

КАК ИЗУЧАТЬ СЕГОДНЯШНЮЮ РОССИЮ*

**РОССИЙСКИЙ МОНИТОРИНГ
ЭКОНОМИЧЕСКОГО ПОЛОЖЕНИЯ
И ЗДОРОВЬЯ НАСЕЛЕНИЯ (РМЭЗ):
ИЗМЕРЕНИЕ БЛАГОСОСТОЯНИЯ**

РОССИЯН В 90-Е ГОДЫ

М.С. Сваффорд, М.С. Косолапов, П.М. Козырева

Журнал «Мир России» продолжает публикацию работы, подготовленной совместно российскими и американскими социологами и призванной содействовать увеличению в нашей стране числа исследователей, использующих в своей научной деятельности исходные данные Российского мониторинга экономического положения и здоровья населения (РМЭЗ) - широкомасштабного опроса домохозяйств и индивидов РФ, проводимого с 1992 г.

В первой части работы, опубликованной в -журнале «Мир России» 1999, № 1-2, были рассмотрены международные принципы и стандарты проведения социологических исследований.

Во второй части работы, предлагаемой Вашему вниманию в этом номере журнала, на конкретном примере лонгитюдного обследования РМЭЗ показывается, как на практике реализуются те международные стандарты качества социологических обследований, которые были описаны в первой части работы. Наиболее подробно здесь рассматриваются два аспекта проводимого мониторинга: а) содержательная часть, описываемая через содержание использованных вопросников, и б) проектирование и реальная процедура выборки. Особый интерес представляет специальное проектирование выборки, позволяющее одновременно проводить как обычный одномоментный, так и панельный микроуровневый анализ данных.

В третьей части работы, которая будет опубликована в следующем номере журнала, авторы намерены представить некоторые содержательные результаты, полученные в исследовании.

Российский мониторинг экономического положения и здоровья населения (РМЭЗ) — обследование, проводимое одновременно для домохозяйств и инди-

* Статьи данного раздела могут быть использованы в качестве учебных материалов к курсам: «Методология социологического исследования», «Социальная стратификация», «Социология труда», «Социология домашнего хозяйства»

М.С. СВАФФОРД, М.С. КОСОЛАПОВ, П.М. КОЗЫРЕВА
Российский мониторинг экономического положения
и здоровья населения (РМЭЗ)

видов и охватывающее различные аспекты их материального благосостояния и здоровья. Это обследование охватывает разные группы и обеспечивает моментальные снимки условий, в которых оказались российские домохозяйства и индивиды, начиная с 1992 г. (за исключением 1997 г.). Однако оно также может быть использовано в качестве панельного обследования, позволяющего исследователям проследить судьбу конкретных домохозяйств и индивидов в условиях переходной экономики.

Хотя содержание вопросников детально описано ниже, мы для знакомства коротко изложим здесь их содержание.

Обследование *домохозяйств* (их число составляет 4000) охватывает:

- состав домохозяйства;
- жилищные условия и право собственности;
- владение предметами длительного пользования;
- продукты питания, полученные на собственных земельных участках, и их использование;
- все виды расходов;
- все категории доходов;
- подробности о невыплатах заработной платы работодателями и потерях, связанных с закрытием банков.

Индивидуальный вопросник предназначен для заполнения *всеми* членами всех 4000 домохозяйств выборки, т.е. более чем 10000 индивидами. Для детей в возрасте до 13 лет включительно вопросники заполняются родителями. Дети в возрасте четырнадцати лет и старше отвечают на вопросы сами (т.е. никому другому за них это делать не разрешается). Индивидуальные вопросники заполнены 97% всех индивидов из участвующих домохозяйств. Основные темы, рассматриваемые в индивидуальных вопросниках, охватывают:

- демографическую статистику, включая возраст и семейное положение;
- работу;
- историю образования;
- самооценку удовлетворенности уровнем благосостояния;
- использование медицинских услуг и лекарств;
- оценку состояния здоровья и измерение роста, веса, объема талии;
- потребление продуктов питания в течение 24 часов, включая показатели пищевой ценности, подсчитанные специалистами по питанию;
- рождение детей, аборт и контроль рождаемости;
- организацию заботы о детях;
- бюджет времени.

Дополнительно *вопросник о месте жительства* обеспечивает данные об инфраструктуре населенного пункта, в котором проживает каждый опрошенный индивид. Эти данные, наряду с любыми другими общедоступными сведениями о районе проживания респондента, могут быть привязаны к данным по индивидам и домохозяйствам.

Как объясняется ниже, эту объемную базу данных может получить любой желающий совершенно бесплатно. Более того, исследователи могут свободно использовать результаты обследования в своих публикациях: разрешения для этого не требуется, достаточно лишь сослаться на источник (приводимый ниже). Более 300 организаций и индивидуальных исследователей со всего мира, в том числе 50 из России, уже пользуются этими данными.

Конкретные процедуры, использованные в обследовании РМЭЗ

В статье «Международные стандарты оценки социологических обследований», опубликованной в предыдущем номере журнала, мы дали обзор некоторых стандартов, установленных для научных социологических обследований. Мы показали, что социологическое обследование должно обеспечивать четкую связь между респондентом и социологом. Такая связь состоит из нескольких звеньев: выборка, вопросник, сбор данных, кодировка, ввод данных и чистка. Связь эта не сильнее самого слабого своего звена, следовательно, каждое звено должно создаваться очень тщательно.

Здесь, в нашей второй статье, мы опишем, как подобные звенья были построены в РМЭЗ. Однако, подробно мы остановимся лишь на выборке и анкете, полагая, что читатели имеют общее представление о смысле таких терминов как «стандартная ошибка» и «вероятностная выборка», которые были рассмотрены в предыдущей статье. В более же детальном рассмотрении четырех последних звеньев, на наш взгляд, нет необходимости; достаточно заметить, что в РМЭЗ сбор данных (включая обучение и контроль работы интервьюеров), кодировка, ввод данных и чистка проводились согласно стандартам, перечисленным в нашей предыдущей статье.

Важно понимать, что РМЭЗ на самом деле представляет собой две различные серии обследований: 1-ая фаза состояла из четырех волн (названных «Волны 1-4»), проведенных в 1992 г. и 1993 г.; 2-ая фаза пока состоит из четырех волн, проведенных в конце 1994 г., 1995 г., 1996 г. и 1998 г. На них порой ссылаются как на «2-ю фазу, волны 1-4»; однако чаще их обозначают «волны 5-8». Последнее вносит некоторую путаницу, так как использование обозначения «волны 5-8» как бы подразумевает, что они являются продолжением волн 1-4, проведенных в ходе первой фазы, тогда как на самом деле они основаны на совершенно новой выборке. Однако структурные сравнения по одномоментным выборкам могут с полным правом быть сделаны для всех восьми волн, и именно в этом смысле волны 2-ой фазы иногда называют «волны 5-8». В данной статье мы будем рассматривать в основном обследования 2-ой фазы. Тем не менее, ниже мы объясним, как получить данные и полную информацию по всем восьми волнам, и сообщим сведения об участниках и спонсорах обеих фаз РМЭЗ.

Наша третья статья, которая должна появиться в следующем номере журнала «Мир России», будет посвящена обзору важнейших результатов о ситуации в России, основанных на данных РМЭЗ.

Разработка вопросника

Уровень жизни определяется взаимодействием индивидов, домохозяйств и их места жительства в контексте стран и всего мира, в котором они расположены. Для обеспечения детального, многоуровневого анализа этих факторов в Российской Федерации в РМЭЗ использовался ряд инструментов: анкета для взрослых, анкета для детей (заполняемая на основе ответов, данных взрослыми), анкета для домохозяйств и анкета по инфраструктуре поселения, включающая также данные по местным ценам. Помимо этого данные по области или району проживания респондента можно связать с данными любого из этих вопросников.

М.С. СВАФФОРД, М.С. КОСОЛАПОВ, П.М. КОЗЫРЕВА
Российский мониторинг экономического положения
и здоровья населения (РМЭЗ) _____

ВОПРОСНИК ДЛЯ ДОМОХОЗЯЙСТВ

- B** Состав домохозяйства (родственные отношения между всеми его членами, дата рождения, семейное положение, пол, количество месяцев, проведенных в семье в течение последних 12 месяцев, причина отсутствия бывших членов домохозяйства)
- C.** Жилищные условия (форма собственности, структура, удобства, коммунальные услуги, подсобные помещения, владение товарами длительного пользования, продажа товаров длительного пользования для того, чтобы заработать деньги на еду и одежду*)
- D.** Сельское хозяйство и разведение животных (владение землей, плата за землю, выращенный урожай и полученные животные продукты [потребление, подарки родственникам, продажа], затраты на производство продукции)
- E.** Расходы (на продукты питания за 7 дней, на одежду за три месяца, на топливо за один месяц: оплата жилья за месяц, неуплата за жилье и коммунальные услуги; крупные покупки за три месяца, покупка различных прочих товаров и услуг за тридцать дней; сбережения, транспортные расходы и другие трансферты; специальные данные по финансовому кризису августа 1998 г., потери, вызванные кризисом)
- F.** Доходы (из всех источников помимо заработной платы, социальные выплаты, оценка общей заработной платы и общего дохода)
- G.** Заметки интервьюера

* Несколько пунктов, подобных данному, были заданы не во всех четырех волнах.

ВОПРОСНИК ДЛЯ ИНДИВИДОВ

ВЗРОСЛЫЕ

ДЕТИ

- | | |
|---|--|
| I. Место рождения, язык, этническая принадлежность | I. Место рождения, язык |
| J. Работа (множество аспектов первичной и вторичной занятости; предпринимательская деятельность и прочая независимая трудовая деятельность; образование; статус безработного и пенсионера; самооценка благополучия, статуса, отношений с другими и удовлетворенности; трудовой опыт; семейное положение) | K. Посещение школы и расходы; физическая активность; чтение и активность в области видео; забота детей в случае данного ребенка |
| L. Медицинское обслуживание (использование услуг и медикаментов, плата за медицинские услуги; страховка) | L. Медицинское обслуживание (использование услуг и медикаментов, плата за медицинские услуги; страховка, прививки и вакцинации) |
| M. Состояние здоровья (включает специальные вопросы для инвалидов; лечение; потребление жидкости; выделения; хронические заболевания; "тест памяти"; употребление спиртных напитков и курение; физические упражнения) | M. Оценка здоровья; лечение; возраст, в котором начались менструации; потребление чая и кофе; недавно прошедшие заболевания |
| N. Рождение детей, выкидыши, аборт и контроль рождаемости; планирование семьи | |
| O. Бюджет времени (предшествующая неделя) | O. Бюджет времени (излагается взрослым) |
| p. Потребление пищи в предыдущий день | P. Потребление пищи в предыдущий день |
| R. Измерение роста, веса и размера талии респондента | R. Измерение роста, веса и размера талии респондента |
| S. Комментарии интервьюера | S. Комментарии интервьюера |
-

Философия, лежащая в основе разработки анкет, нацеливает не на углубленное изучение одной или двух тем, а на широкий охват ситуации. Это разочаровывает специалистов по узким темам, поскольку — в целях высвобождения места для раскрытия большего числа тем — ряд полезных переменных был при разработке анкет исключен. Следует понимать, что в связи с вовлечением в исследование представителей различных институтов окончательные варианты вопросников представляют собой компромисс, при котором все стороны согласились на нечто меньшее, чем свой идеал.

Широта очень важна, так как благодаря ей можно разработать объясняющие модели, учитывающие множество переменных, влияющих на поведение человека. Обычно с методологических позиций проблематично или даже невозможно объединить данные одного хорошего обследования, например, по образованию (проведенного Министерством образования) с данными другого обследования, например, по занятости (проведенного Министерством труда) для того, чтобы изучить соотношение образования и занятости. Очевидно, что для включения обеих переменных в модель требуется собрать данные от одних и тех же респондентов. Более того, статистической ошибкой (называемой экологической ошибкой) является расчет взаимосвязей, основанный на агрегировании по территориальным единицам, как порой поступают с государственными данными. Например, неправомерно изучать взаимосвязь между средним образованием и средним заработком по всем районам с тем, чтобы сделать выводы о влиянии образования на уровень заработка. В силу того, что образование влияет на заработки *индивидов*, данные должны основываться на индивидах, а не на районах.*

Обратимся к нашему примеру, основанному на изучении занятости: в отличие от большинства узких обследований занятости, РМЭЗ позволяет аналитикам включать в анализ множество существующих переменных, таких как: здоровье, образование, участие в уходе за детьми, доход домохозяйств и региональные возможности занятости по рассматриваемой специальности. Кроме того, тот факт, что в нашем случае данные были собраны со всех индивидов домохозяйства, также оказывается весьма полезным. В конце концов, положение респондента в сфере занятости и его действия во многом зависят от положения в сфере занятости и от помощи других членов домохозяйства.

Ниже приводится краткое изложение содержания вопросников. Информация о том, как получить вопросники, приводится в конце данной статьи.

Следует отметить, что поскольку исследование акцентировало внимание не на вопросах национальности, а на вопросах повседневной жизни, оно проводилось на русском языке. В этой связи анкеты составлялись только на русском языке: они не переводились ни на какой другой язык и поэтому обратного перевода не требовалось.

Выборка РМЭЗ

Схема построения выборки, принятая в Российском мониторинге экономического положения и здоровья населения, естественным образом отражает цели самого РМЭЗ. К сожалению, как это почти всегда случается в больших многоце-

* *Правомерным*, однако, является включение в базу данных контекстуальных переменных из домохозяйств, поселений и районов при использовании индивидов как единиц наблюдения в том случае, когда целью является объяснение человеческого поведения. Это требует специальной статистической техники, называемой "иерархическое линейное моделирование"

М.С. СВАФФОРД, М.С. КОСОЛАПОВ, П.М. КОЗЫРЕВА
**Российский мониторинг экономического положения
и здоровья населения (РМЭЗ)**

ИНФРАСТРУКТУРА ПОСЕЛЕНИЯ И ЦЕНЫ НА ПРОДУКТЫ ПИТАНИЯ

1. Численность и площадь поселения
 2. Права на землю и использование зданий для предпринимательства
 3. Расстояние до государственных центров и крупных городов
 4. Имеющиеся типы жилищ
 5. Инфраструктура транспорта и коммуникаций
 6. Учреждения здравоохранения
 7. Общественное питание
 8. Возможности занятости
 9. Институты образования
 10. Банковская система
 11. Пожароохрана и милиция
 12. Такие удобства как вода, канализация, электричество
 13. Государственная социальная поддержка
 14. Цены на примерно 100 видов продуктов питания
-

левых обследованиях, различные цели РМЭЗ требовали бы в идеале различающихся в той или иной степени моделей (методов) выборки. Например, при изучении динамики изменений внутри конкретных домохозяйств требуется *панельный* (panel) метод выборки, в котором те же самые домохозяйства опрашиваются в ходе каждой волны. Тем не менее, в силу того, что часть домохозяйств со временем выпадает из такого обследования, выборка обычно становится все более смещенной с каждой волной. Это противоречило бы второй цели РМЭЗ: обеспечить точное представление о населении в каждый момент времени. Эта цель требовала использования так называемой «структурной модели выборки» или — что то же самое — «модели поперечного сечения {срезы}» (в русском языке нет соответствующего общепринятого термина). На Западе принят термин «cross-sectional design». В реальности РМЭЗ проводился как обследование, использующее модель «повторяемой структурной выборки» (repeated cross-sectional sample design). Эта модификация «структурного» дизайна выборки, как мы объясним ниже, делает возможным проведение панельного анализа.

Другой потенциальный конфликт в требованиях к построению выборки заключается в необходимости одновременной репрезентации как индивидов, так и домохозяйств. Идеальная основа для построения выборки домохозяйств (для анализа распределения доходов или благосостояния домохозяйств) не обязательно является идеальной основой для построения выборки индивидов (например, для измерения уровня безработицы). В самом деле, как хорошо знают специалисты по выборкам, даже если целью обследования являются только, скажем, домохозяйства, наиболее эффективное проектирование выборки для изучения одной переменной (например, дохода домохозяйств) должно отличаться от наиболее эффективного проектирования для изучения другой переменной (например, использования личных садовых участков).

Более того, как это случается при построении любых выборок, данное проектирование выборки разрабатывался в условиях стандартной дилеммы: с одной стороны, — обеспечение точных оценок, с другой, — ограничение на ресурсы. Хотя мы также разработали и другую модель выборки, где количество первичных единиц отбора (ПЕО) в два раза больше, чем в реальной выборке 2-ой фазы, а стандартная ошибка еще меньше, однако применить ее на практике нам не позволили бюджетные ограничения.

Наконец, выборка должна являться результатом компромисса между выше упомянутыми дилеммами и при этом соответствовать, насколько это возможно,

высоким стандартам настоящей вероятностной выборки — выборки, в которой в идеале все домохозяйства и индивиды Российской Федерации имеют рассчитываемую, ненулевую вероятность попадания в выборку.

Полное описание нашей выборочной процедуры превышает 150 страниц. Согласно требованиям, представленным в нашей предыдущей статье, в этом описании задокументированы все этапы построения выборки. В нем также показано, насколько хорошо процедуры выборки были выполнены на практике. Читатели, желающие ознакомиться с полным описанием нашей выборочной процедуры, могут получить ее копию на электронной странице РМЭЗ, адрес которой будет приведен ниже, или же ознакомиться с русской версией, которая будет опубликована в журнале «Социология 4М».* В данной же статье мы детализируем процедуру построения выборки лишь в той мере, которая позволит читателю понять представленные в статье данные.

Построение выборки. В 1994 г. перед первой волной 2-ой фазы для того, чтобы построить выборку жилищ для России, была использована модель многоступенчатой стратифицированной вероятностной выборки.* *

Исходной базой для построения выборки послужили данные переписи, проведенной Госкомстатом СССР в 1989 г. в разрезе административно-территориальных районов. По этим данным, на территории Российской Федерации было 2788 административно-территориальных районов. Эти районы в государственной статистике по умолчанию подразделяются на три типа: а) районы внутри городов с районным делением, б) районы-территории с находящимися в них населенными пунктами и в) районы-города (без районного деления) областного, краевого, республиканского или федерального подчинения, вынесенные вне административно-территориальных районов, на территории которых они физически находятся.

Для использования административно-территориальных районов в качестве единиц отбора на первой ступени (ПЕО) нашей многоступенчатой выборки мы вынуждены были преобразовать государственную статистику, внося города-районы (кроме центров субъектов Федерации и других очень крупных городов), вырванные из территорий, в состав тех административно-территориальных районов, на территории которых они реально находятся. Кроме того, мы объединили районы в городах с районным делением в единое целое. После этих преобразований мы получили список из 2029 районов, которые мы назвали «объединенными» или «выборочными». Это преобразование, во-первых, повышает гетерогенность внутри ПЕО, что требуется теорией выборки, и, во-вторых, превращает все ПЕО в естественные географические территории, что тоже немаловажно для территориальной выборки. Как показано в нашем полном отчете, подобное преобразование снижает величину

* Журнал "Социология 4М" издается Институтом социологии РАН. Его можно купить в киоске Института или заказать, обратившись в редакцию по телефону 125-00-71.

** Мы подчеркиваем, что эта выборка строилась не для того, чтобы обеспечить возможность сравнения между регионами. Для этого нам бы потребовалась значительно большая выборка с существенно большим количеством ПЕО по сравнению с тем, что могло быть построено с учетом финансовых ограничений. Как объяснялось в предыдущей статье "Международные стандарты оценки социологических обследований", точность зависит не только от объема выборки, но и от числа ПЕО в многоступенчатом обследовании. Результаты практически всех статистических пакетов не способны привлечь во внимание кластеризацию (т.е. количество ПЕО), а проверки значимости (или стандартные ошибки), подсчитанные при сравнении регионов, дают чересчур оптимистичные результаты, если количество ПЕО и размер выборки для каждого из сравниваемых регионов невелики.

М.С. СВАФФОРД, М.С. КОСОЛАПОВ, П.М. КОЗЫРЕВА
Российский мониторинг экономического положения
и здоровья населения (РМЭЗ)

стандартной ошибки и сокращает полевые затраты на обследование.

Из объединенных административно-территориальных районов затем было сформировано 38 достаточно однородных страт, образованных преимущественно на основе географических факторов и уровня урбанизации, а также и на основе этнических признаков там, где различия по ним наиболее значительны (выделенные страты представлены в полном отчете по выборке.) Как это принято в большинстве национальных опросов, предполагающих личное интервью, некоторые удаленные районы с малой плотностью населения были исключены с целью снижения издержек. Кроме того, из-за вооруженного конфликта была исключена и Чечня.* Из оставшихся районов, на которые приходится более 95% населения России, три крупнейшие единицы — Москва, Московская область и Санкт-Петербург — образуют *саморепрезентирующие* страты. Другими словами, используя стандартные процедуры, невозможно строить выборку, не включающую эти важнейшие территориальные единицы с огромным населением. По ряду технических причин подобная практика также ведет к снижению величины стандартной ошибки.

Оставшиеся 2026 районов были распределены по 35 стратам приблизительно одинакового размера. Затем в каждой из этих страт с помощью *метода отбора с вероятностью, пропорциональной размеру* (ВПр), случайным образом отбирался один район. Это означает, что вероятность выбора района в данной страте была прямо пропорциональна числу людей, в нем проживающих. Например, если в одном районе данной страты проживают в два раза больше людей, чем в другом, вероятность выбора первого района в два раза больше вероятности выбора второго.

Целевой размер выборки был определен в 4000 жилищ. Они распределились следующим образом: 584 единицы были распределены в саморепрезентирующих стратах — Москва, Санкт-Петербург и Московская область, на которые приходится 14,6% российского населения. Оставшиеся 3416 жилищ были относительно равномерно распределены среди 35 несаморепрезентирующих страт (около 100 жилищ на страту), так как их взяли из примерно одинаковых по размеру страт с использованием ВПр.** Однако, чтобы позволить себе величину недостижимости порядка 15%, мы на самом деле строили выборку, состоящую из 4718 адресов, 940 из которых были распределены среди трех саморепрезентирующих страт (см. табл. 1, столбец 5; на столбцы 6-8 пока можно не обращать внимание). Избыточная выборка была сконструирована преимущественно в мегаполисах, где ожидается самый высокий уровень недостижимости.

* Это означает, что выборка не является строго вероятностной для всей России, в смысле определения, которое давалось выше. Тем не менее, подобный компромисс считается приемлемым в том случае, когда исключаемые районы определяются до построения выборки и их список приводится в общедоступных документах - в нашем отчете они представлены.

** Некоторые читатели могут удивиться, почему примерно одинаковое число респондентов было взято из каждой

ПЕО (за исключением трех саморепрезентирующих страт) В силу того, что в некоторых ПЕО проживает значительно больше людей, чем в других, может казаться, что в них надо привлекать больше респондентов. Действительно, если бы все районы (вне зависимости от того, проживает там много или мало людей) получили бы *равную* вероятность выбора на первом шаге, больше респондентов из более крупных городов было бы привлечено в ходе последнего этапа. Однако, мы использовали метод отбора с вероятностью, пропорциональной размеру (ВПр). На первом этапе более крупные районы получили большую вероятность выбора, чем небольшие. Поэтому, когда сами районы были отобраны, в каждом из них отбиралось одинаковое количество респондентов (с небольшими изменениями в силу предполагаемых различий в величине достижимости). Оба способа дают всем респондентам одинаковую возможность быть отобранными в выборку. Тем не менее, специалисты по выборкам считают ВПр лучше, так как этот метод отбора обеспечивает меньшее значение стандартной ошибки. Он также выравнивает объем работ для различных регионов, облегчая, таким образом, сбор данных.

В силу того, что ни для одного из 38 отобранных ПЕО нет общего списка домохозяйств или жилищ (для административно-территориальных районов их вообще не существует), была, как обычно, введена промежуточная стадия отбора. *Отбор единиц второй ступени* (ВЕО) различался в зависимости от того, было ли население городским (проживающим в городах или ПГТ) или сельским (в деревне). Таким образом, в каждой отобранной ПЕО население стратифицировалось на городское и сельское, и заданный объем выборки пропорционально распределялся между этими подстратами. Например, если 40% населения данного региона является сельским, то 40 из 100 адресов, относящихся к данной страте, берутся из сельских населенных пунктов (СНП).

В сельской местности в качестве ВЕО выступают сами СНП. В городах ВЕО определялись, где это возможно, границами счетных участков переписи 1989 г. Если необходимой информации не было, использовались избирательные участки или почтовые отделения — в порядке убывания предпочтительности. На каждые десять адресов выборки отбиралась приблизительно одна ВЕО.

После того, как были отобраны ВЕО, с помощью натурального обхода и официальных документов о границах участка составлялся полный перечень жилищ.* Подготовка нового списка жилищ позволяла нам избежать соблазна воспользоваться устаревшими списками. Мы постарались составить список жилищ таким образом, чтобы почти всегда в каждом из них жили члены только одного домохозяйства. Например, если в выбранной ВЕО выявлялись дома с коммунальными квартирами, комнаты, занятые отдельными домохозяйствами, заносились в список как отдельные жилища до построения выборки. Наконец, когда окончательный список адресов для ВЕО был составлен, сугубо механическим способом отбиралось требуемое число жилищ, начиная со случайного жилища в списке.

Как в городской, так и в сельской местности от интервьюеров требовалось посетить каждый отобранный адрес как минимум три раза в каждой волне с тем, чтобы получить интервью. Им не позволялось делать какого-либо рода замены жилищ или домохозяйств. Первой задачей интервьюеров было выявление домохозяйств по указанным адресам. («Домохозяйство» определялось как группа людей, живущих вместе и имеющих общие доходы и расходы; в состав домохозяйства включались, согласно определению, не состоящие в браке дети не старше 18 лет, временно проживающие за пределами местожительства во время проведения обследования.) Если интервьюер обнаруживал по данному адресу более одного домохозяйства, случайным образом по специальной процедуре выбиралось одно из них. Затем интервьюер по вопроснику для *домохозяйства* опрашивал наиболее осведомленного о доходах и расходах члена домохозяйства.

Кроме того, интервьюер проводил индивидуальные интервью с максимальным количеством членов домохозяйства в возрасте от 14 лет и старше. (Это порой также требовало нескольких посещений.) Данные о детях, не достигших 14 лет, собирались с помощью взрослых членов домохозяйства и вносились в анкеты для детей. Если взрослые отказывались отвечать на вопросы или их не было дома все время проведения опроса, запрещалось для получения информации о них обра-

* Обратите внимание на то, что использование адресов жилищ подразумевает исключение из выборки заключенных, военнослужащих срочной службы, людей, находящихся в больницах, детских домах, что нормально для большинства национальных опросов. Бомжи также не попадают в выборку.

М.С. СВАФФОРД М.С. КОСОЛАПОВ, П.М. КОЗЫРЕВА
Российский мониторинг экономического положения
и здоровья населения (РМЭЗ)

Таблица 1

Первичные единицы отбора (ПЕО) в 38 стратах в выборке
для 2-ой фазы РМЭЗ*

1	2	3	4	5	6	7	8	9
№ страты	Название района (ЛЕО)	Население страты в целом чел.	Доля населения в страте в целом, %	Количество жилищ, вошедших в выборку N=4718, шт.	Доля жилищ, по которым в 5-ой волне был проведен опрос в каждой ПЕО,	Доля жилищ, по которым в 6-ой волне был проведен опрос N=3755 жилищ (3783 домохозяйств), %	Доля жилищ, по которым в 7-ой волне был проведен опрос, N=3591, %	Количество БЕО,шт.
1	Санкт-Петербург	4 990 749	3,55	222	3,75	2,96	2,45	16
2	Москва	8 875 579	6,32	464	6,64	6,23	5,90	26
3	Московская область	6 647 639	4,73	254	4,98	4,95	5,09	19
4	Республика Коми Сыктывкар	3563417	2,54	117	2,49	2,58	2,51	10
5	Республика Коми Усинский район	3 365 163	2,40	102	2,36	2,29	2,31	10
6	Ленинградская область Волосовский район	3 250 569	2,32	100	2,26	2,34	2,31	10
7	Смоленск	3 694 125	2,63	118	2,67	2,56	2,76	10
8	Калининская область Ржевский район	3 553 950	2,53	109	2,47	2,56	2,64	10
9	Тула	3 297 983	2,35	128	2,47	2,26	2,37	10
10	Калужская область Куйбышевский район	3 320 776	2,37	100	2,39	2,45	2,56	10
11	Нижний Новгород	3 267 563	2,33	123	2,64	2,61	2,37	10
12	Чувашская республика Шумерльский район	3 349 942	2,39	110	2,52	2,69	2,84	10
13	Пензенская область Земетчинский район	3 351 545	2,39	104	2,42	2,40	2,53	10
14	Липецк	3 579 158	2,55	114	2,44	2,40	2,53	10
15	Тамбовская область Уваровский район	3 298 639	2,35	100	2,29	2,18	2,37	10
16	Казань	3 641 742	2,59	108	2,42	2,37	2,51	10
17	Саратов	3 307 126	2,36	125	2,47	2,61	2,64	10
18	Саратовская область Вольский район	3 343 329	2,38	100	2,34	2,32	2,42	10
19	Волгоградская область Руднянский район	3 286 025	2,34	100	2,52	2,58	2,62	10
20	Кабардино-Балкария Зольский район	3 603 117	2,57	100	2,31	2,34	2,59	7
21	Ростовская область Батайск	3571 118	2,54	128	2,52	2,37	2,45	10
22	Краснодар	3 410 747	2,43	108	2,34	2,32	2,12	10
23	Ставропольский край Георгиевский район	3 480 902	2,48	102	2,36	2,37	2,45	10
24	Краснодарский край Куцевский район	3 462 050	2,47	100	2,31	2,56	2,56	6
25	Челябинск	3289013	2,34	108	2,39	2,48	2,45	10
26	Курган	3289012	2,34	106	2,29	2,26	2,06	10
27	Удмуртская республика Глазовский район	3 596 353	2,56	101	2,19	2,40	2,37	11
28	Оренбургская область Орск	3519454	2,51	102	2,44	2,66	2,64	9
29	Пермская область Соликамский район	3 268 053	2,33	102	2,42	2,53	2,64	10
30	Челябинская область Красноармейский район	3 277 244	2,33	100	2,44	2,37	2,28	9
31	Томск	3452815	2,46	106	2,16	2,18	2,48	10
32	Ханты-Мансийский АО Сургутский район	3 283 396	2,34	102	2,54	2,48	2,06	10
33	Алтайский край Бийский район	3 460 148	2,46	108	2,57	2,61	2,56	10
34	Алтайский край Курьинский район	3 389 239	2,41	100	2,42	2,48	2,67	8
35	Красноярск	3 552 244	2,53	115	2,47	2,32	1,95	10
36	Владивосток	3 465 752	2,47	125	2,59	2,85	2,76	10
37	Красноярский край Назаровский район	3 519 543	2,51	107	2,44	2,66	2,62	10
38	Амурская область Тамбовский район	3 525 739	2,51	100	2,29	2,42	2,56	10
	ИТОГО:	140 400	100,0	4718	100,0	100,00	100,00	401

* Население в целом по данной таблице составляет не 147 млн чел., а приблизительно 140 млн. чел.: как мы уже писали, малонаселенные районы и Чечня были исключены при построении выборки. Когда формировалась эта выборка, данные микропереписи 1994 г. были недоступны. Тем не менее, новые веса, ставшие доступными вместе с базой данных, могут быть использованы для того, чтобы привести выборку в соответствие с 1994 г., а не 1989 г.

щаться к другим взрослым членам семьи, т.е. никакой замены не допускалось. Благодаря тому, что было опрошено более 97% всех членов домохозяйств, мы получили достойную вероятностную выборку индивидов без дополнительного — помимо того, что использовалось для жилищ (или домохозяйств) — взвешивания.

Выборка по завершении 5-й волны. После того, как были отобраны жилища и домохозяйства для 5-ой волны, в каждой последующей волне мы старались опросить любое домохозяйство, проживающее в том же жилище. Другими словами, даже если конкретное домохозяйство, проживающее по данному адресу, отобранному для 5-ой волны, отказывалось принять участие в волнах с 6-ой по 8-ую, интервьюеры возвращались в этот дом с целью взять интервью в данном домохозяйстве или в любом новом домохозяйстве, переехавшем в данное жилище. Коротко говоря, изначально для 5-ой волны по описанной выше процедуре была построена так называемая «одномоментная выборка» (cross-sectional sample) жилищ для опроса проживающих в них домохозяйств, которая дает возможность репрезентации домохозяйств и населения России. Затем в каждой последующей волне делалась попытка опросить все домохозяйства, проживающие в тех же самых жилищах, вне зависимости от того, проживали ли они здесь при предыдущем опросе или нет. Данный подход в специальной литературе называется «повторяемая одномоментная выборка» (repeated cross-sectional sample design) домохозяйств. Возобновляемое (повторяемое) обращение к фиксированной выборке жилищ обеспечивает, с одной стороны, достаточно высокую репрезентативность выборки домохозяйств и населения для каждой волны*, и, с другой, — внутреннее сходство между двумя последовательными выборками, что повышает эффективность сравнения агрегированных данных во времени. Т.е. различия в результатах, полученных в разных волнах, не могут быть объяснены как случайные различия, связанные с тем, что использовались различные выборки жилищ для разных волн.

Основной недостаток такой модели выборки для РМЭЗ — это ограничение информационной базы для проведения глубокого микроуровневого панельного анализа лонгитюдных изменений на уровне отдельных домохозяйств и индивидов, так как при действительно панельном дизайне выборки на каждой волне отслеживаются все первоначальные домохозяйства и индивиды и осуществляется попытка их опроса вне зависимости от места их проживания на момент последующих волн. Конечно, и при используемой модели выборки сохраняется возможность такого анализа для тех домохозяйств и индивидов, которые остаются проживать в избранных изначально жилищах. Однако анализ, проводимый на такой «деформированной» панели, может приводить к систематическим смещениям в тех случаях, когда существует зависимость между причинами, объясняющими переезды домохозяйств или отдельных людей в новые места жительства, и переменными, интересующими исследователя.

Даже в том случае, когда в последовательных волнах опрашиваются только домохозяйства, проживающие в изначально выбранных жилищах, задача установления тождественности домохозяйств в разных волнах (а только для таких домохозяйств имеет смысл панельный анализ) зачастую оказывается довольно сложной. Например, пусть семья, включающая два поколения, существовала в

* Один из недостатков заключается в том, что выборка адресов жилищ, построенная в 1994 г., не способна учесть снос жилья и строительство нового с течением времени. Обычно это препятствие не считается достаточно серьезным для того, чтобы требовать корректировки чаще, чем примерно раз в пять лет.

М.С. СВАФФОРД, М.С. КОСОЛАПОВ, П.М. КОЗЫРЕВА
*Российский мониторинг экономического положения
и здоровья населения (РМЭЗ)*

5-ой волне как единое домохозяйство, а к моменту проведения 6-ой волны разбилась на две семьи с отдельными доходами и расходами, оставаясь при этом жить в том же жилище. Очевидно, что, по определению, их уже нельзя считать единым домохозяйством. В подобных случаях опрашивались оба образовавшихся домохозяйства, и в базе данных показывалась их связь с предшествующим домохозяйством из 5-ой волны. Другой возможный вариант — в 5-ой волне существовали два одиноких человека, проживавших отдельно в разных жилищах, входящих в выборку. К моменту проведения 6-ой волны они вступили в брак и стали проживать в одном из этих жилищ, т.е. вместо двух домохозяйств образовалось одно. Как устанавливать тождественность в этом случае? И таких примеров довольно много.

Начиная с 7-ой волны, мы отказались от точного следования исходной модели отбора — «повторяемой одномоментной выборки» и попытались дополнить ее, включив чисто панельный компонент. Соответственно, мы следовали за максимально возможным числом домохозяйств и респондентов, переехавших из первоначально отобранных жилищ. Преимуществом такого подхода является возможность проведения полноценного микроуровневого панельного анализа изменений, происходящих в России. Однако, необходимо понимать, что чисто панельные обследования требуют больших затрат и значительных усилий, связанных с отслеживанием перемещений и дополнительными поездками интервьюеров. Технологически — это очень специфический вид обследований. Один из ведущих специалистов по выборке, руководитель департамента выборки Института социальных исследований (ISR) Мичиганского университета С. Херинга считает, что поддержание настоящей панели домохозяйств и индивидов ведет к невероятному усложнению процедур сбора данных, чистки данных и, в комплексе, процесса управления данными. И по мере роста количества волн в лонгитюдной серии эти проблемы усложняются нелинейно. Естественно, переехавшие домохозяйства и индивиды не являются частью выборки, построенной на адресах (жилищах) и не должны включаться в обычный, не панельный анализ. Далее мы рассмотрим смещения, вызванные отказами от участия в обследовании и переездами.

Достижимость. В ходе 5-ой волны участие в обследовании приняли 84% домохозяйств, проживающих в 4718 жилищах. Другие домохозяйства или отказались участвовать, или их членов невозможно было застать дома, несмотря на повторявшиеся попытки (см. ниже). Достижимость, в волнах с 6-ой по 8-ую составила соответственно 80%, 76% и 74%. Достижимость индивидов в опрошенных домохозяйствах превышала 97% в каждой волне; таким образом, доля ответивших среди всех индивидов по всем отобранным жилищам была несколько ниже, чем соответствующая цифра для домохозяйств, приведенная в предыдущем предложении.

Вместо того, чтобы рассматривать долю ответов, независимо в каждой волне, мы можем подсчитать убывание домохозяйств от волны к волне с тем, чтобы определить качество панели. Из общего количества домохозяйств, участвовавших в 5-ой волне (исключая переехавших, т.е. тех, кого опрашивали за рамками выборки жилищ), в 6-ой волне участвовало 83%, в 7-ой — 80% и в 8-ой — 74%. Однако в 7-ой и 8-ой волнах мы также нашли и опросили сотни домохозяйств, выехавших из местожительств выборки. Принимая это во внимание, можно сказать, что фактические значения сохранения панели на несколько процентов выше

тех, что мы только что привели. Предварительные подсчеты показывают, что сохранение домохозяйств 5-ой волны в 8-ой волне было ближе к 78%, нежели к 74%.

Несомненно, *панельный* анализ представляет уникальные возможности для специалистов, занимающихся статистическим анализом. Однако при этом каждый раз приходится решать множество проблем. Например, что должно быть единицей анализа: домохозяйство или индивид? В анализе бедности, охватывающем различные группы, можно выбрать любую из этих единиц. Это означает, что с полным правом можно подсчитать либо процент домохозяйств, живущих в бедности, либо процент индивидов, живущих за чертой бедности. Тем не менее, в случае панельного анализа домохозяйств задача довольно сложна. По определению, панельный анализ основан на сравнении одних и тех же единиц во времени. Однако состав домохозяйств со временем меняется, и, чем длиннее панель, тем сложнее решить, имеет ли смысл говорить, что группа людей по-прежнему образует все то же домохозяйство. Рассмотрим домохозяйство из 5-ой волны, состоявшее из двух 50-летних родителей и их 30-летнего работающего сына. Предположим, что в 7-ой волне эти же двое родителей занимают то же жилище, но их сын выехал. Одно и то же ли это домохозяйство? Будет ли сокращение доходов в этом домохозяйстве свидетельствовать о том, что его положение ухудшилось? Теперь предположим, что к 7-ой волне родители выехали, а сын живет по тому же адресу с женой и маленьким ребенком. Разумно ли считать два этих домохозяйства все тем же одним домохозяйством при проведении панельного анализа? Любой исследователь, желающий использовать домохозяйства в качестве единицы анализа при панельных исследованиях, в подобных ситуациях должен принимать решения, исходя из каждого конкретного случая: ответы будут различаться в зависимости от того, какие вопросы анализируются, так что в базе данных РМЭЗ нет даже попыток таких ответов. В данных РМЭЗ домохозяйства считаются одними и теми же, если в ходе двух волн хотя бы один человек из данного домохозяйства присутствует в обоих домохозяйствах. Это определение выбрано из утилитарных соображений, так как позволяет связать волны. Таким образом, исследователи должны начинать с этой отправной точки и решать, что разумно для того вида анализа, который они хотели бы провести, используя данные списка домохозяйств.

В настоящее время исследователи с обширным опытом анализа панельных данных обычно рекомендуют использовать в качестве единицы анализа индивидов. Вернемся к примеру о бедности: сколько индивидов, живших в бедных домохозяйствах в 5-ой волне, по-прежнему живут в бедных домохозяйствах в 8-ой волне? В этом вопросе не важно, жили ли они *в одном и том же* бедном домохозяйстве в обеих волнах, просто важно, что они живут в некоем бедном домохозяйстве в обеих волнах, даже если состав их (домохозяйств) поменялся. Тождественность индивидов остается биологической постоянной, так что все эти вопросы сосредоточены вокруг проблемы меняющегося состава домохозяйств.

По этой причине интересно рассматривать статистику сокращения не количества домохозяйств, которая была приведена выше, а количества индивидов от волны к волне. Из общего количества индивидов, принявших участие в 5-ой волне, участие в 7-ой волне приняли 72 %; 82 % тех, кто участвовал в 6-ой волне, участвовали в 7-ой волне. Данные по 8-ой волне пока не доступны.

Доля ответивших, таким образом, является удовлетворительно высокой, осо-

М.С. СВАФФОРД, М.С. КОСОЛАПОВ, П.М. КОЗЫРЕВА
Российский мониторинг экономического положения
и здоровья населения (РМЭЗ)

бенно если учитывать длительность общения с респондентами и большие перерывы между волнами. Тем не менее, как это всегда бывает, недостижимость и истощение панели существуют, так что ниже мы проверим возможности смещений.

Причины недостижимости домохозяйств. Статистика причин недостижимости домохозяйств в волнах с 5-ой по 7-ую изложена в таблице 2. В 12-20% случаев недостижимости жилье явно не использовалось как место проживания. Оно могло быть использовано в коммерческих целях, повреждено пожаром или снесено; или соседи могли сообщить, что помещение остается незанятым в течение длительного промежутка времени. При расчете уровня недостижимости по жилищам из знаменателя (4718) вычитались снесенные или переделанные помещения (около 50), что увеличивало уровень достижимости примерно на один процентный пункт. В 13-19% случаев недостижимости никого не заставляли дома при — как минимум — трех посещениях. И в 60-66% случаев причиной недостижимости являлись различного рода отказы принять участие в обследовании. Перечисленные и прочие менее значимые причины недостижимости приведены в таблице 2.

Таблица 2
Причины недостижимости домохозяйств в волнах 5-7 (%)

Причины недостижимости домохозяйств, в %:	Волна 5	Волна 6	Волна 7
- в настоящее время жилье не используется как место проживания	20,1 13,5	12,4 19,0	14,5 16,5
- никого нет дома после трех повторных посещений	0,8	0,8	1,5
- жилище недоступно (нет прохода в здание)	1,9	1,2	0,7
- респондент нетрезв, т.е. нет четкого отказа (домохозяйство состоит из одного человека)	3,4		1,1 65,7
- респондент слишком болен или немощен (домохозяйство состоит из одного человека)	60,4	1,6	
- прочие формы отказов		64,9	
Абсолютная величина недостижимости (количество домохозяйств) _____	743	965	1126

Потенциальное одномоментное смещение, вызванное недостижимостью домохозяйств. Несмотря на высокую степень достижимости домохозяйств, ее снижение за 4 года составило 10% (с 84% до 74%), т.е. можно сказать, что мы начали с 16% неохваченных домохозяйств и дошли до 26%. До каких пределов результаты, основанные на выборке домохозяйств, разработанной для 5-ой волны, могут быть смещены из-за этого, и в каких пределах выборка стала более смещенной после того, как доля ответивших в последующих волнах сократилась? Наш полный отчет включает ряд комплексных таблиц на эту тему; здесь же мы просто обозначим ряд типичных примеров, которые позволяют предположить, что по отношению к демографическим характеристикам величина смещения очень невелика.

Таблица 3 показывает, что распределение домохозяйств по размерам в выборке как в городской, так и в сельской местности хорошо соотносится с данными микропереписи 1994 г. (Сравните, например, два столбца, «Микроперепись

1994 г.» и «Волна 5».) Наибольшее несоответствие (около 4 %) обнаружено для сельских домохозяйств, в которых проживает один человек. Однако стоит помнить, что все сравнения, имеющие отношение к домохозяйствам с единственным обитателем, подозрительны, так как перепись включает всех так называемых институционализированных людей, тогда как наша выборка явно их исключает. Обратите внимание на то, что проценты практически не меняются во всех трех волнах (сравните значения в каждом ряду, начиная с 5-ой волны).

Прочие переменные также показывают, что «усыхание» выборки лишь слегка смещает распределение домохозяйств по основным демографическим характеристикам. Между четырьмя волнами доля домохозяйств, возглавляемых мужчинами в возрасте от 18 до 59 лет, изменилась лишь с 64,8% до 63,2%; женщинами в возрасте от 18 до 59 лет — только с 10,8% до 11,7%.

Лишь один потенциально способный вызвать смещение тип сокращения числа домохозяйств был выявлен с помощью тщательного изучения таблицы 1 (столбцы 6-8). Доля представления домохозяйств Санкт-Петербурга в общей выборке сократилась с 3,75% в 5-ой волне до 2,45% в 6-ой, а Москвы — с 6,64% до 5,90%. Не удивительно, что уровень уменьшения размера выборки в этих крупных мегаполисах был выше, чем в других местах, и снизил их представительность с уровня, на котором ей следует быть (3,55% и 6,22% в столбце 4). Эта небольшая проблема легко корректируется с помощью взвешивания данных при анализе, хотя ни к какой существенной разнице результатов это не приводит.

Теперь давайте рассмотрим характеристики индивидов, а не домохозяйств. Сместились ли из-за недостижимости данные 5-ой волны, и увеличилось ли это смещение по мере того, как достижимость снижалась с 5-ой волны к 8-ой волне? Используя расчетные данные Госкомстата на начало 1997 г. (т.е. сразу после завершения 7-ой волны), мы видим, что распределение выборки РМЭЗ по полу и возрасту достаточно хорошо с ними сопоставимо, и существенного ухудшения за четыре волны не происходит (табл. 4). Аналогично, и данные по образованию респондентов выглядят довольно ободряюще. Согласно микропереписи 1994 г., 13,3% людей в возрасте от 15 лет и старше получили законченное высшее образование. В РМЭЗ процент респондентов с законченным высшим образованием колебался между 15,1% и 16,4% в ходе четырех волн. Это гораздо лучше, чем результат (на уровне 20%), обычно получаемый российскими компаниями, занимающимися опросами общественного мнения и использующими, зачастую, невероятные или псевдовероятностные выборки.

Возможные смещения в панельных данных из-за недостижимости. Отсут-

ствие смещения внутри каждой волны само по себе не показывает, что панельные данные РМЭЗ можно использовать для изучения изменений между волнами без риска смещения. Вспомним, что между 5-й и 7-ой волнами было потеряно более 20% панели. По многим стандартам это хороший результат, особенно учитывая сложность обследования и само содержание интервью. Те, кто проводят исследования, основанные на панельных обследованиях, вынуждены мириться с таким сокращением числа респондентов. Однако поступать подобным образом можно лишь в тех случаях, когда интересующие исследователя характеристики не коррелируют с теми, которые характеризуют именно выбывающих из панели. Нет причин думать, будто цвет волос сильно коррелирует с тем, остались ли участники 5-ой волны в панели или выпали из нее, так что в этом случае сокращение панели не несет никакой угрозы. Тем не менее, изменения в составе домохозяй-

М.С. СВАФФОРД, М.С. КОСОЛАПОЕ, П.М. КОЗЫРЕВА
Российский мониторинг экономического положения
и здоровья населения (РМЭЗ)

Таблица 3

Размер домохозяйств: данные микропереписи 1994 г. и РМЭЗ
(% домохозяйств каждого размера)

Размер домохозяйства	Микроперепись 1994 г.*	Волна 5	Волна 6	Волна 7
Городские домохозяйства				
1	18,1	16,1	17,3	18,0
2	26,1	26,5	26,0	26,6
3	24,3	25,8	25,6	26,2
4	21,0	21,2	20,6	19,7
5+	10,5	10,3	10,5	9,5
Итого	100,0	99,9	100,0	100
Сельские домохозяйства				
1	22,0	18,1	19,7	19,5
2	26,8	27,3	26,6	26,9
3	18,1	18,8	18,8	18,3
4	19,0	20,7	19,2	20,3
5+	14,1	15,0	15,8	14,8
Итого	100,0	99,9	100,1	100,0

* Типы и состав домохозяйств в России (по данным микропереписи населения 1994 г.). М., Госкомстат РФ, 1995 С 23
Сумма значений по столбцам иногда не равна 100% из-за ошибок при округлении.

Таблица 4

Распределение по возрасту и полу в выборке РМЭЗ и расчетные данные
Госкомстата на 1997 г. (% людей в каждой категории)

	Возраст	Расчетные данные Госкомстата РФ на 01.01.97 г.*	Волна 5	Волна 6	Волна 7
Муж.	0-15	11,2	11,8	11,7	11,4
	16-59	29,9	27,9	27,8	27,7
	свыше 60	5,8	5,9	5,8	6,0
Жен.	0-15	10,7	11,2	11,2	11,0
	16-54	27,4	27,8	27,7	27,8
	свыше 55	15,0	15,5	15,8	16,0
Итого		100,0	99,9	100,0	99,9

* Демографический ежегодник России 1997 М., Госкомстат РФ, 1997 С 35. Официальные цифры относятся к январю 1997г. т.е. примерно на месяц позже завершения 7-ой волны Сумма значений по столбцам иногда не равна 100% из-за ошибок при округлении.

ства часто связаны с переездами, так что изучение изменения состава домохозяйства будет под серьезной угрозой из-за сокращения панели.

Рассмотрим пример, когда нас интересуют полные и неполные семьи. Предположим, что у каждого ребенка, который был потерян, т.е. мы не смогли отследить его для панельного обследования, родители разводятся и, соответственно, разъезжаются. Из-за того, что переехавших невозможно было включить в наши подсчеты в 7-ой волне, панельный анализ будет значительно недооценивать процент детей, которые вместо двух родителей стали иметь одного.

Практически во всех панельных опросах мы не можем показать, каков был бы результат, если между 5-ой и 7-ой волнами панель никоим образом бы не сократилась. Тем не менее, мы можем провести стандартный тест, сравнивающий детей, потерянных в 7-ой волне, с теми, кто в ней остался, используя данные,

МИР РОССИИ. 1999. N3

собранные нами обо всех этих детях в 5-ой волне. Другими словами, мы используем данные 5-ой волны, но мы делим всех детей 5-ой волны на две группы: тех, кто по-прежнему участвует в 7-ой волне, и тех, кого не смогли отследить в ходе 7-ой волны. Цель — увидеть, отличались ли в 5-ой волне те, кто позже был потерян, от тех, кто остался в выборке в 7-ой волне. Если они действительно отличались, то смещение будет представлять собой постоянную реальную угрозу для любого вида панельного анализа, посвященного детям.

Однако результаты, представленные в таблице 5, успокаивают (смотрите два левых столбца, озаглавленные «Тест 1: волна 5»). Поражаешься схожести двух левых столбцов. Разница между долей детей, участвовавших в 5-ой и 7-ой волнах, с одной стороны, и участвовавших в 5-ой волне, но не отслеженных в 7-ой волне, с другой, составляет всего 5% для детей, имеющих двух родных родителей, и 3 % для детей, не имеющих родных родителей. Эти различия представляются более чем скромными, а другие различия, представленные в таблице, еще меньше.

Наш второй тест на смещение использует тот факт, что мы пытались последовать за теми, кто выехал из первоначальной выборки жилищ до 7-ой волны, как это было описано выше. В этом тесте мы сравниваем тех детей, которые не

Таблица 5

Влияние сокращения числа респондентов в панели в РМЭЗ: волны 5 и 7
(процентное распределение не состоящих в браке детей до 18 лет)

	Тест 1: волна 5		Тест 2: волна 7	
	Дети, представленные в волне 5 и принимавшие участие в волне 7	Дети, представленные в волне 5, которых не смогли отследить в волне 7	Дети, представленные в выборке жилищ волны 7	Дети, за которыми последовали по их новым адресам за пределами выборки
Число детей, чел.	2136	807	2583	107
В том числе, имеющих (%):				
- двух родителей	77	72	74	69
- одного родителя и партнера	6	7	6	8
- одного родителя, ра нет	14	15	17	18
- не родных родителей	3	6	3	5
ИТОГО	100	100	100	100

переехали, с теми, которые выехали из жилищ, входящих в первичную выборку к 7-ой волне. Этот тест отвечает на вопрос: с большей ли вероятностью в семьях переехавших детей произошли изменения в составе родителей, чем в семьях тех, кто не переехал? По-видимому, нет, по крайней мере, в семьях примерно 50% всех переехавших детей, о которых мы смогли собрать данные (это те семьи, которые переехали в пределах тех же городов, где были расположены изначальные адреса). Приблизительно 69% переехавших детей по-прежнему оставались с обоими своими родными родителями, а это всего примерно на 5 % меньше, чем

М.С. СВАФФОРД М.С. КОСОЛАПОВ, П.М. КОЗЫРЕВА
*Российский мониторинг экономического положения
 и здоровья населения (РМЭЗ)*

среди переехавших (см. столбцы справа, озаглавленные «Тест 2: волна 7»). Конечно, нужно понимать, что число переехавших невелико, так что проценты в крайнем правом столбце имеют низкую точность (т.е. доверительные интервалы часто превышают 10%). Следовательно, стоит избегать построения серьезных выводов на основании незначительных различий между столбцами. Однако даже при данном уровне точности сравнение этих двух столбцов позволяет предположить, что сильного смещения, вероятно, не было. Возвращаясь к предположению о возможности существенного влияния развода родителей на переезд, о котором мы говорили выше, можно сказать, что тест показывает, что далеко не все переехавшие дети были детьми развода.

Еще один тест проводился для данных по расходам (выраженных в рублях 1992 г. и скорректированных с учетом инфляции). Разделяя участников 5-ой волны на тех, кто участвовал в 7-ой волне, и тех, кто выпал из нее, мы выявили, что средние расходы первой группы составили 464 тыс. руб. в 5-ой волне, а второй группы — 498 тыс. руб. в той же 5-ой волне. Другими словами, домохозяйства, выпавшие из изучения к 7-ой волне, в 5-ой волне тратили примерно на 7% больше тех, кто остался в исследовании — вновь небольшая разница. Другой тест использовал показатель занятости. Быть безработным в 5-ой волне и выпасть из выборки к 7-ой волне — совершенно не коррелирующие переменные.

Мы считаем эти тесты обязательными и чрезвычайно важными. К счастью, различия оказались весьма незначительными. Как мы показали выше, нет способа доказать, что сокращение числа участников панели не привело к какому-либо смещению; тем не менее, результаты, подобные данным, считаются очень убедительными.

Обзор уникальных особенностей обследования РМЭЗ

В заключение мы хотели бы свести воедино все значимые особенности РМЭЗ, обсуждавшиеся в отдельных частях статьи.

— *РМЭЗ охватывает широкий спектр связанных переменных.* Это облегчает построение сложных объяснительных (а не описательных!) моделей, основанных исключительно на сфере человеческого поведения. Тем не менее, ввиду того, что РМЭЗ разрабатывался преимущественно для измерения здоровья и благосостояния россиян, он не является богатым источником сведений по различным вопросам общественного мнения.

— *РМЭЗ измеряет поведение всех индивидов в домохозяйстве, позволяя строить объяснительные модели, учитывающие всех членов домохозяйства.* Действия одного индивида часто зависят от действий других членов домохозяйства, и — благодаря РМЭЗ — аналитики могут включать в модель всех домочадцев.

— *РМЭЗ позволяет провести многоуровневый анализ,* в котором данные по индивидам, их домохозяйствам и поселениям, в которых они проживают, могут быть связаны друг с другом.

— *РМЭЗ позволяет провести панельный анализ,* также как и анализ, охватывающий различные группы. Обследования, охватывающие различные группы населения, позволяют нам год за годом подсчитывать, например, число или долю людей, живущих в бедности. Вместе с тем, только панельный анализ может по-

зволить оценить, сколько одних и тех же людей остается бедными год от года, вместо того, чтобы выходить или входить в состояние бедности из-за неких временных краткосрочных неудач.

—*Данные РМЭЗ распространяются в исходном виде*, в отличие от таблиц, распространяемых Госкомстатом РФ. Это позволяет каждому исследователю принимать свои собственные решения о том, как определять переменные или как свести анализ к ряду случаев, относящихся конкретно к интересующей его сфере. Например, различные исследователи могут опробовать свои собственные определения бедности.

—*РМЭЗ основан на вероятностной выборке* российского населения, разработанной в ходе консультаций с экспертами мирового класса по проблемам выборок. Это дает максимально возможную гарантию того, что результаты несмещены, а также позволяет статистикам подсчитать точность оценок. Начнем с того, что уровень достижимости составлял 70-80%, что является — по мировым стандартам — весьма солидным показателем для обследований такого рода. Выборка полностью задокументирована во всех подробностях, и документация является общедоступной. С другой стороны, РМЭЗ недостаточно велик для того, чтобы обеспечить полноправное сравнение российских регионов.

—*Подготовка интервьюеров, сбор данных, проверка, кодирование, ввод данных и чистка данных проводились согласно стандартам, заданным серьезными академическими организациями*, такими как исследовательский центр Мичиганского университета и центр NORC при Чикагском университете.

—*Исходные данные обследования стали доступными фактически бесплатно* в мировой сети Интернет (World Wide Web). Это было сделано для того, чтобы максимизировать использование этих данных, и для того, чтобы для российских чиновников они служили примером того, как хорошо, когда данные о стране становятся действительно общедоступными, стимулируя информированный общественный диалог. Более 300 исследователей и организаций пользуются данными РМЭЗ, в том числе 50 из Российской Федерации. Одной из целей настоящей статьи является увеличение числа российских пользователей данных.

Как получить данные обследования РМЭЗ

Потенциальные пользователи должны понимать два важных свойства данных.

Во-первых, они объемны. Волны с 1-ой по 4-ую охватывали в среднем 6000 домохозяйств и 15000 индивидов в ходе каждой волны; в вопроснике содержалось примерно 800 переменных. Волны с 5-ой по 8-ую охватывали в среднем 3800 домохозяйств и более 10000 индивидов за волну; количество переменных в вопроснике было примерно тем же. В зависимости от формата хранения данных, каждая волна занимает примерно от 20 до 100 мегабайт. По этой причине на сайте, адрес которого приведен ниже, можно ограничить запрос несколькими подмножествами переменных из каждой волны.

Во-вторых, пользователи должны понимать, что анализ требует комплексного использования базы данных с привлечением программного обеспечения, разработанного для этих целей — чаще всего SPSS*, SAS или STATA. (Программ-

* И русскоязычная, и англоязычная версии SPSS могут быть приобретены в ЗАО SPSS Русь (тел. (095) 125-0069).

М.С. СВАФФОРД, М.С. КОСОЛАПОВ, П.М. КОЗЫРЕВА
*Российский мониторинг экономического положения
и здоровья населения (РМЭЗ)*

ное обеспечение, использующее исключительно электронные таблицы, не подходит в данном случае.) За исключением нескольких переменных, таких как пол и год рождения, практически ни одна переменная в подобных базах не может быть на законных основаниях использована без предварительных действий, направленных на установление соответствия между ней и представлением о ней исследователя. Например, в одном вопросе спрашивается, есть ли у домохозяйства земельный участок, находящийся в частной собственности. Если нет, респондент переходит к другим вопросам, не отвечая на те, что касаются земельного участка. Если в домохозяйстве такой участок есть, и на нем выращиваются какие-то продукты питания, выясняется, сколько килограммов, например, картофеля домохозяйство вырастило на участке. Теперь, предположим, исследователь хочет подсчитать средний размер урожая картофеля (в килограммах), выращенного всеми домохозяйствами. Если он просто подсчитывает среднее всех ответов на вопрос о массе выращенного урожая картофеля, он получит ошибочно высокий ответ, основанный на ответах лишь тех, кто действительно выращивал продукты питания у себя на участках, так как все другие респонденты пропустили вопрос, и в данных о них эта переменная осталась пропущенной. Чтобы получить правильный ответ, исследователь должен вписать «0 килограммов» всем тем, у кого нет участка, и тем, у кого он есть, но кто не выращивает на нем продуктов питания. Это может казаться очевидным. Однако, проводя занятия по работе с РМЭЗ с людьми, имеющими степень кандидата наук, но не имеющими опыта работы с подобными базами данных, мы выяснили, что большинство из них сначала делали грубые ошибки, подобные только что упомянутой. Вот почему проблема анализа сложных баз данных занимает несколько лекционных курсов и годы практического опыта в западных академических программах по социальным наукам. Неопытным пользователям следует позаботиться о том, чтобы работать с опытными программистами.

Основной источник информации о РМЭЗ — сайт www.cpc.unc.edu/projects/rhms/rhms_home.html. С этого сайта можно скачать все отчеты, анкеты (на русском и английском языках) и заархивированные базы данных. (В настоящее время базы данных содержатся в формате SAS, но в будущем также могут быть предложены другие форматы.) Для российских исследователей недостатком этого источника является то, что для скачивания огромных файлов с данными (даже в заархивированном виде) может потребоваться многочасовая работа в Интернете. К тому же большая часть информации представлена на английском языке, включая названия переменных в базе данных.

Можно также обратиться в Институт социологии (тел. (095) 125-8088, 125-0071) за файлами в форматах SPSS и ASCII. В файлах SPSS имеются названия переменных на русском языке; кроме того, эти файлы в шесть раз более компактны по сравнению с файлами SAS. Институт также распространяет бумажные копии вопросников.

Перевод с английского — С. Творогова