
А.М. Мавлетова
(Москва)

СОЦИОЛОГИЧЕСКИЕ ОПРОСЫ В СЕТИ ИНТЕРНЕТ: ВОЗМОЖНОСТИ ПОСТРОЕНИЯ ТИПОЛОГИИ

В статье предлагается основание типологии веб-опросов, опирающихся на представления о концептуальных объектах, на которые экстраполируются результаты исследований. Выделяется пять основных типов веб-опросов, где концептуальный объект приравнивается: к реальным участникам опроса; к посетителям конкретного Интернет-сайта; к определенным социальным группам; к пользователям Всемирной сети; совокупности, состоящей из пользователей и тех, кто не имеет доступ в Интернет.

Ключевые слова: веб-исследования, онлайн-опросы, концептуальный объект, ошибка выборки, ошибка охвата, ошибка неотчетов, панельные исследования, социальные сети, труднодостижимые группы.

Постановка исследовательской задачи

Современные информационные технологии оказывают значительное влияние на развитие методологии социологических исследований. Всемирная сеть порождает множество площадок для высказывания мнений, размещения информации о себе и своих предпочтениях. Вместе с тем веб-исследования пока еще не получили широкого признания в научной среде, и основная причина постановки вопроса о их легитимности – ошибки охвата и использование целевого (невероятностного) отбора респондентов. При этом довольно часто не учитываются как высокая релевантность веб-опросов для решения конкретных исследовательских задач,

Айгуль Маратовна Мавлетова – аспирант ГУ–ВШЭ, факультет социологии.
E-mail: mavletova@yandex.ru.

так и возможности применения вероятностных методов отбора. Сразу уточним, что в статье будут рассмотрены только «количественные» исследования: за рамками выстраиваемой типологии оказались онлайн-фокус-группы, форумные дискуссии, онлайн-сообщества и т.п., поскольку они требуют отдельного анализа.

В целом в отечественной социологии проблематика онлайн-исследований остается сравнительно малоизученной и представлена, прежде всего, в статьях Б.З. Докторова [см., напр., 1] и двух сборниках статей «Онлайн исследования в России» [2; 3]. Наиболее полная на сегодняшний день классификация онлайн-опросов была предложена М. Купером [4], который, отталкиваясь от идей Л. Киша и Р. Гривза, классифицирует веб-опросы на основе типа выборки (случайная/неслучайная). Этот подход не позволяет выстроить типологию, которая помогала бы исследователю видеть возможности и ограничения разных видов онлайн-опросов. Более конструктивной нам представляется классификация веб-опросов, исходя из *концептуального объекта* исследования – идеальной генеральной совокупности. Заметим, что базовой проблемой проектирования реальной генеральной совокупности, возможные решения которой рассматриваются в статье, является отсутствие основы выборки, включающей в себя уникальные, полные и неповторяющиеся идентификаторы Интернет-пользователей (по аналогии с базой телефонных номеров).

Предлагаемая типология веб-опросов позволяет детально рассмотреть методологические основания каждого типа, определить, для решений каких задач применим тот или иной тип исследования, выявить специфику формирования выборок и оценить возможности анализа и корректировки эмпирических данных. Конечно, она не претендует на исчерпывающий характер. Типологию составляют пять основных типов веб-опросов, где 1) концептуальный объект приравнивается к реальным участникам опроса; 2) генеральную совокупность формируют посетители конкретного Интернет-сайта или 3) заданные социальные группы; 4) результаты можно распростра-

нить на пользователей Всемирной сети или 5) на совокупность, состоящую из пользователей и тех, кто не имеет доступ в Интернет.

Итак, рассмотрим перечисленные типы и проиллюстрируем их примерами опросов.

Тип 1. Концептуальный объект = реальные участники опроса

Этот тип веб-исследований, в котором генеральная совокупность сводится к респондентам, принявшим участие в опросе, сегодня является одним из самых популярных. В нем можно выделить четыре подтипа: «развлекательные» опросы, опросы посетителей Интернет-сайтов, пользователей социальных сетей и участников access-панелей.

К подтипу «развлекательных» можно отнести опросы на специально созданных сайтах, например www.misterpoll.com, www.opendebate.com, www.survey.net или www.online-poll.ru, и ежедневные «вопросы дня» на сайтах газет, журналов, радиостанций, видеоканалов и т.д., например www.rbc.ru, www.echo.msk.ru, или сайтах, генерирующих вопросы (www.jetlink.ru). Сомневаясь в качестве подобных голосований на популярных веб-сайтах, У. Крофорд называет все онлайн-опросы «худшими из худших» [5, p. 42], но он не учитывает, что они проводятся в развлекательных целях и не претендуют на валидность и надежность данных. Опросы на сайтах онлайн-голосований открыты для всех желающих: каждый Интернет-пользователь может выступить в роли не только информанта, но и исследователя; основная цель большинства подобных опросов – создание площадки для обмена мнениями как формы развлечения. Так, создатели сайта misterpoll.com прямо заявляют: «Мы поддерживаем директорию наиболее интересных и обсуждаемых голосований для общего развлечения. Ни одно из голосований не является научным, но репрезентирует коллективное мнение участников опроса» [6].

Второй подтип – *опросы посетителей Интернет-сайтов* – предполагает, что исследователи размещают баннеры и всплывающую рекламу с приглашением к участию в опросе на часто посещаемых порталах или тематических сайтах для рекрутирования респондентов. По мнению М. Купера, эти опросы можно было бы отнести к «развлекательным», если бы не утверждения их разработчиков о получении надежных результатов, подкрепляемых большим объемом выборки. В известных проектах Национального географического общества «Опрос 2000» и «Опрос 2001» на сайте www.nationalgeographic.com могли принять участие все Интернет-пользователи, владеющие английским языком. Но вряд ли правомерна позиция организаторов исследования: «Мы получили более 50000 ответов – как минимум, в два раза больше, чем того требует научная валидность» [4, р. 480].

Другой пример: правила голосования российского проекта «Имя России: Исторический выбор 2008», целью которого был выбор личности, оказавшей наибольшее влияние на историю России, были следующие «Устраивайте флэш-мобы, занимайтесь всеми видами рекламы и антирекламы, голосуйте сами (хоть головой, хоть сердцем), зовите на сайт друзей и знакомых. Условия голосования предельно просты: голосуйте неограниченное количество раз за любого персонажа или за нескольких одновременно!» [7]. Отсутствие технических ограничений на повторное участие респондентов в опросе привело к невозможности остановить хакерские атаки (автоматические массовые голосования за определенных кандидатов).

AAPOP («Американская ассоциация исследователей общественного мнения») в разделе «Плохие выборки» приводит следующий пример: у посетителей сайта parade.com интересовались тем, кто виноват в том, что 2005 г. выдался неудачным для известного актера Тома Круза – сам актер или средства массовой информации. Из проголосовавших 84% обвинили в неудачах СМИ, но было выявлено, что более 14000 из 18000 ответов были сделаны всего с 10 компьютеров [8].

Третий подтип формируют *опросы пользователей социальных сетей*, когда исследовательские компании создают специальные сообщества для изучения мнений в Интернете. Например, сеть «Hey!Nielsen» проводит опросы в области телевидения, музыки, кино и Интернета и составляет рейтинги «звезд»: организаторы предлагают оценить некоторые суждения, а зарегистрированные посетители комментируют, создают и публикуют новые. Совмещающая онлайн- и оффлайн-данные, «ACNielsen» составляет сводный рейтинг сайтов, телевизионных передач, музыкальных хитов и известных людей – так называемый «Nielsen Score».

Сеть «Flixster», в которой зарегистрировано более 50 миллионов пользователей, предлагает им оценивать просмотренные фильмы по шкале от «1» до «5», а те, которые только начинают выходить в прокат, – по бинарной шкале: «Хочу увидеть» – «Не заинтересован». Пользователи российской социальной сети «Имхонет» оценивают культурные объекты по 10-балльной шкале – от «Хуже не бывает» до «Лучше не бывает», причем социолог имеет возможность проанализировать предпочтения пользователей в зависимости от социально-демографических характеристик.

Четвертый подтип – *опрос участников access-панелей* – приобретает сегодня огромную популярность и является наиболее интенсивно растущим в маркетинговой индустрии: все больше компаний, занимающихся изучением рынка, создают онлайн-access-панели из добровольно зарегистрированных пользователей, которые согласились принимать участие в опросах на регулярной основе за вознаграждение.

Понятие «access-панель» пришло из традиционных оффлайн-панельных исследований – такой «резервуар» зарегистрированных респондентов имеется у многих маркетинговых компаний. Но если в традиционных панельных опросах индивиды на протяжении длительного времени участвуют в одном исследовании, то члены access-панелей – в разных опросах по различным тематикам.

Формирование панели проходит в два этапа: сначала Интернет-пользователи, увидев рекламу на веб-сайте или получив приглашение по электронной почте или телефону, регистрируются на сайте панели. Затем респондент оставляет базовую социально-демографическую информацию о себе (свою профильную анкету) и только после этого становится участником панели. Основа выборки для каждого исследования в онлайн-панели – зарегистрированные участники, подходящие под определенные критерии (отбор проходит на основе имеющейся информации). В отличие от предыдущих подтипов, в access-панелях вводятся технические ограничения, чтобы избежать многократного заполнения анкеты одним респондентом.

Таким образом, главной особенностью онлайн-исследований первого типа является знание, желание, возможность и готовность респондентов участвовать в опросе, т.е. это разновидность так называемых SLOP-исследований (self-selected opinion polls), основанных на невероятностной выборке добровольцев.

Как же оценить параметры генеральной совокупности в данном типе опросов, если экстраполировать его результаты на группу Интернет-пользователей? Каждый Интернет-пользователь k имеет неизвестную вероятность отбора p_k , где $k = 1, 2, \dots, N_1$. Ожидаемое значение вероятности $p_k = E(r_k)$ называют коэффициентом склонности к участию пользователя k в Интернет-опросе. Если бы у каждого пользователя были равные шансы отбора, то в качестве значения генеральной совокупности выступало бы простое среднее значение, но в случае неравных шансов отбора среднее значение можно оценить по формуле Горвитца-Томпсона [9]:

$$E(\overline{Y}_{HT}) \approx \overline{Y}_I^* = \frac{1}{N_I p} \sum_{k=1}^N p_k I_k Y_k,$$

где $I_k = 1$ при наличии доступа в Интернет у пользователя k , $I_k = 0$ – при отсутствии доступа, p – среднее значение коэффициента склонности для всех Интернет-пользователей, p_k – значение коэффициента склонности для Интернет-пользователя k .

Коэффициенты склонности могут быть оценены апостериорно – с использованием методов ремонта выборки. Один из популярнейших – расчет корректирующих коэффициентов склонности, где в качестве контрольной группы выступает случайная выборка населения или выборка Интернет-пользователей. Тем не менее, AAPOR рекомендует всем исследователям, пользующимся результатами онлайн-опросов, в которых концептуальный объект сводится к реальным участникам опроса, заявлять следующее: «Респонденты для данного онлайн-исследования были отобраны из добровольно зарегистрированных участников. Результаты были перевзвешены в соответствии с социально-демографическими параметрами изучаемого объекта. Исследование подвержено ошибкам выборки, охвата, неотчетов¹, а также взвешивания и корректировки данных» [10].

Тип 2. Концептуальный объект = посетители конкретного веб-сайта

Л. Киш в книге «Статистический дизайн исследования» утверждает, что каждое исследование можно оценить по трем характеристикам: репрезентативность, рандомизация и реализм [11]. Рассмотрим тип веб-опросов, который удовлетворяет этим критериям и результаты которого можно распространить на пользователей определенных сайтов. В нем можно выделить два подтипа: анализ предпочтений индивидов на основе изучения профилей пользователей социальных сетей и опрос посетителей сайта, отобранных по случайной выборке.

Развитие первого подтипа – *анализ информации о предпочтениях на основе изучения профилей пользователей социальных сетей* – определяется огромными возможностями концепции семантического веба, которую развивает основатель Всемирной

¹ Под «ошибкой неотчетов» мы понимаем ошибку, обусловленную отказом выборочной единицы принять участие в опросе.

паутины Т. Бернерс-Ли. Веб 3.0, или семантический веб, воспринимается уже не как просто способ размещения документов в виде гипертекста, но как носитель связанных между собой баз данных и структурированных хранилищ информации, допускающих автоматическую обработку [12].

Т. Бернерс-Ли и соавторы новой концепции Всемирной паутины под семантикой понимают унифицированную систему правил кодирования языковых алгоритмов, конституирующих «взаимопонимание» сетевых данных – оно обеспечивается принятыми логическими метаязыковыми правилами описания имеющейся информации. Речь идет о возможности получения и анализа веб-информации в формате стандартизированных баз данных. Хотя сеть нового поколения вызывает много вопросов и теоретических дискуссий, практическими результатами некоторых проектов WWW Консорциума, внедряющего стандарты семантической паутины, социологам уже можно воспользоваться.

Так, известный проект семантического веба «Друг друга» (Friend of a Friend, или FOAF), стартовавший в 2000 г., нацелен на «введение машинно-читаемых стандартов описания индивида и его отношений с другими людьми» [13]. На операциональном уровне это означает попытку создать такой механизм размещения и оформления информации в социальных сетях, который позволит объединить их в одну глобальную базу персональных данных пользователей. Для этого Д. Брикли и Л. Миллер разработали универсальный машинно-читаемый словарь описания, который включает, как минимум, следующие поля: электронный адрес, географическое местоположение, социально-демографические характеристики, список друзей, хобби/интересы, общие установки, предпочтения в области музыки, кино, телевидения и литературы. Формат данных позволяет компьютерам находить эту информацию, «понимать» и автоматически обрабатывать. Многие социальные сети уже используют этот словарь (например, Livejournal, LiveDoor), поэтому исследователи вполне могут анализировать их структуру.

Наиболее важной нам кажется возможность изучения профильной информации пользователей для определения их основных интересов и факторов, влияющих на их предпочтения, т.е. исследователь не проводит опрос респондентов. «Одна из последних стадий развития текстуальной самопрезентации, – пишет сотрудник Медиа лаборатории Массачусетского технологического института Х. Лю, – это профиль в социальной сети. Виртуальные материалы подобного самопредставления – такие культурные знаки, как внесенная самим пользователем информация о любимых книгах, музыке, фильмах, телевизионных предпочтениях – вместе составляют презентацию вкусов, изучение которых представляет особый интерес» [14].

Исследователи анализируют профили пользователей социальных сетей, выявляя основные интересы индивидов, социально-экономические и эстетические факторы, конституирующие семантику их вкусов. Например, изучая профили пользователей Livejournal в марте 2004 г., Дж. Паолило и И. Райт на основе анализа главных компонент сгруппировали основные предпочтения пользователей, а посредством иерархического кластерного анализа классифицировали их по интересам [15, р. 7–8]. Х. Лю в 2007 г. провел статистический анализ предпочтений по 127477 профилям крупнейшей в мире социальной сети «MySpace». Хотя при регистрации пользователи могут внести любую социально-демографическую и общую информацию (практики курения, потребления алкоголя и т.п.), презентовать любые свои занятия – в свободной форме написать о себе, интересах в музыке, кино, телевидении и литературе, Х. Лю обнаружил, что музыкальные предпочтения являются конституирующим элементом всей профильной информации в этом сообществе, а в качестве общих маркеров вкусов выступают ироничность, сексуальность, комичность и искренность [14].

В данном подтипе онлайн-исследований личная информация индивидов может быть собрана и проанализирована без согласия респондентов, что нарушает базовое методологическое правило

социолога – принцип информированного согласия. В этой связи возникает вопрос о новом этическом кодексе, регулирующем использование личных данных в онлайн-исследованиях.

Отличительной характеристикой *опросов посетителей конкретного сайта, отобранных по случайной выборке*, является отбор респондентов с использованием систематической выборки, применяемой, например, в так называемых **exit polls (опросах избирателей на выходе с избирательных участков)**. Каждому n -му посетителю сайта показывается ссылка, которая появляется в новом окне браузера, т.е. основу выборки составляют посетители определенного Интернет-сайта, что снимает проблему охвата.

Многие западные исследовательские компании, имеющие онлайн-**access-панели**, сталкиваются с **проблемой снижения уровня откликов участников панели**, что определяет рост популярности рекрутирования респондентов через всплывающую рекламу на массовых порталах. Подобную технику сбора информации П. Кумли называет *рандомизированным веб-интервью* – RAWI (Random Web Interviewing), хотя случайной такую выборку можно назвать только применительно к используемым для отбора респондентов сайтам [16].

В этом случае можно оценить параметры генеральной совокупности, рассчитав среднее значение по группе посетителей сайта:

$$E(\bar{Y}_{HT}) = \frac{1}{N_{WS}} \sum_{k=1}^{N_I} a_k \frac{Y_k}{\pi_k},$$

где $a_k = 1$, если элемент k принял участие в опросе, $a_k = 0$ – если k не принял участие, $\pi_k = E(a_k)$ – математическое ожидание вероятности отбора элемента k , N_{WS} – количество посетителей сайта. Для

простой случайной выборки $\pi_k = \frac{n}{N_{WS}}$, тогда: $\bar{Y}_{WS} = \frac{1}{n} \sum_{k=1}^{N_I} a_k Y_k$.

Тип 3. Концептуальный объект = заданные социальные группы

Данный тип веб-исследований предполагает экстраполяцию результатов опроса на 1) труднодоступные группы и 2) группы с высоким Интернет-проникновением с известной основой выборки.

В последнее время Интернет становится популярной и научно признанной площадкой для *опроса труднодоступных совокупностей*, потому что позволяет «найти» редкие группы и соблюсти необходимую степень анонимности, что невозможно при использовании других опросных методов: «Веб-исследования пока не представляют серьезную научную альтернативу превалирующим методам сбора данных, принятых в эмпирических социальных исследованиях. Однако, несмотря на это, онлайн-опросы могут применяться в разведывательных экспериментальных исследованиях труднодостижимых групп» [17, р. 5]. Р. Кумбер одним из первых провел веб-опрос среди наркотических дилеров, чтобы определить основные каналы распространения наркотиков и выявить специфические групповые поведенческие паттерны. Полученные данные были использованы как дополнение к интервью, проведенным в тюрьмах [18].

Исследователи могут рекрутировать группы через специальные тематические сайты или использовать метод «снежного кома». Д. Хекаторн предлагает модифицированный вариант последнего – «выборку, управляемую респондентами» (*respondent-driven sample – RDS*). Он рассматривает цепочку опрашиваемых как цепь Маркова, т.е. моделирует процесс набора респондентов как последовательность случайных событий с конечным или счетным бесконечным числом исходов, характеризующуюся следующим свойством: при фиксированном настоящем будущее независимо от прошлого. Иными словами, предполагается, что смещения, которые могли быть вызваны изначально построенной выборкой, нивелируются в последующих волнах таким образом, что финальная выборка не

зависит от выбранных исследователем первичных респондентов [19]. В 2008 г. Д. Хекаторн и К. Вейнерт провели первый онлайн-опрос, используя данный метод конструирования выборки [20].

Онлайн-опрос с использованием RDS позволяет, с одной стороны, значительно сократить время полевых работ, с другой – увеличить количество опрошенных. К полученным данным можно применять статистические методы (измерить ошибку выборки, вычислить доверительный интервал) и экстраполировать результаты на генеральную совокупность (заданную социальную группу). Ее параметры в RDS оцениваются с помощью статистической программы, в которой моделируется процедура отбора (так называемая оценочная процедура «бутстрэппинга» – *bootstrapping*).

Второй подтип исследований – *опрос групп с высоким Интернет-проникновением* – предполагает, что социолог имеет основу выборки в виде электронных адресов и проводит либо сплошной онлайн-опрос, либо опрос по случайной выборке. К этому подтипу относят опросы студентов высших учебных заведений, если у администрации вуза есть электронные адреса всех студентов, и опросы сотрудников конкретных компаний.

Тип 4. Концептуальный объект = все пользователи Интернета

В данном типе предпринимается попытка сформировать случайную выборку пользователей Всемирной сети. Один из способов ее построения – случайный отбор классическим способом (по базе телефонных или почтовых адресов) и опрос только тех, кто имеет доступ в Интернет. Респондентам, согласившимся принять участие в онлайн-исследовании, называют адрес веб-сайта, на котором можно заполнить анкету или зарегистрироваться в панели.

Ошибка неотчетов – основная проблема данного типа опросов, которая может происходить на нескольких этапах рекрутирования. Участниками панели становятся люди, которые ответили на теле-

фонный звонок, подтвердили, что у них есть доступ в Интернет, предоставили свой электронный адрес, согласились стать участником онлайн-панели и зарегистрировались в ней после получения приглашения по электронной почте. На каждом из перечисленных этапов можно измерить ошибки охвата и неответов на основе имеющейся социально-демографической информации о пользователях и непользователях Интернета, а также об индивидах, выразивших намерение стать участником панели, но отказавшихся зарегистрироваться. По данным компании «Pew Research Center», рекрутирующей панель случайным образом, примерно треть пользователей Интернета называют свои электронные адреса и треть получивших приглашения на почту регистрируются в панели [4, р. 487–488].

Можно оценить параметры генеральной совокупности в данном типе веб-опросов, рассчитав среднее значение по группе Интернет-пользователей по следующей формуле:

$$E(\bar{Y}_{HT}) = \frac{1}{N_I} \sum_{k=1}^N a_k I_k \frac{Y_k}{\pi_k},$$

где $a_k = 1$, если элемент k принял участие в опросе, $a_k = 0$, если k не принял участия, $I_k = 1$ – при наличии доступа в Интернет у пользователя k , $I_k = 0$ – при отсутствии доступа, $\pi_k = E(a_k)$ – математическое ожидание вероятности отбора элемента k .

В случае простой случайной выборки $\pi_k = \frac{n}{N_I}$, где N_I – количество индивидов, имеющих доступ в Интернет, $\bar{Y}_I = \frac{1}{n} \sum_{k=1}^N a_k I_k Y_k$.

Если экстраполировать результаты на все население, то ошибка выборки составит

$$B_{\bar{Y}_{HT}} = E(\bar{Y}_{HT}) - \bar{Y} = \bar{Y}_I - \bar{Y} = \frac{N_{NI}}{N} (\bar{Y}_I - \bar{Y}_{NI}),$$

т.е. она зависит от двух основных показателей – доли населения, не имеющей доступ в Интернет ($\frac{N_{NI}}{N}$), и разницы между индивидами, имеющими и не имеющими доступ в Интернет ($\bar{Y}_I - \bar{Y}_{NI}$).

Тип 5. Концептуальный объект = все население

В ряде случаев онлайн-опросы позволяют конструировать репрезентативные выборки, охватывающие не только пользователей сети, но и не имеющих доступ в Интернет. Речь идет об опросах, в которых Интернет рассматривается как один из возможных способов заполнения анкеты, или же о репрезентативных опросах населения через Интернет.

В первом случае *Интернет выступает как альтернативный способ заполнения анкеты* наряду с традиционными методами. Например, по почте респондентам высылается анкета и указание на веб-сайт, где при наличии доступа в Интернет респондент может заполнить анкету в электронном виде. Д. Дилман и его коллеги из Вашингтонского университета отмечают, что основным трендом XXI в. в области исследований станет проведение массовых опросов, сбор данных в которых будет осуществляться одновременно несколькими методами: респондент сможет выбирать – заполнить анкету, высланную по почте, ответить на вопросы интервьюера по телефону – городскому или мобильному – или же зайдя на сайт [21, р. 1–2].

Преимущества подобных опросов очевидны: сам дизайн исследования помогает уменьшить ошибку охвата и увеличить уровень откликов, снизив ошибку неответов. Эффективным способом увеличения уровня откликов считается использование другого метода сбора данных для тех, кто отказался от участия в опросе на первых этапах, причем метод может меняться несколько раз. Например, изначально исследователь может обратиться к респондентам по почте; затем индивиды, не принявшие участие в почтовом анкетировании, могут

быть опрошены по телефону или Интернету. Однако использование нескольких методов получения информации в одном исследовании может увеличить ошибку измерения – необходимо помнить о различии результатов в зависимости от применяемого метода.

Репрезентативные онлайн-опросы населения через Интернет – единственный онлайн-метод, позволяющий конструировать случайную выборку как пользователей сети, так и не имеющих доступ в Интернет. Онлайн-панель рекрутируется на основе телефонного опроса (используется случайная выборка без возвращения). Независимо от того, имеется в домохозяйстве компьютер и доступ в Интернет или нет, индивидам предлагают зарегистрироваться в панели. В случае отсутствия компьютера и выхода в Интернет потенциальным респондентам предоставляется необходимое оборудование. В итоге появляется возможность репрезентировать все домохозяйства, имеющие домашние телефоны, и снимается основная проблема веб-опросов – ошибка охвата. Ключевой проблемой данного типа веб-исследований является ошибка ответов, которую необходимо фиксировать на четырех основных этапах работы: телефонный звонок, получение согласия участвовать в панели, установка оборудования и заполнение профильной анкеты, участие в опросах.

Пионером здесь является голландский исследовательский институт «Dutch Telepanel» (сегодня – «CenterData»), который в 1988 г. начал строить репрезентативную панель: в голландских домохозяйствах, случайным образом отобранных по телефонным номерам и согласившихся стать участниками панели, Институт устанавливал компьютеры. Участники панели заполняли базовую информацию о себе, затем с периодичностью раз в неделю участвовали в опросах. В 1990-х гг. Институт расширил рамки работы до Всемирной сети: в случае отсутствия компьютера или доступа в Интернет «CenterData» предоставляет телевизионное устройство с выходом в Интернет; если в домохозяйстве нет телевизора, то предоставляется и он [22].

С 1999 г. аналогичной стратегии придерживается американская компания «InterSurvey» (с 2000 г. «Knowledge Networks» – KN). Домохозяйства, чьи телефонные номера отбираются случайным образом и которые имеют почтовые адреса (около 65%), получают по почте приглашение участвовать в панели. Далее по выбранным телефонным номерам, включая те, по которым не определен почтовый адрес, дозваниваются интервьюеры. В случае согласия домохозяйства ему предоставляется прибор для доступа в Интернет, респонденты заполняют профильную анкету и маркируются как «активные участники» панели [23, р. 14–17].

По аналогии с показателями, разработанными AAPOR для телефонных опросов, на каждой стадии рекрутирования респондентов можно рассчитать показатели результативности [24]. На первой стадии телефонного дозвона доля рекрутирования составляет:

$$H_{\text{RECR}}(P_{\text{RECR}}) = \frac{IC}{IC + (R + NC + O) + e(UH + UO)},$$

где IC – согласие стать участником панели; R – отказ от контакта; NC – неустановленный контакт; O – другое; e – доля единиц, которая может удовлетворять критериям отбора в общем количестве единиц, но по которой отсутствует информация; UH – неизвестно, домохозяйство это или нет; UO – неизвестно другое.

Долю неотвечен на данном этапе можно рассчитать по формуле:

$$H_{\text{REFR}}(P_{\text{REFR}}) = \frac{R}{IC + (R + NC + O) + e(UH + UO)}.$$

На стадии заполнения профильной анкеты можно определить долю ответов (*profile rate*) следующим образом:

$$\text{PROR} = \frac{I + P}{I + P + (R + NC + O)},$$

где I – доля заполненных анкет; P – доля незавершенных анкет.

«Активными участниками» панели становятся респонденты, заполнившие профильную анкету; важно отслеживать возможное

смещение в пользу респондентов, изначально согласившихся стать участниками панели, но по каким-либо причинам не заполнивших профильную анкету. Показатель неответов, или отказов от заполнения профильной анкеты (*refusal to profile*), можно рассчитать по формуле:

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

Когда респонденты становятся «активными» участниками панели, на их электронные ящики высылаются приглашения участвовать в опросах. Здесь следует рассчитывать долю доставленных электронных приглашений:

$$\text{Absorption Rate} = \frac{EI - UEI}{EI}$$

где *EI* – количество разосланных приглашений; *UEI* – количество недоставленных приглашений.

Далее рассчитывается процент респондентов, приглашенных участвовать в опросе и заполнивших анкету (*completion rate*), и доля отказов от участия в опросе (*refusal rate*):

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

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

Учитывая уровень сохранения/удержания панели (*retention rate*), общую долю ответов (*cumulative response rate*) в онлайн-опросе можно рассчитать по формуле:

$$\text{CURR} = P_{\text{RECR}} \times \text{PROR} \times \text{RETR} \times \text{COMR}$$

Специалисты компании «KN» фиксируют следующие показатели: 56% респондентов соглашаются участвовать в панели на этапе телефонного обзвона; 75-80% домохозяйств соглашаются на установку адаптеров для выхода в Интернет; профильную анкету для регистрации в панели заполняют 90-93%. Итоговая

доля ответов без учета уровня «истощения» панели составляет 36-42% [25]. Институт «CenterData» приводит иную статистику: уровень контактов – 60%; уровень кооперации – 35-40%; процент домохозяйств, соглашающихся на установку адаптеров для выхода в Интернет, – 50%; итоговый процент рекрутирования – 10-12%. Последний показатель варьирует в зависимости от социально-демографических характеристик респондентов: например, в старшей возрастной группе он составляет 5%, а в средней доходит до 17%. Общая доля рекрутирования различается и в зависимости от наличия компьютера в домохозяйстве (эта переменная коррелирует с доходом) – 16% для домохозяйств, имеющих компьютер, и 5% – для не имеющих. Эти данные позволяют предположить, что установка оборудования не решает проблему охвата респондентов, не имеющих опыта работы с компьютером и выхода в Интернет [26, р. 9–15].

Хотя в действительности представители разных социально-демографических групп имеют разную вероятность быть рекрутированными в панель, базовое допущение репрезентативных онлайн-опросов населения – равная вероятность отбора единиц генеральной совокупности. Чтобы в ошибке неотчетов учесть разные вероятности отбора, применяются процедуры перевзвешивания. Таким образом, фактически это единственный тип веб-исследований, который позволяет конструировать случайные выборки всего населения и является, по мнению Б.З. Докторава, «высшим достижением онлайн-опросной технологии» [27, с. 574].

Итак, предложенный подход к построению типологии онлайн-исследований позволяет оценивать релевантность применения различных вариантов Интернет-опросов и способов формирования выборки в разных исследовательских ситуациях, помогает увидеть сильные и слабые стороны онлайн-опросов вообще и наиболее интересных для исследователей их разновидностей в частности (прежде всего, это опросы труднодостижимых групп, работа с аудиториями социальных сетей и попытки репрезентативных массовых опросов через Всемирную сеть). Конечно, представленная

типология не претендует на исчерпывающий характер и оставляет простор для введения типов и подтипов. Скорее всего, новые подходы к сбору и анализу данных, а также к формированию выборок в онлайн-исследованиях не заставят себя долго ждать – пока же будем исходить из сложившейся типологии.

ЛИТЕРАТУРА

1. Докторов Б. Онлайнновые опросы – обыденность наступившего столетия // http://pseudology.org/Gallup/On_line_Polls.htm.
2. Онлайн исследования в России: Тенденции и перспективы / Под ред. А.В. Шашкина, М.Е. Поздняковой М: ИС РАН, 2006.
3. Онлайн исследования в России 2.0 / Под ред. А.В. Шашкина, И.Ф. Десятко, С.Г. Давыдова. М: РИЦ «Северо-Восток», 2010.
4. Couper M. Web Surveys: A Review of Issues and Approaches // *Public Opinion Quarterly*. 2000. Vol. 64. No. 4. P. 464–494.
5. Crawford W. Survey Says... Or Does It? [Fun with Statistics] // *EContent*. 2004. Vol. 27. No. 11. P. 42–43.
6. <http://www.misterpoll.com/about.htm> (Доступ: 12.07.08).
7. <http://www.nameofrussia.ru/about.html> (Доступ: 13.05.08).
8. <http://www.aapor.org/badandworsesamples> (Доступ: 23.07.08).
9. Bethlehem J. Reducing the Bias of Web Survey Based Estimates: Discussion Paper 07001. Voorburg/Heerlen: Statistics Netherlands. January 2007. P. 1–36 .
10. <http://www.aapor.org/whenrespondentsselectthemselves> (Доступ: 29.07.08).
11. Kish L. *Statistical Design for Research*. N.Y.: John Wiley and Sons, 2004.
12. Berners-Lee T., Hender J., Lassila O. The Semantic Web // *Scientific American Magazine*. 2001. Vol. 284. No. 5.
13. Brickley D., Miller L. FOAF Vocabulary Specification 0.9. May 2007 // <http://xmlns.com/foaf/0.1>.
14. Liu H. Social Network Profiles as Taste Performances // *Journal of Computer-Mediated Communication*. 2007. Vol. 13. No. 1.
15. Paolillo J.C., Wright E. Social Network Analysis on the Semantic Web: Techniques and Challenges for Visualizing FOAF // www.blogninja.com/vsw-draft-paolillo-wright-foaf.pdf.
16. Comley P. Innovation in Online Research – Who Needs Online Panels? // MRS Research Conference: Paper 36. Warrendale: MRS, 2003. P. 1–14.
17. Bandilla W. Web Surveys – an Appropriate mode of Data Collection for the Social Sciences? // *Online Social Sciences* / Ed. by B. Batinic, U-D. Reips, M. Bosnjak. Göttingen: Hogrefe & Huber Publishers, 2002. P. 1–6.

18. *Coomber R.* Using the Internet for Survey Research // Sociological Research Online. 1997. Vol. 2. No. 2 // www.socresonline.org.uk/socresonline/2/2/2.html.

19. *Heckathorn D.* Respondent-driven Sampling: A New Approach to the Study of Hidden Populations // Social Problems. 1997. Vol. 44. No. 2. P. 174–199.

20. *Wejnert C., Heckathorn D.* Web-based Network Sampling: Efficiency and Efficacy of Respondent-driven Sampling for Online Research // Sociological Methods Research. 2008. Vol. 37. P. 105–134.

21. *Dillman D., Phelps G., Tortora R., Swift K., Kohrell J., Berck J., Messer B.* Response Rate and Measurement Differences in Mixed-mode Surveys Using Mail, Telephone, Interactive Voice Response (IVR) and the Internet // Social Science Research. 2009. Vol. 38. No. 1. P. 1–18.

22. *Sarris W.* Ten Years of Interviewing Without Interviewers: The telepanel // Computer Assisted Survey Information Collection / Ed. by M. Couper, R. Baker, J. Bethlehem, C. Clark, J. Martin, W. Nicholls II, J. O'Reilly. N.Y.: John Wiley and Sons, 1998. P. 409–429.

23. *Dennis J., Li R.* Effects of Panel Attrition on Survey Estimates // Paper Presented at the Annual Conference of the American Association for Public Opinion Research. Nashville. May 2003. P. 1–17.

24. *Callegaro M., DiSorga C.* Computing Metrics for Online Panels // Public Opinion Quarterly. 2008. Vol. 72. No. 5. P. 1008–1032.

25. *Rivers D.* Probability-based Web-surveying: An Overview // Paper Presented at the Annual Conference of the American Association for Public Opinion Research. Portland. May 2000.

26. *Hoogendoorn A., Daalmans J.* Nonresponse in the Recruitment of an Internet Panel Based on a Probability Sample: Discussion Paper 08007. Voorburg/Heerlen: Statistics Netherlands, 2008. P. 1–28.

27. *Докторов Б.З.* Реклама и опросы общественного мнения в США: История зарождения. Судьбы творцов. М.: ЦСП, 2008.