

В.Б. ЗВОНОВСКИЙ

РЕАЛИЗАЦИЯ РЕПРЕЗЕНТАТИВНОЙ ВЫБОРКИ В МАССОВОМ ОПРОСЕ

В задачи социологического исследования нередко входит формирование выборочной совокупности, репрезентирующей население той или иной территориальной общности, например города, региона. Как известно, наиболее эффективно эту задачу решает случайная или квазислучайная выборка. В нашей исследовательской практике необходимо было построить выборку для проведения полевого этапа измерения потребительской активности населения Самарской области.

Очевидно, потребительское поведение индивидов, равно как и больших социальных групп, значимо связано с их подвижностью, а следовательно, оно влияет и на их достижимость в случае опроса по месту жительства. Так, люди, чьи возможности ограничены в силу различных причин (физических, имущественных и проч.), судя по всему, относятся к малопотребляющим группам, в то время как те, кто отличается высокой потребительской активностью, реализуют ее вне стен своего жилища, и застать их по месту жительства сложно. Поэтому столь важна была практическая реализация репрезентирующей население области выборки, которая обеспечивала бы точность измерительной процедуры.

В рамках мониторинга индекса потребительской активности населения Самарской области в 2005 г. проведено методологическое исследование с целью выявить преимущества и недостатки основных методов отбора респондентов для участия в опросе — адресной и именной выборки, — а также способы улучшения качества отбора. Помимо этого, исследователи поставили перед собой задачу адаптировать стандарты американской ассоциации исследователей рынка и общественного мнения (AAPOR) к нашей практике, рассчитав предлагаемые количественные показатели результативности контактов.

Дизайн выборки

ИПН Самарской области¹ строится на основе данных опросов общественного мнения, проводящихся один раз в три месяца Фондом социаль-

Звоновский Владимир Борисович — кандидат социологических наук, президент Самарского областного фонда социальных исследований. **Адрес:** 443041 г. Самара, ул. Красноармейская, д. 70. Самарский областной фонд социальных исследований. **Телефон:** (8462) 69-00-26. **Электронная почта:** client@fond.sama.ru

¹ Индекс потребительских настроений (ИПН) представляет собой количественный показатель, отражающий диспозицию населения к наиболее общим формам потребительского поведения в контексте оценок личного материаль-

ных исследований. В основе опросов — специально спроектированная многоступенчатая выборка, репрезентирующая взрослое (старше 18 лет) население Самарской области. Всего в 2005 г. проведено четыре опроса: тестовое измерение ИПН в г. Самаре (объем выборки — 544 респондента); объем выборки в I и II волнах — в рамках всей Самарской области — составлял 1202 и 1154 человека соответственно, в последней, III волне, — 1024 человека. Благодаря многоступенчатости отбора предполагалось точнее воспроизвести структуру населения области. В качестве критериев стратификации были выбраны место жительства, пол, возрастная группа.

Отбор проводился в четыре этапа. На первом отбирались населенные пункты для проведения опроса; на втором — точки опроса, представляющие собой избирательные участки внутри этих населенных пунктов; третий этап включал отбор домохозяйств; четвертый — конкретных респондентов. В основу стратификации по месту жительства положены следующие критерии: размер населенного пункта и его расположение относительно Самары (областного центра) и городов Тольятти и Сызрани. Всего было выделено 8 страт (областной центр; крупный областной город; малые города области; пригородные районы; пригородные ПГТ (поселки городского типа); удаленные ПГТ; районы с дисперсным сельским населением; районы с концентрированным сельским населением). Далее страты, включающие население Самары, Тольятти и малых городов области, были разделены на подстраты с целью максимально точного воспроизведения в выборочной совокупности структуры населения. На втором этапе во всех стратах был произведен отбор кластеров, за которые принимались избирательные участки области²: в каждой из полученных на первом этапе подстрат методом случайного бесповторного отбора были определены избирательные участки, в которых должен проводиться опрос.

Третий и четвертый этапы (отбор домохозяйств) были различными для крупнейших городов области Самары и Тольятти, с одной стороны, и остальных населенных пунктов — с другой. Для последних отбор домохозяйств проводился по маршруту с заданным шагом, интервьюер получал описание избирательного участка³ и обходил его с самого начала

ного положения и экономической ситуации в целом. Данный индекс (Consumer sentiment index) был предложен специалистами Университета Мичигана в 1946 г. В настоящее время это ведущий индекс США для прогнозирования потребительской активности населения.

² Кластеризация имела целью отобразить локализованное на территориях население, представляющее структуру выборочной совокупности.

³ Упорядоченный список домохозяйств или групп домохозяйств, представленный по закону о выборах. Описание избирательных участков соответству-

по порядку, определявшемуся самим интервьюером. В городах и других населенных пунктах, разделенных на два и более избирательных участка, маршрут представлял собой последовательность включенных в этот участок адресов. В малых (один избирательный участок и менее) населенных пунктах интервьюер начинал обход домохозяйств с самой большой улицы. В домохозяйствах респонденты отбирались согласно методике «ближайшего дня рождения». Для этого члены домохозяйства переписывались с указанием основных демографических характеристик — пола, даты рождения — затем из списка отбирался респондент, удовлетворяющий заданному условию.

В Самаре и Тольятти на третьем и четвертом этапах восстанавливалась (составлялась) полная база домохозяйств, принадлежащих отобранному избирательному участку. Для этого проводилась перепись домохозяйств в следующем формате: название района, название улицы, номер дома по улице, номера квартир в этом доме, а также номера нежилых квартир. Из этой базы с помощью модуля SPSS Complex Samples случайным образом отбирались домохозяйства для проведения интервью. В Самаре и Тольятти отбор респондентов в разное время осуществлялся по двум схемам — адресной и именной. Данное методическое исследование посвящено реализации репрезентативной выборки на третьем и четвертом этапах проекта в Самаре и Тольятти⁴.

Адресная схема отбора респондентов

При адресной схеме отбора каждый интервьюер должен был опросить на выделенном ему избирательном участке определенное (указанное руководителем работ) количество респондентов. С этой целью интервьюеру выдавался список адресов участка, число которых вдвое превосходил число требуемых законченных интервью.

Как отмечалось, в домохозяйствах респондент отбирался согласно ближайшему дню рождения. Если отобранный респондент присутствовал дома, интервьюер опрашивал его, если нет, — то интервьюер договаривался о новом визите и переходил в следующее домохозяйство. Респондент также мог длительное время (более 1 недели) отсутствовать дома. В этом случае можно было опросить следующего «по дню рожде-

ет составленному к выборам депутатов Государственной думы ФС РФ 7 декабря 2003 г.

⁴ Реализация аналогичного процесса формирования выборочной совокупности в других населенных пунктах была намечена на последующие этапы проекта ИПН в Самарской области. Но, к сожалению, он не был продлен, и данный материал — пока единственная фактологическая основа нашего исследования.

ния» респондента, живущего здесь же; если такового не находилось, интервьюер отправлялся по другому адресу.

Каждому интервьюеру выдавалось квотное задание, в котором указывалось, сколько респондентов определенного пола и возраста следует опросить на участке. После выбора квоты интервьюер переставал опрашивать тех респондентов, которые должны были быть опрошены согласно отбору по ближайшему дню рождения, и мог опросить другого члена данного домохозяйства, если он не являлся представителем выбранной квоты. Если же все члены данного домохозяйства являлись представителями выбранных квот, то интервьюер переходил к другому адресу. Данная схема отбора респондентов использовалась в I волне ИПН (июнь 2005 г.).

Именная схема отбора респондентов

Для создания выборки респондентов в Самаре и Тольятти была проведена предварительная работа — создана полная база респондентов⁵, проживающих на отобранных избирательных участках. Из этой базы с помощью модуля SPSS Complex Samples случайным образом были отобраны конкретные респонденты для проведения интервью.

Интервьюер для опроса получал список адресов, количество которых превышало необходимое количество законченных интервью в n раз — коэффициент запаса⁶. Когда интервьюер достигал респондента, прежде чем проводить опрос, он должен был сверить правильность написания его/ее имени, даты рождения и адреса с указанными в бланке. Если какие-то сведения не совпадали (например, отчество, или имя, или дата рождения), то нужно было исправить их прямо в бланке, зачеркнув неверное и написав правильное.

Данная схема отбора респондентов использовалась в мартовской (тестовой и проводившейся только в Самаре), II (сентябрь) и III (декабрь) волнах ИПН. Перед проведением сентябрьской волны интервьюерам стали выдавать квотные задания. Это обуславливалось следующим: исследователи опасались того, что интервьюеры, не сдерживаемые квотными заданиями и стремящиеся опросить как можно больше людей за как можно меньшее время, будут обращаться к тем, кого застанут дома, то есть представителям старших возрастных групп. После введения квотных ограничений интервьюер не мог в целях достиже-

⁵ В основе этой базы лежат неофициальные списки жителей городов, распространяемые на свободном рынке, накопленные за несколько последних лет.

⁶ Коэффициент запаса составил от 2 до 3,2 (для июньской волны — 2, для сентябрьской — 3,2, для декабрьской — 2,8). Коэффициент запаса рассчитывался на основе потребности интервьюеров в дополнительных адресах в период опроса по результатам предыдущей волны.

ния количественного плана опрашивать пожилых респондентов и должен был либо работать с имеющимся списком, либо запрашивать у руководителя новый.

Такая техника не является квотированием выборки в чистом виде, поскольку ее состав, включающий отобранных менеджером проекта респондентов, задается не квотами, а процедурой случайного отбора, выполненной программой. Тем не менее с целью реализации случайной выборки респондентов в чистом виде в мартовской и декабрьской волнах исследования была использована исключительно методика случайного отбора без коррекции ее квотными заданиями.

Эффективность выборки

Главный вопрос при проведении любого выборочного исследования — получение точных данных при минимальных затратах, временных, человеческих и материальных. Однако было бы неправильно сводить весь вопрос к банальной стоимости опроса одного респондента для достижения требуемой точности результата. При большой выборке увеличивается время проведения опроса. Также требуется большее количество интервьюеров — следовательно, увеличивается время, требуемое на их подготовку, растут и материальные затраты, связанные с доставкой интервьюеров к месту опроса, типографские расходы и проч. Эффективность выборки определяется методикой ее реализации.

Однако большая выборка означает не только ее большой объем, выраженный числом реально проведенных интервью, но и может быть связана с большим количеством попыток установления контактов. Большое количество неэффективных контактов неизбежно приводит к потерям времени, росту необходимой численности интервьюеров и операционных издержек. Напротив, грамотно составленная выборка позволяет избежать затягивания времени опроса, сократить затраты на его проведение.

Лесли Киш называл эффективной, или экономной, выборкой ту, которая дает более высокую точность при более низких затратах на опрос одного респондента (получение информации от одной единицы отбора)⁷. Поскольку в нашем случае речь идет о выборках одинакового объема и интервью, проведенных в одних и тех же точках опроса, под «затратами» следует понимать затраты времени интервьюера на опрос одного респондента. Так, чем меньшее число контактов требуется установить для получения одного и того же количества законченных интервью, тем более экономной (или эффективной) следует признать данную выборку. Чем больше мы получаем отказов от потенциальных респонден-

⁷ Kish L. Survey sampling. New York: Wiley, 1995. P. 25.

тов, чем меньше законченных интервью возникает в данной реализации выборки, тем она менее эффективна.

Согласно данному подходу к проблеме более эффективной выборки в рамках измерения ИПН, в течение 2005 г. обрабатывалась такая ее реализация, которая позволяет уменьшить количество отказов и при этом сохранить или даже увеличить (по сравнению с предыдущими волнами) долю состоявшихся интервью в общем числе контактов.

Преимущества и недостатки адресной и именной выборки

При проведении опроса по адресной выборке необходимо было реализовать методику «ближайшего дня рождения». Она призвана обеспечить воспроизведение структуры генеральной совокупности (каковой является совокупность жителей населенного пункта) в финальной выборке. Теоретически методика позволяет моделировать «случайность» отбора респондента. На практике интервьюер, не будучи привязанным к конкретному человеку, в целях минимизации затрат своего рабочего времени, может стремиться опросить респондента, чьи социальные характеристики соответствуют методике и квотному заданию. К тому же представители больших социальных групп (молодые, работающие более 10 часов в день, и др.) могут отсутствовать дома целыми домохозяйствами, в случае чего пропадает возможность перевести отбор на четвертый этап (отбор респондента). По этой причине участие в опросе представителей таких социальных групп становится ограниченным. Ошибка, вызванная этим смещением, становится систематической по характеру и значительной по величине. Она может превысить случайную ошибку и тем самым обесценить усилия по уточнению финальной выборки.

В случае именного отбора указанная проблема не возникает, поскольку мы заранее знаем, кого интервьюеры должны опрашивать. С другой стороны, это накладывает ограничения валидности на исходную базу жителей, которая по своим основным демографическим характеристикам должна быть более или менее полной и адекватной. Но и в этом случае вероятность нехватки «мобильных» респондентов достаточно велика, то есть сама по себе именная выборка не решает проблему их охвата. Тем не менее, при использовании именного отбора гораздо проще осуществлять контроль над соблюдением методики опроса — мы заранее знаем о принадлежности респондентов к основным социально-демографическим группам и можем оперативно рассчитать возникающее смещение выборки.

Еще один недостаток адресного отбора — отсутствие личного обращения к респонденту. Можно предположить, что именной отбор порождает гораздо менее отказов, чем адресный, обезличенный. Отсутствие необходи-

мости терять время на выявление подходящего респондента также играет свою роль.

К недостаткам именной выборки относится «дефект базы» — доля брака, которая существует в исходной базе жителей города, то есть домохозяйств, чьи представители принадлежат генеральной совокупности. Эта доля увеличивается со временем: люди умирают, рождаются, взрослеют, переезжают. Решением этой проблемы является периодическое обновление базы.

В таблице 1 показаны преимущества и недостатки, присущие рассмотренным видам отбора, формирующим репрезентативную выборочную совокупность.

Таблица 1

Преимущества и недостатки адресного и именного отбора	
Преимущества	Недостатки
Адресный отбор	
Нет ограничений, связанных с отсутствием в домохозяйстве конкретного респондента	Сложность контроля над выполнением методики отбора
Высокая валидность базы адресов	Большое количество отказов от интервью Смещение выборки в сторону менее «мобильных» респондентов
Именной отбор	
Проще осуществлять контроль над соблюдением методики	Низкая валидность базы жителей
Небольшое количество отказов от интервью	Смещение выборки в сторону менее «мобильных» респондентов

Численное подтверждение

Проверим вышеперечисленные утверждения, проанализировав статистику контактов по проведенным опросам. На текущий момент мы имеем статистику контактов по именной выборке (март), адресной выборке (июнь, I волна ИПН) и именным выборкам (сентябрь, II волна ИПН; декабрь, III волна ИПН) 2005 г.

Коэффициент ответов (процент состоявшихся интервью) при именном отборе был равен 32–34% для всех трех выборок, тогда как при адресном отборе он составил 24% (табл. 2).

Коэффициент отказов при именном отборе (процент отказавшихся участвовать в опросе) составил от 10 до 21%, а при адресном отборе — 35%. И это при том, что дефект базы при адресном отборе составил 2% (возможно, из-за не совсем корректно составленной контактной ведомости), а при именном отборе — от 16 до 26%.

Таблица 2

Сравнение результатов достижимости в схемах именного и адресного отбора, %

Результат контакта	Схемы отбора по датам опросов 2005 г.
--------------------	---------------------------------------

	именной, март (n=2526, m=1669)*	адресный, июнь (n=3568, m=2360)	именной, сентябрь (n=3265, m=2180)	именной, декабрь (n=3126, m=1662)
оконченных интервью	32	28	32	34
респондента нет дома	6	1	5	4
никого нет дома	21	19	16	8
отказ от интервью	10	34	14	20
дефект базы	26	2	18	24

*Примечание: n — число контактов; m — число контактных адресов

Вполне объяснимо, что при адресном отборе количество случаев, когда респондента не было дома, значительно меньше, чем при именном. При именованном отборе интервьюер должен опросить конкретного человека, тогда как при адресном он мог опросить любого человека из указанного домохозяйства, подходящего по методике «ближайшего дня рождения». В результате при именной выборке интервьюер должен был посетить данное домохозяйство снова, а число контактов, соответственно, увеличивалось. В то же время доля случаев, когда никого не было дома, приблизительно одинакова в июньской и сентябрьской, но гораздо меньше в декабрьской волне. Скорее всего это произошло за счет перераспределения результатов контактов, поскольку в последней (декабрьской) волне ИПН количество посещений увеличилось с трех до шести. Интервьюеры в этом случае лучше (полнее) отработывали выданный список потенциальных респондентов.

Многократное посещение

В ходе исследования в первую очередь опрашиваются наименее мобильные респонденты (в основном пожилые люди): интервьюеры в целях экономии сил и времени стремятся опросить как можно больше людей уже при первом посещении, упуская тем самым потенциальных респондентов, которых в данный момент нет дома (людей молодого и среднего возраста). Необходимая в результате такого смещения коррекция финальной выборки осуществляется либо через последующие посещения, либо через применение квот, либо путем совмещения этих двух способов коррекции.

Очевидно, что именно многократное посещение наиболее «мобильных» респондентов является более привлекательным способом, поскольку сохраняет «случайный» характер выборки. Эти группы представляют большой интерес для исследователей как важная часть репрезентируемой генеральной совокупности.

При каждом следующем посещении доля проведенных интервью либо возрастает, либо остается постоянной, а доля адресов, где «никого нет дома», возрастает (табл. 3). Кроме того, уменьшается

доля отказов. Можно заключить, что адреса «отрабатываются» — либо происходит отказ от интервью (чаще всего при первых посещениях, то есть отказываются «немобильные» респонденты), либо интервью проводится, либо в период опроса кого-либо из членов домохозяйства бывает невозможно застать.

Таблица 3

Результаты достижимости респондентов в июньской, сентябрьской и декабрьской выборках 2005 г. (трехкратное посещение), %

Результат контакта	Июнь			Сентябрь			Декабрь		
	1	2	3	1	2	3	1	2	3
оконченных интервью респондента	19	18	16	20	23	27	19	18	17
нет дома	2	3	1	9	15	11	14	16	21
никого нет дома	39	51	50	29	36	45	23	32	37
отказ от интервью	25	17	17	11	9	8	12	12	9
дефект базы	2	0	0	17	7	0	19	8	6

Трехкратное посещение

Как отмечалось выше, в реальной опросной практике при первом посещении опрашиваются в основном люди старшего возраста. При последующих посещениях процент опрошенных этой возрастной категории уменьшается (табл. 4). Напротив, доля опрошенных молодых людей растет при втором и третьем посещениях, в то время как при первом вероятность застать их дома невелика. Людей среднего возраста чаще всего удается опросить при втором посещении: у них более размеренный ритм жизни, и часто бывает достаточно при первом посещении познакомиться с кем-то из домохозяйства, чтобы в следующий раз наверняка застать нужного респондента. Для молодых людей характерен неритмичный и мобильный образ жизни, поэтому главным способом застать их в месте проживания является увеличение числа посещений.

Таблица 4

Распределение опрошенных по возрастным категориям при трехкратном посещении (мартовская, сентябрьская и декабрьская выборки 2005 г.), %

Возрастная группа	Март			Сентябрь			Декабрь		
	1	2	3	1	2	3	1	2	3
18–35 лет	27	35	45	32	32	48	27	36	39
36–55 лет	31	36	20	38	32	38	32	41	28
56 лет и старше	42	29	35	30	36	14	41	23	33

Мобильность молодежи отражают данные, приведенные в таблице 5. Около половины (47–52%) респондентов, сменивших место жи-

тельства, являются именно молодыми людьми, тогда как их доля среди всего наличного населения составляет лишь 35%.

Таблица 5

Распределение по возрасту респондентов, сменивших место жительства (именная выборка), %

Возрастная группа	Сентябрь	Декабрь
18–35 лет	52	47
36–55 лет	34	39
56 лет и старше	14	14

Шестикратное посещение

Увеличение количества посещений с трех до шести дало возможность повысить долю опрошенных респондентов 18–35 лет, что в свою очередь позволило в наибольшей степени приблизить социально-демографические характеристики выборочной совокупности к характеристикам генеральной (табл. 6).

Таблица 6

Распределение опрошенных по возрастным группам при шестикратном посещении (декабрьская выборка), %

Возрастная группа	Посещения			
	1	2	3	4-6
18–35 лет	27	36	39	42
36–55 лет	32	41	28	33
56 лет и старше	41	23	33	25

Шестикратное посещение способствовало увеличению достижимости респондентов, которых ранее не заставали дома. Среди респондентов, которых опросили не с первого раза, молодежи больше, чем пожилых людей, — что подтверждает следующие выводы: количество опрошенной молодежи, то есть наиболее мобильной части населения, увеличивается с каждым посещением, а основная доля пожилых респондентов опрашивается с первого раза (табл. 7).

Таблица 7

Доли возрастных групп респондентов, которых удалось опросить в результате шестикратного посещения, %

Возрастная группа	Схемы выборки по датам опросов 2005 г.			Генеральная совокупность
	адресная, июнь	именная, сентябрь	именная, декабрь	
18–35 лет	40	36	38	35
36–55 лет	37	39	38	38
56 лет и старше	23	25	24	27

О качестве баз данных по жителям города и статистике

Тем не менее, шестикратное посещение не избавляет полностью от смещения в сторону пожилых групп (табл. 8). Здесь следует указать на то, что в декабрьской волне ИПН подготовка списка потенциальных респондентов для полевых работ подразумевала полностью случайную его генерацию, в отличие от июньской, когда отбор в базе производился в соответствии с пропорциями по статистическим данным.

Таблица 8

Доли возрастных групп опрошенных респондентов (по контакт-ным ведомостям), %

Возрастная группа	Схемы выборки по датам опросов 2005 г.			Генеральная совокупность	Исходная база данных
	адресная, июнь	именная, сентябрь	именная, декабрь		
18–35 лет	35	34	32	35	33
36–55 лет	38	38	36	38	34
56 лет и старше	27	28	32	27	33

Вопрос о расхождении (причем значимом при больших выборках) между возрастной структурой населения, предоставляемой органами государственной статистики, и базами данных, чаще всего составленных паспортными столами, входящими в систему МВД, является серьезной проблемой⁸. Поскольку на данный момент точного ответа на этот вопрос нет, возможно, будет корректным усреднить данные о долях возрастных групп в генеральной совокупности, и полученные в результате такой процедуры данные считать целевыми для коррекции выборки.

Тем не менее, можно сделать твердый вывод о необходимости квотирования возрастных групп потенциальных респондентов, отбираемых для опроса из базы жителей. Это позволит скорректировать начальную, а, значит, и финальную выборки в сторону большего соответствия возрастной структуре генеральной совокупности.

Проблема достижимости молодежи при опросах

Поскольку задача репрезентации молодежи в выборочной совокупности даже после шестикратного посещения осталась нерешенной, необходимо устранить возникшее смещение в сторону респондентов среднего и пожилого возраста уже после окончания полевой части исследования. В принципе для решения этой проблемы существует два пути. Первый — можно искусственно увеличить объем выборки, то есть сначала дополнительно взять столько интервью у всех возрастных групп, сколько требуется для достижения планового числа молодежной группы, а затем «отремонтировать»

⁸ Вопрос о происхождении такого смещения мы оставляем за пределами нашего исследования.

выборку — случайным образом исключить из выборочной совокупности излишние анкеты средней и старшей возрастных групп.

Второй путь — это перевзвешивание полученного массива по полу, возрасту и месту проживания. Несмотря на все недостатки перевзвешивания данных, именно этот способ коррекции финальной выборки представляется наиболее предпочтительным, поскольку позволяет сохранить в том или ином виде все собранные валидные материалы полевого этапа.

Суточная динамика посещений

Время посещения домохозяйств является фактором, очевидным образом влияющим на достижимость респондентов и репрезентативность финальной выборки. Рассмотрим суточную динамику посещений респондентов в сентябрьской и декабрьской волнах.

В сентябре время первого посещения чаще всего приходилось на вторую половину дня, с 14 до 20 часов, тогда как в декабре оно переместилось на дневное время, с 11 до 18 часов. Скорее всего это объясняется тем, что зимой ходить по неосвещенным улицам в незнакомом месте по вечерам довольно опасно и некомфортно. Однако реальное положение вещей вынудило интервьюеров переносить опросы на вечернее время, когда шансов застать дома работающую часть населения гораздо больше. Поэтому при втором и третьем посещениях опрос переносится на все более позднее время, фактически такое же, как в периоды сентябрьской и декабрьской волн. В конце концов, время опросов в декабре в значительной степени приблизилось ко времени опросов в сентябре, хотя существенное смещение зимней суточной динамики относительно осенней к более светлому времени суток вполне очевидно.

Определенное влияние на время опроса оказывает день недели. Так, в будни интервью состоялись в основном в вечерние (послерабочие) часы, а в выходные дни время опроса распределялось с 12 до 17 часов. Это особенно явно прослеживается в сентябрьской выборке. Причем если суточная динамика интервью в осенние выходные практически не отличается от аналогичной в зимние, то декабрьская кривая для рабочих дней очевидно смещена в сторону более светлого времени суток.

Растет ли число отказов при увеличении количества посещений?

Выше отмечалось, что использование именной выборки дает, по сравнению с адресной, меньшее число отказов. Теперь необходимо выяснить, не приводит ли это к росту количества посещений, направленных на достижение мобильных респондентов. Действительно, при увеличении числа посещений с трех до шести увеличилось и количество отказов (табл. 9). Однако этот рост отражает повышение уровня достижимости респондентов: в декабрьском опросе было больше контактов непосредственно с респондентами, что существенно

уменьшило количество кодов «никого нет дома», но увеличило долю «дефектных» адресов и число отказавшихся от ответов. Все это указывает на то, что при задании шестикратного посещения интервьюеры более глубоко, «взвездливо» поработали с предоставленной им базой потенциальных респондентов и значительно чаще достигали их. При этом доля состоявшихся интервью не уменьшилась. Иначе говоря, доля отказов возросла, но не за счет уменьшения доли проведенных и законченных интервью, а за счет уменьшения числа «неконтактов» — кодов «никого нет дома» и «нет респондента»⁹.

Таблица 9

Распределение «неконтактов» с респондентами (неопрошенных респондентов), %

Результат контакта	Схема выборки по датам опроса 2005 г.			
	именная, март	адресная, июнь	именная, сентябрь	именная, декабрь
респондента нет дома	6	1	5	4
никого нет дома	21	19	16	8
отказ от интервью	10	34	14	20

Среди отказавшихся респондентов, в том числе и при первом посещении, молодежи меньше, чем пожилых людей (исключение составляет сентябрьская выборка) (табл. 10, 11). Иначе говоря, обращение интервьюера к молодым респондентам не вызывает у последних желания уклониться от участия в опросе, в отличие от старших возрастных групп.

Таблица 10

Распределение отказавшихся от интервью по возрастным группам, %

Возрастные группы	Схема выборки по датам опроса 2005 г.		
	адресная, июнь	именная, сентябрь	именная, декабрь
18–35 лет	27	38	31
36–55 лет	48	37	35
56 лет и старше	25	25	34

Таблица 11

⁹ Часть интервьюеров, стремясь сэкономить время, указывали в контактных листах не код «никого нет дома» или «нет респондента», а «отказ», поскольку это отменяло необходимость посещать данный адрес еще раз. Контрольная проверка, проведенная после окончания полевых работ, показала, что из 24 «отказов» реально отказались лишь 2 респондента. Остальные ничего не знали об опросе, что скорее всего означает отсутствие первичного контакта с интервьюером.

Распределение отказавшихся от интервью после первого посещения по возрастным группам, %

Возрастные группы	Схема выборки по датам опроса 2005 г.		
	адресная, июнь	именная, сентябрь	именная, декабрь
18–35 лет	19	32	28
36–55 лет	40	41	35
56 лет и старше	41	26	37

Среди респондентов, отказавшихся от интервью, которых перед этим не заставали дома, молодых людей больше, чем пожилых. Иными словами, если при первом посещении от интервьюирования уклонялись в основном респонденты пожилого возраста, то с каждым следующим посещением наблюдался рост доли молодых респондентов, отказавшихся от участия в опросе (табл. 12).

Таблица 12

Возрастное распределение респондентов, отказавшихся от интервью, которых не сразу удалось застать дома, %

Возрастные группы	Схемы выборки по датам опросов 2005 г.			Генеральная совокупность
	адресная, июнь	именная, сентябрь	именная, декабрь	
18–35 лет	39	43	36	35
36–55 лет	49	33	36	37
56 лет и старше	12	23	28	28

Это еще раз доказывает необходимость неоднократного посещения адресов для повышения результативности опроса (увеличение количества опрошенных) и восстановления возрастных пропорций (уменьшение доли не опрошенных молодых людей, полученной при первом посещении). Опасения, что неоднократно посещающий домохозяйство интервьюер вызывает негативную реакцию, не подтвердились. Неоднократное посещение позволяет уточнять информацию об адекватности имеющейся базы данных — выявляются люди, сменившие место жительства, умершие, сдающие или снимающие жилье.

Адаптация диспозициональных кодов AAPOR к российской опросной практике

Исследователи предприняли попытку адаптировать диспозициональные коды, стандартизованные AAPOR, к российской опросной реальности¹⁰. Адаптация состояла из трех этапов. Во-первых, интер-

¹⁰ Стандартные определения: систематическое описание диспозициональных кодов и коэффициентов результативности для массовых опросов. (Американская ассоциация исследования общественного мнения. 3-е изд. 2004 г.) / Пер. с англ. Д.М. Рогозина, Е.М. Киселева // Социологический журнал. 2005. № 2. С. 78–120.

претация кодов в российских условиях. Во-вторых, сопоставление их с теми, которые использовались нами в реальном исследовании. В-третьих, спецификация использования кодов при адресной или именной выборках.

В приложении приведены результаты всех трех этапов этой адаптации (табл. 14), а также количественные результаты проведенных Фондом социальных исследований в 2005 г. полевых работ в Самарской области в рамках проекта «Индекс потребительских настроений» (табл. 15.1–15.4).

Далее в соответствии со «Стандартными определениями для массовых опросов» на основе диспозиционных кодов были рассчитаны следующие показатели результативности:

- RR* — коэффициент ответов;
- COOP* — коэффициент кооперации;
- REF* — коэффициент отказов;
- CON* — коэффициент контактов.

При расчете этих коэффициентов использовались следующие обозначения:

- I* — полностью завершенное интервью;
- P* — частично завершенное интервью;
- R* — отказ или прерывание;
- NC* — не установлен контакт;
- O* — другое.

Коэффициент ответов *RR* — это числовое значение, получаемое при делении количества полностью взятых интервью на сумму всех интервью (завершенных и незавершенных) и «неинтервью» (отказы, прерывания, отсутствие и прочее).

$$RR = \frac{I}{(I + P) + (R + NC + O)}$$

Коэффициент кооперации *COOP* — это отношение опрошенных ко всем единицами наблюдения, удовлетворяющим условиям выборки, с которыми удалось связаться.

$$COOP = \frac{I}{(I + P) + R + O}$$

Коэффициент отказов *REF* — это отношение домохозяйств или отдельных респондентов, которые отказались участвовать в опросе или прервали интервью, ко всем потенциально удовлетворяющим условиям выборки единицам наблюдения.

$$REF = \frac{R}{(I + P) + (R + NC + O)}$$

Коэффициент контактов *CON* измеряет долю всех случаев, в которых удалось связаться с ответственным членом домохозяйства, от всех попавших в выборку единиц, удовлетворяющих заданным условиям.

$$CON = \frac{(I + P) + R + O}{(I + O) + (R + O + NC)}$$

Таблица 13

Расчет коэффициентов результативности опроса, %

Коэффициент	Схема выборки по датам опроса 2005 г.			
	именная, март	адресная, июнь	именная, сентябрь	именная, декабрь
RR	32	18	21	34
COOP	67	36	54	54
REF	10	23	10	20
CON	48	50	40	64

Именная выборка, реализованная в ходе опросов в марте, сентябре и декабре 2005 г., дала более высокий уровень коэффициент ответов (21–34%) по сравнению с адресной (18%), а также существенно более высокий коэффициент сотрудничества (кооперации) (54–67% против 36%), более низкий коэффициент отказов (10–20% против 23%) (табл. 13). А шестикратное посещение, использованное в декабре, позволило довести коэффициент контактов, то есть долю состоявшихся контактов с представителями домохозяйства, где проживает потенциальный респондент, до 64%.

Выводы

Проведенное методологическое исследование показало более высокую эффективность использования именной выборки по сравнению с адресной. Во-первых, уменьшается доля отказов при сохранении прежнего количества отказов, что позитивно сказывается на охвате мобильных групп населения, во-вторых, использование квазислучайного способа отбора, в отличие от квотного, позволяет при анализе полученных данных использовать все доступные его виды и инструменты, оценивая статистически значимые различия между группами.

Многokратное посещение позволяет уменьшить смещение, вызванное высокой мобильностью отдельных социальных групп, в первую очередь, молодежи. Это также позитивно сказывается на увеличении охвата различных групп генеральной совокупности и улучшении репрезентации последней в выборочной совокупности. При этом доля отказов не увеличивается.

При организации полевых этапов подобных исследований в дальнейшем следует квотировать первичную выборку, то есть список потенциальных респондентов, генерируемый из баз данных о жителях городов и других населенных пунктов. При формировании квот следует учитывать как данные о составе основных демографических групп, предоставляемые органами статистики, так и данные о жителях, циркулирующие на свободном рынке, источником которых чаще

всего являются информация органов МВД. В случае расхождений этих источников наиболее правильным будет усреднение данных о соотношении основных половозрастных групп в генеральной совокупности. Полученная таким образом первичная выборочная совокупность будет точнее всего представлять население.

Разумеется, для проведения полного анализа контактов, в том числе при оценке собственно результатов исследования, необходим учет всех проведенных контактов примерно в том объеме, в каком это было сделано в данном методологическом исследовании. Даже линейные данные о достижимости, числе отказов, контактов с потенциальными респондентами позволяют делать выводы и уточнять количественную, а, возможно, и качественную информацию, полученную в результате исследования.

Использованные инструменты (первичная выборка, многократное посещение, полный отчет о контактах с потенциальными респондентами) освобождают исследователя от учета суточной динамики проведенных опросов: интервьюер вынужден при выполнении всех инструкций вольно или невольно посещать потенциальных респондентов тогда, когда они находятся в месте проживания, в будни — в вечерние часы, в выходные — в дневные.

Адаптация диспозиционных кодов AAPOR к российской опросной практике

Формулировка AAPOR			Формулировка ФСИ		адресная	именная
проведенные интервью						
интервью завершённое	не менее 80% отве-	1.1	0	проведенное интервью	да	да
интервью незавершённое	от 50% до 80% отве- тов	1.2			да	да
непроведенное интервью, соответствующее условиям отбора						
отказы и прерванные интервью						
отказы						
отказы на уровне домохозяйств		2.111	21	отказы на уровне домохозяйств	да	да
отказы отобранных респондентов		2.112	22	отказы отобранных респондентов	да	да
прерванные интервью	менее 50% ответов	2.12	23	прерванные интервью	да	да
не установлен контакт						
невозможно войти в здание / добраться до дома		2.23	24	невозможно попасть в ДХ (кодовый замок, нет звонка и пр.)	да	да
никого нет дома		2.24	25	никого нет дома	да	да
респондент уехал \ недоступен	в т.ч. сдает жилье	2.25	26	длительное отсутствие (более чем длительность опроса)	да	да
другое						
смерть		2.31	27	смерть	да	да
физически или душевно недееспособен / некомпетентен		2.32	28	не подлежит опросу	да	да
языковые проблемы		2.33	29	языковые проблемы	да	да

Формулировка ААРОР		Формулировка ФСИ		адресная	именная	
проблемы со знанием языка на уровне домохозяйства	2.331					
проблемы со знанием языка у респондента	2.332					
нет интервьюера, знающего язык	2.333					
смешанные причины	2.35	50	другое	да	да	
нет информации о соответствии критериям отбора						
не известно наличие места проживания	3.10					
невозможно добраться \ небезопасная область	3.17		не показывалось интервьюеру	да	да	
невозможно определить адрес	адрес, соответствие которого не удалось установить	3.18	32	неверный адрес	да	да
				смена адреса		да
			33	такой жил, но переехал по конкретному адресу		да
			34	такой жил, но переехал по неизвестному адресу		да
			35	о таком никто из членов ДХ не слышал		да
остановились на уровне домохозяйства \ нет информации о наличии респондента, удовлетворяющего условиям отбора	3.20					
не закончен отбор респондента	3.21		не используется			
другое	3.90			да	да	
нет соответствия критериям отбора						

Формулировка ААРОР			Формулировка ФСИ		адресная	именная
не входит в выборку	адрес, не входящий в выборку адресов	4.10		не показывалось интервьюеру	да	да
не место проживания	не жилое помещение	4.50	41	не жилое помещение	да	да
коммерческая, государственная или иная организация		4.51				
спец. учреждения (тюрьмы, диспансеры)		4.52				
жилое помещение, занимаемое какой-либо группой		4.53				
пустое жилое помещение	находящееся на одном месте жилье	4.60				
постоянное пустое жилье		4.61	42	никто не живет	да	да
сезонное \ временное \ вакантное место жительства		4.62			да	
другое		4.63			да	
респондент не соответствует критериям отбора		4.70		не используется		
квота выбрана		4.80	43	вне квоты	да	да
временные коды						
респондента нет дома			11	респондента нет дома	да	да

ПРИЛОЖЕНИЕ: Результаты полевых работ

15.1

Статистика контактов по опросу в марте 2005 г., %

Код	Расшифровка	Посещения			
		первое	второе	третье	последнее
0	интервью состоялось	24	15	16	32
1	респондента нет дома	11	15	12	6
2	никого нет дома	32	53	60	21
3	смена адреса, новый неизвестен; выезд за пределы н.п.	12	4	4	14
4	о таком не слышали	6	1	2	6
5	длительное отсутствие	3	2	1	3
6	отказ участвовать	7	7	4	10
7	респондент умер	4	1	0	4
8	записан новый адрес	2	0	0	2
9	другое	1	1	0	1

15.2

Статистика контактов по опросу в июне 2005 г., %

Код	Расшифровка	Посещения			
		первое	второе	третье	последнее
0	интервью состоя- лось	19	18	16	28
1	респондента нет дома	2	3	1	0
2	никого нет дома	39	51	50	19
3	длительное отсут- ствие	1	0	3	2
4	отказ	25	17	17	34
5	вне квоты	8	8	8	12
7	респондент умер	0	0	0	0
9	кодовый замок	2	2	4	2
10	неверный адрес	2	0	0	2
12	другое	2	0	1	1
15	помещение сдается	0	0	0	0

15.3

Статистика контактов по опросу в сентябре 2005 г., %

Код	Расшифровка	Посещения			
		первое	второе	третье	последнее
0	интервью со- стоялось	20	23	27	32
1	респондента нет дома	9	15	11	5
2	никого нет дома	29	36	45	16
3	новый адрес неизвестен	9	4	3	10
4	не знают такого	4	2	0	4
5	длительное отсутствие	3	2	2	4

6	отказ респондент	11	9	8	14
7	умер	2	1	1	2
8	записан новый адрес	1	0	1	1
9	кодовый замок не подлежит опросу	2	4	3	2
10	прерванное интервью	1	1	0	1
11	другое	0	0	0	0
12	вне квоты	2	1	0	1
13	неверный адрес помещения	7	3	1	7
14	сдается	1	0	0	1
15		0	0	0	0

15.4

Статистика контактов по опросу в декабре 2005 г., %

Код	Расшифровка	Посещения						
		1-е	2-е	3-е	4-е	5-е	6-е	последнее
0	интервью состоялось	19	18	17	18	10	15	34
11	респондента нет дома	14	16	21	23	25	22	4
21	отказ домохозяйства (ДХ)	3	2	2	1	4	3	4
22	отказ респондента	9	10	7	11	4	4	16
23	прерванное интервью	0	0	0	0	0	1	0
24	невозможно попасть в ДХ	7	10	8	9	8	6	2
25	никого нет дома	23	32	37	31	43	41	8
26	длительное отсутствие	3	3	1	3	4	0	5
27	респондент умер	4	1	1	0	0	0	4
28	не подлежит опросу	1	1	2	0	0	1	1
29	языковые проблемы	0	0	0	0	0	0	0
31	не дошли до адреса	1	0	0	0	0	0	1
32	неверный адрес	2	0	0	0	0	0	2
33	переехал по конкр. адресу	2	1	1	0	1	0	2
34	переехал по неизв. адресу	8	3	3	0	1	3	10
35	о таком не слышали	3	1	1	1	0	1	4
41	нежилое помещение	0	0	0	0	0	1	1
42	никто не живет	1	0	0	0	0	0	1

43	вне квоты	0	0	0	0	0	0	0
50	другое	1	0	0	0	1	0	1