WIR SCHAFFEN TRANSPARENZ.

UNDERSTANDING HEALTH CARE UTILIZATION USING GEOGRAPHICAL WEIGHTED REGRESSION

AGENDA

1. Motivation
2. Measure of Regional Health Care Utilization using GWR
3. First Results

Motivation

Approach by Sundmacher & Ozegowski 2014 / Culyer & Wagstaff 1993

- Need is the "expenditures required to exhaust a person's capacity to benefit“.
  - In Germany every individual “is entitled to the full coverage of any necessary treatment, under the condition that the treatment is provided economically and according to current standards of medical knowledge (Section 70(1) SGB V)”.
- Different risk-structures / need are compensated through the German risk structure compensation scheme (RSCS) showing the average need of a person controlling for age, sex and morbidity on the perspective of statutory health insurances.

➢ Assume that the average need is represented through the RSCS.

Motivation

The German risk structure compensation scheme (RSCS)

$$S_i = \beta_A A_i + \beta_M M_i + \beta_D D_i + \beta_F F_i + e_i = \mathrm{EN_i} + \mathrm{e_i}$$

- Health care costs (S),
- Expected need (EN),
- Age and gender, in 5-year-sex-specific groups (A),
- Morbidity, measured by 80 diseases aggregated into 192 disease groups (M),
- Disability pension as proxy for further need (D),
- Living abroad (F)

Motivation

Measurement of regional inequity in health care

- Inequity is defined as ratio between health care utilisation and need in a predefined region j in terms of the RSCS in

$$I_j = \frac{\sum_{i\ in\ j} S_i}{\sum_{i\ in\ j} EN_i}$$

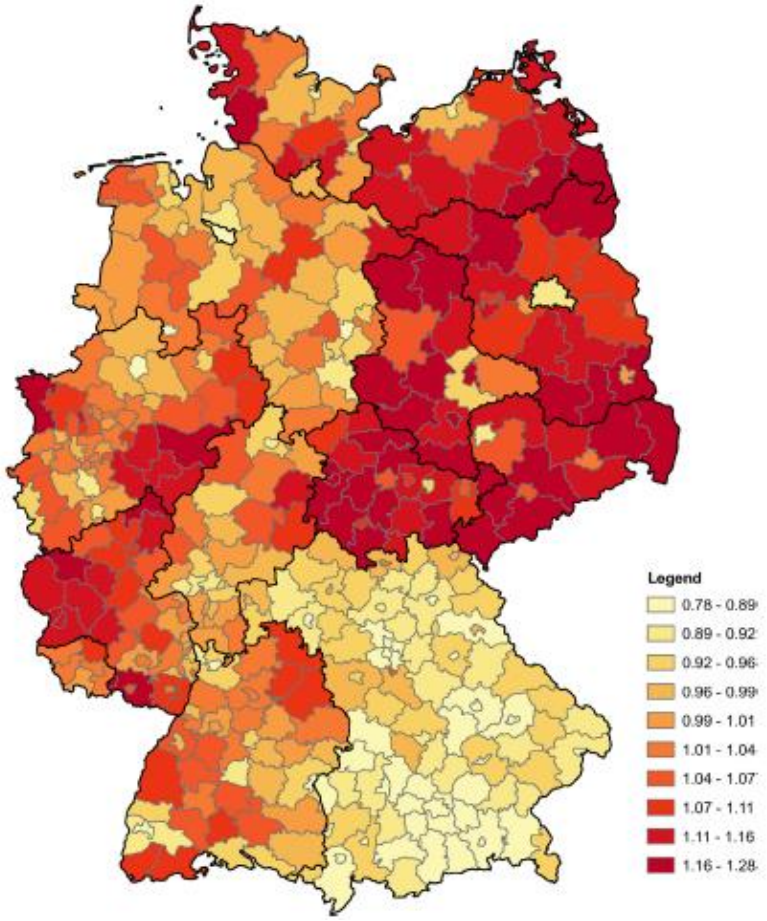
- Overuse: $I_j > 1$ with more health care costs then expected need
- Underuse: $I_j < 1$ with less health care costs then expected need

- **Problem**:
  - Which definition of j (zones) is suitable for this purpose?
  - What happens if we choose arbitrary ones?
  - Are there intrinsically different Problems for health care costs?

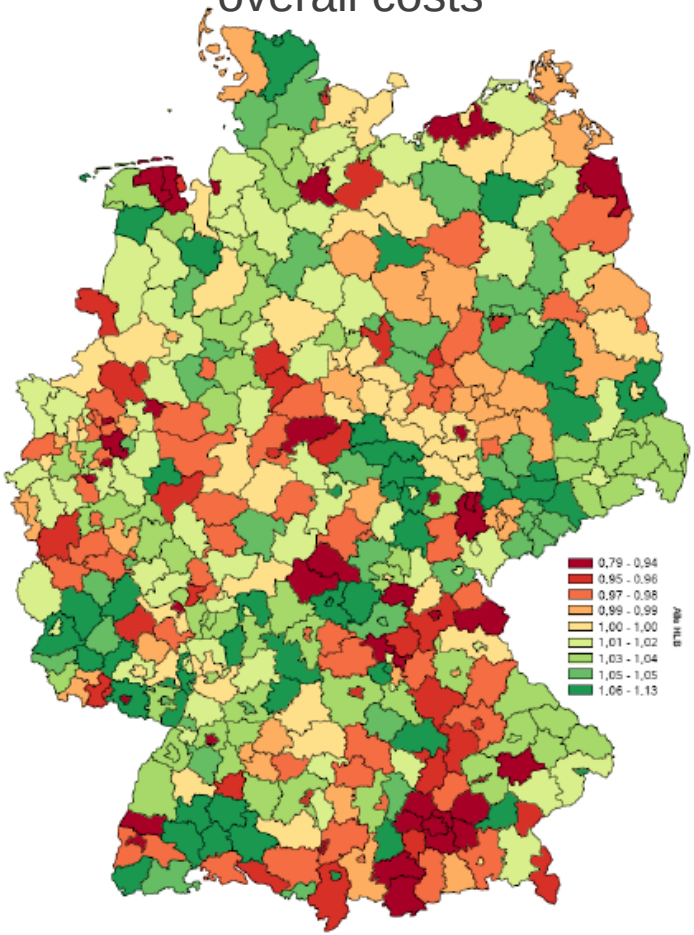
Motivation

Mapping the spatial inequities between need and health care utilisation

Sundmacher & Ozegowski 2014
ambulatory costs



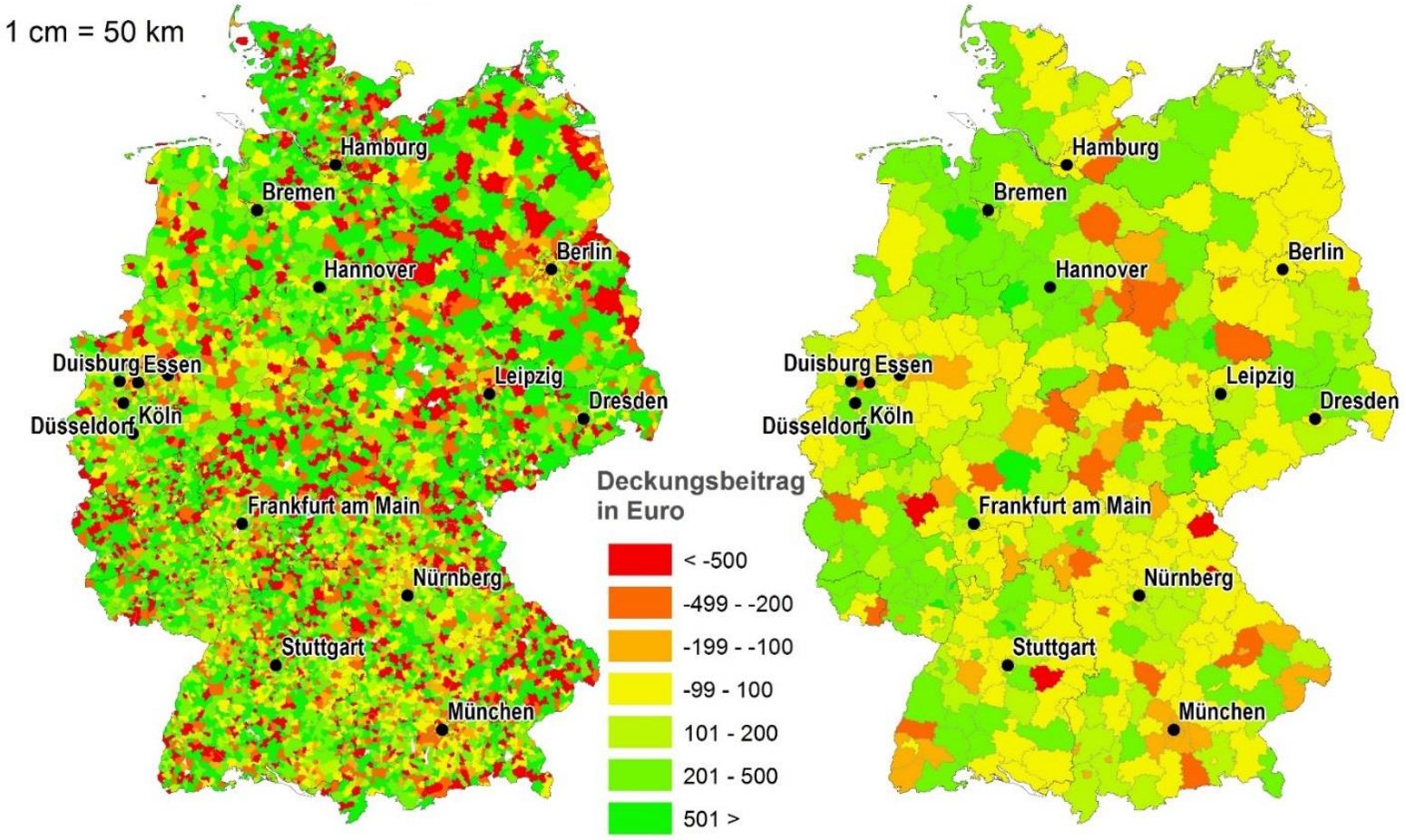
Drösler et al. 2009
overall costs



Motivation

Mapping the spatial inequities between need and health care utilisation

own calculation for overall costs on zip-regions and rural districts



The Zone Definition Problem in Small Area Variation

Problems associated with zone definition

- **The Area to Point Problem (A-, B- and C-type errors)**
  - A: zonal statistics refer to a single point rather to a set of points
  - B: the distance within a zone is assumed to be zero (access to every place at zero costs)
  - C: no windfall gain of supply for other zones and perfectly association of zone and supply
- **The Multiple Area Unit Problem (MAUP)**
  - aggregation-variant of results (the finer the less we see, the wider the less we find)
  - scale-variant of the zones (**boundary problem**)
- **The Yule-Simpson-effect**
  - omitted variable bias through different geographical requirements
  - ecological fallacy while interpreting area results as individual behaviour

AGENDA

1. Motivation
2. **Measure of Regional Health Care Utilization using GWR**
3. First Results

Measure of Regional Health Care Utilization

The German risk structure compensation scheme (RSCS)

$$S_i = \beta_A A_i + \beta_M M_i + \beta_D D_i + \sout{\beta_E E_i} + \beta_K * K_i + e_i = \mathrm{EN_i} + \mathrm{e_i}$$

- Health care costs (S),
- Expected need (EN),
- Age and gender, in 5-year-sex-specific groups (A),
- Morbidity, measured by 80 diseases aggregated into 192 disease groups (M),
- Disability pension as proxy for further need (D),
- ~~Living abroad~~, which is not consider as we are interested in regional differences
- Dying as proxy for high costs in the last month of living (K)

- Using a contemporaneous classification of 2013 (BVA / http://www.bundesversicherungsamt.de/risikostrukturausgleich/festlegungen.html)

Measure of Regional Health Care Utilization

Integrating regional inequities into the model through GWR

$$S_i = I(u_i, v_i) + \mathrm{EN_i} + \mathrm{e_i} \qquad / \qquad \boldsymbol{S} = \boldsymbol{I} + \beta * \boldsymbol{X} + E$$

- Integration a factor of regional requirements / inequities through the point specific $\hat{\boldsymbol{I}}(u_j, v_j)$ where (u_j, v_j) are coordinates in the space of observation.

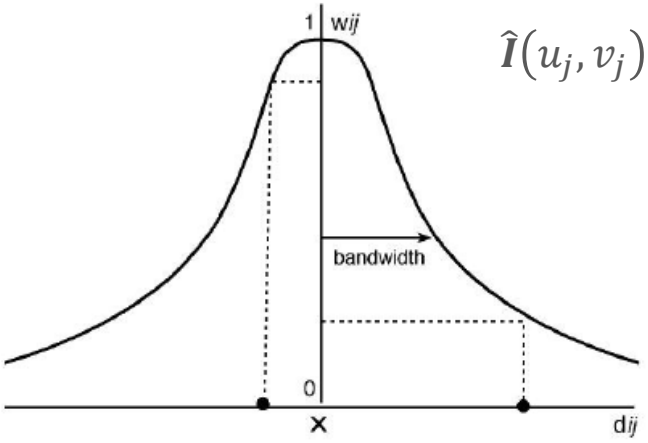
- **Weighted moving window regression method developed by Foterhingham and Brundson (2000, 2002), building on works of Hastie and Tibshirani (1990) and Loader (1999)**
- **Uses weighted least squares approach**

Measure of Regional Health Care Utilization

Integrating regional inequities into the model through GWR

$$S_i = I(u_i, v_i) + \mathrm{EN_i} + \mathrm{e_i} \qquad / \qquad \boldsymbol{S} = \boldsymbol{I} + \beta * \boldsymbol{X} + E$$

$$\hat{\boldsymbol{I}}(u_j, v_j) = \left(\mathbf{1}'\boldsymbol{W}(u_j, v_j)\,\mathbf{1}\right)^{-1}\mathbf{1}'(\boldsymbol{S} - \beta * \boldsymbol{X})$$



x regression point wij is the weight of data point *j* at regression point *i*
• data point dij is the distance between regression point *i* and data point *j*

$\boldsymbol{w_{ii}(u_j, v_j)} = \left[1 - \left(\frac{\boldsymbol{d_{ij}}}{\boldsymbol{b}}\right)^2\right]^2$	if point j is at least as near as the Nth nearest neighbour of i and b is the distance to the Nth nearest neighbour
= 0	otherwise

- Separate regression is run for each observation, using a spatial kernel that centers on a given point and weights observations subject to a distance decay function.
- An adaptive kernel is used as data is not evenly distributed

WIG2 Wissenschaftliches Institut für Gesundheitsökonomie und Gesundheitssystemforschung

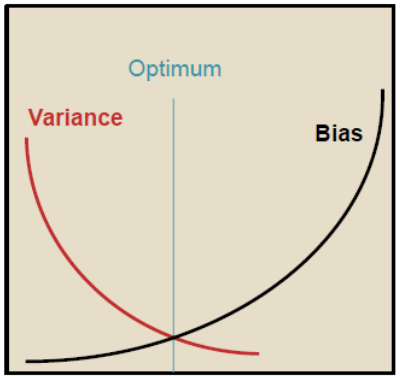
Bias and variance trade-off

- The smaller the bandwidth (N), the more variance but the lower the bias, the larger the bandwidth, the more bias but the more variance is reduced
- This is because we assume there are many betas over space and the more it is like a global regression, the more biased it is.
- BIC minimization provides a way of choosing bandwidth that makes optimal tradeoff between bias and variance.

$$BIC = n * \ln\left(\widehat{\sigma_e^2}\right) + tr(\boldsymbol{S}) * \ln(n)$$

Where tr(S) is the trace of the hat matrix

and n is the number of observations



AGENDA

1. Motivation
2. Measure of Regional Health Care Utilization using GWR
3. **First Results**

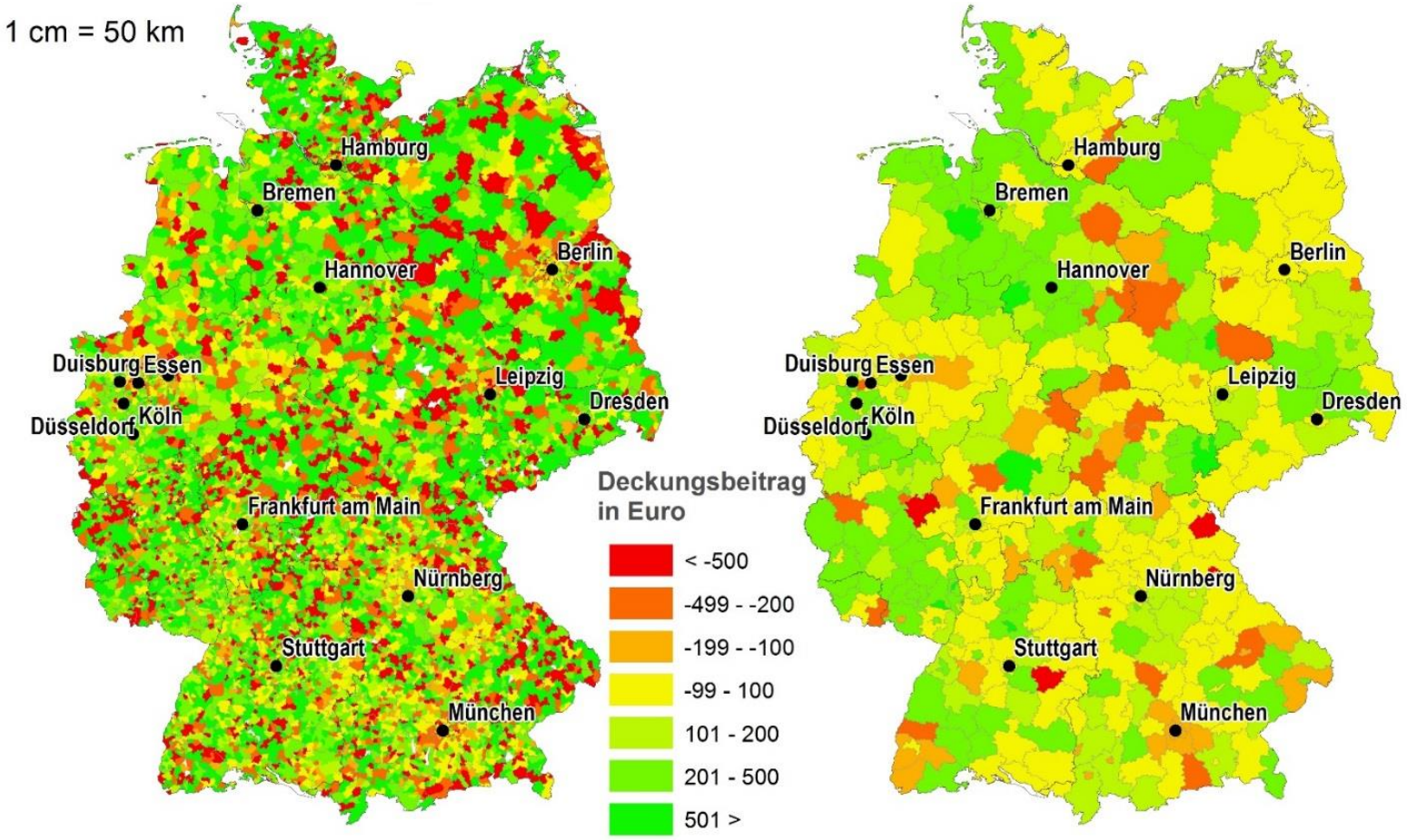
Data considerations

- Information of German statutory health insurances including 1.8 million individuals.
- A sample of 1.5 million individuals is drawn randomly controlling for age, sex and morbidity on the administrative level of rural districts.
- All individuals are georeferenced on their place of residence to calculate UTM coordinates (u_i, v_i).
- A grouping of the diseases into disease-groups is done through the German RSCS using the contemporaneous classification of 2013.
- To control for extreme values a “Huber M-Estimator” is used instead of a weighted regression.
- Estimates or done on a 5x5km grid of Germany for ambulatory, stationary, pharmaceutic and overall costs.

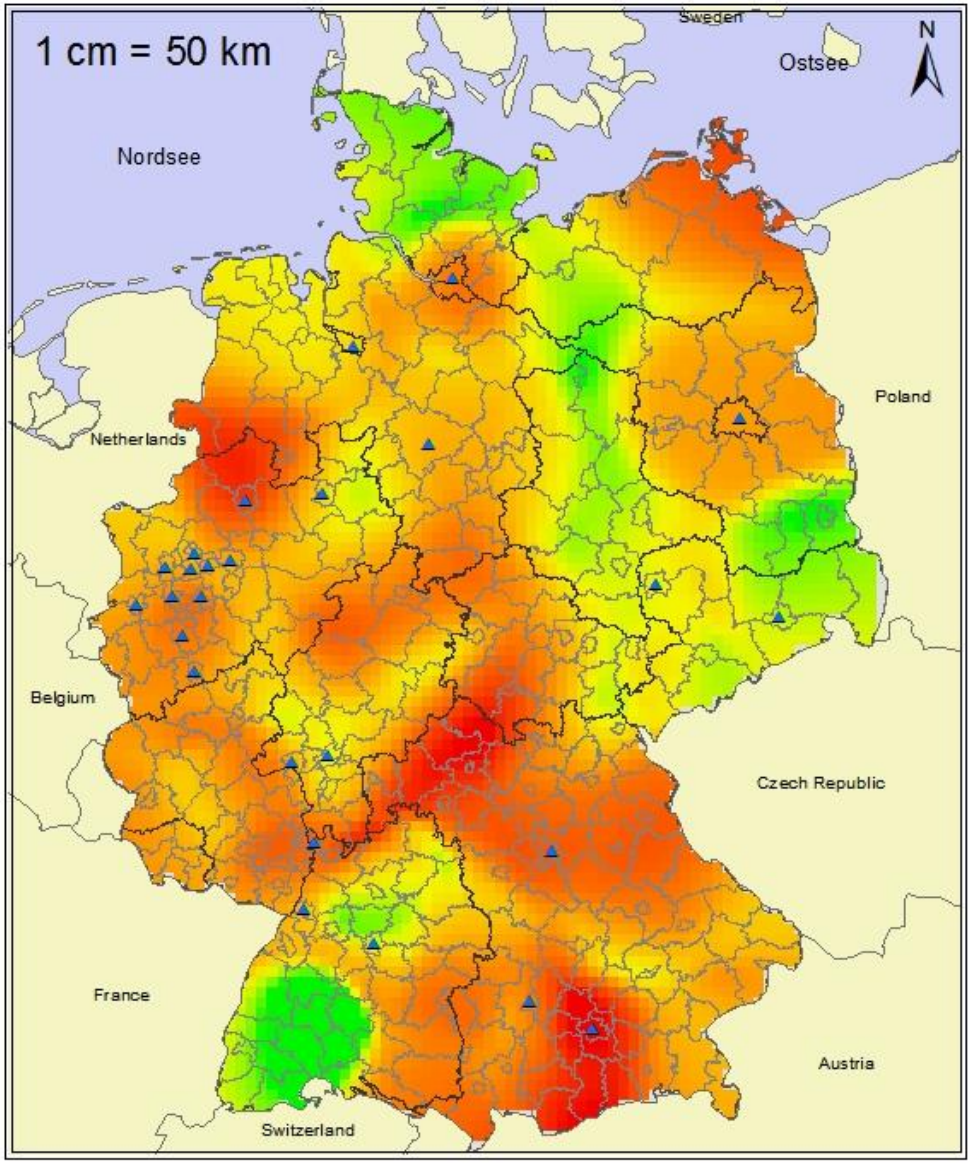
First Results

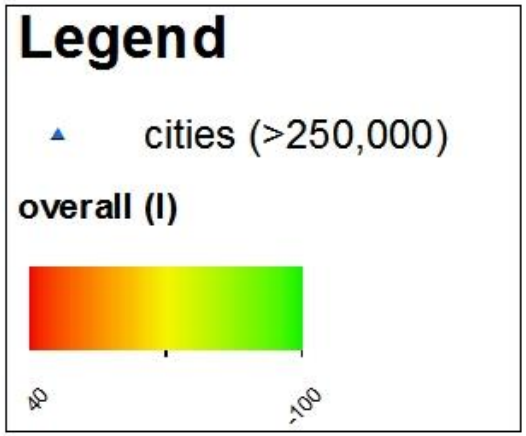
Where we come from:

own calculation for overall costs on zip-regions and districts



Overall inequity between need and utilisation





Red colour
means more utilisation than expected need. It refers to an oversupply for the sick and a negative marginal return for statutory sickness funds. A number of 10 means 10 Euro more utilisation than expected.

Green colour
means more expected need then the actual utilisation. It refers to an undersupply for the sick and a positiv marginal return for statutory sickness funds. A number of -10 means 10 Euro less utilisation than expected.

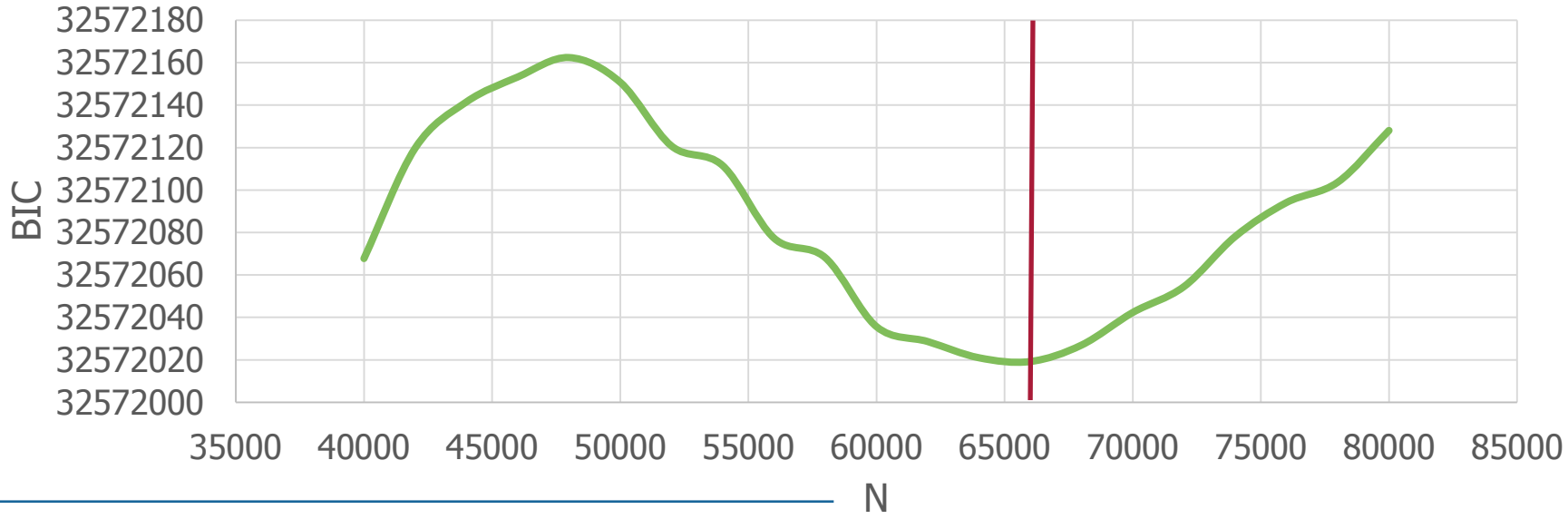
data: cost p.c. of 1,5 million observations
method: GWR
kernel: adaptive bi-squared
parametrisation: N=66,000

date: ETRS 1989 LAEA
author: Danny Wende

First Results

The bandwidth selection

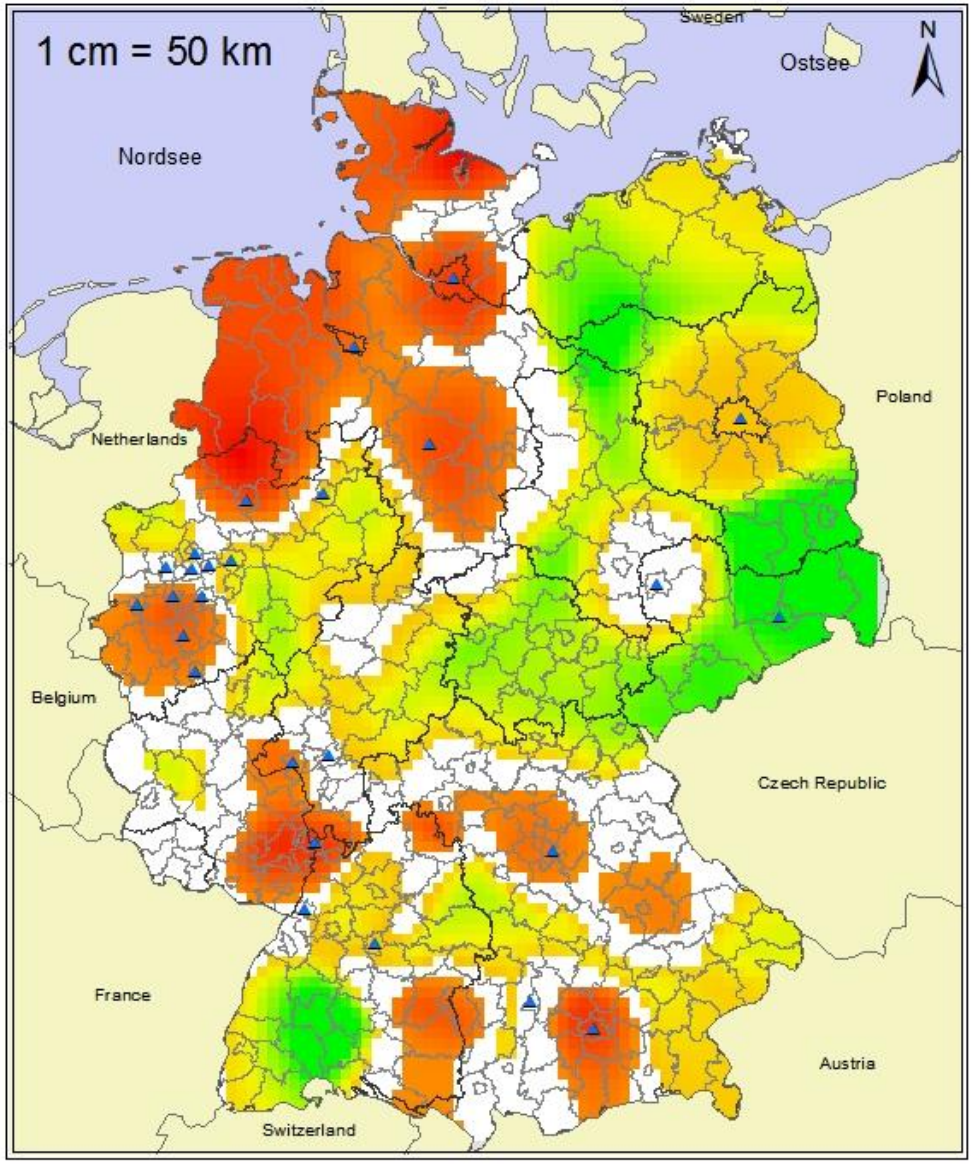
optimal bandwidth
for overall costs

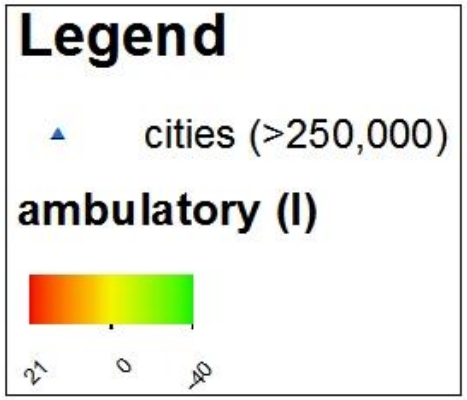


Costs	optimal bandwidth (N)
ambulatory	64,000
stationary	82,000
pharmaceutic	60,000
overall	66,000

WIG2 Wissenschaftliches Institut
für Gesundheitsökonomie und
Gesundheitssystemforschung

Ambulatory inequity between need and utilisation





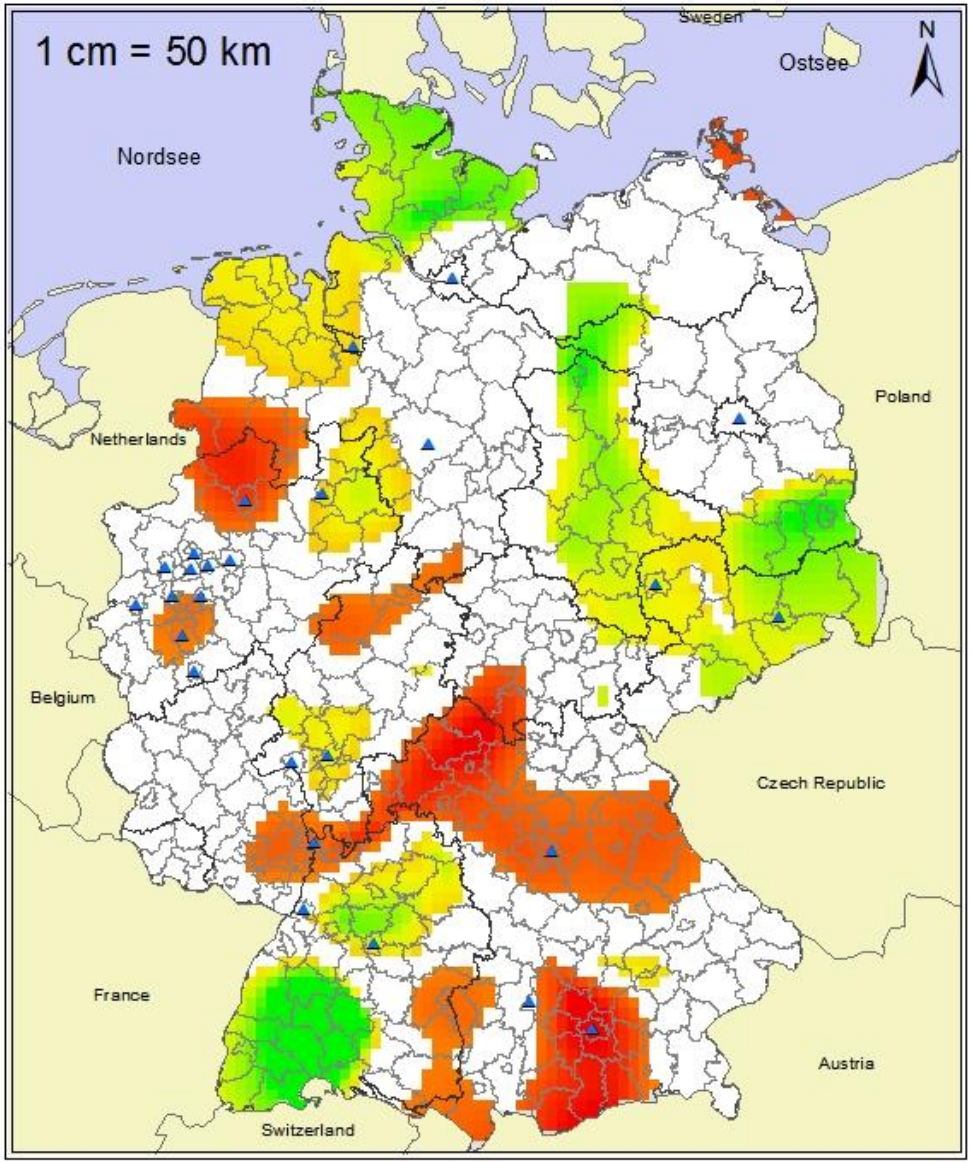
Red colour
means more utilisation than expected need. It refers to an oversupply for the sick and a negative marginal return for statutory sickness funds. A number of 10 means 10 Euro more utilisation than expected.

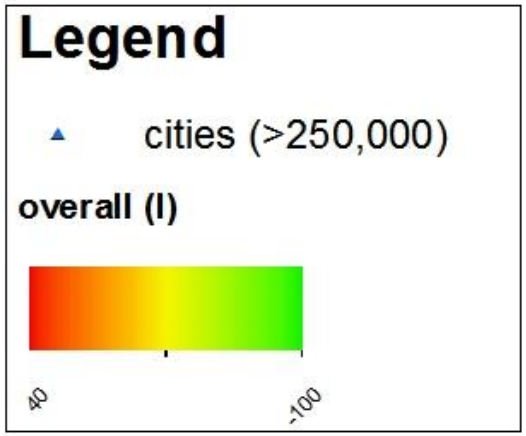
Green colour
means more expected need then the actual utilisation. It refers to an undersupply for the sick and a positiv marginal return for statutory sickness funds. A number of -10 means 10 Euro less utilisation than expected.

data: cost p.c. of 1,5 million observations
method: GWR
kernel: adaptive bi-squared
parametrisation: N=60,000

date: ETRS 1989 LAEA
author: Danny Wende

Overall inequity between need and utilisation





Red colour
means more utilisation than expected need. It refers to an oversupply for the sick and a negative marginal return for statutory sickness funds. A number of 10 means 10 Euro more utilisation than expected.

Green colour
means more expected need then the actual utilisation. It refers to an undersupply for the sick and a positiv marginal return for statutory sickness funds. A number of -10 means 10 Euro less utilisation than expected.

data: cost p.c. of 1,5 million observations
method: GWR
kernel: adaptive bi-squared
parametrisation: N=66,000

date: ETRS 1989 LAEA
author: Danny Wende

Spatial nonstationarity

The same stimulus provokes a different response in different parts of the study region. (Fotheringham et al. 2002 p. 9)

Because of:

- Sampling variation
- Relationships intrinsically different across space (difference in preferences, supply, or different administrative, political or other contextual effects produce different responses to the same stimuli)
- Model misspecification (missing variables, misspecification of the relationship)

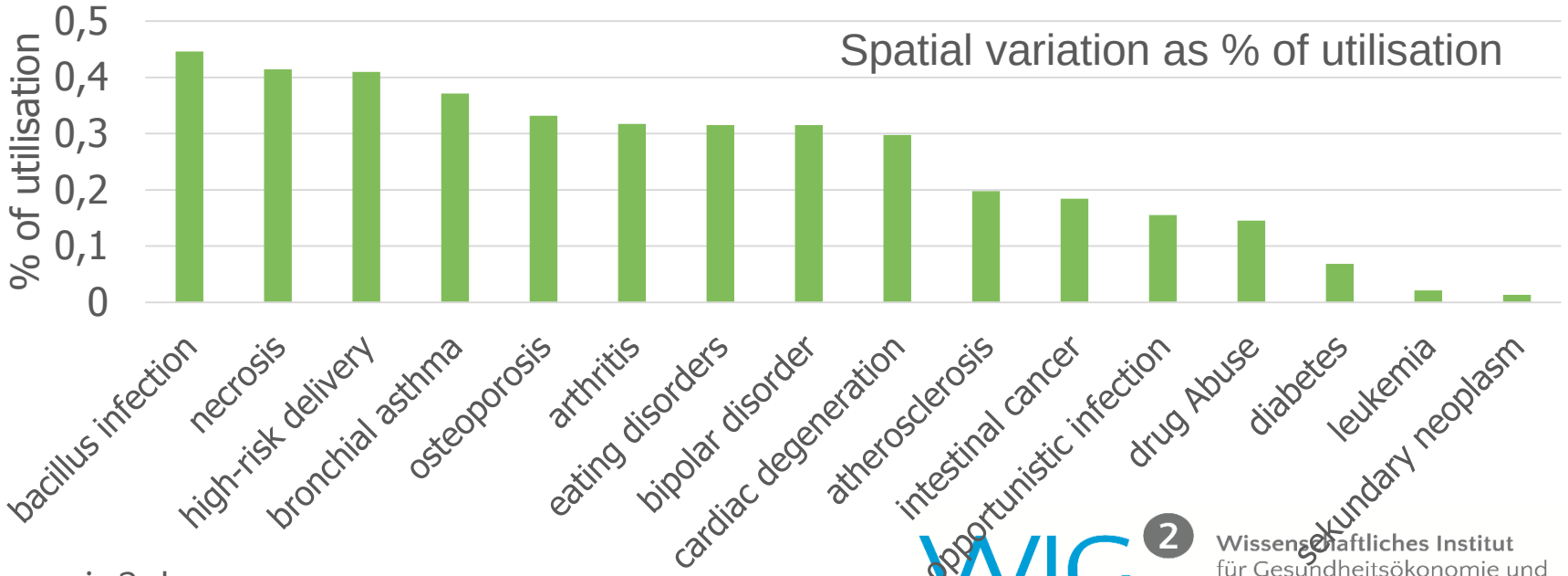
First Results

Expanding the model for spatial nonstationarity in treatment

$$S_i = \mathrm{EN_i}(u_i, v_i) + \mathrm{e_i} \qquad / \qquad \boldsymbol{S} = \beta(u_i, v_i) * \boldsymbol{X} + E$$

$$\widehat{\boldsymbol{\beta}}(u_j, v_j) = \left(\boldsymbol{X}'\boldsymbol{W}(u_j, v_j)\,\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{S}$$

- We treat each disease as treatment for spatial varying utilization.



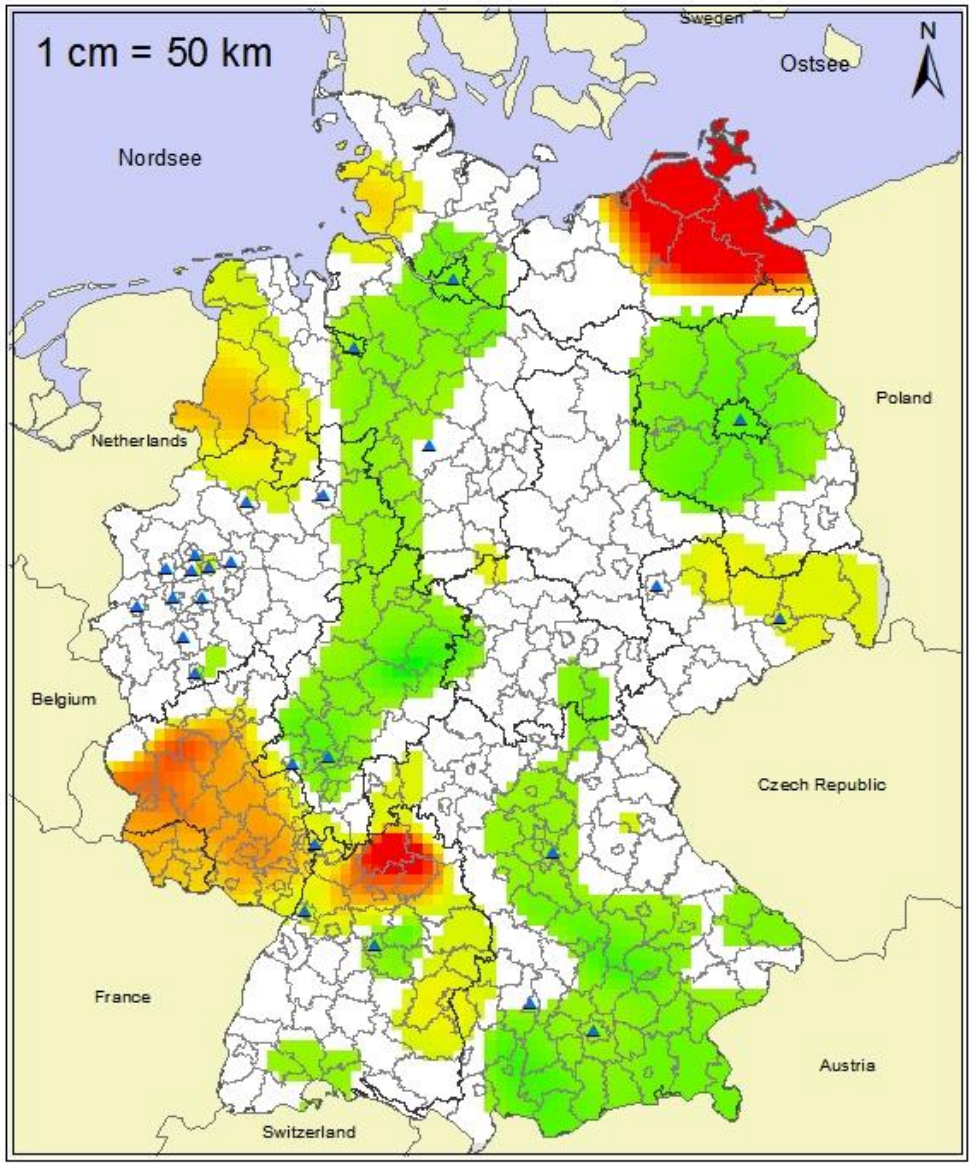
Concluding Remarks

- Using distances instead of zones helps to address MAUP.
  - Relevant spatial structures could be found and measured.
  - Windfall gains of supply could include in the mapping.
  - Treatment effects of diseases vary up to 0.5% of the total utilisation which could be over 100 Euro per capita.
- The common definition of administrative zones is far to determinate for analysing regional differences in health care utilisation.
  - regional effect going to be lost, as we assume independent observations for each zone.
- The definition of diseases could be to conservative in the RSCS, as we see spatial clustering which could come up from unobserved diseases and deprivation

References

- **Culyer, Anthony J., and Adam Wagstaff**. "Equity and equality in health and health care." *Journal of health economics* 12.4 (1993): 431-457.
- **Drösler, S., Hasford, J., Kurth, B.-M., Schaefer, M., Wasem, J., & Wille, E. (2011)**: *Evaluationsbericht zum Jahresausgleich 2009 im Risikostrukturausgleich*.
- **Jacobs, K., P. Reschke und J. Wasem (1998)**: Zur funktionalen Abgrenzung von Beitragssatzregionen in der gesetzlichen Krankenversicherung. Baden-Baden.
- **Jacobs, K. (2007)**: Solidarität geht über Ländergrenzen. *Gesundheit Und Gesellschaft*, *14*(11), 14–15.
- **Göpffarth, D., Greß, S., Jacobs, K., & Wasem, J. (Hrsg.) (2007)**: Jahrbuch Risikostrukturausgleich 2007. Asgard-Verlag, Sankt Augustin 2007.
- **Göpffarth, D. (2011)**: Regionalmerkmale im Risikostrukturausgleich. Ein Beitrag zum funktionalen Wettbewerb und zu bedarfsgerechter Versorgung? *BARMER GEK Gesundheitswesen Aktuell*, *2011*, 16–127.
- **Fan, J. and Gijbels, I**. (1996). Local Polynomial Modelling and Its Applications, Chapman & Hall. London.
- **Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2003)**. *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons
- **Mei, C. L., He, S. Y., & Fang, K. T. (2004)**. A Note on the Mixed Geographically Weighted Regression Model*. *Journal of Regional Science*, *44*(1), 143-157.
- **Nakaya, T., Fotheringham, A. S., Charlton, M., & Brunsdon, C. (2009).** Semiparametric geographically weighted generalised linear modelling in GWR 4.0.
- **Ozegowski, Susanne, and Leonie Sundmacher (2014)**. "Understanding the gap between need and utilization in outpatient care—The effect of supply-side determinants on regional inequities." *Health policy* 114.1: 54-63.
- **Wasem, J., S. Greß & K. Jacobs (2007)**: Gesundheitsfonds und Regionaldebatte I: Ordnungspolitischer Rahmen. In: D. Göpffarth, S. Greß, J. Jacobs und J. Wasem (Hrsg.): Jahrbuch Risikostrukturausgleich 2007. Sankt Augustin. S. 139-162.
- **Wille, E., & U. Schneider (1997)**: Zur Regionalisierung in der gesetzlichen Krankenversicherung. In: Recht und Politik im Gesundheitswesen. 2 (4). S. 141-158.
- **Yan, Xin. (2009).** Linear regression analysis : theory and computing. Published by World Scientific Publishing Co. Pte. Ltd. 5 Toh Tuck Link, Singapore 596224.

pharmacy inequity between need and utilisation



Legend

cities (>250,000)

pharmacy (l)

16 0 8

Red colour
means more utilisation than expected need. It refers to an oversupply for the sick and a negative marginal return for statutory sickness funds. A number of 10 means 10 Euro more utilisation than expected.

Green colour
means more expected need then the actual utilisation. It refers to an undersupply for the sick and a positiv marginal return for statutory sickness funds. A number of -10 means 10 Euro less utilisation than expected.

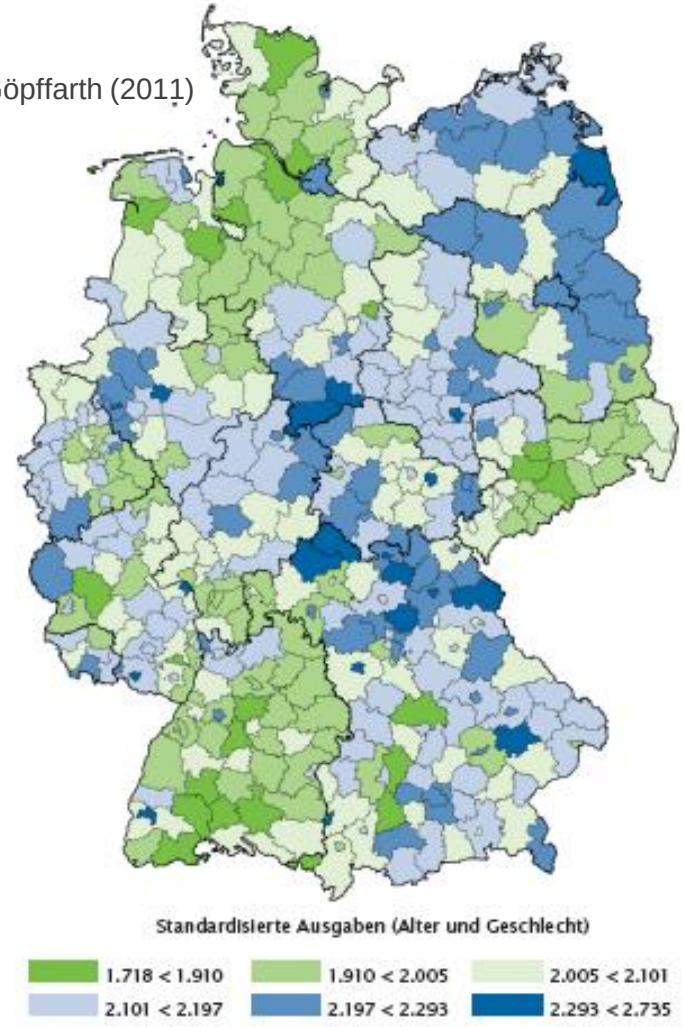
data: cost p.c. of 1,5 million observations
method: GWR
kernel: adaptive bi-squared
parametrisation: N=60,000

date: ETRS 1989 LAEA
author: Danny Wende

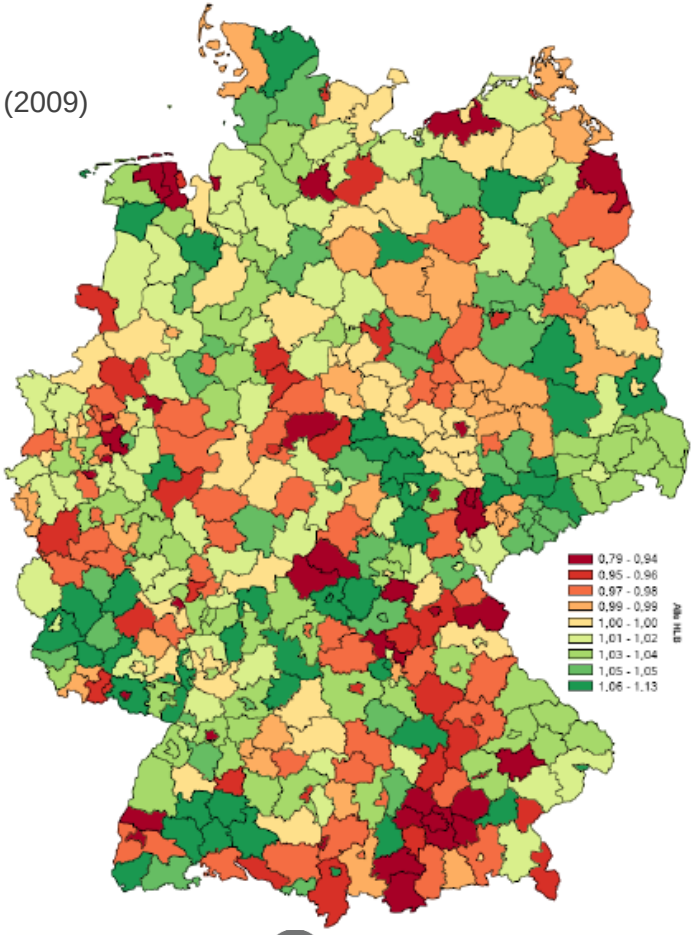
The Zone Definition Problem in Small Area Variation

Examples of MAUP / standardised health care costs and coverage ratios

Göpffarth (2011)



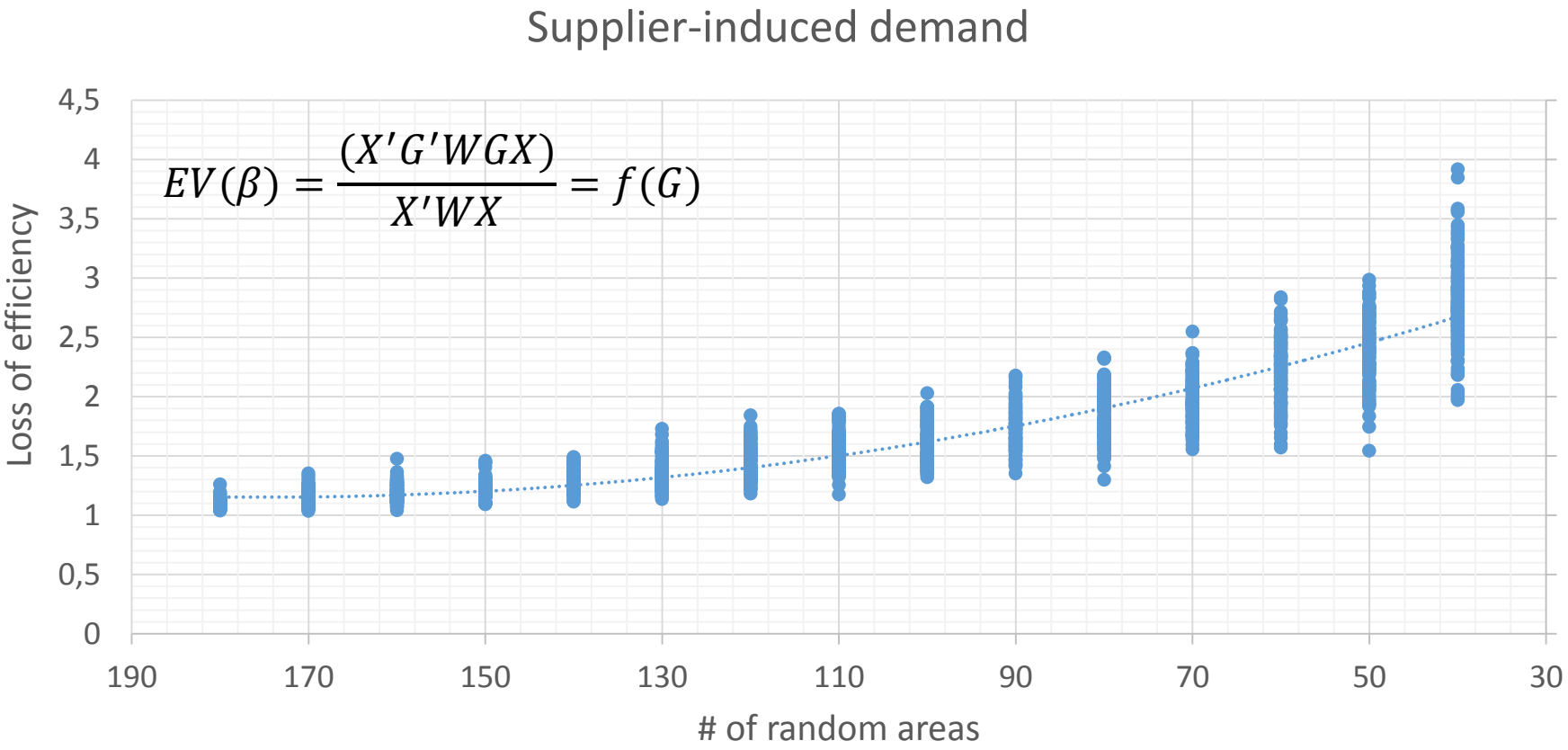
Drösler et al. (2009)



WIG2 Wissenschaftliches Institut für Gesundheitsökonomie und Gesundheitssystemforschung

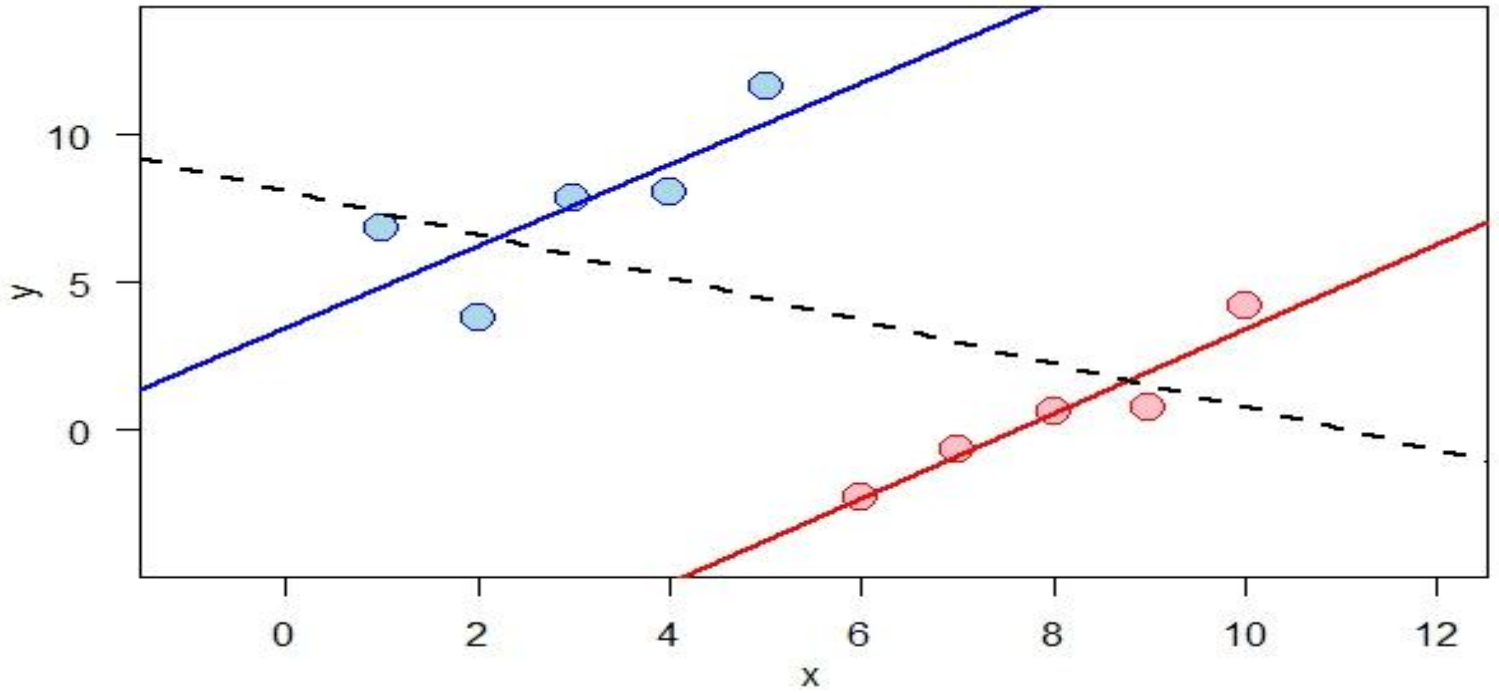
The Zone Definition Problem in Small Area Variation

The Multiple Area Unit Problem (MAUP)



The Zone Definition Problem in Small Area Variation

The Yule-Simpson-effect



$$E[Y_{red}|X = x, Z = z] = E[Y_{blue}|X = x, Z = z]$$

$$E[Y_{red}|X = x] \neq E[Y_{blue}|X = x]$$

WIG2 Wissenschaftliches Institut für Gesundheitsökonomie und Gesundheitssystemforschung