  |
| --- |
| **Пример нарушения статистической валидности** |
| Если регрессионная модель не включает в число объясняющих показателей важный, значимый фактор, то все остальные регрессионные коэффициенты, вообще говоря, будут смещенными и несостоятельными, то есть модель даст некорректные результаты. В частности, при проведении регрессионного анализа **следует избегать оценки парных регрессий**, используя множественную регрессию с включением константы и основных факторов, влияющих на зависимую переменную.    С другой стороны, включение нерелевантных показателей может увеличить коэффициент детерминации регрессионной модели, но это также повлечет увеличение стандартных ошибок и расширение доверительных интервалов для регрессионных коэффициентов, так что действительно важные показатели могут стать незначимыми. Это приведет к тому, что будут сделаны некорректные выводы относительно взаимосвязей между входящими в модель показателями |