



УДК 303.425.2

DOI: 10.19181/snsp.2024.12.3.9

EDN: XFHTMI

Научная статья

АНАЛИЗ ДОСТОВЕРНОСТИ ДАННЫХ МАССОВЫХ ОПРОСОВ (НА ПРИМЕРЕ ОЦЕНОК РАСПРОСТРАНЁННОСТИ КУРЕНИЯ В РОССИИ)

Николай Сергеевич Бабич¹

Алексей Александрович Михайлов²

¹ Институт социологии ФНИСЦ РАН,
Москва, Россия,

¹ sociolog@mail.ru,

ORCID 0000-0001-8697-3038

² Государственный академический
университет гуманитарных наук,
Москва, Россия,

² mix-al.mix@yandex.ru,

ORCID 0000-0001-8874-0377

Для цитирования: Бабич Н. С., Михайлов А. А. Анализ достоверности данных массовых опросов (на примере оценок распространённости курения в России) // Социологическая наука и социальная практика. 2024. Т. 12, № 3. С. 198–213. DOI 10.19181/snsp.2024.12.3.9. EDN XFHTMI.

Аннотация. Основанные на опросных данных оценки распространённости курения в России, представляемые различными вполне уважаемыми организациями на один и тот же временной период, колеблются в диапазоне от менее 20% до более 50%. Такой разброс существенным образом влияет на общую картину ситуации с курением в России и оценку успешности антитабачной политики и не может объясняться различиями в концептуализации показателя (кого считать курильщиками). Проблемная ситуация заставляет поставить вопрос о достоверности соответствующих опросных данных, который может быть решён, как минимум, двумя способами: анализом и учётом факторов расхождений или выбором одного из исследований в качестве эталонного с тщательной проверкой достоверности данных. В статье обосновывается выбор второго способа и Российского мониторинга экономического положения и здоровья населения НИУ ВШЭ (далее РМЭЗ) как эталона. Для проверки достоверности его данных использованы три критерия: анализ критических публикаций о погрешностях исследования, методическая триангуляция (сравнение с данными, полученными другими методами) и анализ корреляции с последствиями курения. Все три критерия показали высокий уровень достоверности данных РМЭЗ. Выдвинутая в критических публикациях гипотеза о погрешности исследования не находит эмпирического подтверждения.

© Бабич Н. С., 2024

© Михайлов А. А., 2024

Сравнение с другими предположительно надёжными данными, как опросными, так и не опросными (официальная статистика, наблюдение) показывает расхождение ниже уровня статистической погрешности. Корреляция данных РМЭЗ о курении со смертностью от заболеваний системы кровообращения оказывается исключительно высокой. Следовательно, РМЭЗ целесообразно рассматривать как эталонный источник данных о распространённости курения в России, а существенно отличающиеся от него показатели других опросов — как потенциально сомнительные в своих методических решениях.

Ключевые слова: распространённость курения, массовые опросы, достоверность опросных данных, методическая триангуляция, РМЭЗ

Введение

Курение — один из важнейших поведенческих факторов, повышающих смертность в России и во всём мире, поэтому борьба с ним относится к приоритетам социальной политики. Важный её инструмент — измерение распространённости курения, основой которого служат массовые опросы, проводимые по репрезентативным национальным выборкам. Но разброс оценок доли курящего населения, получаемых с помощью опросов, оказывается очень велик.

По данным выборочных обследований Росстата на 2022 год среди россиян в возрасте старше 14 лет было 19,2% курильщиков¹. ВЦИОМ в июльском опросе того же 2022 года получил величину 33% курящих среди совершеннолетнего населения страны². Разница почти в 14% явно не укладывается в обычную статистическую погрешность. Но ещё более драматический отрыв наблюдается в данных международной исследовательской компании NielsenIQ. По результатам регулярного опроса «Индекс потребительского оптимизма» в первом квартале 2022 г. доля курящих на регулярной основе каждый день россиян составила 37%, а курящих время от времени — 14%, итого 51%³. Эти данные подтверждаются результатами опроса, проведённого компанией «Росгосстрах. Жизнь», согласно которому в мае 2022 г. курили 55% жителей России⁴. Показатели Росстата служат основой для осуществления государственной

¹ Выборочное наблюдение состояния здоровья населения. 2022 // Федеральная служба государственной статистики (Росстат) : [сайт]. URL: https://rosstat.gov.ru/free_doc/new_site/zdor22/PublishSite_2022/index.html (дата обращения: 22.04.2024).

² Курение в России – мониторинг // ВЦИОМ : [сайт]. 12 июля 2022 г. URL: <https://wciom.ru/analytical-reviews/analiticheskii-obzor/kurenie-v-rossii-monitoring-2022> (дата обращения: 22.04.2024).

³ Тренды индустрии: рынок табачных изделий // NIQ : [сайт]. 25 июля 2022. URL: <https://nielseniq.com/global/ru/insights/analysis/2022/trendy-industrii-rynok-tabachnyh-izdeliy/> (дата обращения: 22.04.2024).

⁴ Здесь не курят: в России наметился тренд на отказ от вредной привычки // Росгосстрах Жизнь : [сайт]. 30 мая 2022. URL: <https://rgsl.ru/news/zdes-ne-kuryat-v-rossii-nametilsya-trend-na-otkaz-ot-vrednoy-privyчки/> (дата обращения: 22.04.2024).

политики в России, ВЦИОМ является одним из основных поставщиков информации о массовом сознании россиян, данным NielsenIQ доверяют крупнейшие международные корпорации по всему миру, а страховые компании типа «Росгосстрах. Жизнь» вообще существуют благодаря качеству такой статистики как распространённость курения и его последствий. Тем не менее, оценки распространённости курения, представленные этими респектабельными организациями, колеблются в диапазоне от менее 20% до более 50%.

Такой разброс оценок обосновывает необходимость постановки методического вопроса о том, можно ли вообще полагаться на опросные данные для определения распространённости курения, а если да — то на какие? В статье предпринимается попытка ответить на этот вопрос посредством проведения насколько возможно полной проверки (аудита) методик сбора опросных данных о курении.

Задача проведения аудита разрешима при помощи двух подходов. С одной стороны, можно собрать максимальную по объёму совокупность оценок распространённости курения, сопроводив их различными сопутствующими данными, такими как композиция выборки, способ опроса, формулировка вопросов и т. п. Затем, детально анализируя различия между методиками, сопоставлять их различия в результатах. Таким путём (обычно называемым «метаанализом») выявить более и менее достоверные методики, отбросить недостоверные и в конце концов, через обобщения достоверных дойти до интегрированной, наиболее правдоподобной оценки. С другой стороны, можно взять одно, наиболее подходящее на роль некоего «эталона» исследование, и проверить разными способами достоверность именно его данных. Тогда при положительном исходе проверки, достоверность данных всех остальных опросов может быть легко проверена путём сопоставления с полученным «эталонным».

Первый подход к решению задачи затруднён тем, что процитированные исследования имеют разную степень открытости данных. Например, Росстат размещает на своём сайте файлы с микроданными, ВЦИОМ по вопросу курения сообщает только обобщённую для национального уровня статистику, а для данных NielsenIQ и Росгосстраха неизвестны даже точные параметры выборки. Второй подход может столкнуться с ещё большими препятствиями так как положительный исход проверки не гарантирован и потребует наличия ярко выделяющегося «эталонного» опроса — такого, который выполнен на очень высоком методическом уровне, желательно имеет характер многолетнего мониторинга и широко используется в исследованиях на соответствующую тему. Анализ позволил авторам выявить в российской социологической опросной практике такой опрос, что обусловило выбор в пользу второго пути проведения аудита.

РМЭЗ как «эталонный» опрос о курении

«РМЭЗ» или «RLMS» — «Российский мониторинг экономического положения и здоровья населения НИУ ВШЭ», проводимый НИУ «Высшая школа экономики» и ООО «Демоскоп» при участии Центра народонаселения Университета Северной Каролины в Чапел Хилле и Института социологии ФНИСЦ РАН. Подробное описание исследования и детальные открытые данные за период 1994–2022 гг. представлены на сайтах ВШЭ¹ и Университета Северной Каролины². Объём репрезентативной для населения России выборки исследования с учётом взвешивания колеблется от около 8 (2005 год) до около 17 тыс. человек (2012 год). Исчерпывающие данные о выборке по каждому году доступны на сайте исследования.

Основной представляющий для авторов интерес вопрос «Вы курите в настоящее время?» с вариантами ответа «да» и «нет» задавался в индивидуальном вопроснике, как и другой показательный вопрос «Сколько примерно штук сигарет или папирос Вы обычно выкуриваете за день?». В вопроснике для домохозяйств задавался вопрос о приобретении сигарет за последнюю неделю.

Исследование РМЭЗ было спроектировано с учётом высоких методических требований и международного опыта [1; 2] и в настоящее время является одним из основных источников информации о распространённости курения, его факторах и последствиях в России [3; 4; 5; 6].

Однако до сих пор данные РМЭЗ по доле курильщиков не подвергались подробной и многосторонней проверке. Чтобы восполнить этот пробел и в случае успеха получить возможность опираться на «эталонную» долю курильщиков при сравнении неоднородных опросных данных, потребуется предпринять четыре шага. Во-первых, необходимо представить значения искомой доли в анализируемом исследовании для того, чтобы можно было оценивать их качество. Во-вторых, следует рассмотреть имеющуюся критику этих показателей. Каким бы совершенным ни было исследование, полностью нивелировать все его недостатки невозможно, поэтому необходимо проанализировать те из них, которые уже были обнаружены ранее. В-третьих, может быть применена одна из наиболее убедительных стратегий контроля достоверности результатов исследования — методическая триангуляция. Сравним распространённость курения по РМЭЗ с аналогичными показателями, получаемыми иными методами в других исследованиях. Но данные, прошедшие два

¹ Российский мониторинг экономического положения и здоровья населения НИУ ВШЭ // НИУ ВШЭ: [сайт]. 20 мая 2022. URL: <http://www.hse.ru/rlms> (дата обращения: 22.04.2024).

² Russia Longitudinal Monitoring Survey of HSE // University of North Carolina at Chapel Hill : [сайт]. 20 мая 2022. URL: <https://rlms-hse.cpc.unc.edu> (дата обращения: 22.04.2024).

испытания — анализ существующей критики и триангуляцию — ещё не могут быть признаны «эталонными» так как распространённость курения, конечно, интересует всех не сама по себе, а в качестве переменной, тесно связанной с общественным здоровьем. Поэтому последним испытанием для данных РМЭЗ должно быть установление их корреляционной связи с показателем, характеризующим объективные последствия курения.

В авторском исследовании использованы данные РМЭЗ с 1994 по 2022 гг. Основной вопрос о курении задавался в рамках мониторинга в так называемом «взрослом» опроснике, который использовался только для респондентов, начиная с неполных 14 лет. Отметим, что темп падения доли курящих с сокращением возраста (см. рис. 1) показывает, что для групп младше 13 лет эта доля может быть принята близкой к нулю. Таким образом, доля курящих во всей выборке индивидов может вполне обоснованно рассматриваться как показатель распространённости курения среди всего населения. Пропущенные значения переменной «Вы курите в настоящее время?», соответствующие детскому опроснику, будут соотноситься со значением «не курю».

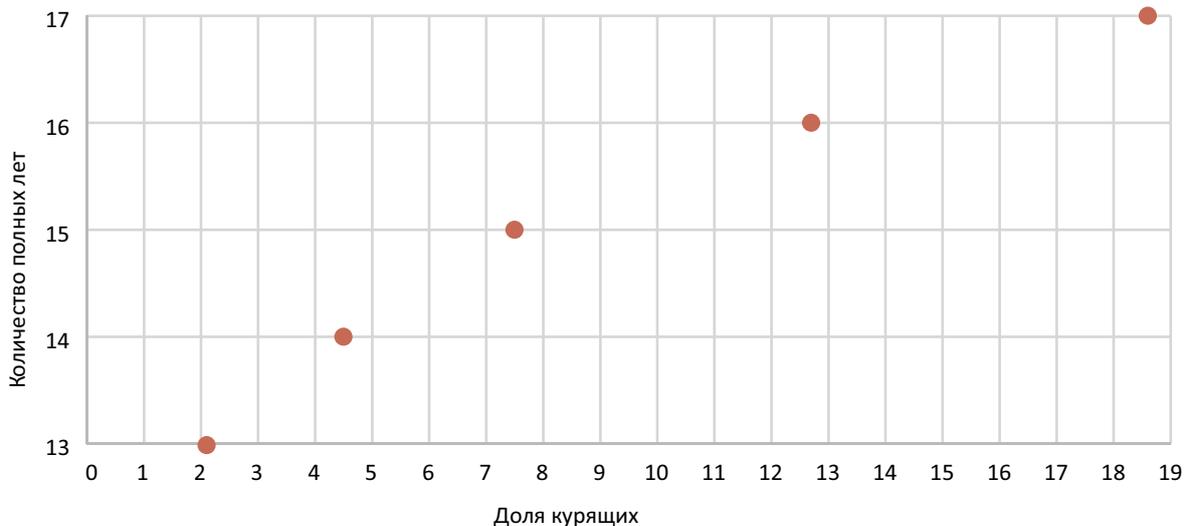


Рис. 1. Распределение доли курильщиков по возрастам 13–17 лет в выборке РМЭЗ (взвешенные данные по репрезентативной выборке индивидов 1994–2022, N=313996), % от давших ответ в возрастной группе

Для сопоставления с данными других организаций приведём данные о распространённости курения для населения в возрасте от 18 лет (см. табл. 1).

Итак, данные, представленные в таблице 1 — это и есть основные показатели курения по РМЭЗ, которые подлежат проверке в качестве потенциального «эталона» достоверных оценок.

Таблица 1

*Распределение доли курильщиков по годам в выборке РМЭЗ
(взвешенные данные по репрезентативной выборке индивидов 1994–2022),
доля опрошенных в соответствующей группе, %*

Годы	Распространённость курения	
	Доля в совокупности всех опрошенных	Доля в совокупности опрошенных, достигших 18 лет
1994	23,9	32,3
1995	24,6	33,3
1996	25,0	34,0
1998	25,4	34,0
2000	26,2	35,0
2001	27,3	36,3
2002	27,7	37,0
2003	27,3	36,2
2004	29,2	35,9
2005	28,8	35,5
2006	29,6	36,4
2007	29,5	36,3
2008	28,7	35,4
2009	27,9	34,7
2010	26,6	33,6
2011	26,8	32,4
2012	27,0	32,7
2013	26,2	31,9
2014	25,5	31,2
2015	24,2	29,7
2016	23,7	29,1
2017	23,2	28,5
2018	22,9	28,0
2019	21,9	27,0
2020	21,1	26,1
2021	21,1	26,0
2022	21,1	26,1

Анализ критики РМЭЗ

Открытость данных РМЭЗ и их частое использование сторонними исследователями служат гарантией постоянной перепроверки всех основных показателей со стороны независимых профессиональных наблюдателей. Поэтому имеет смысл прежде всего обратить внимание на усилия других

исследователей, которые уже осуществили критический анализ данных в ходе своей работы.

Для обнаружения критических замечаний к методике РМЭЗ в области измерения доли курильщиков авторы провели обзорный поиск публикаций на русском и английском языках в базах данных Google Академия, РИНЦ, Web of science и Scopus по полному тексту статей и ключевым словам «курение» и «smoking» в сочетании с «РМЭЗ» и «RLMS». Дополнительный поиск был осуществлён по ключевым словам «курение» и «smoking» в базе данных публикаций РМЭЗ¹. Из более чем тысячи публикаций обнаружены две, содержащие серьёзные критические замечания.

В одной из них В. Школьников с соавторами обнаружили, что из-за изменения выборки между волнами исследования в возрастной группе 25–55 лет наблюдались сильные флуктуации доли курильщиков в 2011–2012 гг. [7, с. 1–11]. Это замечание, хотя и важное для использования данных РМЭЗ применительно к конкретным проблемным годам, носит сугубо локальный характер и не ставит под сомнение общую достоверность временного ряда измерений.

Во второй обнаруженной публикации критика имеет больший размах. Специалисты ЦНИИ организации и информатизации здравоохранения Минздрава России пришли к выводу, что «в опросах домохозяйств, таких как исследования Росстата и РМЭЗ, как правило, систематически занижается распространённость социально нежелательных явлений, поскольку в таких исследованиях опрос членов семьи проводится в присутствии друг друга, в результате чего показатели табакокурения значительно ниже, чем при индивидуальных опросах» [8, с. 58]. К сожалению, в указанной статье нет подробной аргументации этого важного методического вывода. Однако высказанная её авторами гипотеза о влиянии присутствия родственников может быть легко проверена при помощи параданных, доступных в файле микроданных РМЭЗ. Как показано в таблице 2, в подвыборке индивидов, опрошенных в присутствии других членов домохозяйства, доля утвердительных ответов о курении оказывается даже выше, чем при интервью, проведённых тет-а-тет.

Аналогичная картина наблюдается с количеством выкуриваемых сигарет: респонденты, дававшие ответы в присутствии родственников, указывали в среднем 15,76 сигарет в день, а без присутствия родственников — только 15,22 сигареты.

Таким образом, доступные параданные прямо опровергают гипотезу влияния присутствия других членов домохозяйства, якобы существующего в РМЭЗ. Против неё свидетельствуют и данные некоторых других опросов.

¹ Публикации РМЭЗ НИУ ВШЭ (RLMS-HSE) // НИУ ВШЭ : [сайт]. 20 мая 2022. URL: <https://www.hse.ru/rlms/bdp> (дата обращения: 22.04.2024).

Таблица 2

Сопряжённость курения и присутствия других членов домохозяйства во время интервью РМЭЗ (взвешенные данные по репрезентативной выборке индивидов 1994–2022, N=313996), в % от типа ситуации

Факт курения *	Присутствие других членов домохозяйства в момент опроса **	
	Присутствовали	Не присутствовали
Да	25,2	23,7
Нет	74,8	76,3

* Формулировка вопроса: Вы курите в настоящее время?

** Формулировка вопроса: «Присутствовал ли при проведении интервью, хотя бы некоторое время, кто-либо ещё из членов домохозяйства?»

В частности, в 2012–2013 гг. проходило исследование ЭССЕ-РФ, в рамках которого было опрошено более 18 тыс. человек в возрасте 25–64 лет в 12 регионах России. Формулировка вопроса о курении совпадала с РМЭЗ, однако к курящим относили также тех, кто бросил менее 1 года назад (то есть, способ подсчёта давал завышенные относительно РМЭЗ величины). При этом сам опрос проходил в рамках посещения лечебно-профилактических учреждений, так как сопровождался инструментальным обследованием показателей здоровья. Хотя ЭССЕ-РФ не вполне сопоставимо с РМЭЗ по способу концептуализации измеряемой величины, эта погрешность сопоставимости завышает, а не занижает долю курильщиков. Однако в ЭССЕ-РФ на 2012 год доля курильщиков среди населения 25–64 лет составила всего 23,5% [9], что существенно ниже показателей РМЭЗ для всего совершеннолетнего населения (32,7%). А ведь опрос в рамках ЭССЕ-РФ проходил не просто тет-а-тет, но в поликлинических условиях, то есть строго конфиденциально и даже в такой обстановке, которая скорее располагает к описанию вредных привычек — врачу рассказать о них гораздо проще и логичнее, чем случайному постороннему человеку.

Против гипотезы социального давления свидетельствуют и данные индивидуальных опросов ВЦИОМ. Дело в том, что мониторинг общественного мнения, проводимый ВЦИОМ на еженедельной основе, менял формат интервью. До 2017 года это исследование проходило в режиме поквартирного опроса, при котором в значительном количестве случаев могли присутствовать другие члены семьи. По крайней мере, любой член семьи, находившийся дома, мог услышать вопросы и ответы интервью. А начиная с 2017 года ВЦИОМ перешёл на телефонные интервью, в силу очевидных технических особенностей гораздо более конфиденциальные. Если бы присутствие других членов семьи при опросе оказывало существенное влияние на определение доли курильщиков, то в данных ВЦИОМ между 2016 и 2017 гг. должен был бы наблюдаться какой-то ощутимый рост

Таблица 3

Распространённость курения среди совершеннолетнего населения
по данным ВЦИОМ¹

Дата	Ответы респондентов на вопрос о курении *, % от общего числа опрошенных					
	Одну пачку сигарет в день или больше	Несколько сигарет почти каждый день	Иногда несколько сигарет в неделю или в месяц	Бросили курить и не курят уже некоторое время	Никогда не курили	Затруднились ответить
01.05.2011	22	10	2	12	53	1
07.04.2013	22	11	2	11	53	0
06.07.2014	20	12	3	13	52	0
24.05.2015	22	9	3	11	55	0
22.05.2016	17	12	2	12	58	0
22.01.2017	18	11	3	19	48	0
04.03.2018	15	9	3	20	54	0
31.05.2020	17	9	2	19	52	0
29.05.2022	13	10	4	23	49	0

* Формулировка вопроса: «Вы курите или нет? Если да, то как часто?» (в данных может присутствовать погрешность до 1% в результате округления).

этой доли. Однако, как мы видим из таблицы 3, колебания доли курящих между 2016 и 2017 гг. не превышают одного процента. Существенные колебания наблюдаются в пропорции никогда не куривших с бросившими, но эти категории в сумме между годами различаются всего на 3%.

Таким образом, гипотеза давления членов семьи при опросе в домохозяйствах не согласуется с имеющимися данными, и соответствующую критику РМЭЗ следует счесть необоснованной.

Методическая триангуляция

Близость данных, полученных разными способами, может свидетельствовать о незначительном влиянии характерных для каждого из способов погрешностей, а значит — об общей относительно высокой достоверности данных. Начать сравнение целесообразно с наиболее полной альтернативы имеющимся данным РМЭЗ — тоже опросной оценки распространённости курения, но осуществляемой на основе иной методики опроса. Такую альтернативу предоставляет ВЦИОМ. Предельная ошибка выборки регулярных телефонных опросов ВЦИОМ составляет от 2,5% до 3,1%, а квартирных опросов — 3,5%. Ошибка выборки РМЭЗ может быть оценена как заведомо

¹ ВЦИОМ-Навигатор : официальный сайт. Москва, 2024. URL: <https://bd.wciom.ru> (дата обращения: 22.04.2024).

Таблица 4
Сравнение распространённости курения в исследованиях РМЭЗ и ВЦИОМ,
доля от общего числа совершеннолетних респондентов, %

Годы	РМЭЗ	ВЦИОМ	Модуль разности
2011	32,4	34	1,6
2013	31,9	35	3,1
2014	31,2	35	3,8
2015	29,7	34	4,3
2016	29,1	31	1,9
2017	28,5	32	3,5
2018	28,1	27	1,1
2020	26,1	28	1,9
2022	26,1	27	0,9

не превышающая те же величины (в силу большего объёма выборки). Следовательно, консервативная оценка значимых различий между двумя исследованиями может быть определена на уровне 7%. Меньшие по масштабу различия могут рассматриваться как случайные статистические флуктуации.

Как видим, во всех девяти взятых точках наблюдения расхождения между данными РМЭЗ и ВЦИОМ значительно меньше статистической погрешности, причём это относится как к технологии поквартирных опросов, так и к телефонным опросам.

Помимо массовых опросов, распространённость курения может измеряться другими методами, сравнение с которыми представляет особую ценность в триангуляции. Один из них — прямое наблюдение за покупками в магазинах. Хотя частота покупки сигарет необязательно прямо соотносится с распространённостью курения (покупки могут совершать некурящие члены семьи для своих близких, курильщики могут покупать сигареты не в каждый визит в магазин, и т. д.), эти два показателя могут совпадать при соблюдении достаточно простых условий равенства частот покупок для разных категорий товаров и визитов в магазин для различных категорий покупателей. Поэтому близость наблюдаемой частоты покупок с опросными данными служит хотя и косвенным, но подтверждением достоверности последних. Наблюдение за посетителями супермаркетов, проведённое специалистами Nielsen с 10-го по 20-ое июня 2018 года показало, что сигареты в прикассовой зоне покупает каждый третий¹, что не только приблизительно соответствует величине в 28,0%, полученной в тот год для совершеннолетних участников РМЭЗ, но и косвенно

¹ 4 способа повысить продажи в прикассовой зоне // Retail.ru: [сайт]. 1 августа 2018. URL: <https://www.retail.ru/articles/4-sposoba-povysit-prodazhi-v-prikassovoy-zone/> (дата обращения: 22.04.2024).

свидетельствует против более поздних данных той же компании Nielsen (см. Введение к статье), согласно которым курит более половины населения России. Соответствие между наблюдением и показателями РМЭЗ можно ещё усилить, если взять долю домохозяйств, в которых сигареты покупались за неделю: по данным РМЭЗ на 2018 год эта доля составляла 34,5%, то есть, практически ровно треть.

Имеется и возможность сопоставить показатели РМЭЗ с достаточно строго собираемой официальной статистикой. Данные волны 2021 года собирались в четвёртом квартале, по которому доклад Минпромторга России о состоянии рынка табака в 1 квартале 2022 года сообщает, что в государственной информационной системе маркировки товаров была зафиксирована продажа 1,7 млрд пачек сигарет¹. Доля нелегальных сигарет при этом составляла порядка 12%². Таким образом, общий объём продаж в четвёртом квартале 2021 года можно оценить в величину $1,7/0,88=1,93$ млрд пачек. Это точный показатель, с которым можно сопоставить оценки РМЭЗ. По данным анкеты домохозяйств, в четвёртом квартале 2021 года среднее количество купленных пачек составляло 2,5 штуки в неделю. Исходя из того, что в России на 2021 год насчитывалось около 66 млн домохозяйств³, имеем: $2,5*66*12=1980$ млн пачек. Разница с итогами государственного учёта составляет всего 50 млн или 2,6% от действительного количества проданных пачек.

Итак, методическая триангуляция показывает весьма высокую достоверность показателей распространённости курения, измеряемых в рамках РМЭЗ.

Корреляционный анализ

Если в некотором исследовании достаточно хорошо измеряется самооценка курения респондентов, то его результаты должны демонстрировать тесную статистическую связь с объективными проявлениями или последствиями курения — заболеваемостью и смертностью. При этом смертность представляет собой более надёжный показатель, так как гораздо меньше зависит, во-первых, от интенсивности обращения к врачам, во-вторых, от вероятности обнаружения заболевания. Иными словами, смертность — это отражение максимально объективных последствий курения, так как измерению подвергается не просто самочувствие или образ жизни человека, но само его физическое существование.

¹ Обзор состояния розничных рынков продаж табачной продукции в субъектах Российской Федерации за I квартал 2022 года : URL: https://dpt.admin-smolensk.ru/files/491/obzor_tabak-1kv_2022.pdf (дата обращения: 22.04.2024).

² Там же. С. 24.

³ Частные домохозяйства по размеру домохозяйства // Федеральная служба государственной статистики (Росстат) : [сайт]. URL: https://rosstat.gov.ru/storage/mediabank/Tom8_tab2_VPN-2020.xlsx (дата обращения: 22.04.2024).

Также очевидно, что необходимо использовать данные о смертности от заболеваний, тесно связанных с курением. В России по некоторым оценкам более 60% смертности от курения приходится на болезни системы кровообращения [10]. Следовательно, именно эта причина может использоваться для верификации данных о распространённости курения.

Для такого сопоставления был использован корреляционный анализ распространённости курения по РМЭЗ с данными Росстата¹. Хотя число умерших в ней берётся от всего населения в целом, смертность от курения явно должна приходиться на возрастные группы с определённым стажем этой вредной привычки, поэтому в качестве оценки РМЭЗ целесообразнее всего взять долю курящих от совершеннолетнего населения из таблицы 1. Коэффициент Пирсона показывает весьма тесную, более того, близкую к единице корреляцию между смертностью от болезней системы кровообращения и долей курильщиков (см. табл. 5).

Но для наблюдаемой корреляции можно выдвинуть и очевидное возражение. Выявленная взаимосвязь двух временных рядов легко может возникать потому, что оба они связаны не между собой, а, скажем, с течением времени. Альтернативное объяснение может состоять в том, что, благодаря профилактике курения постепенно снижается его распространённость, и, одновременно, благодаря развитию медицины, независимо от показателей курения по РМЭЗ, постепенно снижается смертность. И действительно, как видно из матрицы корреляций, и смертность и доля курящих значимо и достаточно тесно коррелируют с годами, в которые производилось измерение.

Таблица 5
 Корреляция доли курильщиков, смертности и переменной времени
 (коэффициент r Пирсона)

Показатели		Годы	Доля курящих	Смертность
Годы	Корреляция Пирсона	1	-0,781	-0,777
	Значимость (2-сторонняя)		0,000	0,000
	N	27	27	27
Доля курящих	Корреляция Пирсона	-0,781	1	0,926
	Значимость (2-сторонняя)	0,000		0,000
	N	27	27	27
Смертность	Корреляция Пирсона	-0,777	0,926	1
	Значимость (2-сторонняя)	0,000	0,000	
	N	27	27	27

¹ Число умерших по основным классам и отдельным причинам смерти в расчёте на 100000 населения за год // ЕМИСС : [сайт]. URL: <https://www.fedstat.ru/indicator/31270> (дата обращения: 22.04.2024).

Для проверки этого контраргумента рассчитаем частную корреляцию между долей курильщиков по РМЭЗ и смертностью, исключив влияние года наблюдения. Как видно из таблицы 6, корреляция смертности и доли курящих сохранилась на статистически значимом уровне, теснота связи при этом снизилась незначительно.

Таблица 6

Частная корреляция доли курильщиков по РМЭЗ и смертности с исключением переменной года наблюдения

Показатели		Смертность	Доля курящих
Смертность	Корреляция	1,000	0,813
	Значимость (2-сторонняя)	–	0,000
	Число степеней свободы	0	24
Доля курящих	Корреляция	0,813	1,000
	Значимость (2-сторонняя)	0,000	–
	Число степеней свободы	24	0

Если же рассчитать частную корреляцию смертности и даты измерения, исключив влияние доли курящего населения, то теснота связи падает весьма существенно, и корреляция оказывается статистически незначимой (см. табл. 7). Таким образом, вывод о тесной связи данных РМЭЗ и объективной контрольной переменной смертности может быть признан достаточно достоверным.

Таблица 7

Частная корреляция переменных года наблюдения и смертности с исключением доли курильщиков по РМЭЗ

Показатели		Годы	Смертность
Годы	Корреляция	1,000	-0,228
	Значимость (2-сторонняя)	–	0,263
	Число степеней свободы	0	24
Смертность	Корреляция	-0,228	1,000
	Значимость (2-сторонняя)	0,263	–
	Число степеней свободы	24	0

Этот корреляционный анализ имеет свои ограничения. Во-первых, при установлении зависимости по принципу «год к году» получается всего 27 наблюдений, что, хотя и является формально достаточным для получения статистически значимого результата, всё же представляет собой относительно небольшую выборку. Во-вторых, любой корреляционный анализ несёт в себе риск существования «третьей переменной», зависимость от которой объясняет наблюдаемую взаимосвязь, в таком случае, оказывающуюся ложной. Хотя исключено в качестве такой «третьей переменной»

само время, это не единственный из мыслимых вариантов, а перебрать все их не представляется возможным. В-третьих, влияние курения на смертность, конечно, не ограничивается годом измерения, но распространяется на все последующие периоды, постепенно угасая до исчезающе малых величин. Поэтому полученный результат, по мнению авторов, не следует рассматривать с содержательной точки зрения как свидетельствующий о конкретном уровне корреляционной связи между курением и смертностью от определённой причины. В этом смысле величина коэффициента представляется достаточно условной. Однако исследовательским интересом является не столько содержательная интерпретация показателя, сколько именно формальное отражение им отношений между сравниваемыми рядами данных. Для этой цели предпринятый анализ вполне подходит, и он показывает, что оценки распространённости курения по РМЭЗ тесно связаны с распространённостью объективных последствий курения. А это означает, что данные РМЭЗ успешно проходят и третью проверку достоверности.

Заключение

Проведённый анализ показал, что среди всех опросных измерений распространённости курения может быть выделен один «эталонный», который целесообразно рассматривать как наиболее достоверный. Это РМЭЗ (RLMS). Его данные успешно проходят проверку критическими публикациями, сопоставлением с другими методами измерения распространённости курения, в том числе с расчётами через статистику потребления и с непосредственным наблюдением за покупкой сигарет. Кроме того, данные РМЭЗ демонстрируют весьма высокую корреляцию с объективными последствиями курения, такими как смертность от заболеваний системы кровообращения. Всё это означает, что данные РМЭЗ можно рассматривать как в высокой степени достоверные, и, если данные других опросов от него отличаются, это следует расценивать как повод для сомнений в принятых в этих других опросах методических решениях.

СПИСОК ИСТОЧНИКОВ

1. *Сваффорд М. С., Косолапов М. С., Козырева П. М.* Российский мониторинг экономического положения и здоровья населения (РМЭЗ): измерение благосостояния россиян в 90-е годы // Мир России. Социология. Этнология. 1999. Т. 8, № 3. С. 153–172. EDN [KJRTZB](#).
2. *Kozyreva P., Kosolapov M., Popkin B.* Data Resource Profile: The Russia Longitudinal Monitoring Survey — Higher School of Economics (RLMS-HSE). Phase II: Monitoring the Economic and Health Situation in Russia, 1994–2013 // International Journal of Epidemiology. 2016. Vol. 45, № 2. P. 395–401. DOI [10.1093/ije/dyv357](#).

3. *Русинова Н. Л., Озерова О. В., Сафронов В. В.* Курение в России: социальные различия и тенденции в 1990-е и 2000-е гг. // Социологические исследования. 2013. № 3. С. 104–113. EDN **QASAYB**.
4. *Кузнецова П. О.* Курение как фактор сокращения ожидаемой продолжительности жизни в России // Демографическое обозрение. 2019. Т. 6, № 3. С. 31–57. DOI **10.17323/demreview.v6i3.9854**. EDN **BQHGSX**.
5. *Назарова И. Б.* Мониторинг состояния здоровья населения и факторов риска (к методологии изучения здоровья) // Вестник Российского университета дружбы народов. Серия: Социология. 2022. Т. 22, № 3. С. 616–629. DOI **10.22363/2313-2272-2022-22-3-616-629**. EDN **KDWVEN**.
6. Tobacco, Alcohol and Diet as Mortality Risk Factors: the secondary analysis of the 25-year Cohort Study / A. Egorova, B. Idrisov, R. Redman [et al.] // Journal of Health Research. 2023. Vol. 38, № 1. DOI **10.56808/2586-940X.1062**.
7. Time trends in smoking in Russia in the light of recent tobacco control measures: synthesis of evidence from multiple sources / V. M. Shkolnikov, E. Churilova, D. A. Jdanov [et al.] // BMC Public Health. 2020. Vol. 20, № 1. DOI **10.1186/s12889-020-08464-4**. EDN **KTPZMM**.
8. *Зубкова Т. С., Замятина Е. С., Халтурина Д. А.* Система индикаторов поведенческих факторов риска России на национальном и региональном уровнях // Общественное здоровье. 2021. Т. 1, № 4. С. 56–67. DOI **10.21045/2782-1676-2021-1-4-56-67**. EDN **JCXXTR**.
9. Распространённость поведенческих факторов риска сердечно-сосудистых заболеваний в российской популяции по результатам исследования ЭССЕ-РФ / Ю. А. Баланова, А. В. Концевая, С. А. Шальнова [и др.] // Профилактическая медицина. 2014. Т. 17, № 5. С. 42–52. EDN **TQOSLL**.
10. *Масленникова Г. Я., Оганов Р. Г.* Медицинский и социально-экономический ущерб, обусловленный курением табака в Российской Федерации: болезни системы кровообращения // Профилактическая медицина. 2011. Т. 14, № 3. С. 19–27. EDN **PUGUZD**.

Сведения об авторах

Н. С. Бабич

кандидат социологических наук,
старший научный сотрудник
SPIN-код: **8164-5646**

А. А. Михайлов

аспирант
SPIN-код: **9137-2318**

Авторы внесли одинаковый вклад в данную статью.
Авторы заявляют об отсутствии конфликта интересов.

Статья поступила в редакцию 23.04.2024; одобрена после рецензирования 13.08.2024; принята к публикации 27.08.2024.

Original article

DOI: 10.19181/snsp.2024.12.3.9

ANALYSIS OF THE MASS SURVEY DATA RELIABILITY (USING THE EXAMPLE OF SMOKING PREVALENCE IN RUSSIA ESTIMATES)

Nikolay Sergeevich Babich¹
Alexey Aleksandrovich Mikhailov²

¹Institute of Sociology of FCTAS RAS,
Moscow, Russia,

¹*sociolog@mail.ru*,

ORCID 0000-0001-8697-3038

²State Academic University of Humanities,
Moscow, Russia,

²*mix-al.mix@yandex.ru*,

ORCID 0000-0001-8874-0377

For citation: Babich N. S., Mikhailov A. A. Analysis of the mass survey data reliability (using the example of smoking prevalence in Russia estimates). *Sociologicheskaja nauka i social'naja praktika*. 2024;12(3):198–213. (In Russ.). DOI 10.19181/snsp.2024.12.3.9.

Abstract. Survey-based estimates of smoking prevalence in Russia, given by various quite respectable organizations for the same period, range from less than 20% to more than 50%. This scatter cannot be explained by differences in the conceptualization of the indicator (who is considered a smoker) and significantly affects the overall picture of smoking in Russia and the assessment of the success of anti-tobacco policy. This raises questions about the reliability of the relevant survey data. It can be solved in at least two ways: by analyzing and eliminating discrepancy factors or by choosing one of the studies as a reference with a thorough check of the reliability of its data. The article substantiates the choice of the second path and the Russia Longitudinal Monitoring Survey (RLMS) as a reference. To check its reliability, three criteria were used: analysis of critical publications about research errors, methodological triangulation (comparison with data from other methods) and analysis of correlation with the consequences of smoking. All three criteria showed a high level of reliability of the RLMS data. The hypothesis about the research error put forward in critical publications does not find empirical confirmation. Comparison with other supposedly reliable data (both survey and non-survey) shows discrepancies below the margin of statistical error. The correlation of RLMS data on smoking with mortality from diseases of the circulatory system turns out to be extremely high. Therefore, it is advisable to consider the RLMS as a reference source of data on the prevalence of smoking in Russia, and indicators from other surveys that differ significantly from it as potentially questionable in their methodological decisions.

Keywords: smoking prevalence, social surveys, reliability of survey data, methodological triangulation, RLMS

REFERENCES

1. Swafford M. S., Kosolapov M. S., Kozyreva P. M. Russian longitudinal monitoring survey (RLMS): measuring the well-being of Russians in the 90s. *Universe of Russia. Sociology. Ethnology=Mir Rossii. Sociologiya. Etnologiya*. 1999;8(3):53–172. (In Russ.).

2. Kozyreva P., Kosolapov M., Popkin B. Data resource profile: The Russia longitudinal monitoring survey — Higher School of Economics (RLMS-HSE) Phase II: Monitoring the economic and health situation in Russia, 1994–2013. *International Journal of Epidemiology*. 2016;45(2):395–401. DOI [10.1093/ije/dyv357](https://doi.org/10.1093/ije/dyv357).
3. Rusinova N. L., Ozerova O. V., Safronov V. V. Smoking in Russia. Social differences and trends in 1990-th to 2000-th. *Sotsiologicheskie Issledovaniia=Sociological Studies*. 2013;(3):104–113. (In Russ.).
4. Kuznetsova P. O. Smoking as a factor of reduced life expectancy in Russia. *Demograficheskoe obozrenie=Demographic Review*. 2019;6(3):31–57. (In Russ.). DOI [10.17323/demreview.v6i3.9854](https://doi.org/10.17323/demreview.v6i3.9854).
5. Nazarova I. B. Monitoring of the population health and health risk factors (research methodology). *RUDN Journal of Sociology=Vestnik Rossijskogo universiteta družby` narodov. Seriya: Sociologiya*. 2022;22(3):616–629. (In Russ.). DOI [10.22363/2313-2272-2022-22-3-616-629](https://doi.org/10.22363/2313-2272-2022-22-3-616-629).
6. Egorova A., Idrisov B., Redman R. [et al.]. Tobacco, Alcohol and Diet as Mortality Risk Factors: the secondary analysis of the 25-year Cohort Study. *Journal of Health Research*. 2023;38(1). DOI [10.56808/2586-940X.1062](https://doi.org/10.56808/2586-940X.1062).
7. Shkolnikov V. M., Churilova E., Jdanov D. A. [et al.] Time trends in smoking in Russia in the light of recent tobacco control measures: synthesis of evidence from multiple sources. *BMC Public Health*. 2020;20(1). DOI [10.1186/s12889-020-08464-4](https://doi.org/10.1186/s12889-020-08464-4).
8. Zubkova T. S., Zamiatnina E. S., Khalturina D. A. The system of indicators of behavioral risk factors in Russia at the national and regional levels. *Public health=Obshchestvennoe zdorov'e*. 2021;1(4):56–67. (In Russ.). DOI [10.21045/2782-1676-2021-1-4-56-67](https://doi.org/10.21045/2782-1676-2021-1-4-56-67).
9. Balanova Iu. A., Kontsevaia A. V., Shal'nova S. A. [et al.] Prevalence of behavioral risk factors for cardiovascular disease in the Russian population: results of the ESSE-RF epidemiological study. *Profilakticheskaya Meditsina=The Russian Journal of Preventive Medicine*. 2014;17(5):42–52. (In Russ.).
10. Maslennikova G. Ia., Oganov R. G. Medical and socioeconomic damage caused by smoking in the Russian Federation: diseases of circulatory system. *Profilakticheskaya Meditsina=The Russian Journal of Preventive Medicine*. 2011;14(3):19–27. (In Russ.).

Information about the Authors

N. S. Babich

Candidate of Sociology,

Senior Researcher

ResearcherID: [I-5630-2016](https://orcid.org/0009-0001-5630-2016)

Scopus AuthorID: [57201994829](https://orcid.org/0009-0001-5630-2016)

A. A. Mikhailov

Postgraduate student

The authors contributed equally to this article.

The authors declare no conflicts of interests.

The article was submitted 23.04.2024; approved after reviewing 13.08.2024; accepted for publication 27.08.2024.