

DOI: 10.14515/monitoring.2014.3.02

УДК 316:303

А.А. Ипатова¹

НАСКОЛЬКО РАЗУМНА НАША ВЕРА В РЕЗУЛЬТАТЫ ОПРОСОВ, ИЛИ НАРУШЕНИЕ ИССЛЕДОВАТЕЛЬСКОЙ ЭТИКИ В СОЦИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

НАСКОЛЬКО РАЗУМНА НАША ВЕРА В
РЕЗУЛЬТАТЫ ОПРОСОВ, ИЛИ НАРУШЕНИЕ
ИССЛЕДОВАТЕЛЬСКОЙ ЭТИКИ В
СОЦИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

*ИПАТОВА Анна Алексеевна — старший научный
сотрудник Центра методологии федеративных
исследований РАНХиГС. E-mail:
ipatova_anna@mail.ru*

Аннотация. Зачастую у заказчика нет возможности самостоятельно проверить достоверность предоставляемых опросной компанией данных и оценить проведенное исследование с точки зрения нарушения научной этики. В таких случаях приходится полагаться на личную и профессиональную честность менеджеров и руководителей опросной компании, особенно если речь идет о закрытом исследовании.

В данной статье описана обратная ситуация, когда по требованию заказчика почти вся полевая документация была предоставлена. Опрос проводился в августе 2013 г. одной из лидирующих компаний на рынке маркетинговых исследований. Таким образом, в статье анализируются анкеты общероссийского опроса, маршрутные листы интервьюеров, описание выборки, предоставленный опросной компанией методический отчет, база приписок интервьюеров к точкам выборки.

На основе имеющихся данных автор статьи поднимает вопрос о качестве социологических данных, а также описывает различные виды нарушения научной этики, например, фабрикацию и фальсификацию. В результате

IS OUR TRUST IN SURVEY RESULTS RATIONAL, OR
BREAKING THE ETHICS IN THE SOCIAL RESEARCH

*IPATOVA Anna Alekseevna — senior
researcher, Centre for Federative Research
Methodology, Russian Presidential Academy for
National Economy and Public Administration. E-
mail: ipatova_anna@mail.ru*

Abstract. Often it is impossible for the employer to verify the received data as well as the quality of the investigation from the perspective of scientific misconduct. In such cases, the only way is to rely on the business integrity of company managers and executives.

The article describes the reverse situation, when the employer had an access to a great part of the field documentation. The survey was conducted in August, 2013 by one of the leading companies in the area of marketing research. Therefore the author analyses the completed questionnaires, interviewer's routes, sample description, allocation of addresses for interviewers and methodological report made by the company.

This leads to the issue of the reliability of data and such types of scientific misconduct as fabrication and falsification. As a result, the author establishes the consistency of only 74% of data. Moreover, two cases of falsification have been disclosed. The author concludes that only 21% of interviewer's routes comply with the requirements.

¹ Автор выражает глубокую признательность коллективу лаборатории методологии федеративных исследований Института социального анализа и прогнозирования РАНХиГС при Президенте РФ, а также группе «Методический цех» за помощь в организации исследования и многократные обсуждения его результатов.

анализа всей полевой документации автор находит преобладание только в 74% данных, обнаруживает 2 случая явной фальсификации и заключает, что без ошибок (несуществующих домов, нарушений шага, наличия нежилых домов) отработан только 21% маршрутов.

Ключевые слова: полевая документация, полевой этап, нарушение научной этики, фабрикация, фальсификация, качество социологических данных, процедура опроса, личное интервью.

Keywords: field documentation, fieldwork, scientific misconduct, fabrication, falsification, quality of survey data, survey procedure, face-to-face interview

В мае 2009 г. в открытом электронном журнале PLoS ONE вышел систематический обзор Даниэля Фанелли о фабрикации и фальсификации данных [4]. Автор задается вопросом: насколько велико число ученых, которые подделывают результаты своих исследований? Увеличивающееся с каждым годом число публикаций недостоверных данных, даже в самых известных изданиях и на самые сложные темы², наводит его на мысль, что случаи фальсификации, которые становятся достоянием гласности, — лишь вершина айсберга, и таких случаев гораздо больше. Действительность далека от образа «чистой» науки, где научное сообщество руководствуется мертоновскими принципами беспристрастности и организованного скептицизма. Схожую мысль еще в 1991 г. выразил Дэвид Гудстейн, назвав мифом о благородном ученом представление, согласно которому ученый более добродетелен и честен, чем обычные люди [2, с. 95].

Проблема усложняется и тем, что не всегда можно достоверно установить, имела место намеренная фальсификация или же исследователь допустил ошибку в результате небрежности и невнимательности. Более того, даже если ошибка допущена по злему умыслу, исследователь всегда может сослаться на ее непреднамеренность и случайность. Зачастую истинное положение дел известно только небольшому кругу причастных лиц, которые, по понятным причинам, не заинтересованы в обнаружении намеренных ошибок и фабрикаций.

Необходимо внести терминологическую ясность. Прежде всего это касается схожих по значению и часто подменяемых друг другом форм нарушения исследовательской этики: фабрикации и фальсификации. Фанелли определяет их довольно кратко, но четко: фабрикация — «выдумка данных или случаев», фальсификация — «намеренное искажение данных или результатов» [4, р. 1], причем фальсификация представляется более сложной и проблематичной категорией ввиду отсутствия объективных факторов, доказывающих злой умысел. В российских исследованиях этот вопрос наиболее последовательно изложен в профессиональном врачебном сообществе, где он стоит более остро³. Так, на сайте «Медицинской газеты» приведена статья Василия Власова о плагиате и других видах нарушения норм научной работы, в числе которых указаны фальсификация и фабрикация [1]. Фабрикация определяется им как «любое улучшение (изменение в угоду каким-либо представлениям) данных исследования, записей о них и отчетов (выдумывание результатов, подделка записей, сообщение искаженных данных)». В то же время фальсификация представлена как «манипуляция исследовательскими материалами, оборудованием,

² В качестве примера Фанелли приводит скандал, вызванный публикацией южнокорейских ученых о клонировании человека. Имеется в виду вышедшая в 2004 г. в журнале «Science» статья Hwang и соавт. «Evidence of a Pluripotent Human Embryonic Stem Cell Line Derived from a Cloned Blastocyst» (Science. 2004. Nr 12). Подробнее см. http://art.russ-med.ru/full_genetics_05_russ-med.html.

³ К слову, большинство проанализированных Фанелли исследований (14 из 21) также относятся к области медицины.

процессами или изменением данных, исключение данных или результатов таким образом, что действительный результат не отражается точно в отчете (изменение материалов, оборудования, протоколов, данных, результатов)». Как видим, оба эти определения значительно шире, и в одно из них даже включена цель: «улучшение данных в угоду чему-либо», что, на наш взгляд, не до конца верно, поскольку можно гипотетически предположить ситуацию, когда данные намеренно ухудшаются. Кроме того, не совсем понятна разница между двумя понятиями, так как определения частично пересекаются, а определение фабрикации настолько широко, что может включать и фальсификацию. На наш взгляд, основное отличие заключается в следующем: фабрикация — это создание, придумывание (выдумывание) заведомо ложных данных, создание фиктивных данных, а фальсификация — изменение уже имеющихся данных. Грубо говоря, для фабрикации не всегда необходимо проводить настоящее исследование, в то время как фальсификация — это некое редактирование того, что есть. Оба эти вида намеренного (сознательного) нарушения исследовательской этики могут присутствовать в рамках одного проекта и пересекаться, потому не всегда есть смысл их разграничивать.

По Фанелли, существует и третий вид нарушения исследовательской этики — плагиат, но он в отличие от предыдущих двух в особом представлении не нуждается и в меньшей степени «влияет на научное знание, хотя несет в себе серьезные последствия для карьеры вовлеченных в него людей» [4, р. 1]. Установить текстовый плагиат в ряде случаев не составляет большого труда, на сегодняшний день разработаны многочисленные программы по его выявлению. К этой же категории стоит отнести и автоплагиат, когда автор или компания воспроизводит свои же собственные наработки. Является ли автоплагиат нарушением исследовательской этики? Например, если опросная компания сделала хорошую и дорогую выборку, а потом провела по ней несколько опросов для разных заказчиков? Выявить такой вид плагиата довольно сложно, поскольку данные закрыты для общественности. Является ли это нарушением? Ответ не так однозначен, как может показаться на первый взгляд, а сама проблема требует более пристального внимания научного сообщества.

В последнее время все чаще говорят об открытости социологических данных, их качестве, а также о доверии к ним со стороны общества. Примером тому может служить IV международная социологическая конференция «Продолжая Грушина», на которой один из основных тезисов звучал так: «Качество социологических данных — основа доверия к социологии» (<http://wciom.ru/conference2014>). При обсуждении результатов электоральных опросов в период политических изменений проблема доверия к данным выходит на первый план. Эти данные в той или иной мере становятся достоянием общественности, возможно, даже инструментом влияния на общественное мнение, потому их открытость (пусть и частичная) необходима. Иначе обстоит дело с закрытыми (или относительно закрытыми) исследованиями, цель которых не потрясти общественность, а стать инструментом для принятия решений или формирования рекомендаций для тех, кто эти решения вправе принимать. С такой ситуацией зачастую сталкиваются научно-исследовательские институты, выполняющие государственные заказы или же проводящие внутренние научно-исследовательские работы. Здесь не всегда можно самостоятельно организовать полевую работу, особенно для больших исследований, поэтому зачастую полевой этап отдается на аутсорсинг, а внутри ведется работа с полученным по всем требованиям массивом данных. По сути, разделение труда, или, как это сейчас модно говорить, коллаборативность, когда каждый профессионал выполняет свою работу и не вмешивается в работу коллег, несет в себе определенные и неоспоримые преимущества. Риски тоже высоки: когда случается сбой

в самом начале, на полевом этапе или при его планировании, далее число ошибок нарастает по цепочке, как снежный ком, создавая целый пласт ложных по своей сути данных и интерпретаций. У заказчика нет возможности самостоятельно проверить достоверность полученных данных, оценить проведенное исследование на предмет нарушения исследовательской этики, ведь в лучшем случае он получает методологический отчет (зачастую краткий), в котором из соображений нераскрытия личной информации третьим лицам отсутствуют все привязки данных к конкретным респондентам. Более того, не всегда опросные компании рады предоставить «грязный», или первоначальный, неотредактированный массив, считая, и, возможно, по праву, это внутренним документом. Иногда методологический отчет не требуется самому заказчику, если тот доверяет исполнителям, на что у него есть свои основания, особенно, когда полученные данные не вызывают вопросов и/или несильно противоречат Росстату. Мы уже писали о том, что авторитетные журналы, имеющие разработанные стандарты для публикаций, общий перечень которых можно найти, например, на сайте COPE (Комитета по этике публикаций) (<http://publicationethics.org/resources/international-standards>), тем не менее публикуют статьи, содержащие недостоверные данные. И это происходит в сфере, где есть действительный контроль над авторами и их публикациями. А как же может обстоять дело там, где публичного контроля нет? Получается, что для внутренней документации решение таких вопросов отдается на откуп полевым менеджерам и руководителям, что, как это понятно, полностью зависит от их личной и профессиональной честности.

С этой проблемой мы столкнулись при анализе данных общероссийского опроса, сделанного по нашему заказу одной из крупнейших на рынке опросной компанией. Мы не ставим цель уличить кого-либо в недобросовестной работе, подлоге, фальсификации и фабрикации данных, мы хотим лишь поднять вопрос о важности экспликации методологического этапа сбора социологических данных. Именно из этих соображений мы не раскрываем название проекта, его непосредственного заказчика и исполнителя, а также фамилии интервьюеров. Открытыми остаются лаборатория методологии федеративных исследований Института социального анализа и прогнозирования РАНХиГС, поскольку в рамках данного института была проведена аналитическая работа по описываемому в статье проекту. Автор не видит смысла скрывать названия административных округов, районов и улиц Москвы, где проходило исследование.

По предварительной договоренности с опросной компанией мы имели доступ к полевой документации, кроме того, нам были отданы на руки заполненные анкеты и маршрутные листы по всем регионам. Сразу отметим, что связующее звено между этими документами отсутствовало: в маршрутном листе не было телефонов и адреса респондента, а на анкетах не было данных о пройденном маршруте. Были указаны только фамилия интервьюера, время и дата опроса, его продолжительность, номер анкеты в базе.

Согласно полученным отчетным материалам, по Москве было пройдено 70 маршрутов силами 46 интервьюеров. Обозначим те материалы, на которых мы основываем наш анализ:

- распределение выборки по округам и районам Москвы (далее — спроектированная выборка) — 73 маршрута на 803 анкеты, из них мы позже исключили Новомосковский АО, Троицкий АО и г. Зеленоград, т.е. для Москвы это 70 маршрутов с запланированными 770 анкетами;

- база приписок интервьюеров к точкам выборки (далее — база приписок) — документ, в котором обозначены все интервьюеры (их фамилии и идентификаторы), взятые интервью, имена и адреса респондентов, их пол и возраст — всего 783 позиции по Москве;
- отчет по маршрутным листам, где для каждого интервьюера перечислены стартовые точки маршрутов (т.е. адрес, случайно сгенерированный из адресов попавшего в выборку района и выданный на руки интервьюеру вместе с квотным заданием; с этой точки интервьюер должен начать прохождение маршрута; на маршруте запланировано 11 интервью, не более 3-х интервью в одном доме с шагом в 11 квартир), а также указаны результаты всех обращений по маршруту;
- маршрутные листы (в наличии был 61 маршрутный лист из 70; позже выяснилось, что 61 маршрутный лист включал 64 маршрута);
- отчет о работе интервьюеров (в котором указано, что проверке подверглись 52% данных).

Мы разделили работу на три этапа: до поля (спроектированное исследование), в поле (реализованное исследование) и после поля (что позже получило название отчетное исследование). Каждому этапу свойственны свои виды нарушения исследовательской этики, например, спроектированная выборка может дублировать выборку для другого исследования (своего рода автоплагиат), а реализованная выборка может сильно не совпадать по своим характеристикам с спроектированной. Конечно, нельзя оставить без внимания фальсификации интервьюеров, а также подбивку данных под требуемые. К сожалению, все эти виды были обнаружены нами при анализе.

Первый этап (до поля: проектирование полевого исследования и подготовка интервьюеров)

Вначале мы не знали, что в отчетных документах, предоставленных опросной компанией, отсутствует преемственность. Например, это касалось самого «чистого» этапа — описания спроектированной выборки. Для всероссийского исследования с общим объемом выборки в 9500 респондентов была предложена многоступенчатая стратифицированная районированная (кластерная) репрезентативная выборка с отбором домохозяйств маршрутным способом и контролем половозрастных квот на этапе выбора респондента в домохозяйстве. Статистическая погрешность выборки указана как $\pm 1\%$ при вероятности 0,95. В Москве общий объем выборки в 803 интервью распределился пропорционально 12 административным округам (табл. 1).

Таблица 1 Распределение выборки по Москве

Административный округ, г. Москва (после 1 июля 2012 г.)	Население, тыс. чел.	Доля, %	Расчет на 803 анкеты	Выборка, чел.	Маршруты, шт.
Центральный	696,6	6,5	52,2	55	5
Северный	1112,9	10,3	82,7	88	8
Северо-Западный	805,4	7,5	60,2	66	6
Северо-Восточный	1249,4	11,6	93,1	88	8
Южный	1573,4	14,6	117,2	121	12
Юго-Западный	1243,8	11,5	92,3	88	8
Юго-Восточный	1160,7	10,7	85,9	88	8

Западный	1098,5	10,2	81,9	77	7
Восточный	1403,5	13	104,4	99	9
Зеленоградский	218,8	2	16,1	11	1
Новомосковский и Троицкий	235,8	2,2	17,7	22	2
Итого:	10 798,8 тыс	100%	803	803	74

Однако уже в отчете мы обнаружили, что «для репрезентации Москвы как отдельного субъекта РФ город выделен в отдельную территориальную единицу с объемом выборки, пропорциональным численности населения Москвы – 794 респондента для общей выборки исследования». Кроме того, в массиве данных Москва представлена числом респондентов в 808. Возник первый вопрос: почему в рамках одного исследования имеются такие, пусть и незначительные, расхождения и, что более важно, почему это не объяснено и/или не оговорено исполнителем? Эта ошибка может быть следствием небрежности при составлении отчетной документации и, возможно, результатом копирования описания выборки или же таблицы из другого отчета. Впрочем, это лишь гипотеза.

Следующий шаг – распределение маршрутов внутри административных округов. На 9 административных округов Москвы приходится 70 маршрутов и 770 анкет, если исходить из 11 анкет на один маршрут. Нам были предоставлены таблицы с описанием попавших в выборку районов по каждому административному округу. Рассмотрим такую таблицу на примере Центрального административного округа (табл. 2): из 10 районов в выборку попали 5 наиболее населенных: Басманный, Пресненский, Таганский, Тверской, Хамовники. Далее на каждый выбранный округ попадает один маршрут, для чего случайным образом из списка улиц определяется стартовая точка.

Таблица 2 Распределение выборки по районам Москвы, ЦАО

№	Районы ЦАО	Население, тыс. чел.	Маршруты, кол-во	Стартовая точка, кол-во
1	Арбат	28 536	0	0
2	Басманный	109 086	1	2
3	Замоскворечье	56 098	0	0
4	Красносельский	47 839	0	0
5	Мещанский	58 514	0	0
6	Пресненский	124 015	1	2
7	Таганский	117 228	1	1
8	Тверской	76 064	1	2
9	Хамовники	103 861	1	1
10	Якиманка	26 791	0	0
Итого:			5	8

В предоставленном нам списке стартовых точек, состоящем из 70 наименований, к ЦАО относятся целых 8 (табл. 2, 3): дважды указана Спартаковская улица, даны по две стартовые точки в Пресненском и Тверском районах. Очевидно, что продублированный Басманный район – это ошибка, тем не менее он присутствует именно в таком виде во всех отчетных материалах. Несоответствие этих данных указывает прежде всего на неточности в оформлении документации, а также явные логические и фактические противоречия, которые могут быть следствием намеренной фальсификации.

Таблица 3 Стартовые точки по районам ЦАО

№ из списка	АО	Район	Отчетная стартовая точка	Респонденты в массиве данных (1 – да, 0 – нет)	Маршрутный лист (1 – да, 0 – нет)
37	ЦАО	Басманный	Спартакoвская, д.6/1	1	1
38	ЦАО	Пресненский	Красная Пресня, д.9	1	0
40	ЦАО	Пресненский	Большая Грузинская, д. 37	0	1
51	ЦАО	Таганский	Калитниковская, д. 18	1	0
9	ЦАО	Тверской	Тверская ул., д. 27	1	1
20	ЦАО	Тверской	ул. Лесная, д. 7, к. 5	0	0
28	ЦАО	Хамовники	Ефремова, д. 21	1	1
62	ЦАО	Басманный	Спартакoвская, д. 6	1 (дубль)	1 (дубль)
Итого:				5	4

Удивительно, но если посмотреть на полевую документацию — на маршрутные листы, то они предоставлены только для 4-х маршрутов. Например, адреса респондентов с маршрута «Красная Пресня» внесены в базу, но маршрутный лист для них отсутствует. Для респондентов с маршрута «Большая Грузинская», напротив, есть маршрутный лист и заполненные анкеты, но эти респонденты отсутствуют в массиве данных. Не ясно, как и почему это произошло и зачем тогда были переданы эти анкеты. В итоге отчетная документация совпадает только для 3-х из 5 маршрутов. Просмотрим эти же показатели по другим административным округам (табл. 4).

Таблица 4 Спроектированная выборка по округам

АО	Спроектированная выборка		Отчетные стартовые точки		Есть ли ошибка в районах (1 – да, 0 – нет)
	Маршруты (кол-во)	Респонденты (кол-во)	количество	прогноз: кол-во респ.	
ЦАО	5	55	8	88	0
САО	8	88	7	77	0
СВАО	8	88	8	88	0
ВАО	9	99	8	88	1
ЮВАО	8	88	8	88	0
ЮАО	11	121	11	121	0
ЮЗАО	8	88	7	77	1
ЗАО	7	77	7	77	0
СЗАО	6	66	6	66	0
Итого:	70	770	70	770	2

На данном этапе можно обнаружить, что для 3-х административных округов из 9 было разработано неправильное количество стартовых точек, хотя общее число маршрутов и

респондентов совпадает. Неправильное распределение стартовых точек по округам приводит к ошибке в 55 адресов, что составляет порядка 7% от общего числа респондентов (за 100% берутся 770 адресов на 70 маршрутах). Кроме того, еще в 2-х административных округах мы обнаруживаем несовпадения по попавшим в выборку районам и разработанным стартовым точкам. В ВАО отсутствуют стартовые точки для попавших в выборку районов Ивановское и Измайлово, тогда как для района Северное Измайлово, который не был выбран для опроса, имеется стартовая точка. Третий округ с ошибкой — ЮЗАО, здесь отсутствуют стартовые точки для районов Академический и Коньково, указанных в спроектированной выборке, но разработана стартовая точка для маршрута в районе Котловка, который в выборку не попал. Перепутаны еще 6 районов, что приводит к числу в 66 адресов, а это 8% ошибок от общего числа респондентов. Если сложить эти неточности, то только на первом обозначенном нами этапе обнаруживаются 15% домохозяйств, включенных в спроектированную выборку, по которым выявлены несоответствия в документации. Неизвестно также, с какой документацией шла работа в поле. Мы выделили первый этап логически, поскольку выборка проектируется до поля, но вполне вероятно, что описанная в отчете «спроектированная» выборка стряпалась на лету уже после завершения полевого этапа и сугубо для отчета. Причем, судя по всем несовпадениям, это отчет вдумчиво не читался самими менеджерами опросной организации, и уж тем более они не предполагали, что кто-либо будет этот отчет читать. Как тут не вспомнить случай, описанный И. Коном, когда имеет место производство никому не нужной документации: «В Академии наук рассказывали о каком-то физике, который в середине толстого отчета написал: "Если кто-нибудь дочитает до этой страницы, пусть позвонит по такому-то телефону и получит бутылку лучшего коньяка". Коньяк остался невостребованным» [3, с. 132].

Второй этап (полевой)

Второй этап — сбор информации самими интервьюерами — несет в себе многочисленные возможности для нарушений исследовательской этики. В отчете указано, что проверке подверглись более 52% анкет. К сожалению, у нас не было возможности проверить работу интервьюеров опросной компании в процессе сбора информации, поэтому сейчас мы можем анализировать их работу лишь по маршрутным листам. Какие же ошибки мы обнаружили? В первую очередь это нарушение методики проведения опроса. Интервьюеры не соблюдали шаг, опрашивали все квартиры сплошняком, иногда даже забывали менять дом и собирали все 11 интервью в одном доме. Иногда просто ходили в те квартиры, в какие им хотелось, не демонстрируя видимой логики своих перемещений. В целом нарушение шага было обнаружено в большинстве (39) предоставленных маршрутных листов, причем практически каждый интервьюер нарушал по-своему. Получается, с соблюдением методики опроса (мы не проверяли переход от дома к дому), у нас только 25 маршрутных листов, что составляет 36% от планируемых 70 маршрутов. Все это делает просто невозможным как-либо методически оценивать реализованную выборку.

Не всегда понятен переход интервьюеров от одного дома к другому. Иногда это соседние дома, а иногда расстояние между ними составляет более километра. Были обнаружены и заходы в соседние районы, например, интервьюер отработывал 2 маршрута в районах Южное и Северное Орехово-Борисово. Работая практически на пересечении этих районов, он взял 9 интервью в первом и 13 во втором. Очень много вопросов вызывает регистрация интервьюерами диспозиционных кодов: например, у одних интервьюеров для 11

интервью потребовалось 28, или 19 обращений, у другого — 160. Удивляет, и когда у интервьюера по улице Шоссейная идут следующие коды: с 49 кв. по 56 кв. — никого нет дома, с 57 кв. по 69 кв. — половозрастная квота, с 70 кв. по 89 кв. никого нет дома и т.д. У другого интервьюера по улице Михайлова: с 16 кв. по 20 кв. — никого нет дома, с 21 кв. по 27 кв. — отказ от интервью, с 28 кв. по 32 кв. — категорический отказ от интервью, с 33 кв. по 40 кв. никого нет дома. При этом видимого нарушения шага у этих интервьюеров нет. А вот интервьюер по маршруту на Большой Грузинской улице перестарался с шагом — у него каждое обращение, вне зависимости от результата, идет через 11 квартир: 1 кв., 12 кв., 23 кв., 34 кв., 45 кв... Это говорит о явном непонимании процедуры опроса большинством интервьюеров. У кого-то шаг в 5 квартир, у кого-то — в 10 квартир. В самом же отчете указаны шаги в 10 и в 11 квартир, что опять же свидетельствует о небрежном отношении к этому документу со стороны исполнителя.

Не всегда в маршрутным листам маршруты доведены до конца, хотя для этих маршрутов в базе приписок указано большее число интервью, совпадающее по числу с запланированными. Неизвестно, что является источником ошибки: забывчивость интервьюера заполнять маршрутные листы или же добавление полевым отделом необходимых квартир. И если в первом случае опрос действительно мог быть, то во втором это явная фабрикация.

Всего по имеющимся у нас маршрутным листам было зафиксировано 701 интервью, причем в 13 случаях число интервью в маршрутном листе и число адресов респондентов в отчетной документации не совпадает. Как минимум, в 35 маршрутных листах неправильно подсчитаны коды достижимости, соответственно, велика вероятность, что они неправильно перенесены в базу данных. Сами маршрутные листы иногда могли включать только половину или часть маршрута, но были и такие, где отрабатывались сразу два. В общей сложности по последним подсчетам у нас оказались в наличии 64 маршрутных листа, т.е. не хватать должно только 6. Но самое удивительное, что маршрутные листы и разработанные для них стартовые точки не всегда совпадают. Так, у нас появились 7 новых адресов, не указанных нигде ранее, а всего из имеющихся маршрутных листов стартовые точки изначально были указаны лишь для 57 маршрутов. Если смотреть на адреса респондентов: о на те дома, в которых проходил опрос, можно обнаружить большое число ошибок, например, в маршрутном листе ясно и четко зафиксирована улица Жукова, которой в Москве нет, а есть проспект Маршала Жукова. Соответственно, в базе приписок обозначена тоже улица Жукова, что ошибочно. Некоторые маршруты не совпадают с попавшими в выборку районами, особенно это относится к новым стартовым точкам. Спроектированное и реализованное число маршрутов совпадает только по 3-м округам из 9, причем в одном из них имеется ошибка в районе (табл. 5).

Таблица 5 Реализованная выборка по округам (исходя из маршрутных листов)

АО	Спроектированная выборка		Реализованная выборка		Есть ли ошибка в районах (1 — да, 0 — нет)
	Маршруты (кол-во)	Респонденты (кол-во)	маршруты	интервью	
ЦАО	5	55	4	44	1
САО	8	88	6	66	1
СВАО	8	88	5	57	1
ВАО	9	99	10	108	1
ЮВАО	8	88	8	88	0
ЮАО	11	121	10	112	0

ЮЗАО	8	88	8	88	1
ЗАО	7	77	7	77	0
СЗАО	6	66	5	49	0
Итого:	70	770	64	700	5

Если проанализировать дома, в которых проходил опрос, по всем обозначенным в 64 маршрутных листах респондентам (без учета номера квартиры), то получится, что в 30 из них обнаруживаются либо несуществующие, либо нежилые дома. Так, в маршруте по бульвару Яна Райниса указан дом №7 по улице Планерная, который является торговым центром. Аналогичная ситуация с адресом улица Гарибальди, дом № 23 — это торговый центр «Панорама», в котором, очевидно, нет жилых квартир. Всего на маршрутах были найдены в виде опрошенных домов небольшой магазин (ул. Люблинская, д. 27 — ларек), офисное здание (ул. Партизанская, д. 27), медицинское учреждение (ул. Абрамцевская, д. 16, к. 1 — ДЦ СВАО № 5), бытовое помещение (ул. Судостроительная, д. 11) и вовсе промышленная зона (8-я ул. Соколиной Горы, д. 26а). В данном случае мы имеем дело с прямыми фальсификациями и фабрикациями интервьюеров, не проверенными опросной компанией. На наш взгляд, напрямую оценить число ошибок, а также всех случаев намеренного нарушения исследовательской этики не возможно.

Третий этап (после поля)

Перейдем к последнему этапу: проверке полевой работы, вводу данных и их редактированию исполнителем. В первую очередь мы обнаружили большое число ошибок в названии улиц: Новокузнецкая вместо Новошукшинская, Алышов и Альков переулков вместо Алымов, Дубнинская вместо Дубининская, улица Болотников вместо Болотниковской и т.д. В базу приписок дважды внесены адреса маршрута по Осташковскому шоссе. Были отмечены также исправления номеров домов (например, интервьюер не соблюдал шаг — в отчете изменены номера квартир так, чтобы шаг был не менее 11 квартир, или по маршрутному листу опрос был всего в двух домах, а в отчете указано минимально необходимое число — 4). Более чем в 20 случаях изменен возраст респондента. Можно ли это списать на случайные ошибки? Вряд ли, особенно когда меняется шаг. Обнаруживаются и респонденты, не указанные в имеющихся маршрутных листах, т.е. их наличие не подтверждается полевой документацией.

Далее, по адресам респондентов (из списка 783 адресов) были выявлены еще 2 маршрута по СВАО, которые присутствуют в набивке: по Янтарному проезду (район Лосиноостровский) и по улице Лескова (район Бибирево). По обоим этим маршрутам отсутствуют маршрутные листы, и мы не можем проверить их достоверность. В любом случае, по этим данным, в каждом из этих районов были реализованы 2 маршрута вместо одного, и все эти 4 маршрута занесены в массив.

Есть 2 маршрутных листа, которые отсутствуют в базе: по улице Енисейская (Бабушкинский район) и по улице Вешняковская (район Вешняки). При более детальном изучении этих маршрутных листов, а также в виду того что не так много адресов в базе приписок остались не подтвержденными, мы решили сопоставить имена и возраст респондентов. Оказалось, что имена и данные всех респондентов из этих двух маршрутных листов в базе присутствуют, но их адреса изменены на адреса районов Митино и Строгино (оба из СЗАО). Удивительно, что совпадают даже номера квартир и имена интервьюеров.

Получается, что имеет место прямая фальсификация, когда запасные интервью с одного административного округа переписываются на другой, причем исходные данные не меняются, т.е. подлог обнаружить не так сложно (ведь нам были даны маршрутные листы). Это говорит не только о халатности полевого отдела, но даже об определенной наглости и уверенности в том, что подлог раскрыт не будет, что еще раз подтверждает уверенность в том, что эти все документы нужны лишь для проформы.

В целом отчетная документация по третьему этапу содержит 71 маршрут по Москве, в котором обозначены 783 респондента (табл. 6). Эти данные не сильно противоречат тому, что планировалось в спроектированной выборке: 70 маршрутов на 770 респондентов. Можно предположить, что ошибка объясняется как раз двойной набивкой респондентов по одному маршруту. Если же разбить эти маршруты на округа, то полное совпадения между спланированными респондентами и опрошенными обнаружится только в 4-х АО (табл. 6), причем только в одном из оставшихся 5 расхождение незначительно. Если же свести всю отчетную документацию по всем трем этапам вместе, мы получим совпадения уже только в 53 маршрутах и 570 респондентах и только 2 чисто отработанных АО. Например, по СВАО совпали только 3 маршрута и 35 респондентов, что дает всего 40%. Таким образом, в самих отчетных материалах обнаруживается преимущество только для 74% данных, что не может не шокировать.

Таблица 6 «Отчетная» выборка

АО	Спроектированная выборка		Отчетная выборка		Совпадает полностью		% совпадений
	маршруты (кол-во)	респонденты (кол-во)	маршруты (кол-во)	респонденты (кол-во)	маршруты (кол-во)	респонденты (кол-во)	
ЦАО	5	55	5	55	3	33	60
САО	8	88	7	75	5	55	63
СВАО	8	88	10	112	3	35	40
ВАО	9	99	8	88	7	75	76
ЮВАО	8	88	8	88	8	88	100
ЮАО	11	121	11	122	10	111	92
ЮЗАО	8	88	9	100	6	67	76
ЗАО	7	77	7	77	7	77	100
СЗАО	6	66	6	66	4	39	59
Итого:	70	770	71	783	53	570	74

Наиболее важной ошибкой третьего этапа является отсутствие преимущества всей отчетной документации, которую предоставил исполнитель. Именно ввиду обнаруженного нами несоответствия при анализе маршрутных листов возник вопрос о чистоте предоставленных данных, что привело к такому тщательному и даже в чем-то дотошному анализу материалов. Очевидно, что именно третий этап необходим, для того чтобы привести данные в презентабельный вид, по возможности сгладив все шероховатости и исправив недочеты, а при явных несоответствиях воспользоваться возможностью сфабриковать и сфальсифицировать. Мы отдаем себе отчет в том, что если бы опросная компания уделила достаточное внимание предоставляемым данным, скрыла бы все неточности, обнаружить подлог было бы крайне сложно. И то, что обнаружили мы, по идее, должно было быть обнаружено и исправлено полевыми менеджерами именно на третьем, послеполевым этапе.

Игнорирование исполнителем необходимости корректировки и приведения в соответствие с требованиями не только итогового массива, но и менее важных документов, по которым можно судить о проведенном опросе, говорит скорее о том, что эта задача для компании не только не является первостепенной, но и, вполне возможно, расценивается как никому не нужный труд. В противном случае стала бы она рисковать своим имиджем, предоставляя внутренне противоречивую информацию о своей работе, более того, с очевидными доказательствами значительных фальсификаций и фабрикаций?

Мы сделали и следующий шаг — дополнительно проанализировали те 53 «правильных» маршрута, которые совпадают во всей документации по трем этапам. Во-первых, мы просмотрели эти маршрутные листы на предмет ошибок. Во-вторых, проверили наличие домов, в которых был проведен опрос, при помощи Яндекс-карт⁴ с панорамами улиц. В итоге из 53 маршрутов в 23 были найдены несуществующие и нежилые дома, что сужает число «чистых» маршрутов до 30, а это уже всего 43% от общей выборки по Москве. Далее, из этих 30 маршрутов в половине обнаруживается нарушение шага, в ряде случаев достаточно грубое. Таким образом, без нарушений или ошибок остается всего 15 маршрутов, или 21% от общего числа. Это радикальное снижение числа четко отработанных маршрутов, где на выходе остается всего пятая часть от общего числа запанированных анкет, мы метафорично назвали воронкой правды (рис. 1)⁵.

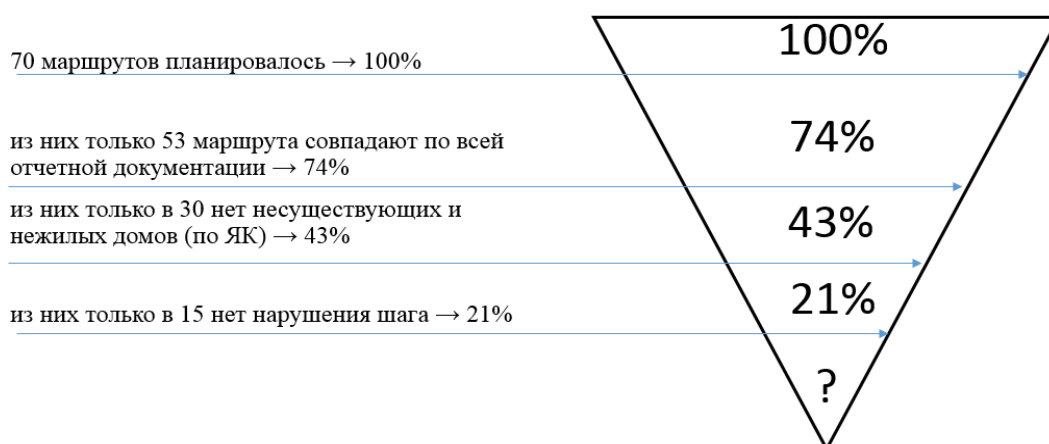


Рисунок 1 — Воронка правды

Оставшиеся 15 маршрутов, конечно же, должны быть проверены более детально. Например, в 4-х из них большие вопросы вызывают передвижения интервьюера по маршруту, а в 3-х — количество обращений.

Выводы

Кратко обозначим результаты, к которым мы пришли в ходе проведенного анализа:

- стартовые точки маршрутов не всегда совпадают с районами спроектированной выборки (последние могут заменяться, дублироваться или отсутствовать). Число районов не совпадает с указанным в спроектированной выборке. Для 15%

⁴ <http://maps.yandex.ru/?ll=37.617671%2C55.755768&spn=1.290894%2C0.367081&z=10&l=map>.

⁵ Автор хотела бы выразить благодарность Д.М. Рогозину за помощь в визуализации результатов.

- домохозяйств, включенных в выборку, обнаружены несоответствия в отчетной документации;
- в методическом отчете опросной компании указываются разные объемы выборки для Москвы: в одном месте — 794 респондента, ниже на три страницы — 803 респондента. В массив данных внесено 808 анкет;
 - на предварительном инструктаже интервьюерам не разъяснились шаг и процедура регистрации обращений. Было обнаружено всего 25 маршрутных листов, в которых соблюдается методика опроса, что составляет 36% от планируемых 70 маршрутов;
 - в 30 из 64 маршрутных листов (без учета номера квартиры) обнаруживаются либо несуществующие, либо нежилые дома;
 - обнаружено большое число ошибок и прямых фальсификаций при вводе данных: ошибки в названии улиц, дублирование адресов, исправления номеров домов, изменения возраста респондента, добавление респондентов, не указанных в маршрутных листах, множественные ошибки при подсчете диспозиционных кодов;
 - были обнаружены 2 случая прямой фальсификации (для 22 анкет, или 2 маршрутов);
 - только по отчетным материалам обнаруживается преемственность всего для 74% данных;
 - всего 15 маршрутов из 70 запланированных (21%) не имеют тех или иных ошибок.

Вместо заключения

Является ли полевая документация, которая должна служить для оценки и подтверждения проведенной работы, пережитком, реликтом исследовательской работы? Какие данные на самом деле собираются? Как организована работа внутри большой опросной компании? Есть ли спрос на методологию исследования у заказчика? Как работают опросные компании, если по факту у них нет нормального контроля интервьюеров? И самое главное — почему данные, полученные с таким нарушением процедуры, совпадают с данными Росстата?

Эти вопросы скорее риторические. Можно говорить об отсутствии спроса на методологию в научном сообществе, о чем свидетельствуют небольшие описания методологии исследования в большинстве отечественных статей; можно вспомнить и про препятствия к проведению опросов, про сложность реализации спроектированной выборки в мегаполисе, про краткие сроки для подготовки материалов (к слову, краткими они не были), про ошибки из-за человеческого фактора, про текучесть кадров среди интервьюеров, непривлекательность данной работы и большие соблазны подделать анкеты для личной выгоды. На последнем пункте остановимся подробнее. В большинстве исследований интервьюеры работают временно, по совместительству и в основной своей массе не участвуют в разработке дизайна исследования, выборки, анкеты; более того, не всегда они имеют соответствующую подготовку, в том числе они не знакомы с исследовательской этикой. Фальсификация интервьюера в данном случае означает намеренное нарушение требований проведения опроса или инструкций, полученных от начальства [5]. Принципиально важно,

чтобы интервьюер понимал, что он нарушает исследовательскую процедуру, которая может привести к искажению данных, что возможно только в том случае, если он является участником процесса исследования, а не просто сборщиком данных. Сам факт того, что интервьюеры в большинстве своем нарушают шаг, опрашивая квартиры подряд, беря все 11 интервью в одном доме, при этом исправно заполняя протокол, говорит о том, что они не понимают, что это прямое нарушение. И это не вина интервьюеров. Для соблюдения методики опроса и улучшения качества исследования как такового интервьюеры должны быть вовлечены в нечто большее, чем просто сбор данных, о чем однозначно говорит Американская ассоциация исследователей общественного мнения: «Исследователи общественного мнения (опросные компании) принимают на себя обязательство достоверно собирать данные и докладывать о них. В дизайне исследования, сборе информации, обработке данных опроса они должны быть привержены принципам целостности исследовательских данных. Для эффективности эти обязательства должны быть распространены не только на администрацию, но и на весь персонал, включая интервьюеров» [5]. До тех пор пока осознание этого факта не дойдет как до заказчиков исследований, так и для их исполнителей, мы будем иметь масштабные нарушения не только процедуры опроса, но и исследовательской этики в целом. Конечно, данный путь — не панацея, и мы понимаем, что нарушение исследовательской этики — явление неизбежное, но не хотелось бы вместо сбора и анализа данных заниматься их ненамеренным производством.

Литература

- 1 Власов В. Вокруг плагиата // Медицинская газета. 2007. № 41. URL: <http://www.mgzt.ru/article/387>.
- 2 Гудстейн Д. Обман в науке / пер. с англ. Ж. С. Журавлевой и Л. Н. Крыжановского // Успехи физических наук. 1993. Т. 163, № 1.
- 3 Кон И. 80 лет одиночества. М. : Время, 2008.
- 4 Fanelli, D. How many scientists fabricate and falsify research? Systematic review and meta-analysis of survey data // PLoS ONE. 2009. No.4. [Online] <<http://www.plosone.org/article/info:doi/10.1371/journal.pone.0005738>> [Date of access] 9.04.2014
- 5 American Association for Public Opinion Research (AAPOR). (2003) Interviewer falsification in survey research: Current best methods for prevention, detection and repair of its effects / Paper presented at Ann Arbor Falsification Summit on Interviewer Falsification. [Online] <<http://www.amstat.org/sections/srms/falsification.pdf>> [Date of access] 9.04.2014