

Пространственный анализ данных по заболеваемости туберкулезом: факторы риска

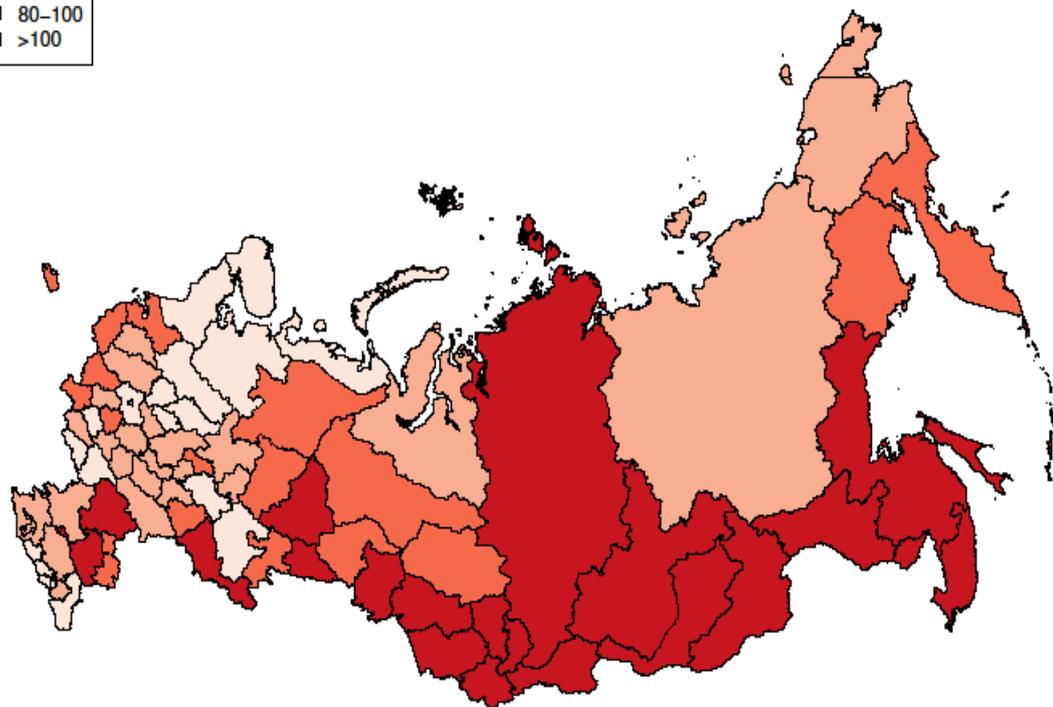
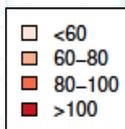
О.А. Мельниченко
МГУ, факультет ВМК

Текущее место работы
Новартис Фарма

Туберкулез в России

- Россия входит в список 22 стран с наибольшим бременем туберкулеза.
- Средняя заболеваемость туберкулезом 80 случаев на 100 тыс населения.
- Межрегиональные различия, 2009 год: Ингушетия – 38, Тува – 211 случаев на 100 тыс населения.

Заболееваемость ТБ, 2009 год



Вопросы:

- Действительно ли существует пространственный паттерн?
- Какие факторы могут объяснить наблюдаемый паттерн?

Пространственная автокорреляция: матрица пространственных весов

Соседи:

- общая граница
- регулярные потоки рабочей силы (Москва и окрестности)

Матрица W :

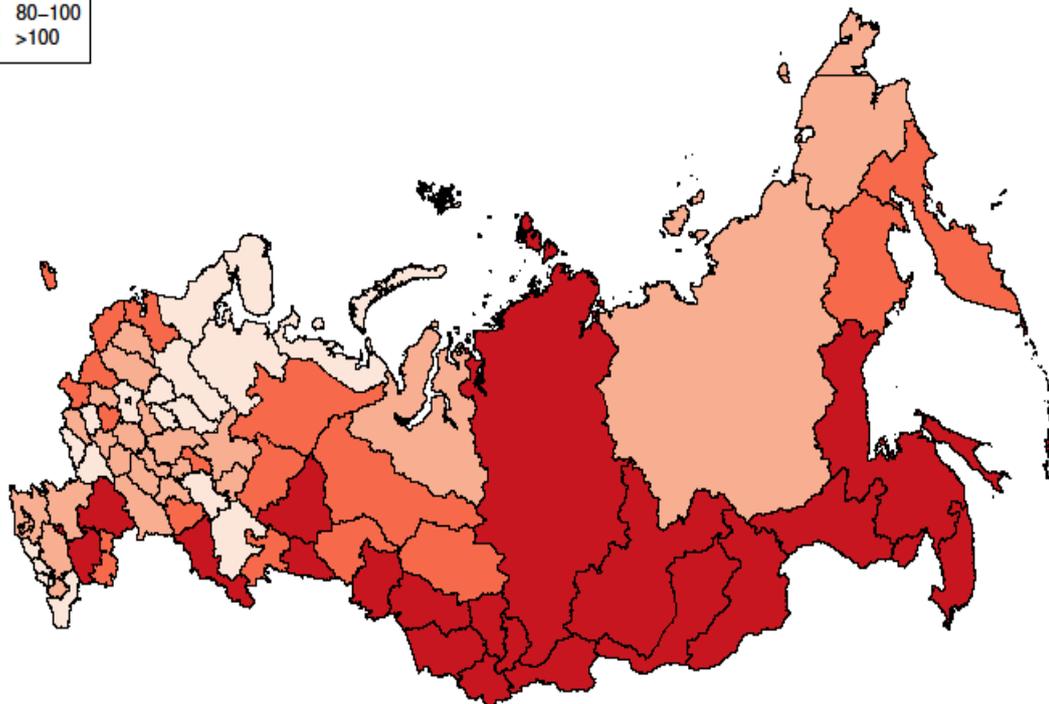
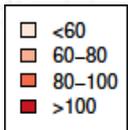
- $w_{ij} = \begin{cases} 1, & \text{если } i \text{ и } j \text{ регионы — соседи,} \\ 0, & \text{иначе.} \end{cases}$
- элементы нормируются по строкам

$$w_{ij} = \frac{w_{ij}}{\sum_{k=1}^n w_{ik}}$$

Пространственная автокорреляция: индекс Морана

Moran's I

$$I = \frac{n}{\sum \sum w_{ij}} \times \frac{\sum \sum w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{\sum (X_i - \bar{X})^2}$$



Заболеваемость ТБ,
2009 год

$I = 0.59, p < 0.01$

Обзор литературы: факторы риска

- Низкий доход
- Скученность населения (недостаток жилплощади)
- Высокий уровень безработицы
- Миграция из регионов с высокой заболеваемостью ТБ
- Низкий уровень образования

Результаты получены для Англии, Германии, Бразилии, Индии, Китая,...

Социально-экономические показатели регионов РФ

- Уровень безработицы
- Средний месячный доход на душу населения
- Жилплощадь на душу населения
- Уровень бедности (доля населения с доходом ниже прожиточного минимума)
- Уровень преступности
- Уровень загрязнения воздуха

Социально-экономические показатели: корреляция с заболеваемостью ТБ, 2009 год

Variables	Mean	SD	Correlation coef.
TB incidence	85.13	34.76	
Unemployment rate (%)	9.95	6.23	0.06
Average monthly income (Russian rubles)	15393.62	7562.13	-0.13
Poverty level (%)	16.51	5.53	0.34*
Housing space (m ²)	22.31	3.43	-0.28*
Crime level (per 100000)	2017.96	632.65	0.54*
Emission of air pollutants (ths of tonnes)	269.53	532.34	0.09

* $p < 0.01$

Линейные модели

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \varepsilon_i$$

Тест Морана

Если индекс Морана для остатков модели остается значимым, то:

- не включены в модель переменные, характеризующиеся высокой пространственной корреляцией,
- не учтена пространственная корреляция ошибок измерения.

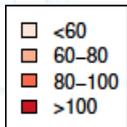
Линейные модели: результаты, 2009 год

	Model 1	Model 2	Model 3
Intercept	85.13 (**)	85.13 (**)	85.13 (**)
Poverty level		11.87 (**)	9.18 (**)
Housing space			-7.03 (*)
Crime level	18.59 (**)	18.72 (**)	18.98 (**)
R^2	0.28	0.39	0.42
AIC	801.59	788.86	785.76
Moran's I (model)	0.33 (**)	0.2 (**)	0.17 (**)
* $p < 0.01$, ** $p < 0.001$			

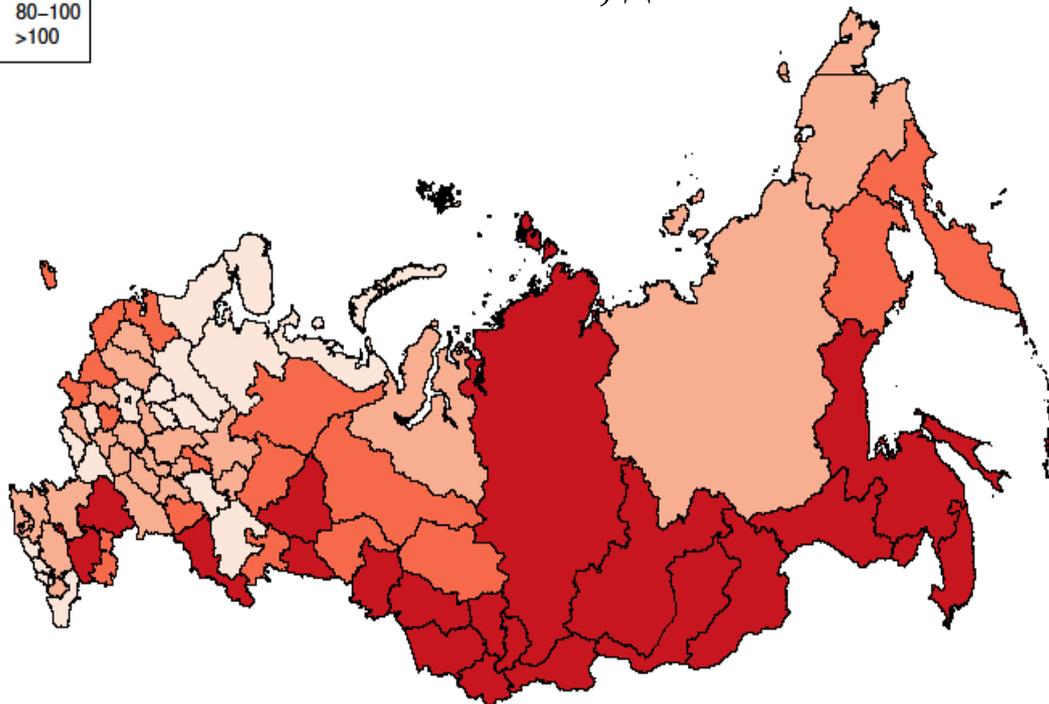
Выбранная модель:

TB incidence ~ Poverty level + Housing space + Crime level

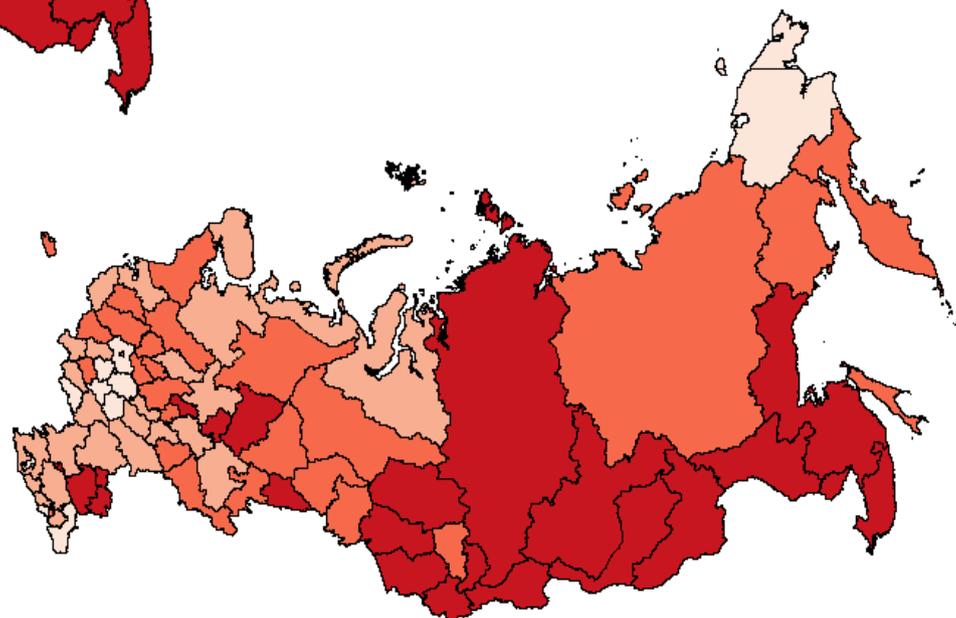
Линейные модели: результаты, 2009 год



Заболееваемость ТБ, данные



Заболееваемость ТБ, модель 3



Возможное влияние особенностей эпидемического процесса

- Туберкулез имеет медленную динамику.
- Социально-экономические факторы в предыдущие годы могут оказывать влияние на заболеваемость в текущем году.
- Необходимо рассмотреть социально-экономические показатели за предыдущие годы.

Линейные модели: результаты, 2000 и 2005 гг.

	2005			2000		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Intercept	85.13 (**)	85.13 (**)	85.13 (**)	85.13 (**)	85.13 (**)	85.13 (**)
Poverty level		11.55 (**)	9.35 (**)		11.34 (**)	7.95 (*)
Housing space			-5.39			-8.89 (**)
Crime level	17.14 (**)	19.24 (**)	18.57 (**)	19.84 (**)	21.83 (**)	23.35 (**)
R ²	0.23	0.33	0.35	0.32	0.41	0.46
AIC	806.45	795.82	795.22	796.87	785.09	779.09
Moran's I (model)	0.47 (**)	0.39 (**)	0.34 (**)	0.47 (**)	0.38 (**)	0.27 (**)

* p < 0.01, ** p < 0.001

Пространственные модели

- Модель с пространственным лагом

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \lambda \sum_j w_{ij} y_j + \varepsilon_i$$

- Модель с пространственной корреляцией ошибок

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \rho \sum_j w_{ij} \varepsilon_j + \xi_i$$

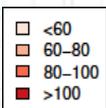
Модель 3, тест множителей Лагранжа:

LM-Err 4.946 (*)

LM-Lag 20.95 (**)

* $p < 0.05$, ** $p < 0.001$

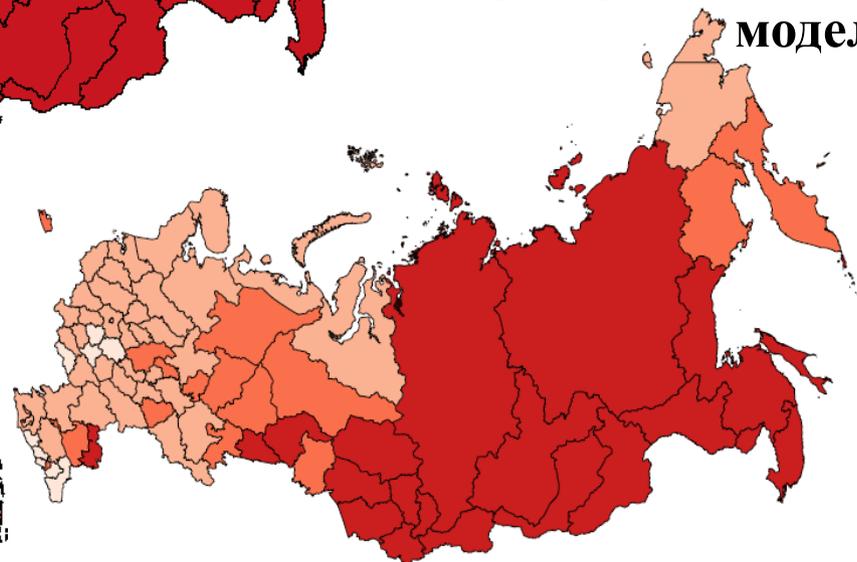
Пространственные модели: результаты, 2009 год



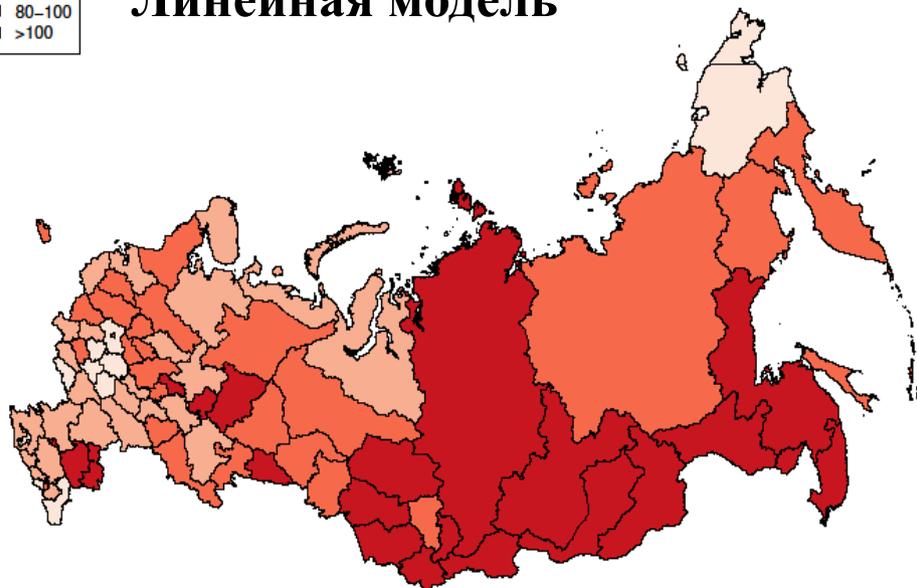
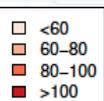
Данные



Пространственная
модель



Линейная модель



Выводы

- Наиболее значимыми факторами риска являются:
 - уровень преступности – характеристика внешней среды,
 - уровень бедности и жилплощадь – индивидуальные характеристики
- Линейная модель объясняет существенную долю пространственной автокорреляции заболеваемости ТБ.
- Пространственная модель объясняет пространственную автокорреляцию полностью, но предсказания модели плохо согласуются с данными.

Дальнейшие планы

- Включить в рассмотрение дополнительные данные (качество работы ЛПУ, национальный состав, климатические).
- Провести пространственно-временной анализ данных.
- Получить детальные эпидемиологические данные (для районов субъектов РФ).

Благодарности

Институту демографических исследований
Макса Планка (г.Росток, Германия) за курс
“Spatial Demography” и материальную
поддержку.

Спасибо за внимание!
Вопросы?

День математического моделирования в Новартис

- 14 ноября 2012 Новартис совместно с ИВМ РАН проводит научную конференцию, посвященную использованию математического моделирования в современной фармакологии и медицине.
- Мы приглашаем принять участие всех желающих, ведущих активную научную работу в данной области.
- Тезисы докладов можно присылать на следующие адреса: olesya.melnichenko@novartis.com
kirill.peskov@novartis.com