

**ИСПОЛЬЗОВАНИЕ КЛАСТЕРНОГО АНАЛИЗА  
ДЛЯ ВЫЯВЛЕНИЯ ОКРУГОВ ПО УРОВНЮ ЗДОРОВЬЯ  
И СОЦИАЛЬНЫМ ПОТРЕБНОСТЯМ<sup>1</sup>**

**Беллис М.А. (а), Джарман Я. (б), Даунинг Дж. (а), Перкинс К. (а), Бейнон К. (а), Хафс К. (а), Лисбоа П. (б)**

(а). Центр общественного здоровья, Ливерпульский университет им. Джона Мурса, Великобритания

(б). Центр компьютерных и математических наук, Ливерпульский университет им. Джона Мурса, Великобритания

*Разработка программ, содействующих улучшению здоровья населения, требует понимания не только взаимосвязи между различными показателями здоровья, но и того, как они соотносятся с более широкими вопросами образования, уголовным судопроизводством и другими социально значимыми проблемами. Для того чтобы способствовать информационному обеспечению политики, необходимы методы, позволяющие лучше использовать междисциплинарные данные. Применив кластерный анализ к 30 переменным здоровья и социального благополучия, мы выделили 5 типов административных единиц, причем плохие показатели по большинству переменных оказались сконцентрированы в одном кластере. Основные различия кластеров проявлялись в значениях следующих переменных: детская бедность, кормление грудью, детский кариес, подростковая беременность, здоровое питание, психические заболевания и смертность от туберкулеза и курения. В кластер с наихудшими показателями в 63% попала территория Северной Англии, население которой в генеральной совокупности составляет 15,7% (n=324). Концентрация проблем в определенных областях создает неравномерную нагрузку, которая может превосходить кумулятивный эффект индивидуальных проблем. Такие различия кластеров по здоровью поднимают вопрос о возможности применения одних и тех же эффективных социальных политик в регионах, относящихся к разным кластерам.*

**ИСТОРИЯ ВОПРОСА**

Различие в уровнях состояния здоровья населения между странами и внутри стран является не только результатом бедности и сильной дифференциации по степени доступа к медицинским услугам. Это также связано со структурным неравенством, влияющим на условия, в которых люди живут и работают [Murray et al., 2007; Commission on Social Determinants of Health, 2008; Department of Health, 2009; Marmot, 2005, 2010]. Развитые страны обладают более богатым опытом борьбы с плохим состоянием здоровья и больше стремятся к повышению продолжительности жизни, чем развивающиеся, которым тяжело справиться с различными факторами, такими как социальная исключенность, плохое обеспечение питанием, неразвитость транспорта

<sup>1</sup> Перевод статьи *Bellis M. A., et al. Using clustering techniques to identify localities with multiple health and social needs. Health & Place (2011), doi:10.1016/j.healthplace.2011.08.003.* Научный редактор перевода — Синявская О. В., к.э.н., заместитель директора НИСП.

и отсутствие политики в области занятости [Wilkinson и Marmot, 2003]. Таким образом, состояние здоровья населения должно волновать лиц, принимающих политические решения в различных областях, а не только тех, кто работает в области здравоохранения [Marmot, 2005]. В дополнение к данным, описывающим плохое состояние здоровья, специалисты, работающие в сфере здравоохранения и политики, в настоящее время часто используют данные о поведении населения, состоянии окружающей среды, экономической ситуации, образовании и даже данные правоохранительных органов [Stachenko, 2008; Bellis et al., 2005; Rutherford et al., 2007]. Исторически сложилось так, что для определения мер по предотвращению таких явлений как ожирение, психические отклонения, беременность в подростковом возрасте, уровень преступности и снижение продолжительности жизни, эти данные изучаются по отдельности. Однако более поздние работы рассматривают их взаимосвязь [Felitti et al., 1998; Harden et al., 2009; Hillis et al., 2010; Duarte et al., 2010; Weiler et al., 2010; Barton and Pretty, 2010]. Именно сочетание этих факторов влияет на благосостояние сообществ, порождая различные медицинские и социальные проблемы и препятствуя эффективности возможных вмешательств (например, меры, побуждающие жителей к частым прогулкам и другим упражнениям на свежем воздухе, в районах с высоким уровнем преступности и насилия работать не будут) [Bellis et al., 2008].

В то время как необходимость основывать политические изменения на результатах научных исследований не вызывает сомнений, растущее количество социальных индикаторов и показателей здоровья требует интегрированного подхода к их анализу и интерпретации, привносящего скорее понимание, нежели дополнительную сложность в локальные стратегии в области здравоохранения. Взаимосвязи между различными показателями должны выявлять связи между целями различных стейкхолдеров и тем самым информационно содействовать совместной работе. Кроме того, можно выделить различные районы в зависимости от профиля здоровья, одним из которых необходимо внешнее вмешательство для устранения проблем, а другие могут справиться сами, перенимая определенные стратегии и активно взаимодействуя между собой. Примерами таких стратегий на международном уровне являются: использование межсекторальных данных для определения наиболее отсталых стран [Klugman, 2010]; использование метода кластерного анализа в субнациональных регионах Европейского Союза [I2SARE, 2010] для выявления районов со схожими уровнями состояния здоровья; кластеризация малочисленных популяций внутри страны в преимущественно социально-экономических целях [Office for National Statistics, 2001; Vickers and Rees, 2007]. Подобный метод использовался для сравнения и выявления связей между аналогичными районами в рамках судебной реформы [Association of Police Authorities, 2006].

В 2005 г. Ассоциация исследований здравоохранения (по заказу Министерства здравоохранения Англии) подготовила набор показателей здоровья в нескольких районах Англии [Association of Public Health Observatories, 2006]. Этому предшествовал процесс систематизации, направленный на выявление как значимых баз данных, относящихся к здравоохранению и здоровью населения, так и такой информации, которая может быть собрана с надлежащим качеством и регулярностью на местном административном уровне. Доступные в форме единичных профилей здоровья для каждой административной

единицы ( $n=324$ ), последние (на 2010 г.) результаты этой работы включают 31 переменную, каждая из которых рассчитана в соответствии с методиками надлежащего качества, находящимися в открытом доступе и прошедшими экспертизу. Показатели включают информацию о рисках для здоровья и мерах по его поддержанию, заболеваемости и смертности, об образовании, экономике и окружающей среде [Department of Health and the Association of Public Health Observatories, 2010]. Стандартные методы кластерного анализа являются объективной методологией для комбинирования показателей здравоохранения с целью лучшей идентификации отдельных групп районов с похожими профилями здоровья населения и для установления связей между трендами каждого показателя, которые могут быть визуализированы по когортам [Friedman and Rubin, 1967; Lisboa et al., 2008]. В работе авторами применяется уже известный метод К-средних<sup>2</sup> на системе показателей здоровья, рассчитанных для профилей административных единиц 2010 г., с тем, чтобы выявить, насколько по-разному ряд факторов влияет на жизнь людей в различных районах Англии; определяется, как распределение по кластерам соотносится с каждым показателем здоровья и с независимо созданными переменными неравенства (располагаемый доход и депривация); наконец, обсуждаются преимущества использования такого подхода в разработке национальной политики и межведомственной деятельности на местном уровне.

## МЕТОДЫ

Вся информация о подсчете, выборке и качестве моделей, используемых для создания профилей здоровья, подробно изложена в других источниках [Department of Health and the Association of Public Health Observatories, 2010]. Для создания профилей из 31 переменной здоровья не использовался только индекс множественной депривации, так как это составной показатель, который содержит внутри себя индексы и показатели других профилей. Все значения переменных профилей здоровья были проанализированы на предмет пропущенных значений и пригодности для проведения кластерного анализа. Во всем массиве было обнаружено пять переменных с пропущенными данными: о курении во время беременности [8,6%], о кормлении грудью [6,8%], о наличии постоянного места проживания [6,5%], о болезнях зубов у детей [4,3%] и о случаях заболевания туберкулезом [1,2%]. В каждую пропущенную ячейку были добавлены медианные значения.

Данные по каждому показателю были перенесены на одну шкалу (от 0 до 100), так чтобы ни один индикатор не доминировал в процессе кластеризации. Затем стандартизованные значения анализировались методом К-средних. На первом этапе кластеризации случайным образом генерируется небольшое количество кластерных центров с определенными значениями (т.е. координатами) для каждой используемой переменной из данных профилей здоровья.

<sup>2</sup> Кластеризация методом К-средних предполагает распределение  $N$  наблюдений по  $K$  кластерам (число  $K$  задается в начале процедуры) таким образом, чтобы каждое наблюдение попало в кластер с ближайшей средней. Данный метод часто применяется, когда исследователь уже имеет некоторые гипотезы относительно числа кластеров и хочет получить кластеры, максимально отличные друг от друга. Счетные процедуры, используемые в данном методе, позволяют применять его в том числе при большом количестве наблюдений. *Примеч. науч. ред.*

Затем все административные единицы были отнесены к ближайшему кластерному центру, который, в свою очередь, был переопределен как среднее значение для каждой переменной профиля здоровья в административных единицах, попавших в этот кластер. Повторная кластеризация проводилась до тех пор, пока все точки не были распределены вокруг обновленного кластерного центра, а затем значения центральной точки снова пересчитывались. Процесс прекращается, когда достигается точка, в которой никаких изменений больше не происходит. Путем аналогичного повторения всего процесса, начиная с широкого диапазона различных переменных и различных заданных значений кластерных центров, были определены наиболее устойчивые (робастные) кластерные модели, характеризующиеся наибольшей дифференциацией между кластерами [Vassiu et al., 2009]. Для кластеризации использовались программы SPSS (версия 17.0.0, SPSS, 2008) и MATLAB (версия 7.8.0 R2009a; MATLAB, 2009) 64bit.

Различия в показателях были измерены при помощи расчета медианных тестов для всех кластеров. Для визуализации модели данные представлены в виде диаграммы рассеяния. Здесь оси X и Y (обозначенные как «главные разделяющие оси») представляют собой комбинацию перешкалированных переменных профилей здоровья, использованных в процессе кластеризации. Данная система координат сопоставима с методом главных компонент [Jolliffe, 2002]. Однако здесь главные разделяющие оси отбираются по принципу максимального различия между кластерами, с целью демонстрации их отличия в двумерном пространстве [Lisboa et al., 2008]. Для того, чтобы определить стабильность принадлежности административных единиц к кластерам (анализ чувствительности), был проведен повторный кластерный анализ на подвыборке переменных профиля здоровья и проанализирован вклад каждой переменной профиля здоровья в каждую главную разделяющую ось и, соответственно, в разделение кластеров.

Индекс множественной депривации впоследствии был использован в анализе после завершения кластеризации с тем, чтобы выявить связь этого показателя с критериями разделения на кластеры. Для этих же целей использовался валовый располагаемый доход домохозяйства. Информация о валовом располагаемом доходе домохозяйства представлена в «Номенклатуре территориальных единиц» (NUTS) по Англии для уровня 3, определенного как округа (графства) или группы округов унитарной системы местного самоуправления; n=92 [Office for National Statistics, 2009]. Медианный валовый располагаемый доход домохозяйства для региона уровня 3 (NUTS3) был использован для каждого муниципалитета, входящего в этот регион. Индекс множественной депривации и валовый располагаемый доход домохозяйства были выбраны для анализа кластеров в качестве устойчивых сводного и дискретного показателей неравенства соответственно. Данные также были представлены графически с использованием границ административных единиц и регионов, а также в виде шестиугольных графиков, площадь которых соотносится с численностью той или иной административной единицы. Например, Бирмингем был представлен 11 шестиугольниками, в то время как Рутланд, менее крупный район, был репрезентирован половиной шестиугольника [Map 1a, b; Thomas and Dorling, 2004]. Для каждого муниципалитета данные о численности населения и национальном составе были определены в соответствии с данными

Управления национальной статистики [Office for National Statistics, 2008; Office for National Statistics, 2011].

## РЕЗУЛЬТАТЫ

На *рисунке 1* представлена двухмерная визуализация кластеров (C1–C5), где каждая административная единица нанесена на прямоугольную систему координат согласно значениям двух составных показателей профилей здоровья, рассчитанных таким образом, чтобы показать максимальное отличие кластеров друг от друга (см. раздел «Методы»). На этой диаграмме рассеяния все кластеры разделены по первой главной разделяющей оси (PSA1). Кластеры с 1 по 5 представляют 11,3, 12,3, 11,5, 5,3 и 10,7 млн чел. соответственно (на середину 2008 г.) [Office for National Statistics, 2008].

Корреляция всех измерений с PSA1 (*рис. 1*) представлена в *таблице*. Переменные, отражающие данные по смертности от курения, по беременности в подростковом возрасте, ожидаемой продолжительности жизни для мужчин и женщин, по ранней смертности от болезней сердца и инсульта, по выплате пособий в связи с потерей трудоспособности из-за психических заболеваний, по уровню детской бедности, преждевременной смерти от рака и по часто курящим взрослым, сильно коррелируют с осью PSA ( $R > 0,8$  или  $R_0 = -0,8$ ). Каждое значение для административной единицы на оси PSA1 также коррелирует с показателями неравенства, с индексом множественной депривации ( $R = -0,85$ ,  $R < 0,001$ ) и с валовым располагаемым доходом домашних хозяйств ( $R = 0,65$ ;  $P < 0,001$ ). Ось PSA2 главным образом отделяет кластер C4, в котором 95,5% (21 из 22) административных единиц базируется в Лондоне (*рис. 2а*).

Рассмотрение профилей здоровья в координатах PSA1 и PSA2 позволило выделить следующие переменные, вносящие основной вклад в разделение осей (т.е. переменные профиля здоровья с наибольшей нагрузкой в абсолютных значениях по оси PSA1 или PSA2): уровень детской бедности, грудное вскармливание, детский кариес, подростковая беременность, правильно питающиеся взрослые, психические заболевания, новые случаи заболеваемости туберкулезом и смертность от курения. При повторении кластеризации для данного набора переменных только 12,35% территорий меняют принадлежность к кластеру. Исключение возрастающей доли из подвыборки указанных переменных постепенно изменяет результаты кластеризации.

В *таблице* представлены медианы и квартильные ранги, приписанные к каждой переменной в кластере и отсортированные по убыванию — от наилучших к наихудшим показателям благополучия. Каждый цвет в таблице означает кластер; наилучшие значения переменных 1–12 наблюдаются в 1-м кластере, в то время как наилучшее значение 13-й переменной (здоровое питание) находится в 4-м кластере и т.п. Уровень значимости различий медианных значений всех кластеров высокий для всех переменных ( $P < 0,001$ ), за исключением физической активности детей ( $P = 0,078$ ). Большинство показателей имеет наилучшие значения в первом кластере (наиболее богатом) и наихудшие — в пятом кластере (наиболее депривированном). Так, в 5-м кластере были самые низкие средние значения по 20 переменным, а в 1-м — лучшие значения по 19 переменным. Четвертый кластер (представленный в основном Лондоном) является своего рода аномалией, поскольку, несмотря

**Таблица 1. Значения кластерных центров (медиан) по переменным, отсортированным от наилучших к наихудшим значениям, и значения коэффициентов корреляции с осью PSA1.**

№	Название переменной	R (PSA1)	Наилучшее значение	>	Медианное значение	>	Наихудшее значение
1	Количество смертей от курения	-0,925	151,9 (140,5-163,8)	182,9 (169,5-193,1)	219,1 (185,6-257,2)	227,0 (216,1-238,3)	281,3 (260,4-301,0)
2	Подростковая беременность (до 18 лет)	-0,921	23,8 (21,0-28,0)	35,1 (33,3-38,5)	46,3 (41,5-50,4)	48,0 (40,3-59,2)	53,9 (49,0-58,4)
3	Ожидаемая продолжительность жизни (мужчин)	0,908	79,9 (79,3-80,4)	78,8 (78,4-79,3)	77,5 (75,9-78,7)	77,2 (76,8-77,7)	75,7 (75,3-76,2)
4	Преждевременная смертность (болезни сердца, инсульты)	-0,889	55,9 (51,3-59,8)	65,4 (59,8-70,4)	79,7 (75,3-85,9)	90,1 (82,0-100,1)	97,0 (89,7-105,9)
5	Выплаты по инвалидности психически больным	-0,876	15,2 (13,4-17,6)	21,9 (17,9-25,6)	28,9 (25,1-35,7)	29,5 (24,1-33,0)	40,7 (34,6-47,7)
6	Детская бедность	-0,866	10,8 (8,9-12,6)	15,7 (13,3-18,1)	21,9 (19,2-24,5)	29,6 (24,7-33,3)	38,7 (31,3-44,9)
7	Преждевременная смертность (онкология)	-0,857	98,5 (94,6-103,4)	108,1 (104,1-112,9)	113,1 (106,1-127,4)	119,7 (115,7-123,4)	134,5 (127,4-143,8)
8	Количество курящих (взрослых)	-0,845	16,7 (15,3-18,4)	20,5 (18,8-22,1)	21,8 (18,6-27,0)	24,4 (22,4-26,5)	28,0 (26,0-29,9)
9	Ожидаемая продолжительность жизни (женщин)	0,838	83,5 (83,1-83,9)	82,5 (82,0-83,1)	82,3 (81,2-83,1)	81,5 (81,0-82,0)	80,1 (79,6-80,6)
10	Число дней пребывания в больнице по причинам, связанным с употреблением алкоголя	-0,791	1167,2 (1004,8-1334,2)	1339,9 (1218,0-1523,3)	1596,5 (1466,2-1852,9)	1766,1 (1444,4-1897,0)	2113,7 (1904,3-2388,3)
11	Преступления, связанные с применением насилия	-0,687	9,2 (7,6-12,2)	12,8 (10,6-15,2)	17,1 (14,9-21,7)	18,1 (14,8-23,0)	27,1 (23,9-30,7)
12	Результаты выпускных экзаменов в средней школе <sup>3</sup>	0,671	58,3 (54,9-62,6)	52,0 (46,8-54,6)	50,3 (47,7-55,3)	46,4 (43,0-51,7)	44,9 (41,0-47,9)
13	Здоровое питание у взрослых	0,670	36,9 (33,6-39,3)	32,0 (29,7-33,6)	29,6 (27,9-31,6)	25,4 (23,7-27,8)	22,5 (21,1-24,5)
14	Детский кариес (у детей до 5 лет)	-0,659	0,7 (0,5-0,9)	0,9 (0,7-1,1)	1,1 (0,8-1,4)	1,2 (1,0-1,8)	1,7 (1,4-1,9)
15	Младенческая смертность	-0,623	3,3 (2,4-4,2)	4,4 (3,5-4,9)	5,0 (4,2-5,7)	5,1 (4,0-5,8)	5,8 (5,0-6,4)
16	Дети с ожирением	-0,617	7,6 (6,8-8,5)	9,1 (8,3-9,8)	9,5 (8,8-10,6)	10,2 (9,2-11,5)	12,1 (11,3-13,3)

<sup>3</sup> Имеются в виду результаты тестирования по различным предметам на получение Общего сертификата среднего образования (General Certificate of Secondary Education — GCSE), которое традиционно проводится среди школьников 14-16 лет в Англии. Результаты тестирования не ниже определенного уровня — часто необходимое условие для продолжения образования. Например, ведущие университеты Великобритании ожидают от поступающих оценок на уровне A или A\* по большинству сданных ими предметов. См.: [http://en.wikipedia.org/wiki/General\\_Certificate\\_of\\_Secondary\\_Education](http://en.wikipedia.org/wiki/General_Certificate_of_Secondary_Education). *Примеч. науч. ред.*

Окончание таблицы 1

№	Название переменной	R (PSA1)	Наилучшее значение	>	Медианное значение	>	Наихудшее значение
17	Люди, у которых диагностирован диабет	-0,608	3,7 (3,5-4,0)	4,2 (4,0-4,6)	4,4 (4,1-4,7)	4,5 (3,9-5,1)	4,7 (4,3-5,0)
18	Физически активные взрослые	0,564	13,0 (11,7-14,2)	11,2 (10,3-12,4)	10,5 (9,5-11,8)	10,3 (9,4-11,0)	9,2 (7,4-11,1)
19	Взрослые с ожирением	-0,523	21,2 (18,4-23,4)	22,5 (20,5-23,5)	24,8 (23,2-26,1)	25,4 (23,6-27,3)	26,8 (25,1-28,4)
20	Грудное вскармливание	0,475	88,0 (83,8-90,6)	76,1 (73,3-79,9)	73,4 (71,7-76,5)	71,9 (66,3-74,3)	59,7 (54,1-67,3)
21	Случаи заболевания злокачественной меланомой	0,416	5,5 (4,4-7,4)	10,3 (8,6-11,8)	12,4 (10,7-14,2)	13,6 (10,8-15,7)	14,8 (11,7-18,4)
22	Новые случаи заболеваний туберкулезом	-0,414	3,5 (2,5-6,0)	4,8 (2,7-8,3)	6,5 (4,1-12,1)	9,8 (4,0-23,9)	50,8 (38,3-62,2)
23	Курение во время беременности	-0,408	6,9 (5,1-8,9)	13,3 (8,8-5,1)	15,0 (13,1-17,5)	15,4 (14,0-18,3)	20,8 (18,3-24,5)
24	Травмы и смертельные случаи в результате дорожно-транспортных происшествий	0,391	41,8 (36,6-51,7)	43,7 (35,2-56,6)	47,6 (42,5-58,3)	52,7 (40,9-60,9)	60,3 (48,1-75,3)
25	Узаконенная бездомность	-0,376	1,3 (0,6-1,7)	1,7 (1,1-2,2)	2,0 (1,4-3,0)	2,0 (1,0-3,8)	3,9 (2,7-4,7)
26	Пьянство среди взрослых	-0,327	14,1 (10,5-19,6)	18,3 (16,5-20,8)	18,7 (16,0-21,5)	20,9 (18,8-23,0)	24,4 (20,4-27,3)
27	Случаи перелома шейки бедра после 65 лет	-0,264	438,5 (409,0-476,1)	449,8 (415,0-487,7)	462,9 (427,0-514,4)	487,5 (460,2-526,1)	523,8 (489,4-549,4)
28	Выбросы углекислого газа	0,246	5,4 (4,7-6,3)	6,1 (5,6-7,5)	6,6 (6,1-7,4)	6,9 (6,0-8,1)	7,3 (6,2-8,3)
29	Физически активные дети	0,157	51,30 (46,2-56,3)	51,25 (46,6-56,9)	48,9 (43,9-55,4)	48,5 (43,7-55,8)	48,4 (45,0-52,8)
30	Повышенная смертность в зимнее время года	0,147	13,7 (11,9-17,6)	14,2 (11,6-16,7)	15,5 (12,4-18,3)	16,3 (12,7-18,9)	16,3 (13,2-19,3)

Кластер 1   
  Кластер 2   
  Кластер 3   
  Кластер 4   
  Кластер 5

Примечания:

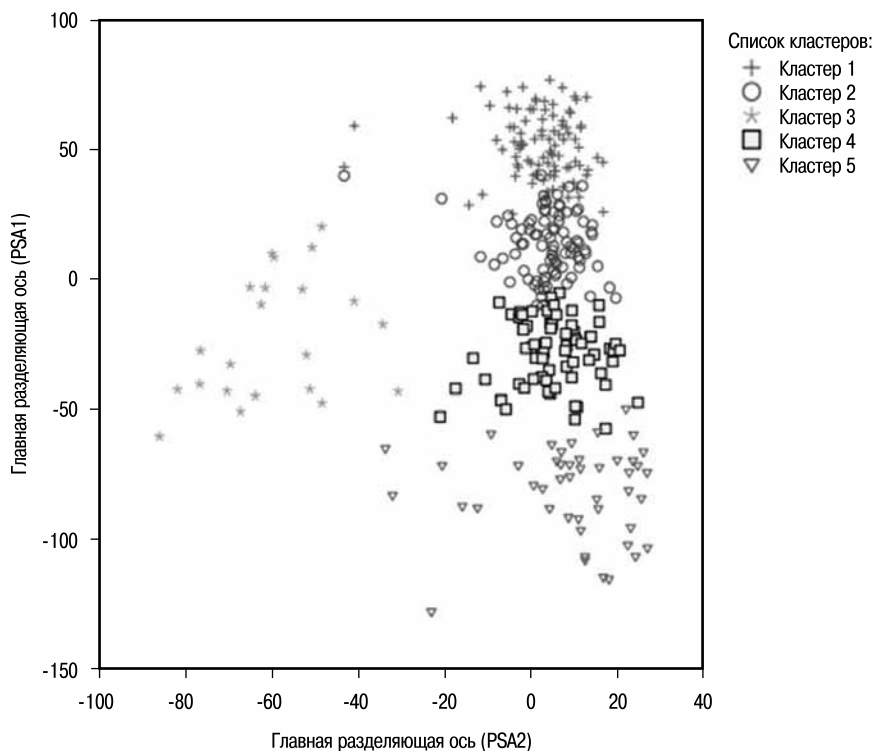
1. Смертельные исходы, связанные с курением среди населения от 35 лет и старше; стандартизация по возрасту; 2006–2008 гг.
2. Число случаев беременности в возрасте до 18 лет на 1 тыс. женщин 15–17 лет; 2006–2008 гг.
3. Ожидаемая продолжительность жизни при рождении для мужчин; число лет; все возрасты; 2006–2008 гг.
4. Смертельные исходы, связанные со всеми болезнями системы кровообращения на 100 тыс. чел. населения до 75 лет; стандартизация по возрасту; 2006–2008 гг.
5. Получающие пособия (пенсии) по потере трудоспособности из-за психических заболеваний на 1 тыс. человек населения трудоспособного возраста; 2008 г.
6. Процент детей, живущих в семьях, получающих пособия по бедности (один из «Индексов депривации 2007»).
7. Количество смертей от раковых заболеваний всех видов на 100 тыс. чел. населения в возрасте до 75 лет; стандартизация по возрасту; 2006–2008 гг.

8. Процент курения среди взрослых; сгенерированные оценки по данным «Медицинского обследования населения Англии»; 2006–2008 гг.
9. Ожидаемая продолжительность жизни при рождении для женщин; число лет; все возрасты; 2006–2008 гг.
10. Численность людей, попавших в больницу из-за употребления алкоголя на 100 тыс. чел. населения; стандартизация по полу и возрасту; 2008–2009 гг.
11. Зарегистрированные случаи насилия и преступлений против личности на 1 тыс. чел. населения, 2008–2009 гг.
12. Число учащихся, сдавших 5 и более экзаменов на оценки А, В или С по окончании первой ступени среднего образования (GCSE), в том числе по математике и английскому языку; в процентах на момент окончания продвинутого уровня (два последних года обучения, включая время экзаменов), по данным Местного управления образования; 2008–2009 гг.
13. Доля взрослых людей, которые употребляют здоровую пищу; сгенерированные оценки по данным «Медицинского обследования населения Англии»; 2006–2008 гг.
14. Среднее количество поврежденных кариесом или выпавших из-за кариеса зубов у детей в возрасте 5 лет; 2007–2008 гг.
15. Число умерших младенцев на 1 тыс. живорожденных; 2006–2008 гг.
16. Процент школьников, страдающих ожирением; на момент приема в 2008–2009 гг.
17. Процент людей с диагнозом «диабет», учтенных в реестре медицинских работников; 2008–2009 гг.
18. Доля физически активного населения в возрасте 16 лет и старше; 2006–2008 гг.
19. Процент взрослого населения, страдающего ожирением; сгенерированные оценки, по данным «Медицинского обследования населения Англии»; 2007–2008 гг.
20. Доля матерей, кормящих грудью, где это было возможно выявить; 2007–2008 гг.
21. Стандартизованный коэффициент заболеваемости меланомой на 100 тыс. чел. населения в возрасте до 75 лет; 2004–2006 гг.
22. Число новых случаев заболевания туберкулезом на 100 тыс. человек населения; 2006–2008 гг.
23. Число матерей, куривших во время беременности, где это было возможно выявить; 2008–2009 гг.
24. Число людей, погибших или сильно пострадавших в автокатастрофах на 100 тыс. чел. населения; 2006–2008 гг.
25. Официальное число индивидуумов и семей без определенного места жительства (бездомных) на 1 тыс. домохозяйств; 2008–2009 гг.
26. Процент пьющих среди взрослого населения; сгенерированные оценки по данным «Медицинского обследования населения Англии»; 2007–2008 гг.
27. Число случаев перелома тазобедренного сустава на 100 тыс. чел., попавших в больницу; стандартизация по возрасту; 2008–2009 гг.
28. Доля выбросов углекислого газа на душу населения; 2007 г.
29. Процент учащихся 1–13 классов, которые, по меньшей мере, 3 часа в неделю занимаются физкультурой; 2008–2009 гг.
30. Отношение числа смертей в зимнее время года (наблюдаемая смертность в зимнее время года минус предполагаемая смертность, не учитывающая смертность в зимнее время года) к предполагаемой общей смертности за период 01.08.2005 г.— 31.07.2008 г. R (PSA1) — коэффициент ранговой корреляции Спирмена по каждому замеру с осью PSA 1.

на показатели депривации выше среднего, он демонстрирует лучшие медианные значения для девяти показателей здоровья населения (кормление грудью, ожирение среди взрослых и смертность в зимнее время года) и худшие — для пяти показателей (включая, например, преступления, связанные с применением насилия, и уровень детской бедности). Четвертый кластер также отличается разнообразным этническим составом (процент белых людей с 1-го по 5-й кластер — 92,6%, 93,0%, 91,3%, 64,5% и 86,6% соответственно,



**Рисунок 1. Двумерная визуализация кластеров профилей здоровья.  
Распределение профилей здоровья  
в координатах главных разделяющих осей PSA1 и PSA2**



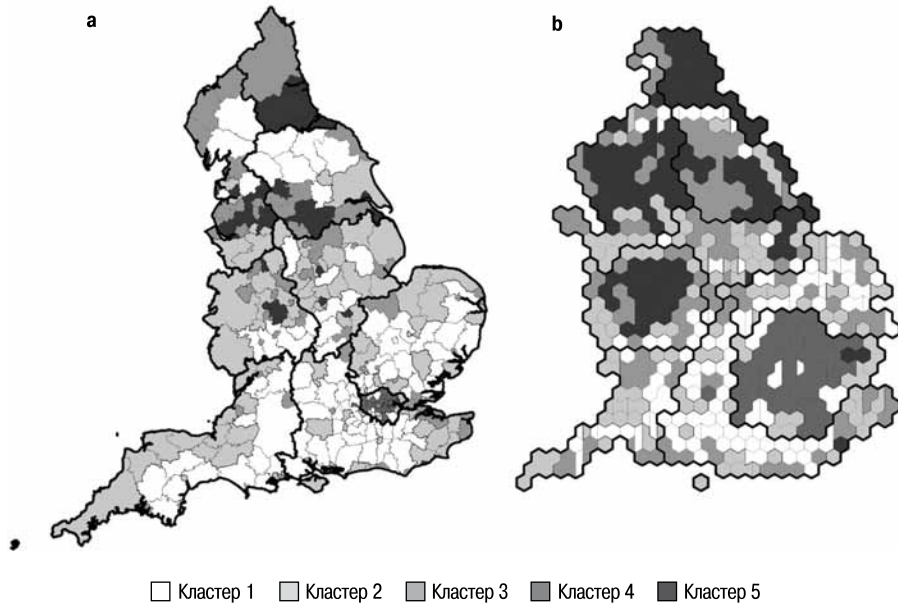
*Примечание.* В кластерный анализ были включены все административные единицы, по которым были разработаны профили здоровья (N=324). Профили здоровья не были разработаны для Лондонского Сити и архипелага Силли.

в середине 2007 г.). На долю северо-западных и северо-восточных территорий приходится 63% административных единиц, попавших в пятый кластер, с наихудшими показателями, и лишь 15,7% всех административных единиц Англии, включенных в профили здоровья. В то же время восточные и юго-восточные территории Англии составляют 35,2% всех административных единиц и 58,2%, включенных в 1-й кластер с наилучшими показателями здоровья населения (*кластер 1, рис. 1*).

## ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Применив кластерный анализ методом К-средних, мы распределили административные территории по пяти различным кластерам, отражающим состояние здоровья населения и здравоохранения. Несмотря на большое количество использованных переменных, подавляющее большинство из них тесно связано с единственной разделяющей осью (PSA1, *рис. 1*), демонстрируя сильные взаимосвязи между переменными. Выявление тесной взаимосвязи между различными показателями здоровья не является чем-то новым [Bellis et

Рисунок 2. Распределение кластеров профилей здоровья на карте по (а) стандартному географическому принципу, по (б) шестиугольникам численности населения



*Примечание.* Толстой линией обозначены региональные границы, тонкой — границы округов под началом местных органов власти. Шестиугольники населения изменяют размеры округов и представляют административные единицы в шестиугольниках в соответствии с численностью проживающего в них населения.

*Источник:* Thomas and Dorling, 2004.

al., 2011; Schuit et al., 2002; Poulton et al., 2002]. Тем не менее тот факт, что столь широкий набор показателей может быть использован при анализе географии общественного здоровья, еще раз подчеркивает необходимость применения междисциплинарного подхода к изучению здоровья и благополучия, особенно в тех районах, где больше всего проблем в области здравоохранения (например, в 5-м кластере). Более того, то, что агрегированные показатели здравоохранения явно коррелируют с индексом множественной депривации и даже с одиночным показателем благосостояния (валовой располагаемый доход домохозяйств), способствует увеличению внимания к проблемам неравенства и необходимости обращения к базовым детерминантам плохого состояния здоровья населения на национальном и международном уровнях [Commission on Social Determinants of Health, 2008; Department of Health, 2010a; Marmot, 2010].

Отличия 4-го кластера от всех остальных подтверждают результаты предыдущих исследований, где при помощи метода кластерного анализа различные области Великобритании были классифицированы по состоянию окружающей среды. В них также был выделен отдельный кластер, в который была включена значительная часть Лондона [Richardson et al., 2010; Shortt et al., 2011]. В целом эти данные свидетельствуют: об уникальном сочетании вызовов в сфере здравоохранения и экологии, с которыми сталкиваются все

мультикультурные мегаполисы (численность населения которых превышает 10 млн чел.); о необходимости специальных индивидуальных решений в данных обстоятельствах; о потенциальной опасности экстраполяции решений, примененных в этих городах, в другие условия. В самом деле, необходимость выявить и описать районы Лондона как демографически и экономически различные объекты возникла за рамками исследований здравоохранения [Vickers and Rees, 2007]. В то же время, однако, наиболее депривированные северо-западные и северо-восточные территории, свыше 60% которых попадают в 5-й кластер [с наихудшими показателями — Ред.], также сталкиваются с уникальным сочетанием проблем в сфере здравоохранения. Результаты исследования выявили, что в таких районах состояние здоровья очень плохое с самых ранних лет и в течение всей жизни человека. Это начинается с курения во время беременности, низких показателей грудного вскармливания и раннего развития кариеса, продолжается плохой успеваемостью в школе, подростковыми беременностями, и уже во взрослом возрасте человек страдает от ожирения, курения, алкоголизма, ухудшения здоровья, включая психическое, что в итоге приводит к ранней смерти. Отсюда следует, что преобразования в области улучшения здоровья необходимо осуществлять на всех этапах жизни человека [Lynch and Davey Smith, 2005; Braveman and Barclay, 2009; Department of Health, 2010a]. Возможности ребенка в последующей жизни во многом обусловлены средой, в которой он вырос (например, школой, наличием рабочих мест и доступного жилья) [Dorling et al., 2001]. Действительно, все больше появляется свидетельств того, что источники большинства проблем со здоровьем во взрослом возрасте коренятся в плохих социальных и экологических условиях детства, например, столкновение с жестоким обращением и безнадзорностью в детстве приводит к антисоциальному поведению и более высоким рискам алкогольной или наркотической зависимости, депрессии и даже развития онкологии [Felitti et al., 1998; Anda et al., 2006].

На индивидуальном уровне взаимосвязь бедности с каждым показателем профиля здоровья хорошо изучена, но их взаимоусиливающее влияние на протяжении всей жизни и комплексное взаимодействие между всеми показателями понимается плохо. Тем не менее политика по оздоровлению взрослых и детей путем побуждения населения к регулярной физической активности наталкивается на барьер высокого уровня преступности, вынуждающей людей оставаться дома [Bellis et al., 2008]. Ограниченный доступ к медицинской информации обусловлен низким уровнем грамотности [Sanders et al., 2009], а усилия по улучшению родительского ухода за детьми зависят от плохого здоровья родителей [Waylen and Stewart-Brown, 2010]. Даже восстановление человека после болезней, таких как, например, рак, неизбежно осложняется общим состоянием здоровья пациента и той общественной и семейной поддержкой, на которую он может рассчитывать [Falagas et al., 2007; Kroenke et al., 2006; Woods et al., 2006]. В результате при одновременном комплексном воздействии факторов состояние здоровья населения гораздо хуже, чем в случае влияния отдельных из них, поскольку способности людей и сообществ преодолевать недостаточное благосостояние оказываются сломлены. Следовательно, административные единицы, попавшие в 5-й кластер по характеристикам здоровья населения, требуют не только дальнейшего изучения комплексного взаимодействия между угрозами благополучию населения, но

и поддержки, необходимость в которой растет в геометрической прогрессии, а не прямо пропорционально каждому из имеющихся в них вызовов здоровью.

Существуют различные методики по выявлению характеристик на индивидуальном уровне, традиционно разрабатываемые для целей социального маркетинга или анализа состояния здоровья населения (например, P2 People and Places, Mosaic, Acorn) [Dedman et al., 2006]. Однако данные гео-демографические методики объединяют людей в группы с различными характеристиками, но не помогают выявлять районы, в которых требуется коррекция политики в области здравоохранения. Метод кластерного анализа позволяет решить эту проблему, в нем применяется механизм, при помощи которого можно выделить области с наибольшим количеством проблем со здравоохранением и обеспечить органы местного самоуправления районов с близкими профилями здоровья информацией о том, какие действия будут эффективными, а какие нет. В Англии полиция также проводит сравнительные исследования ситуации в похожих районах, чтобы ставить себе более содержательные и реалистичные задачи [Association of Police Authorities, 2006]. Кроме того, эти методы позволяют уйти от практики измерения ситуации в районе по одной переменной или по небольшому набору переменных, в результате которого можно получить либо чрезмерно положительные, либо излишне отрицательные характеристики этого района.

Для Англии совмещение данных о состоянии здоровья (например, о курении, употреблении алкоголя и ожирении) с данными, находящимися, в основном, в центре внимания местных органов власти (например, о насильственных преступлениях и загрязнении воздуха), очень своевременно, в таком виде их гораздо легче интерпретировать. После почти 40-летнего централизованного управления сферой здравоохранения (с 1974 г.) правительство постановило передать ответственность за здоровье населения обратно в местные органы власти [Министерство здравоохранения, 2010b]. Лучшее понимание того, как различные переменные здоровья взаимосвязаны и группируются в конкретных административных единицах, должно помочь в расстановке приоритетов и развитии политики на национальном уровне. Данные такого типа также могут использоваться для распределения ресурсов между административными единицами и внутри них, принимая во внимание не только информацию о конкретных проблемах в области здравоохранения и социальной сфере, но и дополнительные данные об их взаимодействии.

Используемые в нашей работе показатели здоровья были выявлены в результате независимого исследования и прошли экспертизу с целью включения в состав национальных профилей здоровья [Department of Health and the Association of Public Health Observatories, 2010]. Однако, как и экологические переменные, они неизбежно сглаживают неравенство между небольшими районами каждой административной единицы, а валидность каждого показателя в различных административных единицах зависит от того, насколько аккуратным был сбор данных для этого показателя и соответствующего показателя численности населения, применяемого в качестве знаменателя. Используемые нами данные относятся к разным годам, но всегда — к наиболее поздней из имеющихся дат. Даже на основе существующих данных могут быть выбраны иные показатели и применены другие методы исследования. Тем не менее сильная корреляция между переменными предполагает, что даже если будут использованы другие методы, добавлены или заменены отдельные

показатели, концентрация множественных проблем со здоровьем в определенной группе административных единиц сохранится. В случае применения метода кластеризации с целью информирования органов власти о состоянии здоровья в длительной перспективе потребуются изучение частоты, с которой кластеры должны пересматриваться, чтобы отобразить изменения в статусе различных административных единиц.

## ЗАКЛЮЧЕНИЕ

Метод кластерного анализа, который был использован для проведения данного исследования, позволил выявить, как большой набор различных находящихся в открытом доступе характеристик здоровья населения концентрируется в определенных районах. Последствия подобной концентрации плохо изучены, но вполне вероятно, что комплексное влияние проблем в области здоровья и благополучия будет способствовать непропорциональной нагрузке на общество, выраженной в необходимости предоставления определенных услуг и других ресурсов. Подобные вопросы обычно не рассматриваются при формировании политики, находящейся в ведении отдельных ведомств. Необходим межведомственный подход к изучению той среды, в которой будут реализовываться программы в сфере здравоохранения, образования, права, социальная или экономическая политика.

Кластеры плохого здоровья населения (см. *табл.*) требуют комплексных действий со стороны местных руководителей здравоохранения (например, директоров департаментов) по целому ряду направлений, связанных со здравоохранением, а образовательные, транспортные программы, программы по повышению уровня физической активности населения и программы социального обслуживания должны быть скорректированы с учетом государственных приоритетов в области здравоохранения. Коллеги этих местных руководителей в соответствующих национальных министерствах должны оказывать аналогичное влияние на другие государственные департаменты. Важно отметить, что подобные действия должны учитывать, что примеры успешного государственного вмешательства в районах с относительно благополучной ситуацией в здравоохранении из 1-го и 2-го кластеров невозможно просто скопировать в районах, попавших в 5-й кластер; либо для достижения аналогичных результатов могут потребоваться более значительные ресурсы. Тем не менее метод кластерного анализа может быть использован для выявления районов, между которыми возможен успешный обмен опытом реализации программ и государственных интервенций в сфере здравоохранения. С увеличением массива данных, имеющих отношение к здоровью и благосостоянию, такие механизмы помогут перейти от рассмотрения отдельных независимых показателей здоровья населения к комплексным проблемам здравоохранения, описывающим набор вызовов, с которыми сталкиваются различные сообщества, привлекающим большее количество заинтересованных людей и организаций к поиску решений.

## ЛИТЕРАТУРА

- Anda, R.F., Felitti, V.J., Bremner, D., Walker, J.D., Whitfield, C.H., Perry, B.D., Dube, S.H.R., Giles, W.H.*, 2006. The enduring effects of abuse and related adverse experiences in childhood. *European Archives of Psychiatry and Clinical Neuroscience* 256, 174—186.
- Association of Police Authorities, 2006. *iQuanta: a police authority guide* second ed. Association of Police Authorities, London.
- Association of Public Health Observatories, 2006. *Health Profiles Project: Project brief*. Association of Public Health Observatories, York.
- Bacciu, D., Jarman, I.H., Etchells, T.A., Lisboa, P.J.G.*, 2009. Patient stratification with competing risks by multivariate Fisher distance. *International Joint Conference on Neural Networks*, 213—220.
- Barton, J., Pretty, J.*, 2010. What is the best dose of nature and green exercise for improving mental health? A multi-study analysis. *Environmental Science and Technology* 44, 3947—3955.
- Bellis, M.A., Hughes, K., Tocque, K., Hennell, T., Humphrey, G., Wyke, S.*, 2005. Assessing and communicating the health and judicial impact of alcohol use. *Public Health* 119, 253—261.
- Bellis, M.A., Hughes, K., Anderson, Z., Tocque, K., Hughes, S.*, 2008. Contribution of violence to health inequalities in England: demographics and trends in emergency hospital admissions for assault. *Journal of Epidemiology and Community Health* 62, 1064—1071.
- Bellis, M.A., Hughes, K., Wood, S., Wyke, S., Perkins, C.*, 2011. National five-year examination of inequalities and trends in emergency hospital admission for violence across England. *Injury Prevention* 10.1136/ip.2010.030486.
- Braveman, P., Barclay, C.*, 2009. Health disparities beginning in childhood: a life-course perspective. *Pediatrics* 124, S163—S175.
- Commission on Social Determinants of Health, 2008. *Closing the gap in a generation: health equity through action on the social determinants of health*.
- Final Report of the Commission on Social Determinants of Health. World Health Organization, Geneva.
- Dedman, D., Jones, A., Tocque, K., Bellis, M.A.*, 2006. *Population targeting: Tools for social marketing*. Synthesis Report, Issue 6. Liverpool: North West Public Health Observatory.
- Department of Health, 2009. *Tackling health inequalities: 10 Years On — A review of developments in tackling health inequalities in England over the last 10 years*. Department of Health, London.
- Department of Health, 2010a. *Healthy lives, healthy people: Transparency in outcomes*. Department of Health, London.
- Department of Health, 2010b. *Equity and excellence: Liberating the NHS*. The Stationary Office, London.
- Department of Health and the Association of Public Health Observatories, 2010. *The indicator guide Health Profiles 2010*. Department of Health, London.
- Dorling, D., Smith, G., Noble, M., Wright, G., Burrows, R., Bradshaw, J., Joshi, H., Pattie, C., Mitchell, R., Green, A.E., McCulloch, A.*, 2001. How much does place matter? *Environment and Planning A* 33, 1335—1369.
- Draper, P., Glenhorn, G., Best, G.*, 1976. The organization of health care: A critical review of the 1974 reorganization of the National Health Service. in: Tuckett, D. (Ed.), *An Introduction to Medical Sociology* Tavistock Publications Ltd., London, pp. 254—289.
- Duarte, C.S., Sourander, A., Nikolakaros, G., Pihlajamaki, H., Helenius, H., Piha, J., Kumpulainen, K., Moilanen, I., Tamminen, T., Almqvist, F., Must, A.*, 2010. Child mental health problems and obesity in early adulthood. *The Journal of Pediatrics* 156, 93—97.

- Falagas, M.E., Zarkadoulia, E.A., Ioannidou, E.N., Peppas, G., Christodoulou, C., Rafailidis, P.R.*, 2007. The effect of psychosocial factors on breast cancer outcome: a systematic review. *Breast Cancer Research* 9, R44.
- Felitti, V.J., Anda, R.F., Nordenberg, D., Williamson, D.F., Spitz, A.M., Edwards, V., Koss, M.P., Marks, J.S.*, 1998. Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The Adverse Childhood Experiences (ACE) Study. *American Journal of Preventive Medicine* 14, 245–258.
- Friedman, H.P., Rubin, J.*, 1967. On some invariant criteria for grouping data. *Journal of the American Statistical Association* 62, 1159–1178.
- Harden, A., Brunton, G., Fletcher, A., Oakley, A.*, 2009. Teenage pregnancy and social disadvantage: systematic review integrating controlled trials and qualitative studies. *British Medical Journal* 339, b4254.
- Hillis, S.D., Anda, R.F., Dube, S.R., Felitti, V.J., Marchbanks, P.A., Macaluso, M., Marks, J.S.*, 2010. The protective effect of family strengths in childhood against adolescent pregnancy and its long-term psychosocial consequences. *The Permanente Journal* 14, 18–27.
- I2SARE, 2010. Development of a regional typology — aim, methods and results. Available at: <http://www.i2sare.eu/Default.aspx> Accessed on 1st June 2011.
- Jolliffe, I.T.*, 2002. *Principal Component Analysis* second ed. Springer, New York.
- Klugman, J.*, 2010. *Human development report 2010. The real wealth of nations: pathways to human development.* United Nations Development Programme, New York.
- Kroenke, C.H., Kubzansky, L.D., Schernhammer, E.S., Holmes, M.D., Kawachi, I.*, 2006. Social networks, social support and survival after breast cancer diagnosis. *Journal of Clinical Oncology* 24, 1105–1111.
- Lisboa, P., Ellis, I., Green, A., Ambriogi, F., Dias, M.*, 2008. Cluster-based visualization with scatter matrices. *Pattern Recognition* 29, 1814–1823.
- Lynch, J., Davey Smith, G.*, 2005. A life course approach to chronic disease. *Annual Review of Public Health* 26, 1–35.
- MATLAB version 7.8.0. R2009a, 2009. Natick, Massachusetts: The MathWorks Inc.
- Marmot, M.*, 2005. Social determinants of health inequalities. *The Lancet* 365, 1099–1104.
- Marmot, M.*, 2010. *Fair society, healthy lives. Strategic review of health inequalities in England post-2010. The Marmot review final report.* University College London, London.
- Murray, C.J.L., Lopez, A.D., Black, R., Mathers, C.D., Shibuya, K., Ezzati, M., Salomon, J.A., Michaud, C.M., Walker, N., Vos, T.*, 2007. Global burden of disease 2005: call for collaborators. *The Lancet* 370, 109–110.
- Office for National Statistics, 2001. National statistics 2001 area classification. Режим доступа: [http://www.statistics.gov.uk/about/methodology\\_by\\_theme/area\\_classification/default.asp](http://www.statistics.gov.uk/about/methodology_by_theme/area_classification/default.asp). Дата обращения: 9 июня 2011 г.
- Office for National Statistics, 2008. Mid-year population estimates. Режим доступа: <http://www.statistics.gov.uk/statbase/product.asp?vlnk=15106>. Дата обращения: 14 февраля 2011 г.
- Office for National Statistics, 2009. Regional household income. Режим доступа: <http://www.statistics.gov.uk/statbase/Product.asp?vlnk=14651>. Дата обращения: 14 февраля 2011 г.
- Office for National Statistics, 2011. Population estimates for ethnic group (experimental), 2011. Режим доступа: <http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=14238&More=Y>. Дата обращения: 14 февраля 2011 г.
- Poulton, R., Caspi, A., Milne, B.J., Thomson, W.M., Taylor, A., Sears, M.R., Moffitt, T.E.*, 2002. Association between children's experience of socioeconomic disadvantage and adult health: a life-course study. *The Lancet* 360, 1640–1645.
- Richardson, E.A., Mitchell, R., Shortt, N.K., Pearce, J., Dawson, T.P.*, 2010. Developing summary measures of health-related multiple environmental deprivation for epidemiological research. *Environment and Planning A* 42, 1650–1658.

- Rutherford, A., Zwi, A.B., Grove, N.J., Butchart, A., 2007.* Violence: a priority for public health? (part 2). *Journal of Epidemiology and Community Health* 61, 764—770.
- Sanders, L.M., Federico, S., Klass, P., Abrams, M.A., Dreyer, B., 2009.* Literacy and child health: a systematic review. *Archives of Pediatrics and Adolescent Medicine* 163, 131—140.
- Schuit, A.J., van Loon, A.J.M., Tijhuis, M., Ocke, M.C., 2002.* Clustering of lifestyle risk factors in a general adult population. *Preventive Medicine* 35, 219—224.
- Shortt, N.K., Richardson, E.A., Mitchell, R., Pearce, J., 2011.* Re-engaging with the physical environment: a health related environmental classification of the UK. *Area* 43, 76—87.
- SPSS for windows, release 17.0.0, 2008. Chicago: SPSS Inc.
- Stachenko, S., 2008.* Challenges and opportunities for surveillance data to inform public health policy on chronic non-communicable diseases: Canadian perspectives. *Public Health* 122, 1038—1041.
- Thomas, B., Dorling, D., 2004.* Advances in the human cartography of the UK. *Cartographic Journal* 41, 109—115.
- Vickers, D., Rees, P., 2007.* Creating the UK national statistics 2001 output area classification. *Journal of the Royal Statistical Society, Series A* 170, 379—402.
- Waylen, A., Stewart-Brown, S., 2010.* Factors influencing parenting in early childhood: a prospective longitudinal study focusing on change. *Child: Care, Health and Development* 36, 198—207.
- Weiler, R., Stamatakis, E., Blair, S., 2010.* Should health policy focus on physical activity rather than obesity? Yes. *British Medical Journal* 340, c2603.
- Wilkinson, R., Marmot, M., 2003.* The solid facts Second Edition World Health Organization, Geneva.
- Woods, L.M., Rachet, B., Coleman, M.P., 2006.* Origins of socio-economic inequalities in cancer survival: a review. *Annals of Oncology* 17, 5—19.