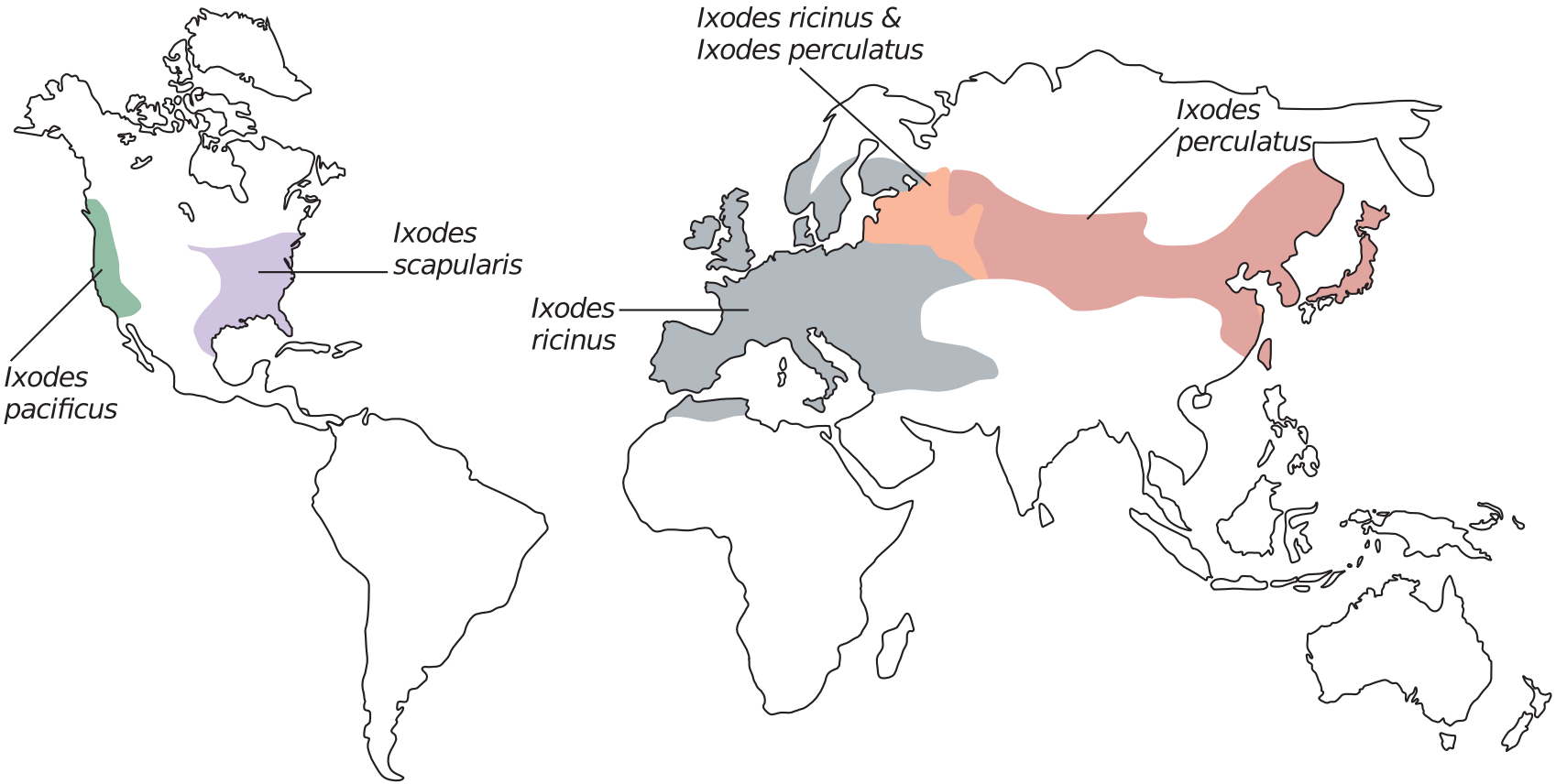
Spatial and environmental factors of vector spreading and distribution

Solymosi Norbert

3/10/2018

Environmental effects

- Geographical distribution
  - Vector (poikilotherm)
  - Agent of vector-borne disease



Stanek et al., *The Lancet*

Epidemiology is the study of disease in populations and of factors that determine its occurrence; the key word being **populations**.

Epidemiology is concerned with the prevention and control of disease in human and animal **populations**. Veterinary epidemiology additionally includes the investigation and assessment of other health-related events, notably **productivity**.

Ojectives of epidemiology:

- determination of the origin of a disease whose cause is known
- investigation and control of a disease whose cause is either unknown or poorly understood
- acquisition of information on the ecology and natural history of a disease
- planning, monitoring and assessment of disease control programmes
- assessment of the economic effects of a disease, and analysis of the costs and economic benefits of alternative control programmes

Incidence

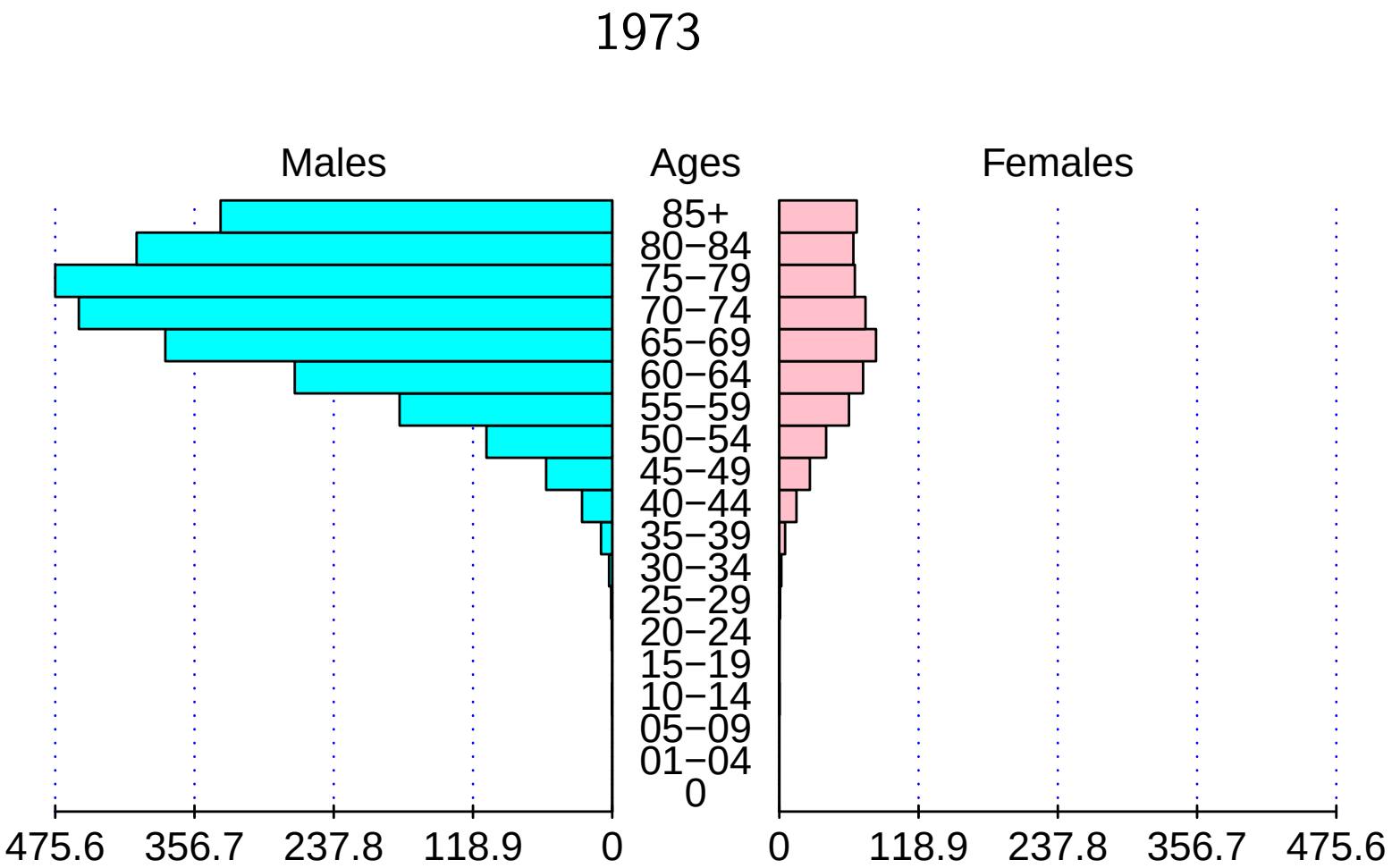
Incidence is the number of **new cases** that occur in a known population over a specified period of time. **Cumulative incidence**: the ratio between the number of animals that conctracted the disease in a certain period and the number of healthy animals at risk in the population at the start of that period.

Week	No. of new cases	CI
1	20	0.20
2	15	0.35
3	10	0.45
4	5	0.50
5	1	0.51

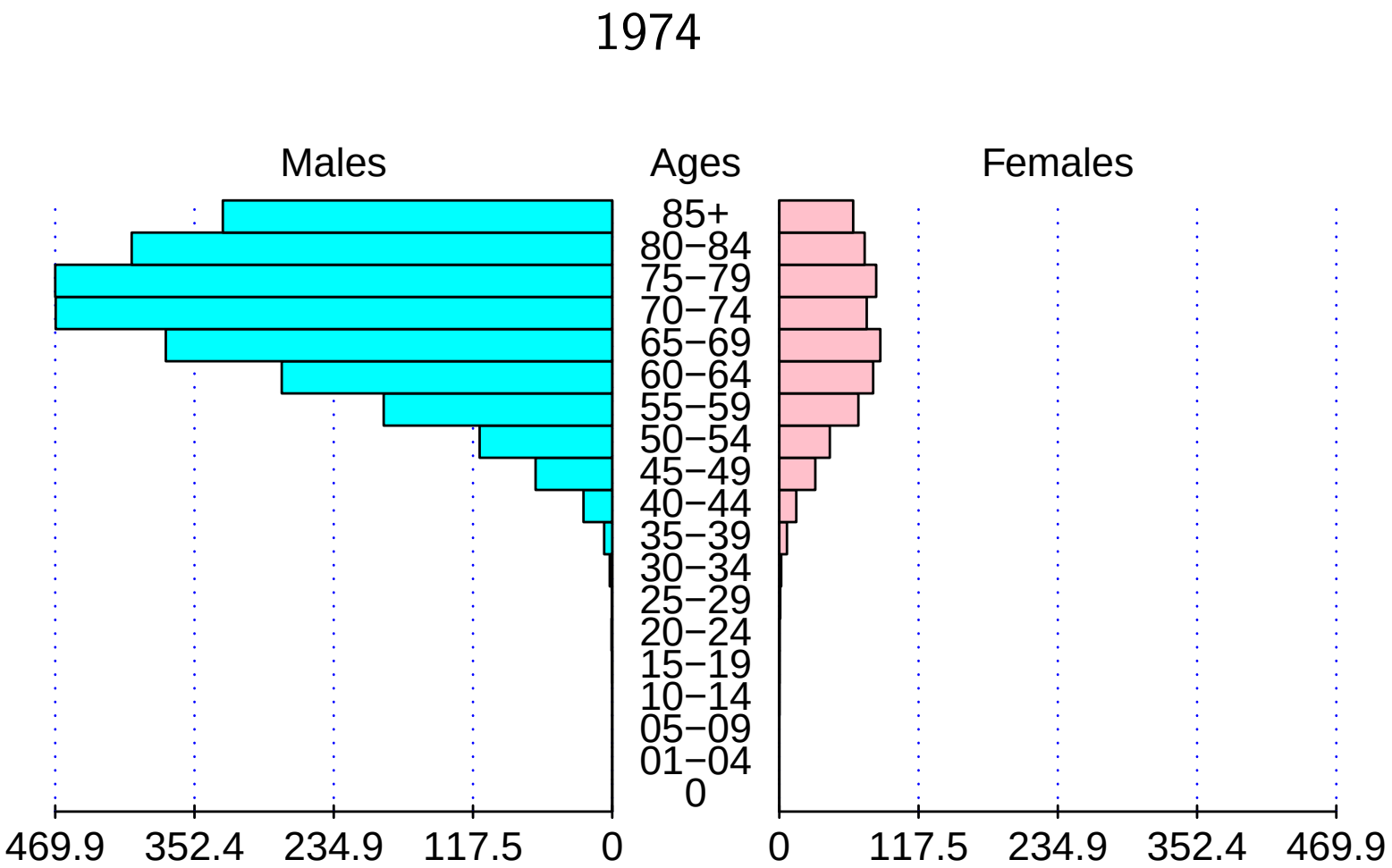
Incidence rate: the ratio between the number of new cases of disease in a population during certain period and the sum of the time-units at risk for all animals in the population at risk.

$$\frac{51}{(20 * 0.5 + 15 * 1.5 + 10 * 2.5 + 5 * 3.5 + 1 * 4.5) + (49 * 5)} = 0.157$$

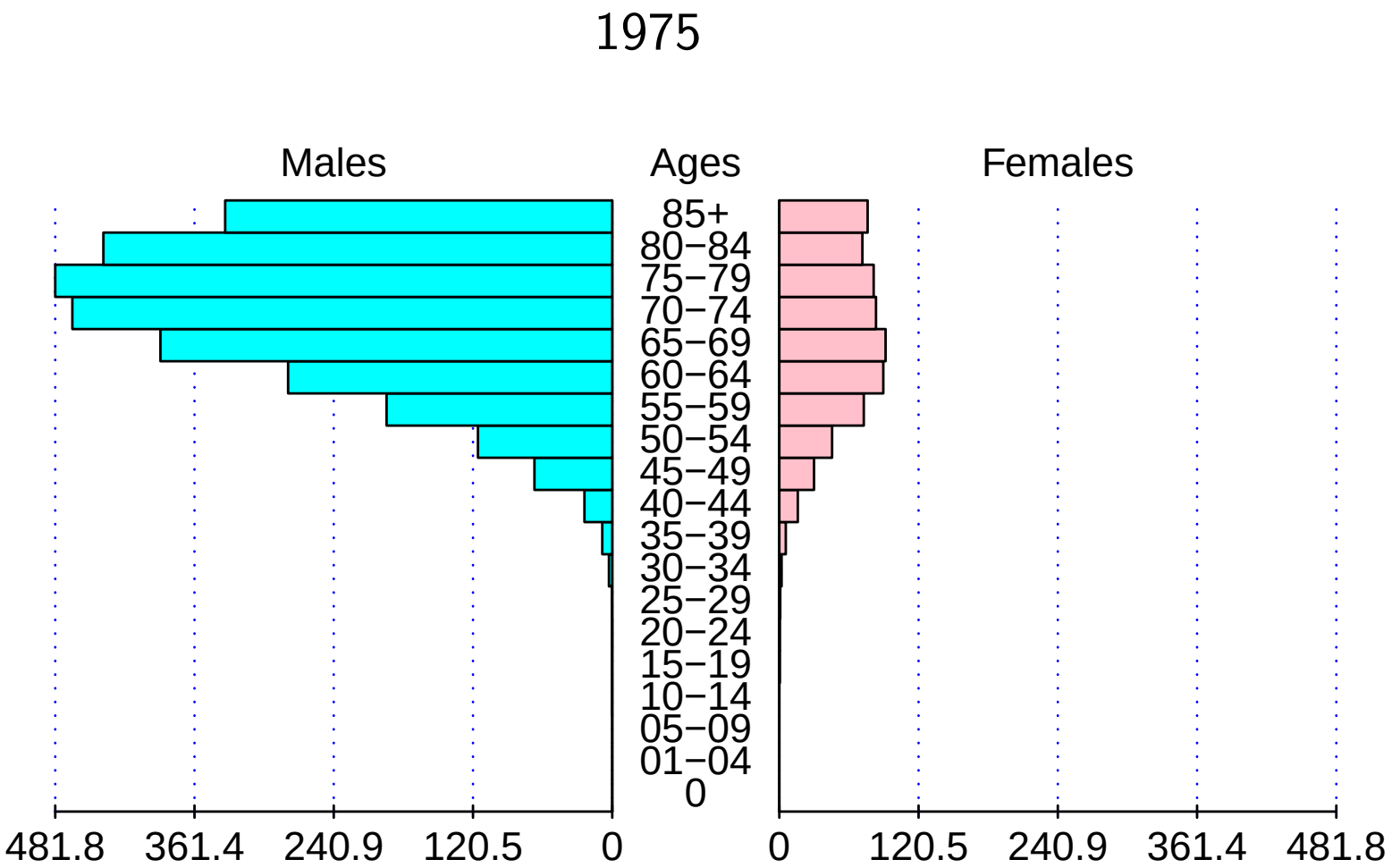
Incidence


1973
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
475.6
356.7
237.8
118.9
0
0
118.9
237.8
356.7
475.6

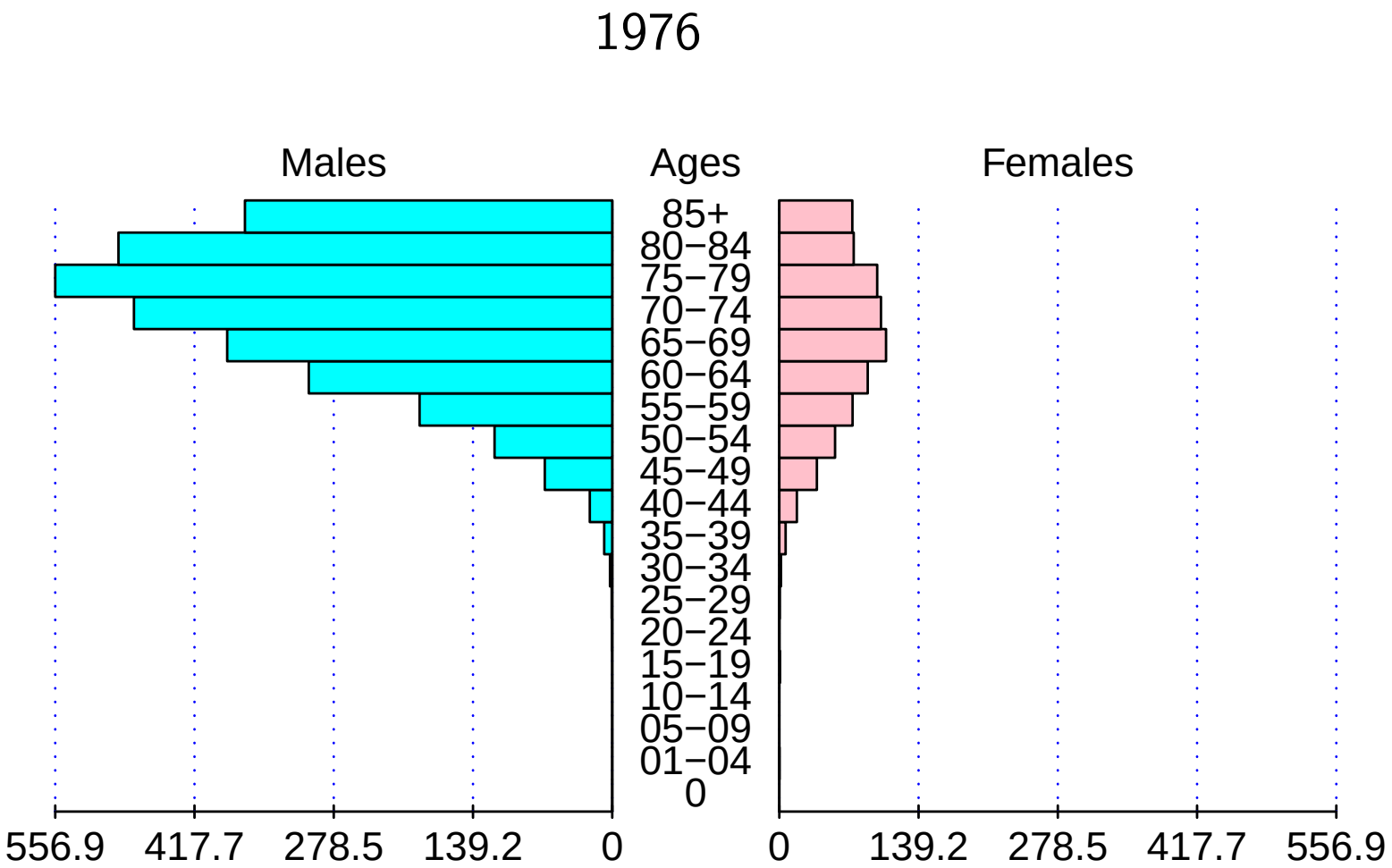
Incidence


1974
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
469.9
352.4
234.9
117.5
0
0
117.5
234.9
352.4
469.9

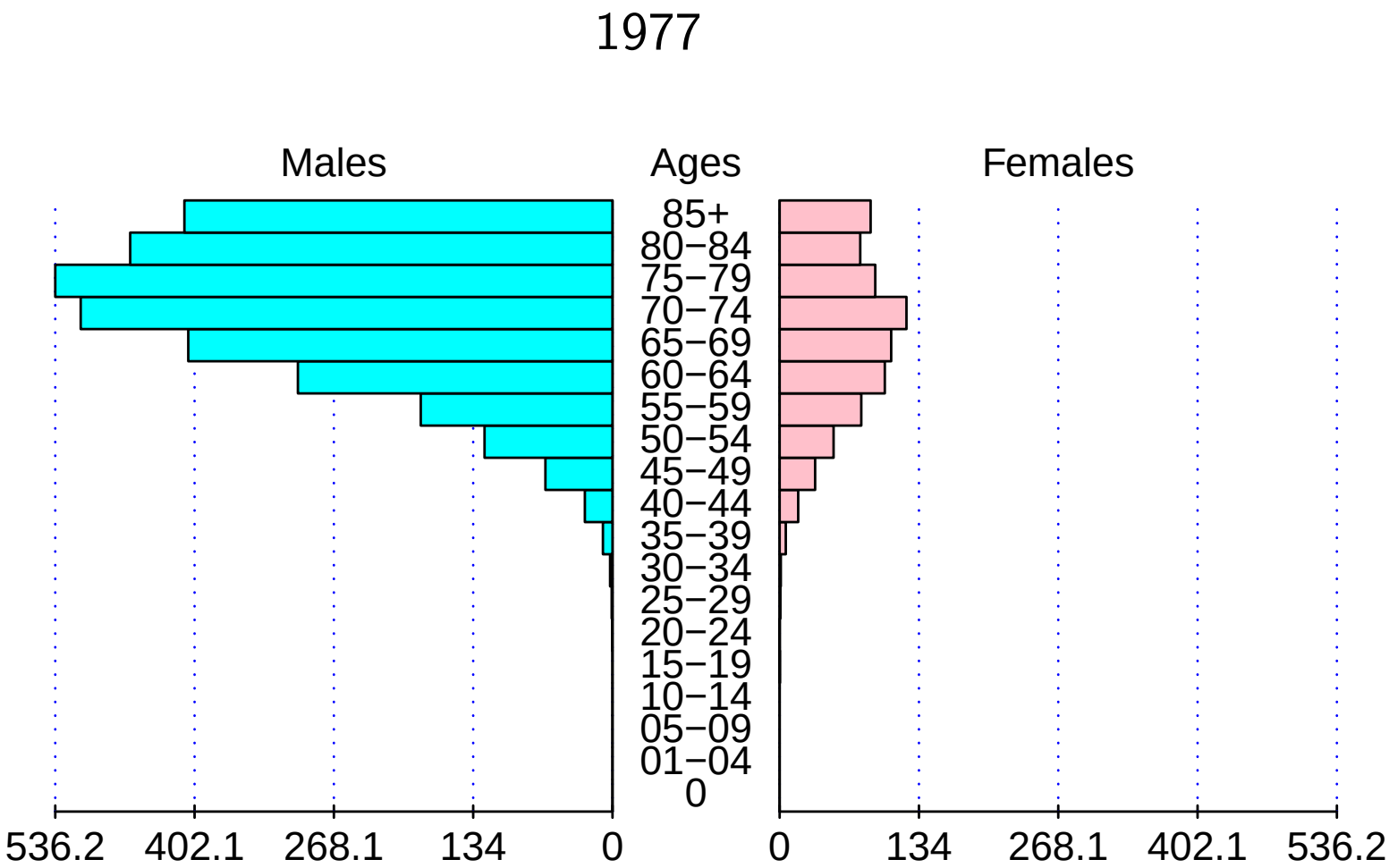
Incidence


1975
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
481.8
361.4
240.9
120.5
0
0
120.5
240.9
361.4
481.8

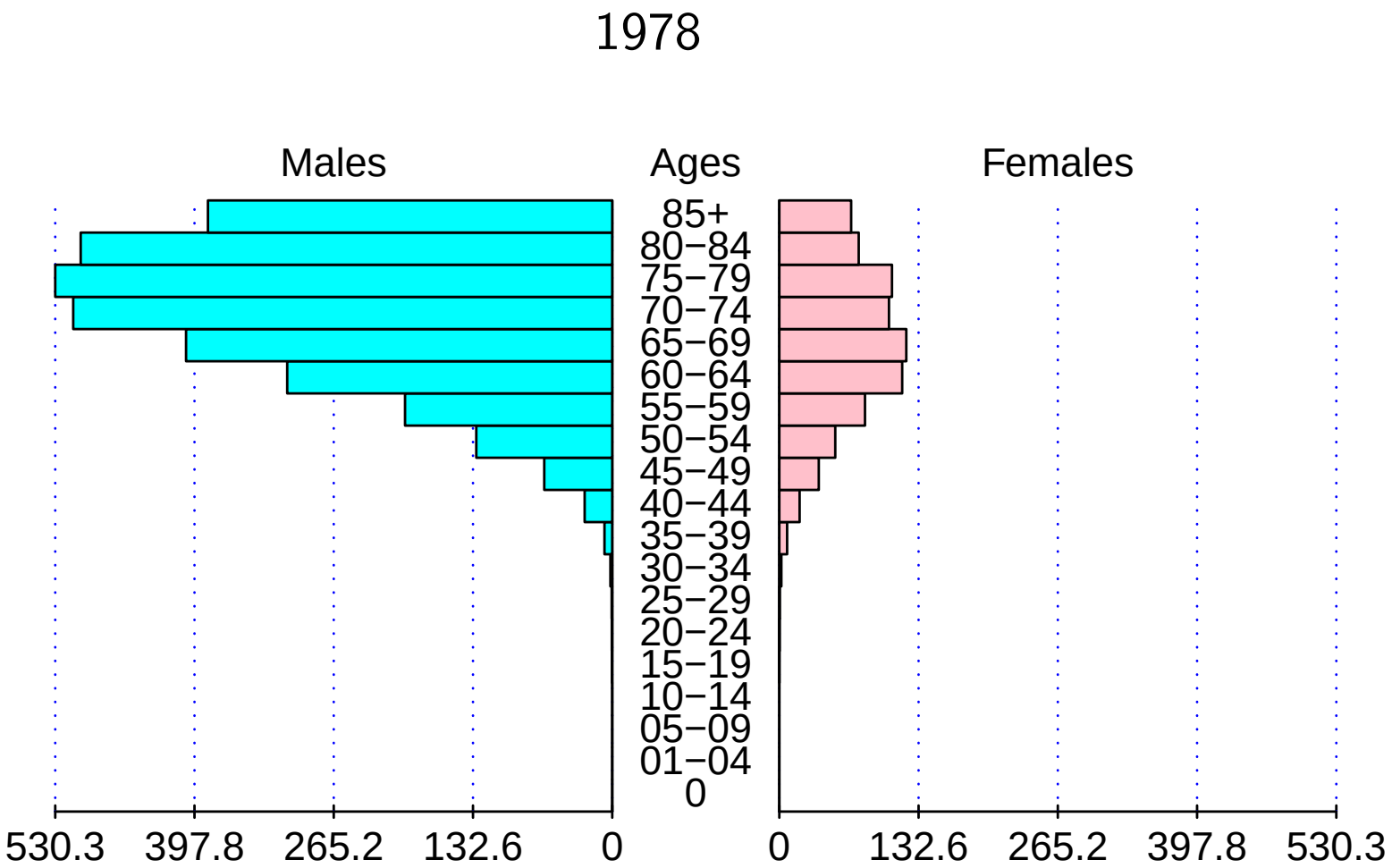
Incidence


1976
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
556.9
417.7
278.5
139.2
0
0
139.2
278.5
417.7
556.9

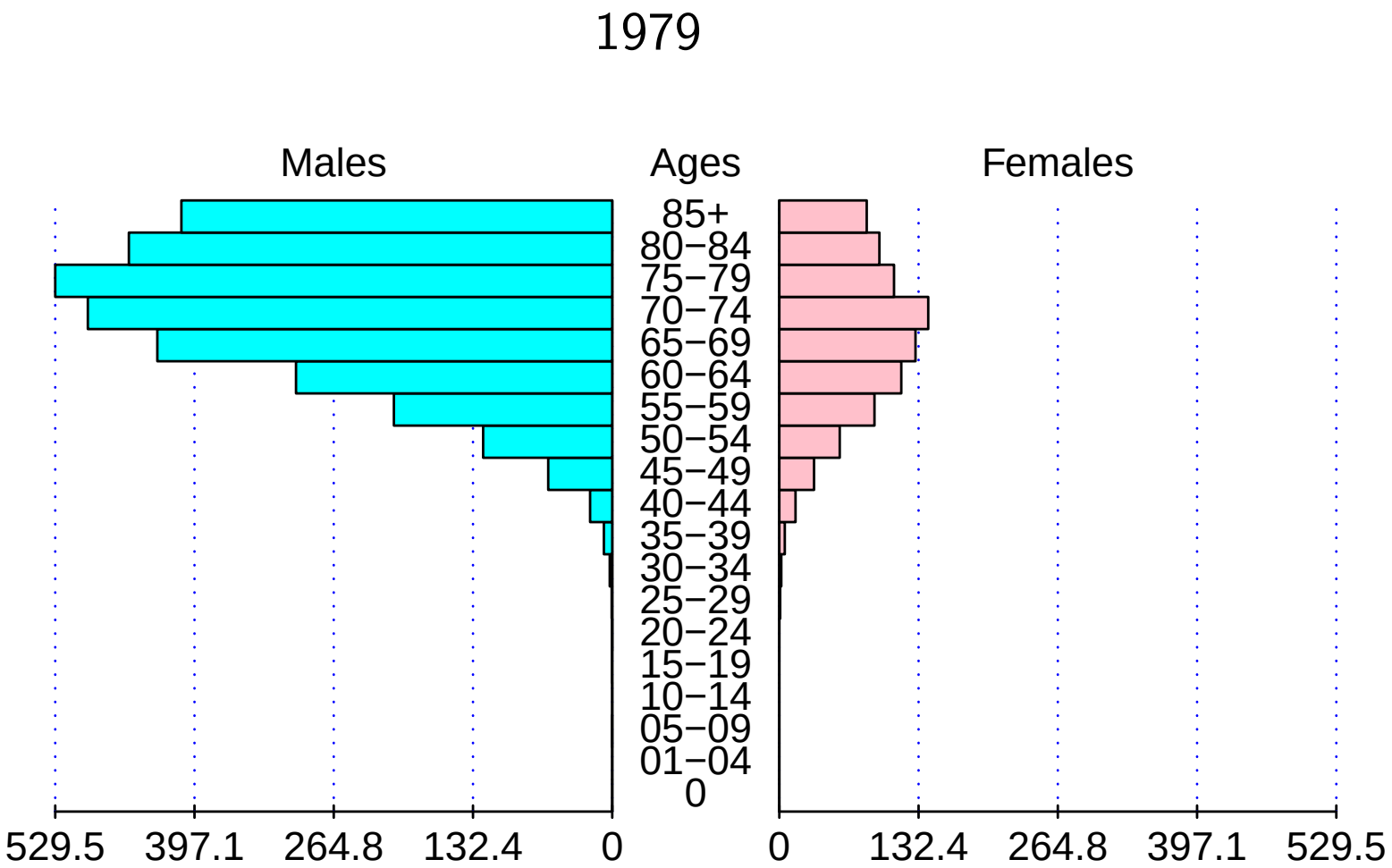
Incidence


1977
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
536.2
402.1
268.1
134
0
0
134
268.1
402.1
536.2

Incidence


1978
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
530.3
397.8
265.2
132.6
0
0
132.6
265.2
397.8
530.3

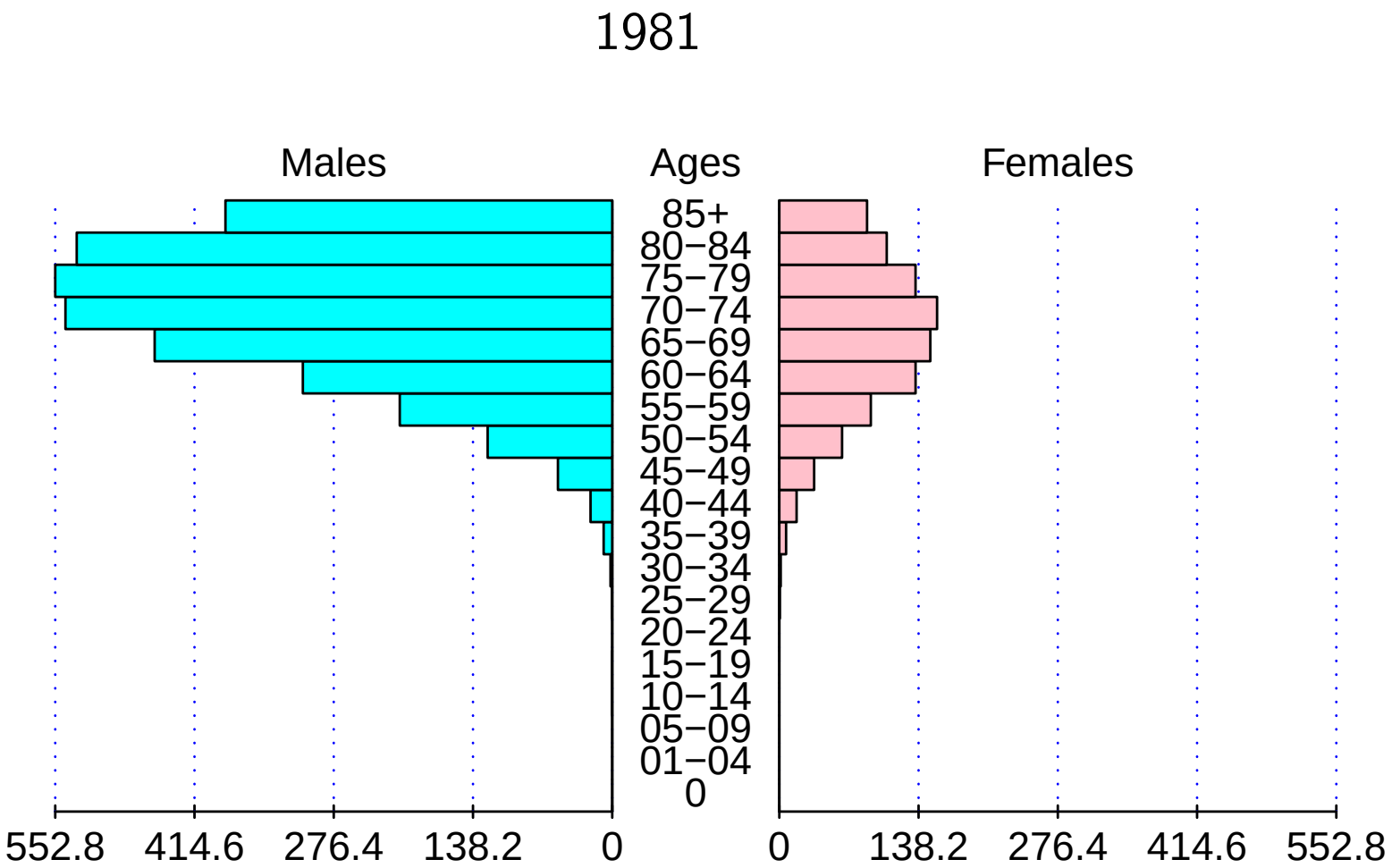
Incidence


1979
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
529.5
397.1
264.8
132.4
0
0
132.4
264.8
397.1
529.5

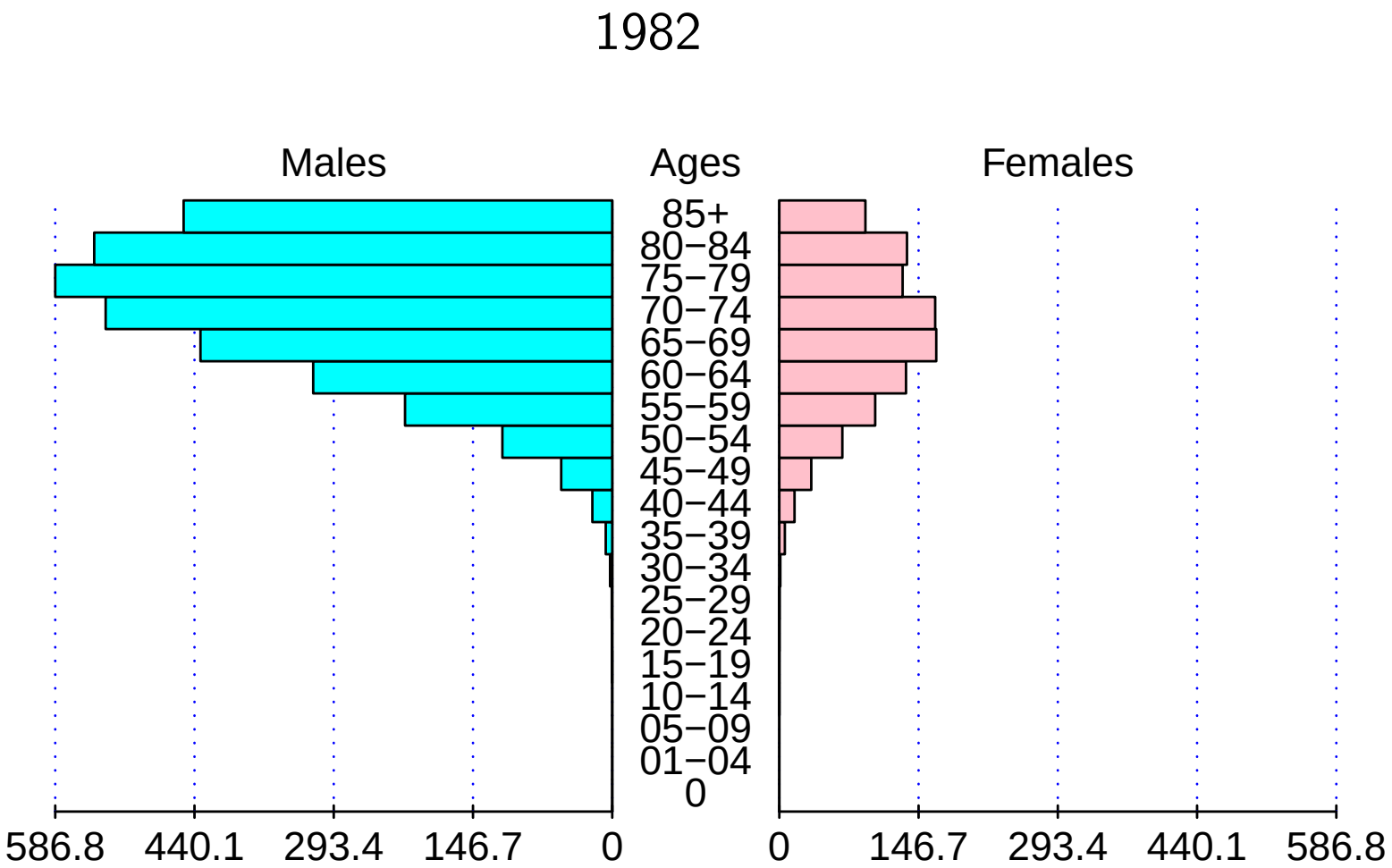
Incidence


1980
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
561.2
420.9
280.6
140.3
0
0
140.3
280.6
420.9
561.2

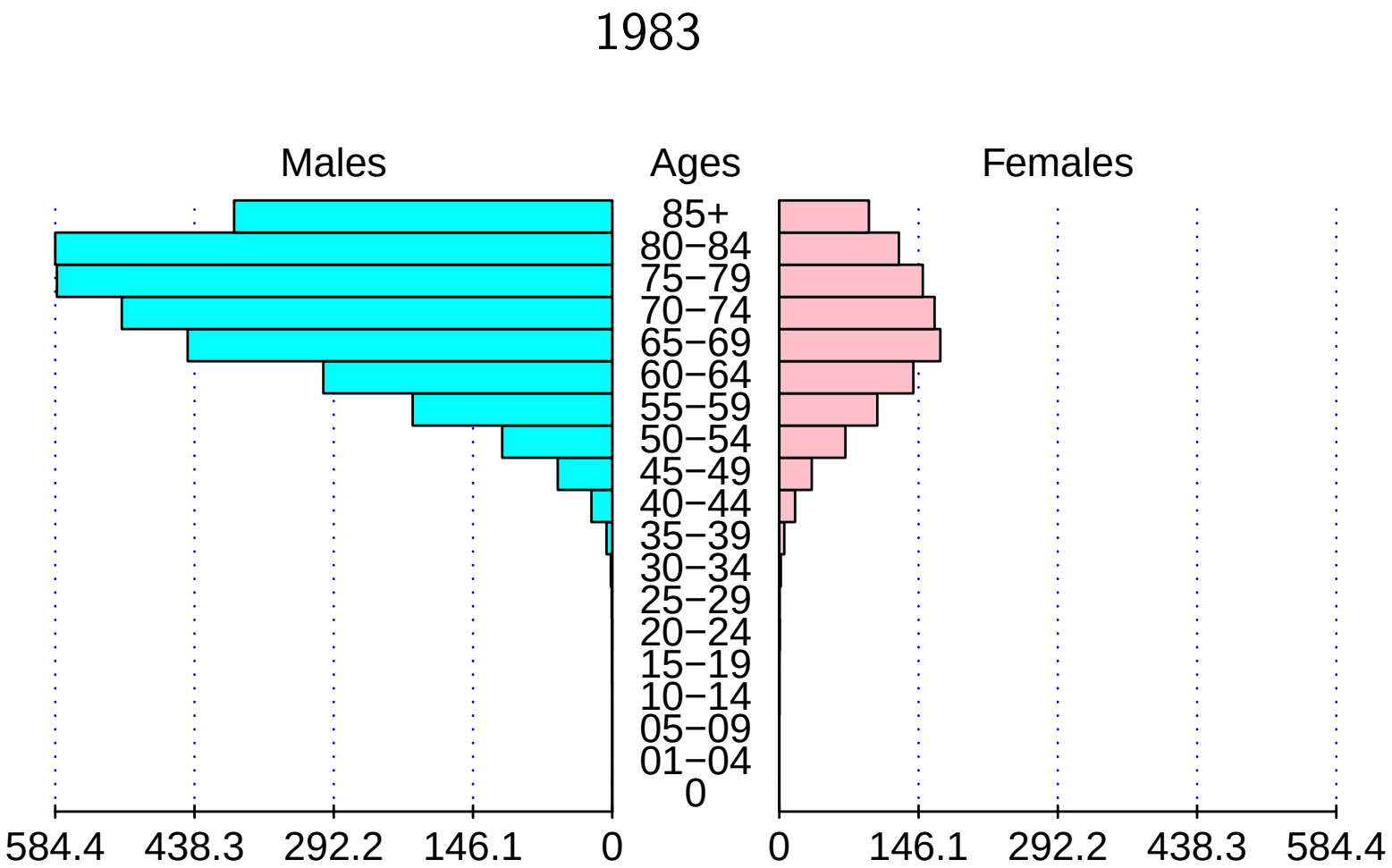
Incidence


1981
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
552.8
414.6
276.4
138.2
0
0
138.2
276.4
414.6
552.8

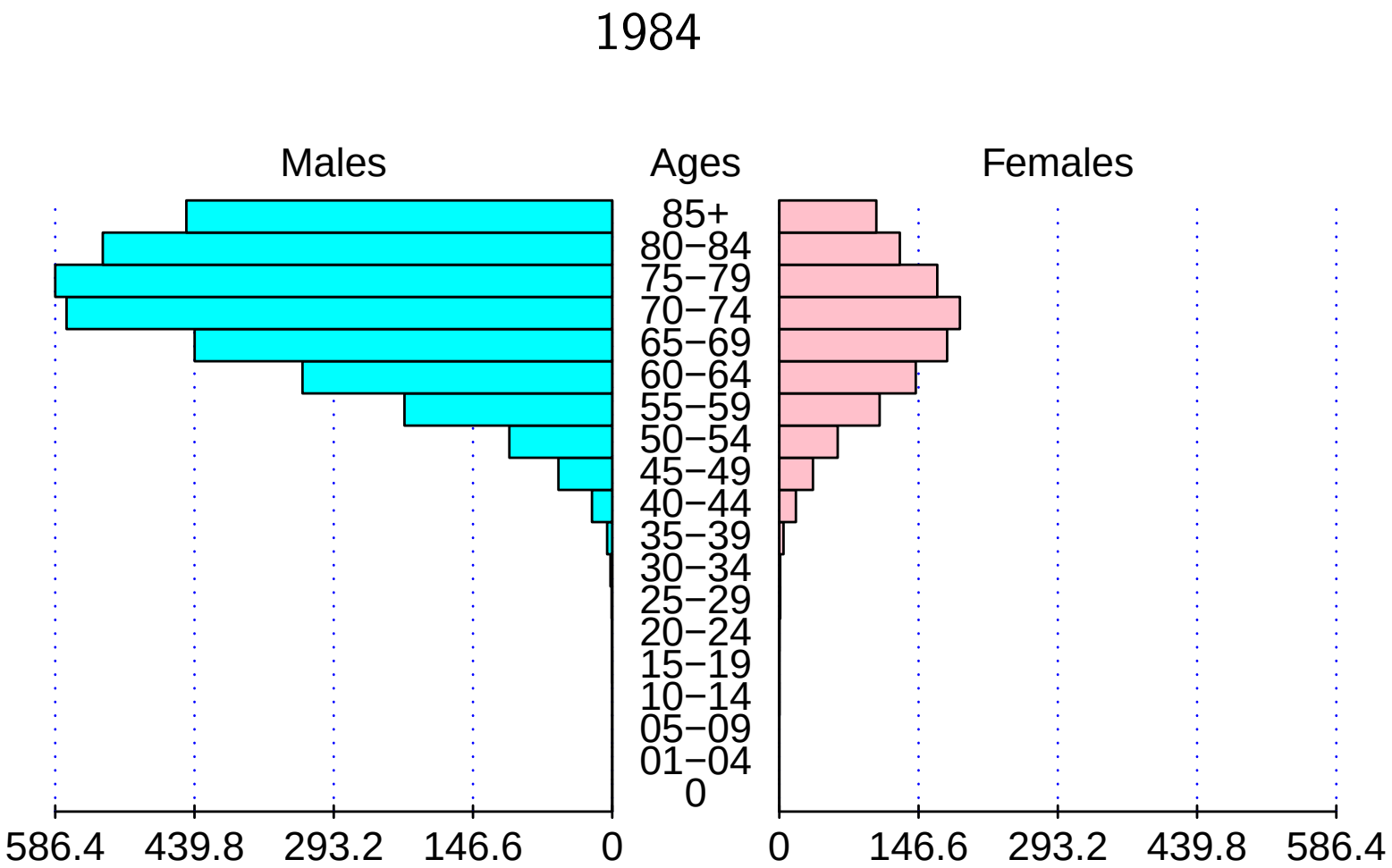
Incidence


1982
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
586.8
440.1
293.4
146.7
0
0
146.7
293.4
440.1
586.8

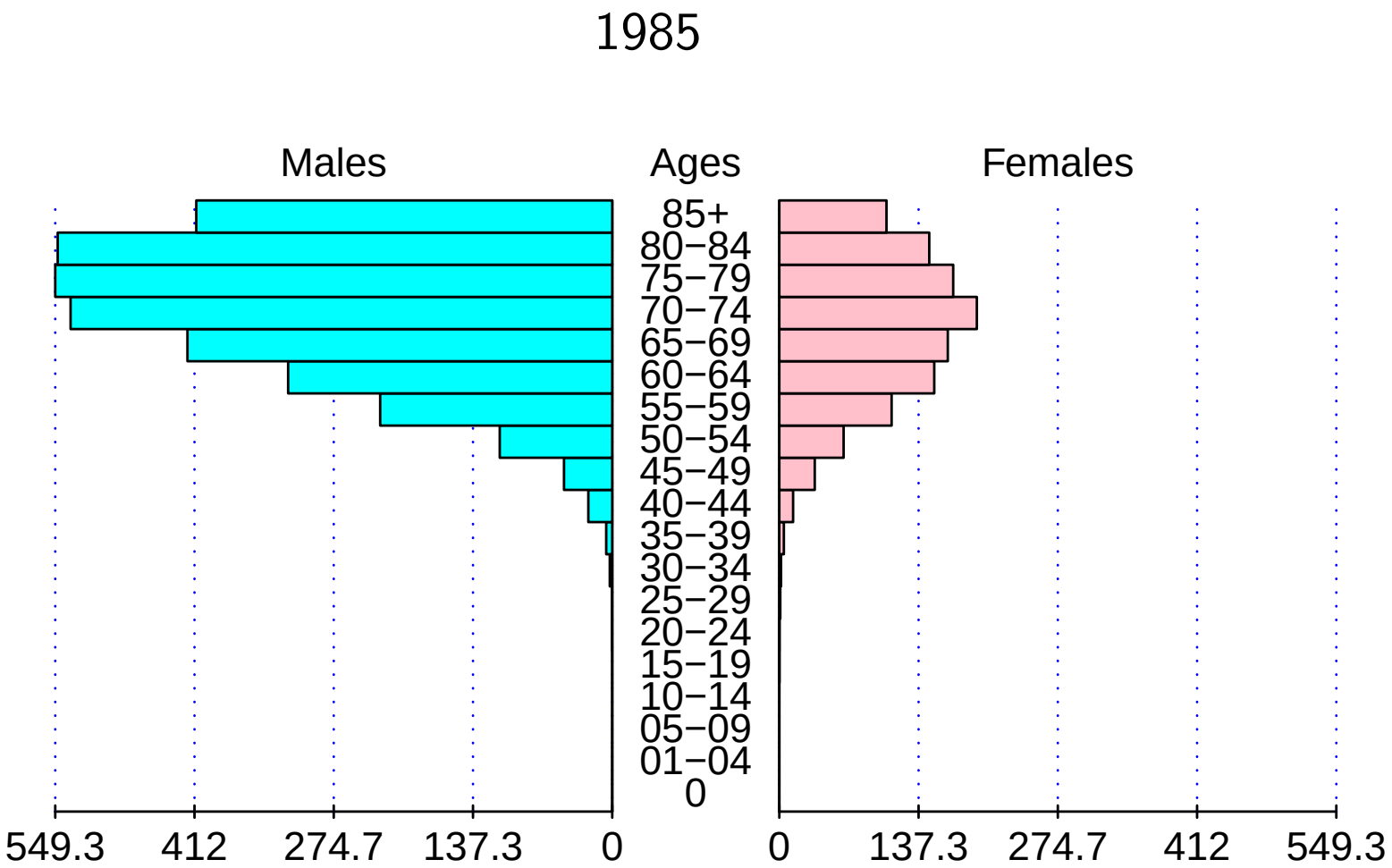
Incidence


1983
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
584.4
438.3
292.2
146.1
0
0
146.1
292.2
438.3
584.4

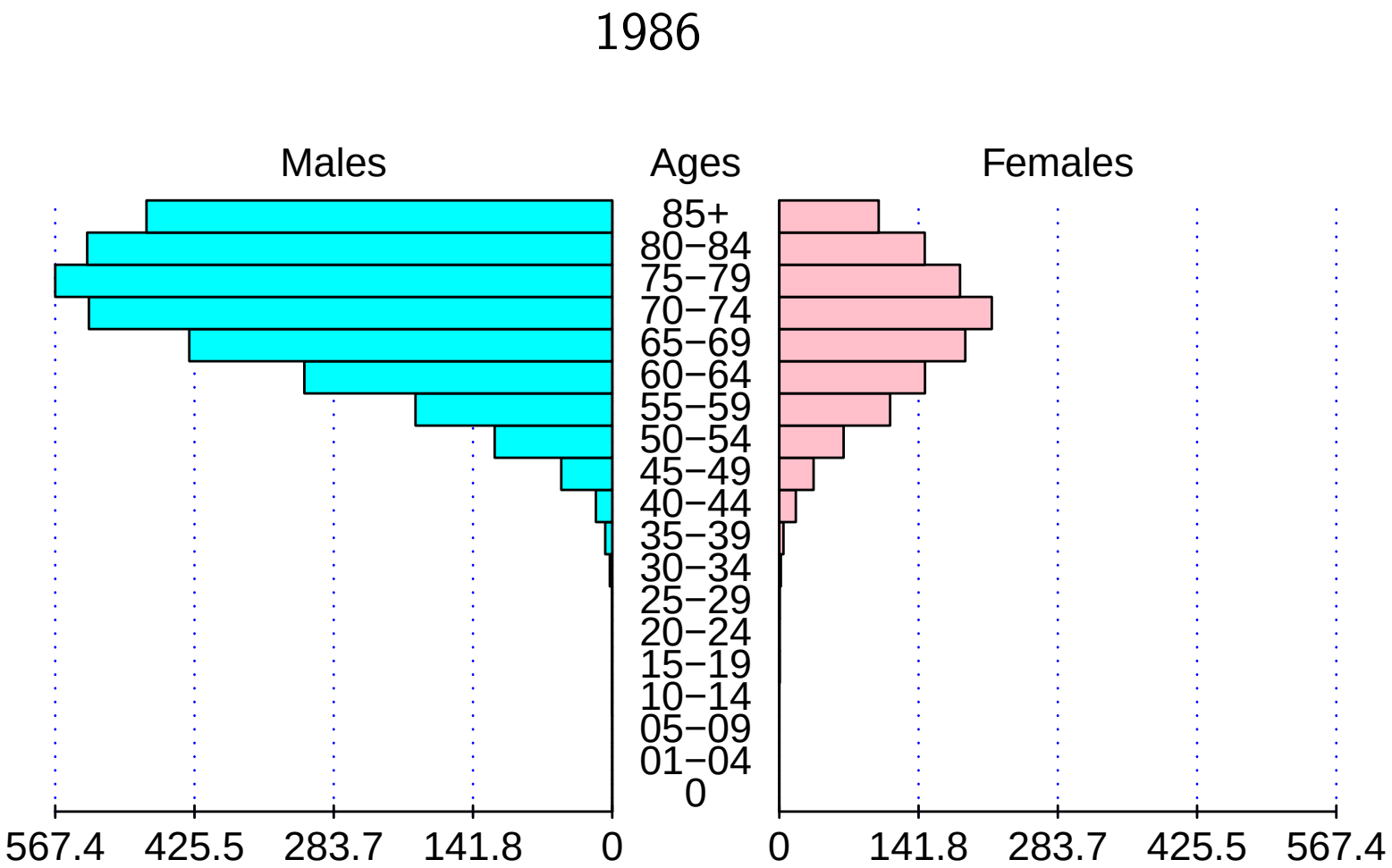
Incidence


1984
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
586.4
439.8
293.2
146.6
0
0
146.6
293.2
439.8
586.4

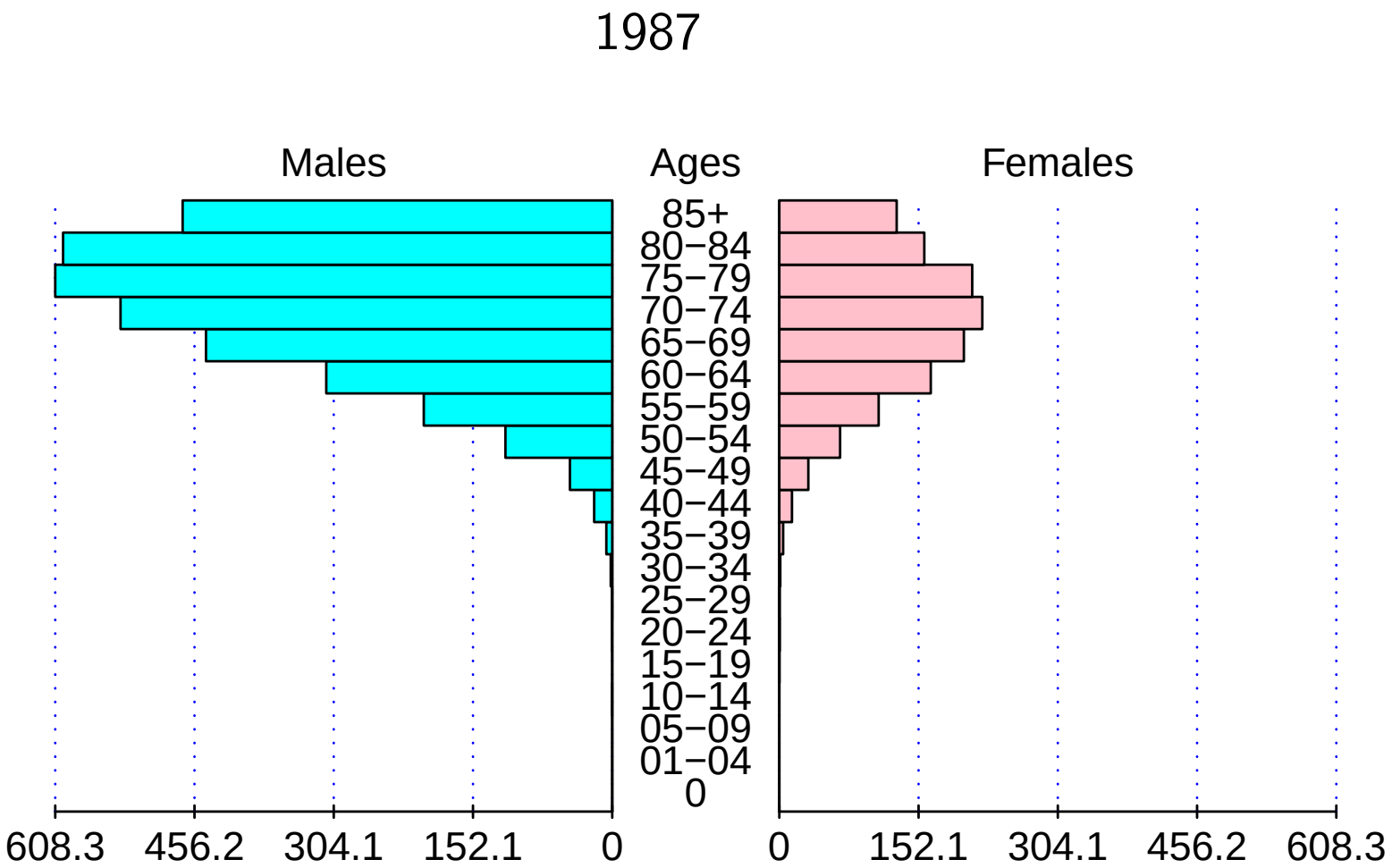
Incidence


1985
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
549.3
412
274.7
137.3
0
0
137.3
274.7
412
549.3

Incidence


1986
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
567.4
425.5
283.7
141.8
0
0
141.8
283.7
425.5
567.4

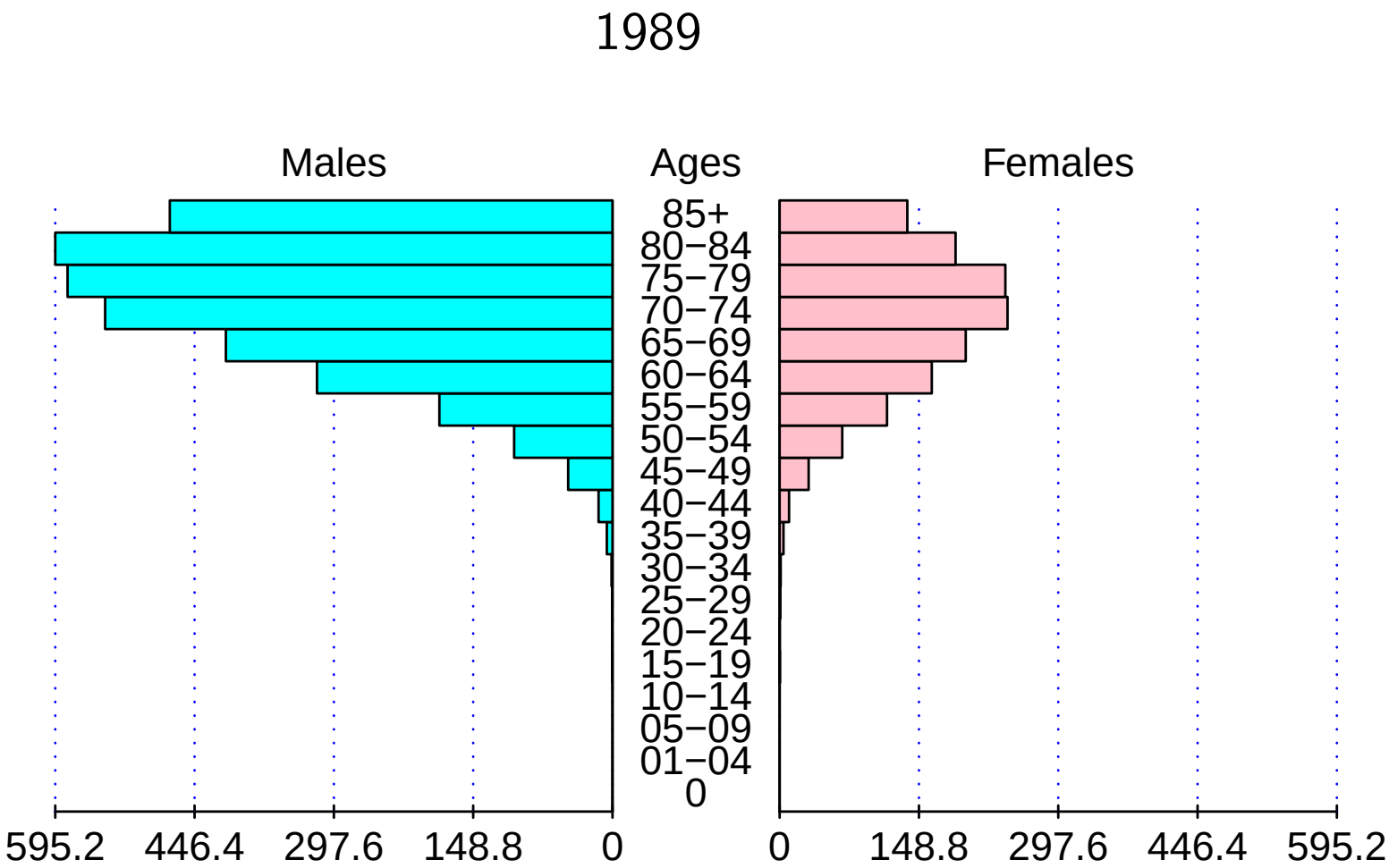
Incidence


1987
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
608.3
456.2
304.1
152.1
0
0
152.1
304.1
456.2
608.3

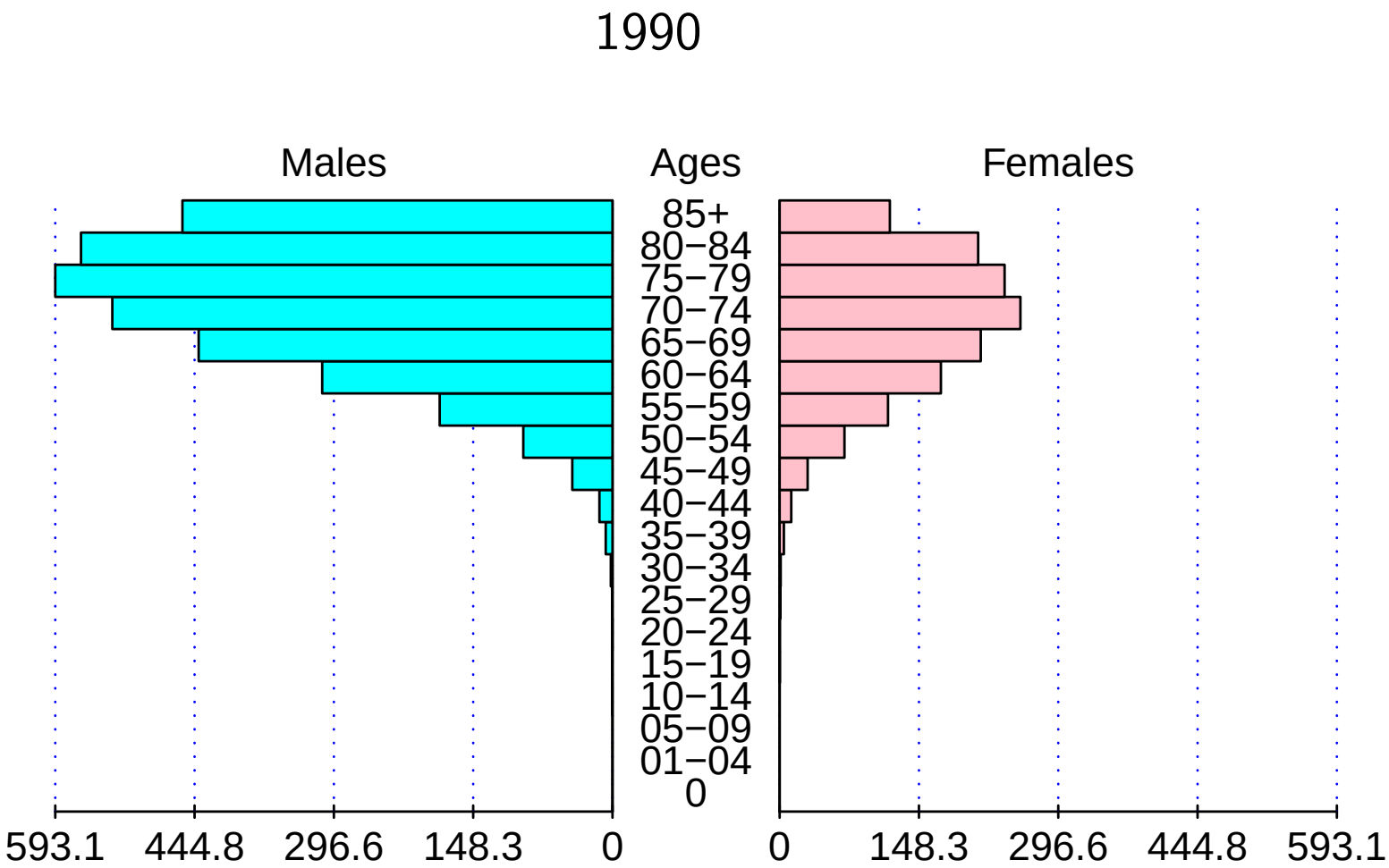
Incidence


1988
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
580.2
435.1
290.1
145
0
0
145
290.1
435.1
580.2

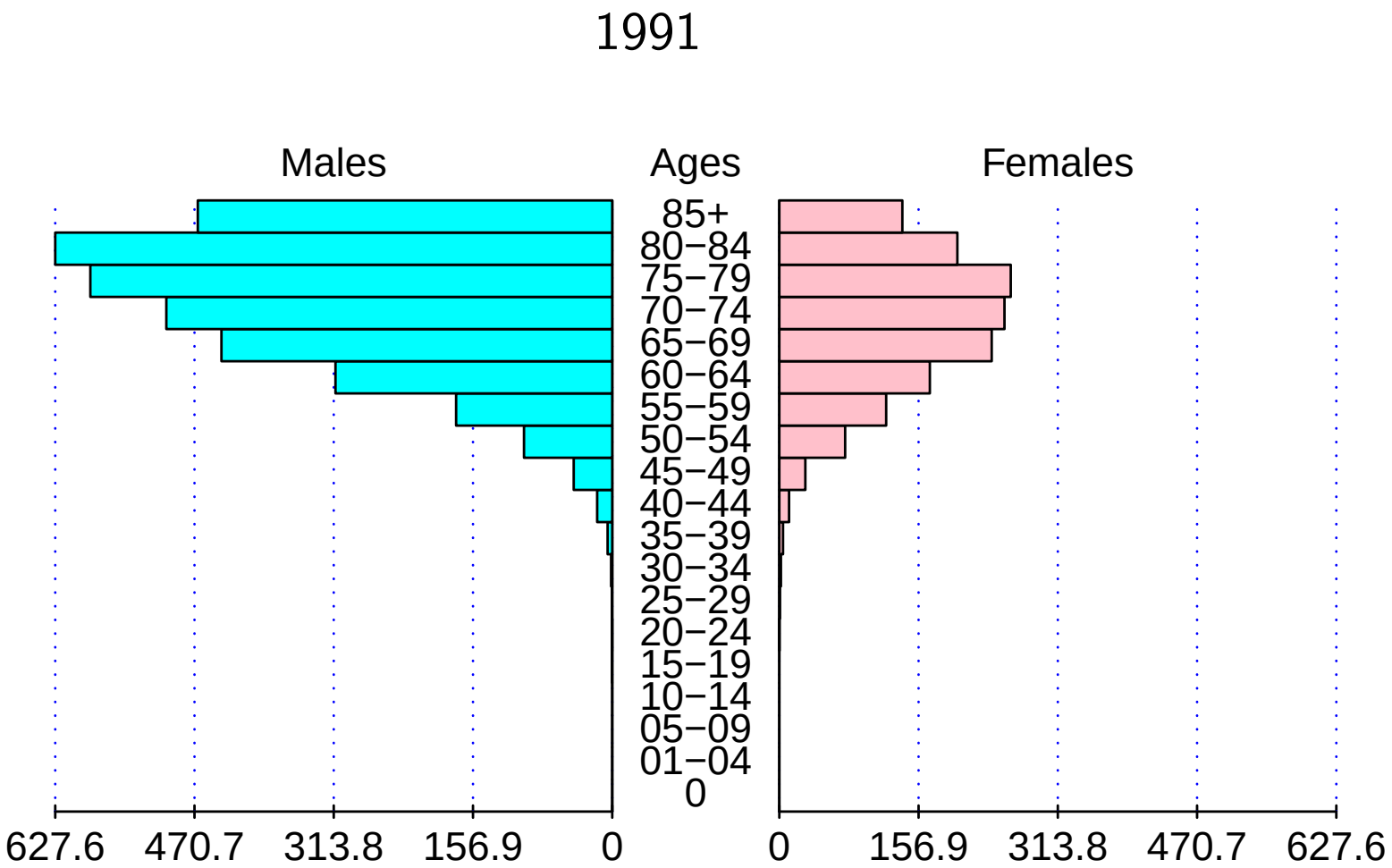
Incidence


1989
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
595.2
446.4
297.6
148.8
0
0
148.8
297.6
446.4
595.2

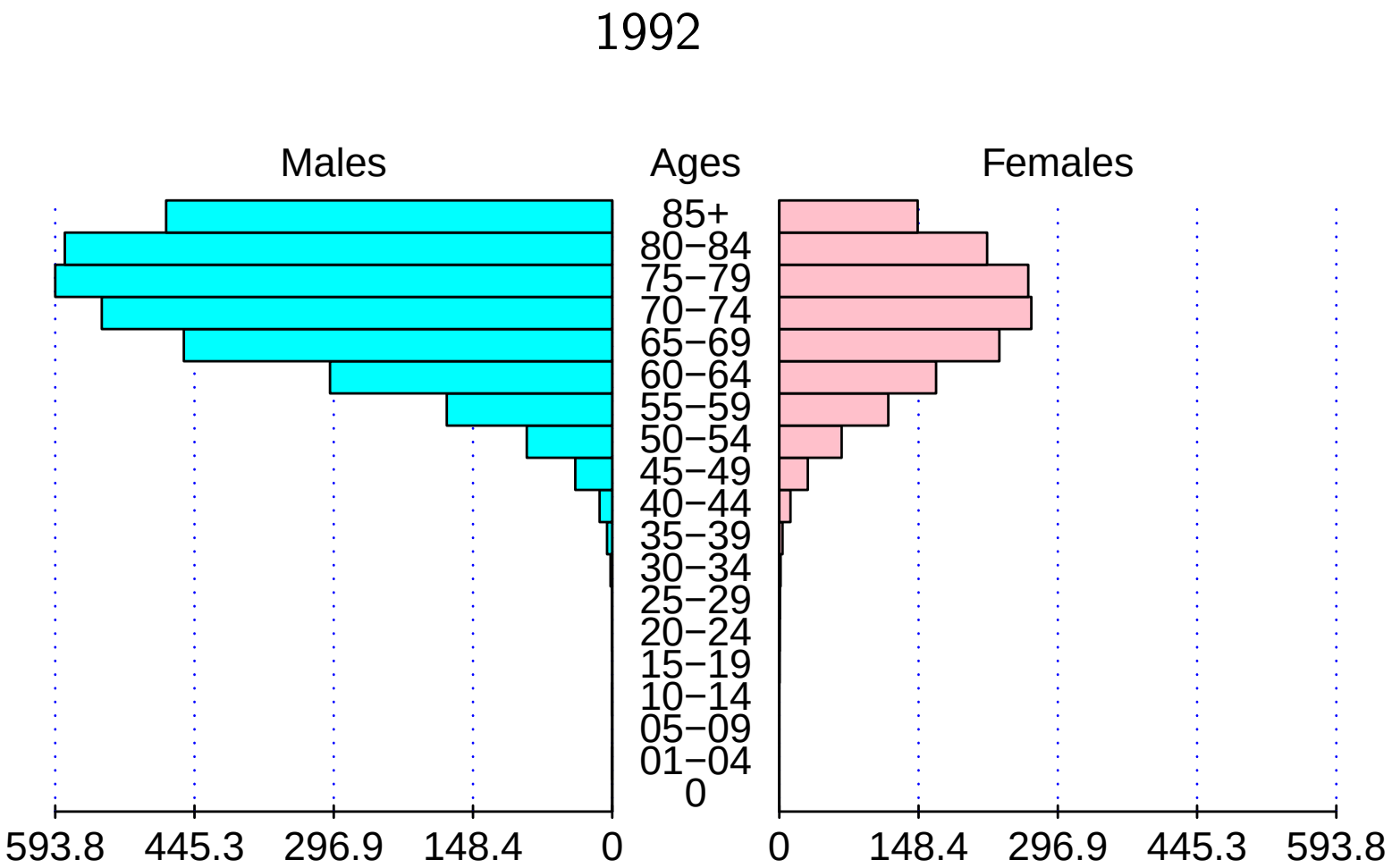
Incidence


1990
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
593.1
444.8
296.6
148.3
0
0
148.3
296.6
444.8
593.1

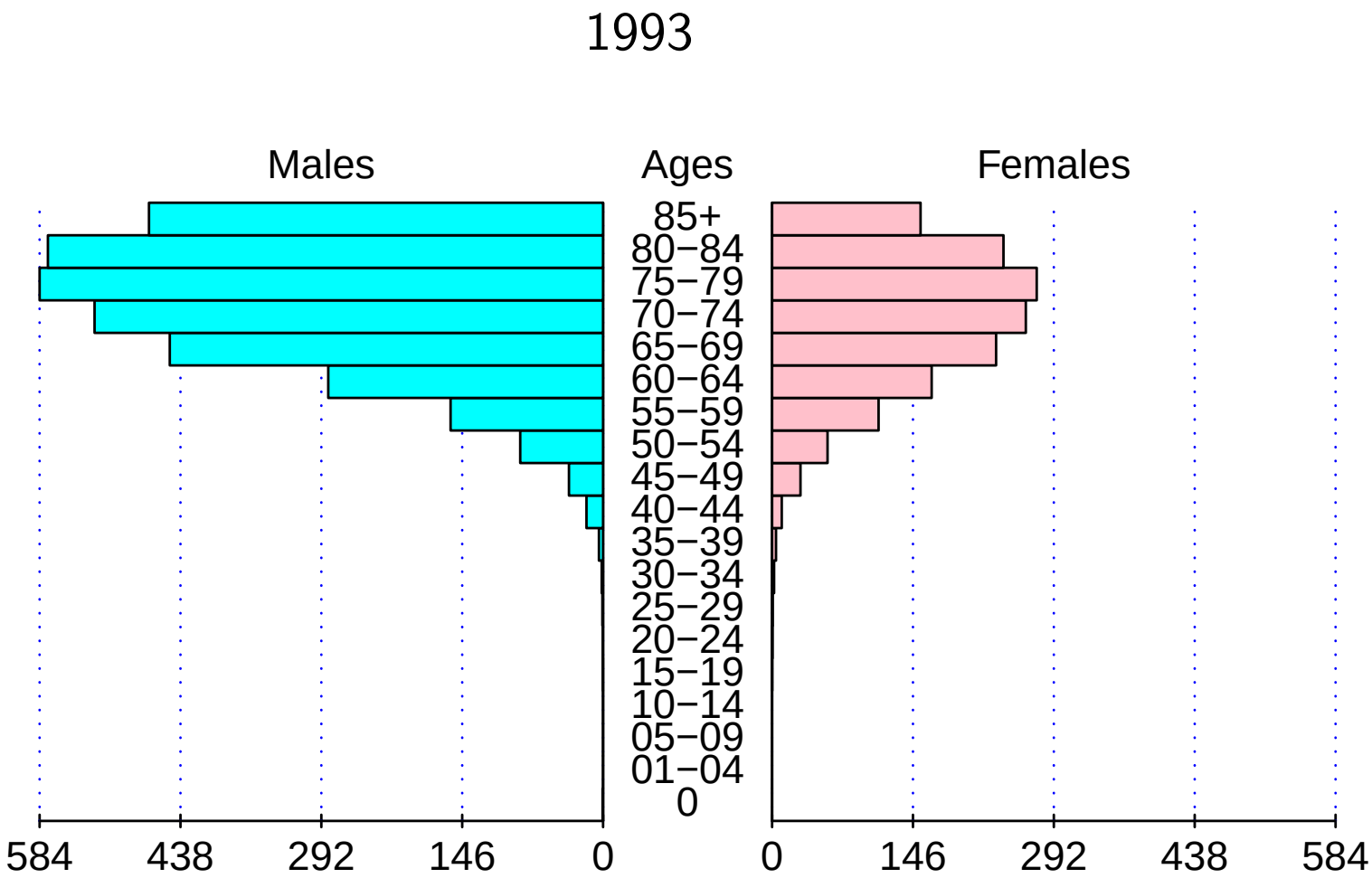
Incidence


1991
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
627.6
470.7
313.8
156.9
0
0
156.9
313.8
470.7
627.6

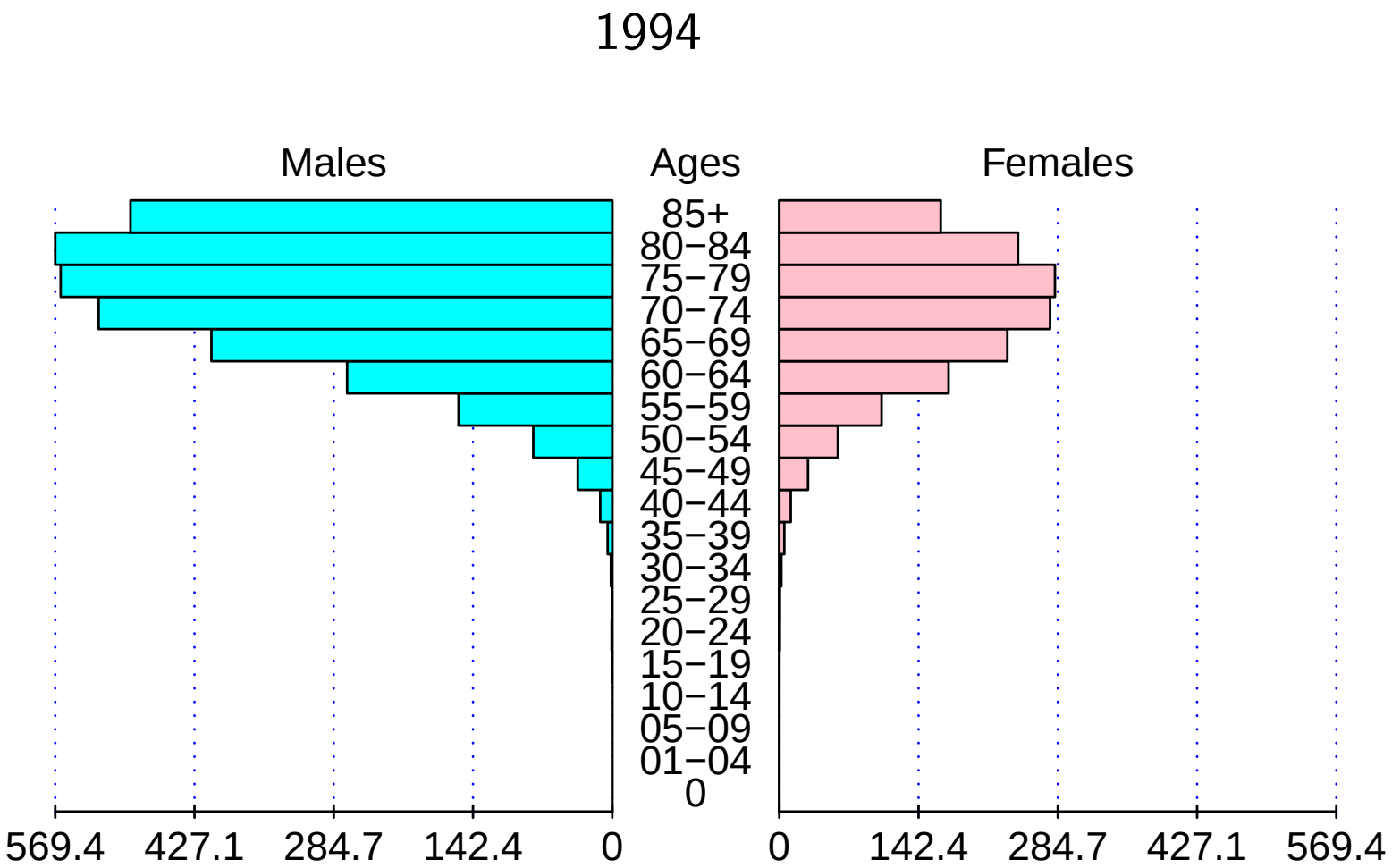
Incidence


1992
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
593.8
445.3
296.9
148.4
0
0
148.4
296.9
445.3
593.8

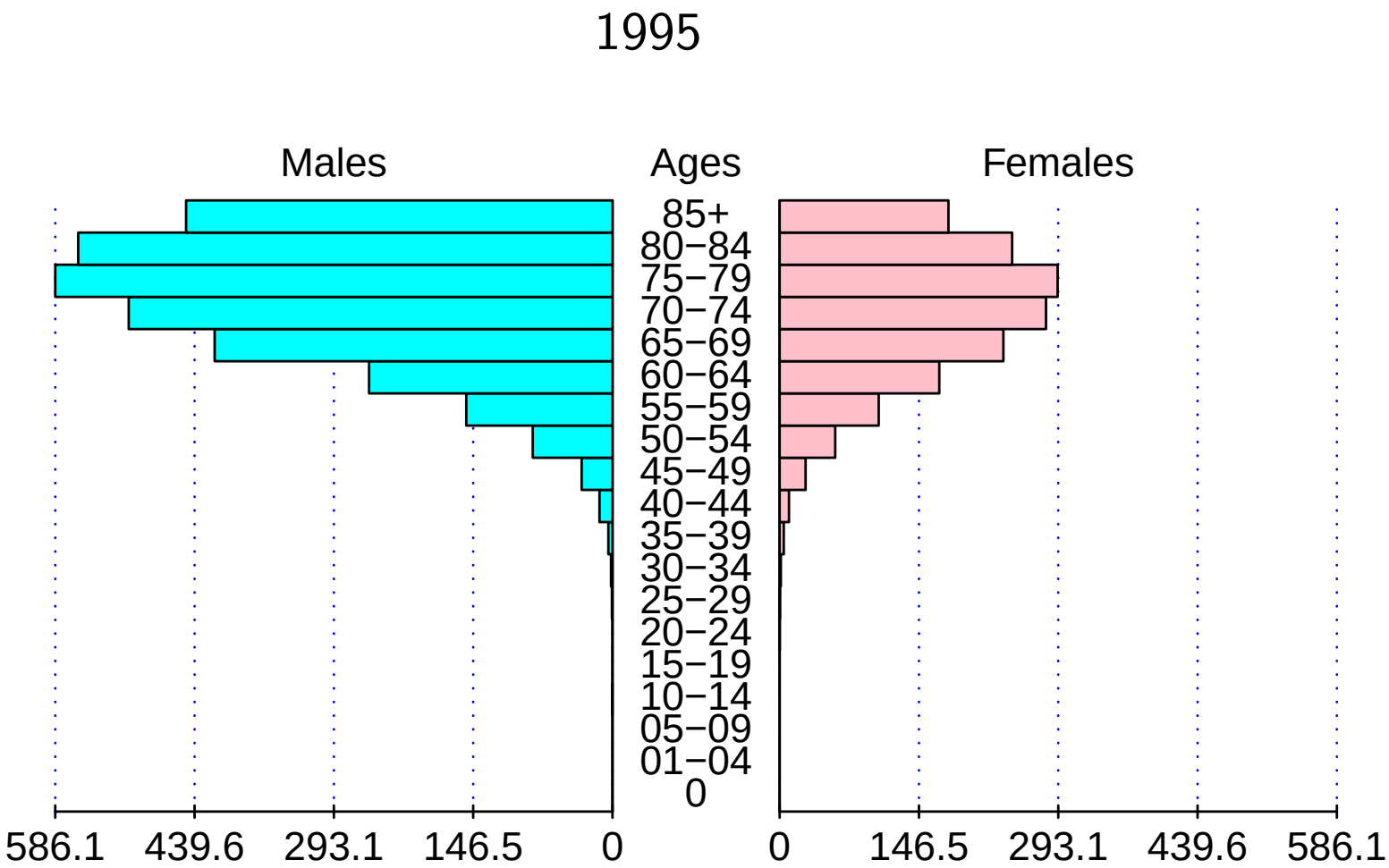
Incidence


1993
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
584
438
292
146
0
0
146
292
438
584

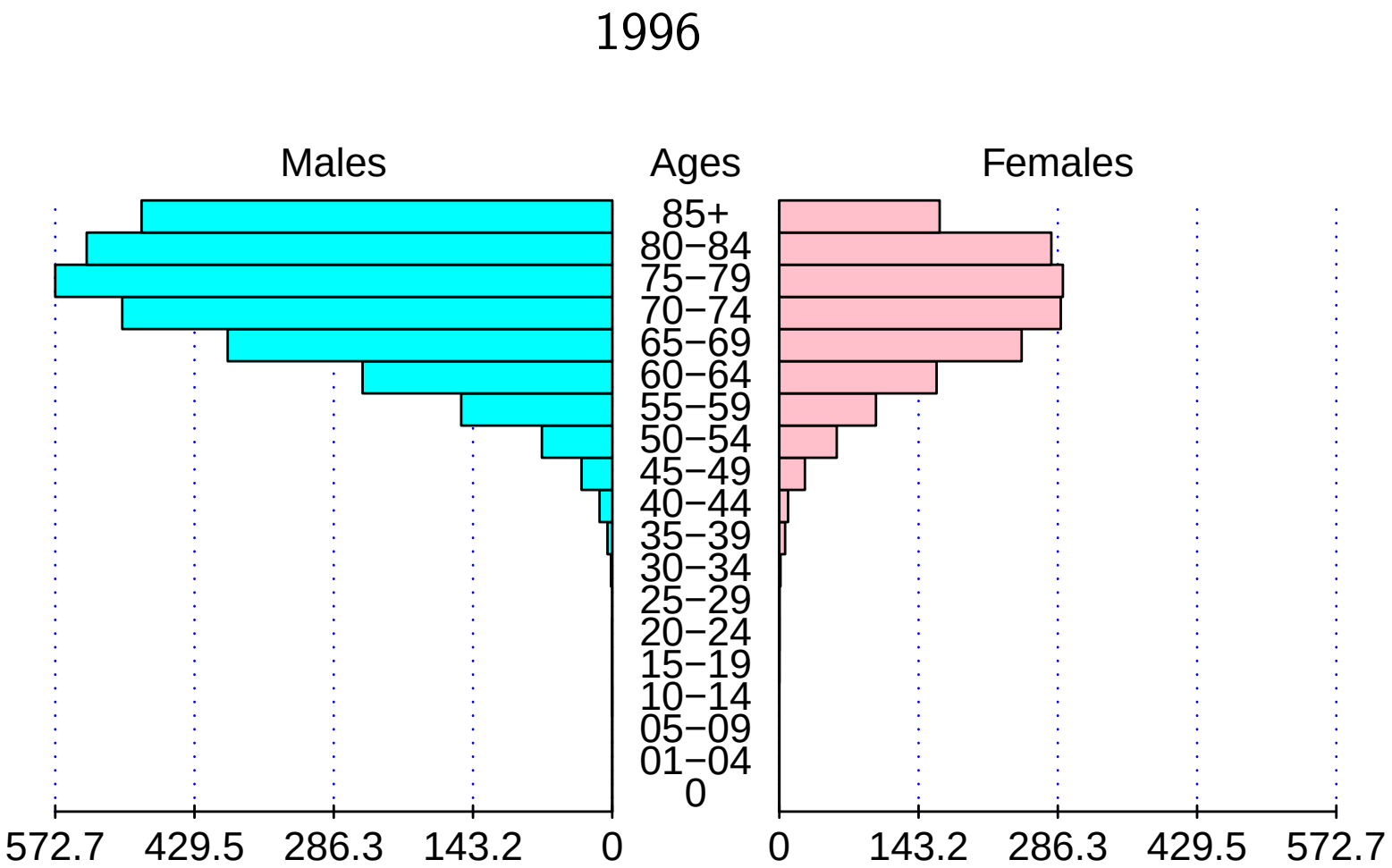
Incidence


1994
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
569.4
427.1
284.7
142.4
0
0
142.4
284.7
427.1
569.4

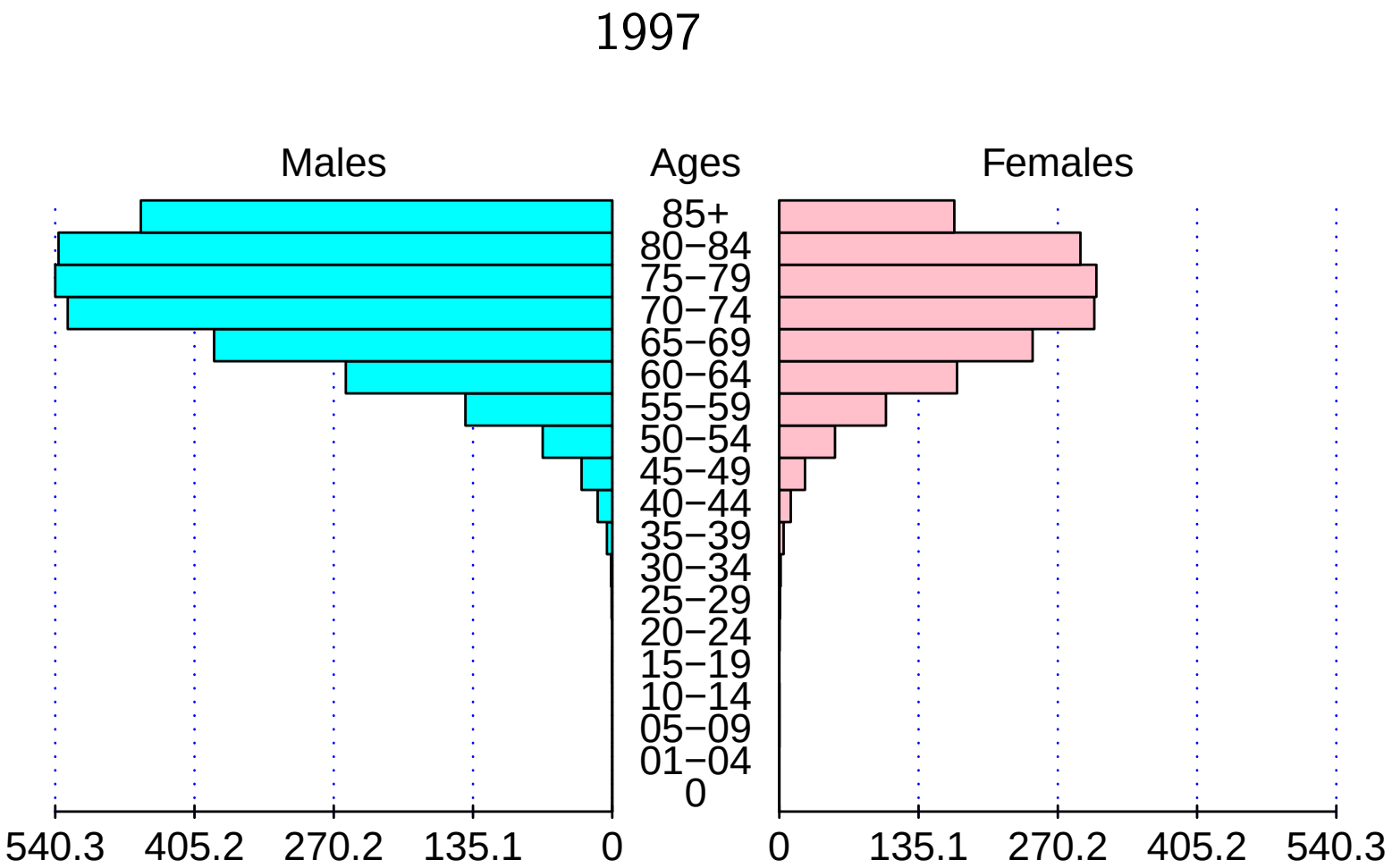
Incidence


1995
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
586.1
439.6
293.1
146.5
0
0
146.5
293.1
439.6
586.1

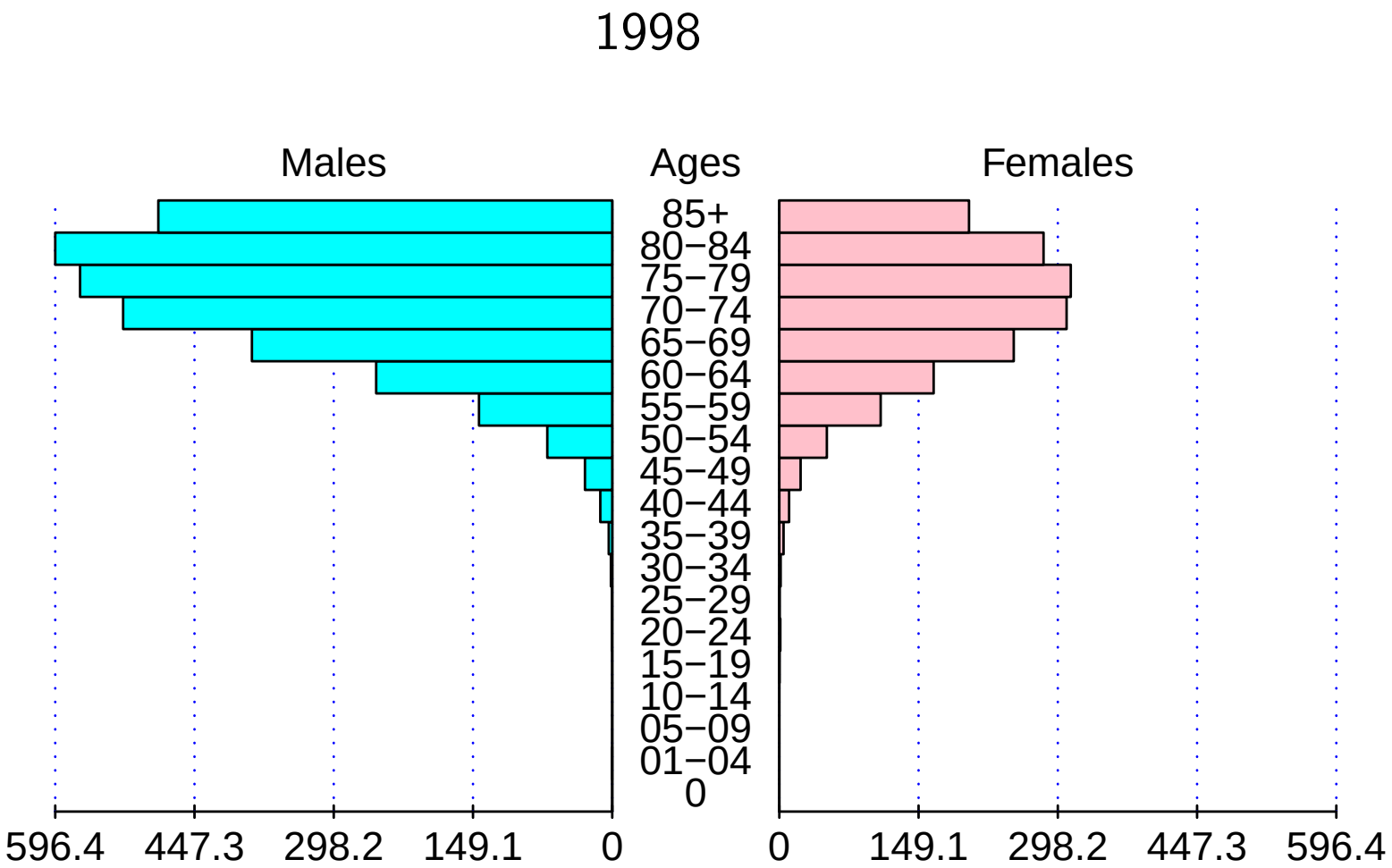
Incidence


1996
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
572.7
429.5
286.3
143.2
0
0
143.2
286.3
429.5
572.7

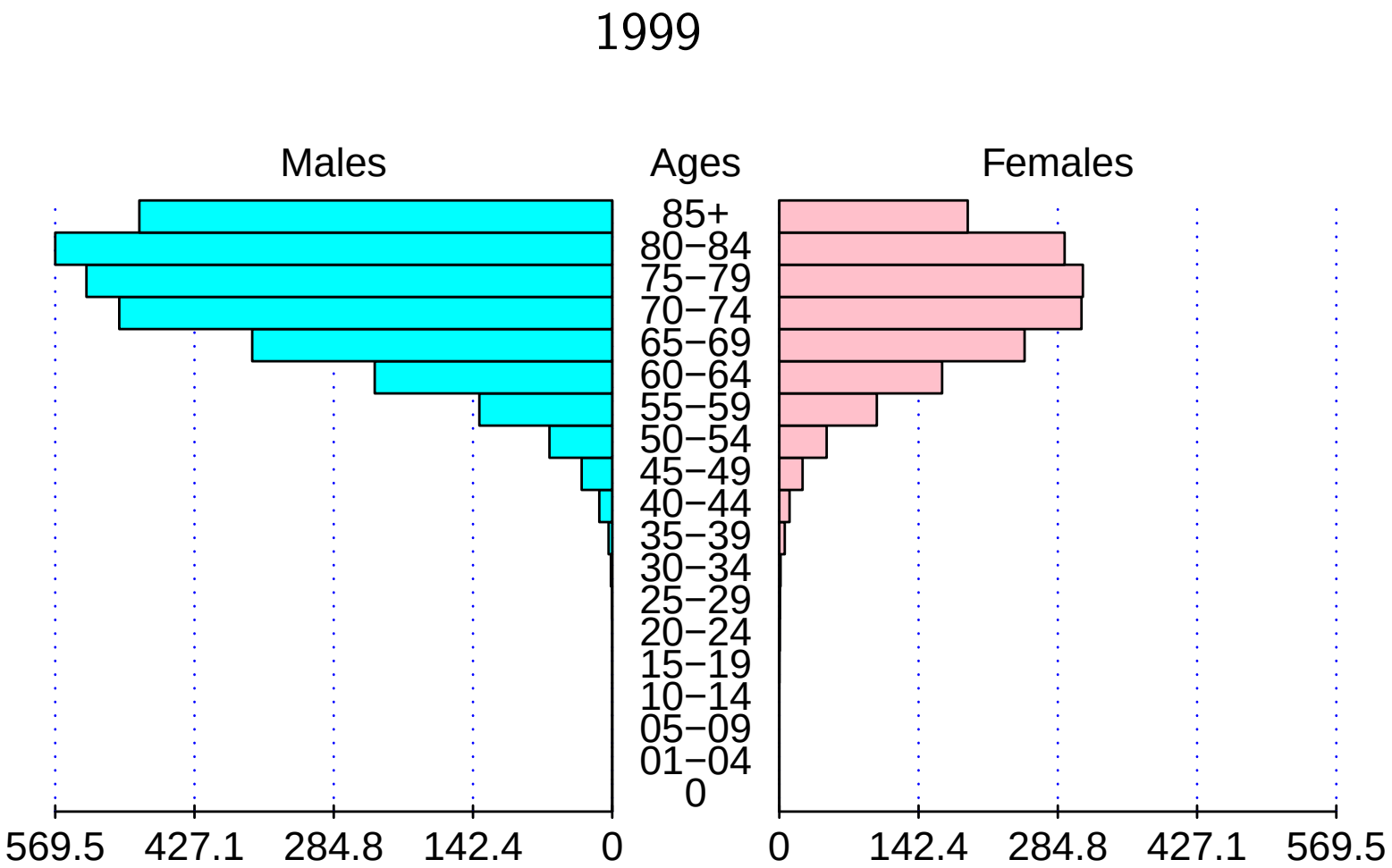
Incidence


1997
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
540.3
405.2
270.2
135.1
0
0
135.1
270.2
405.2
540.3

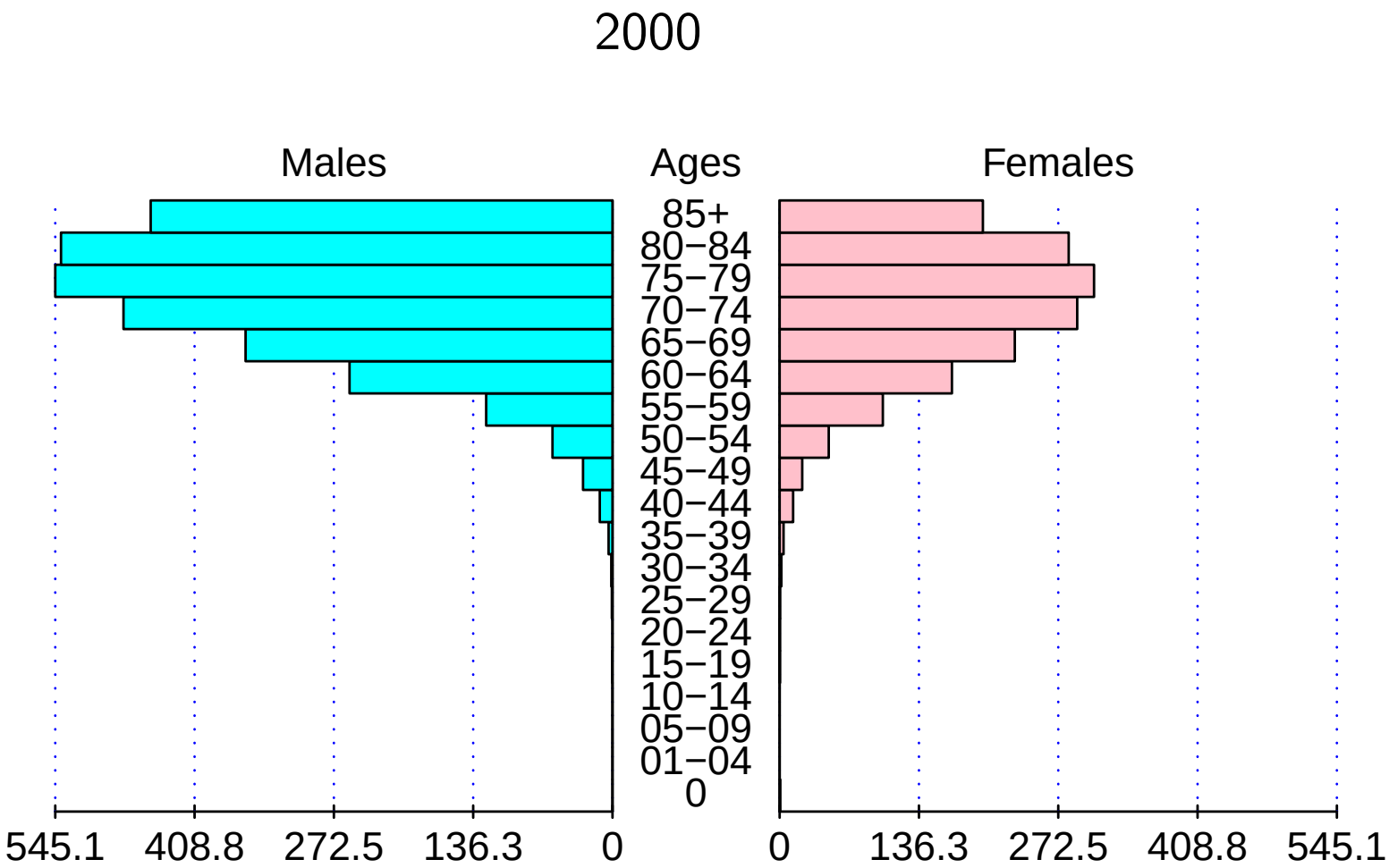
Incidence


1998
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
596.4
447.3
298.2
149.1
0
0
149.1
298.2
447.3
596.4

Incidence


1999
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
569.5
427.1
284.8
142.4
0
0
142.4
284.8
427.1
569.5

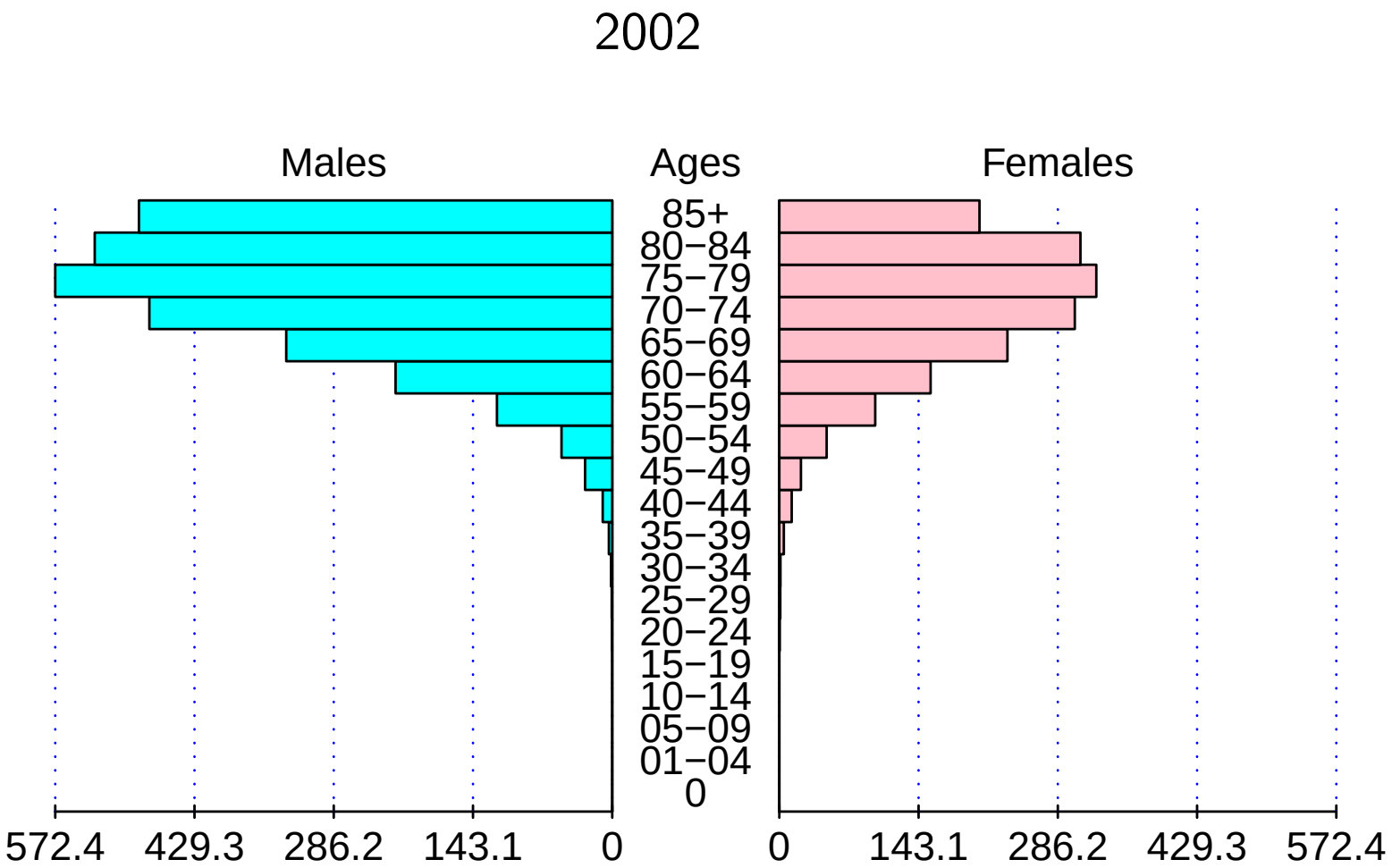
Incidence


2000
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
545.1
408.8
272.5
136.3
0
0
136.3
272.5
408.8
545.1

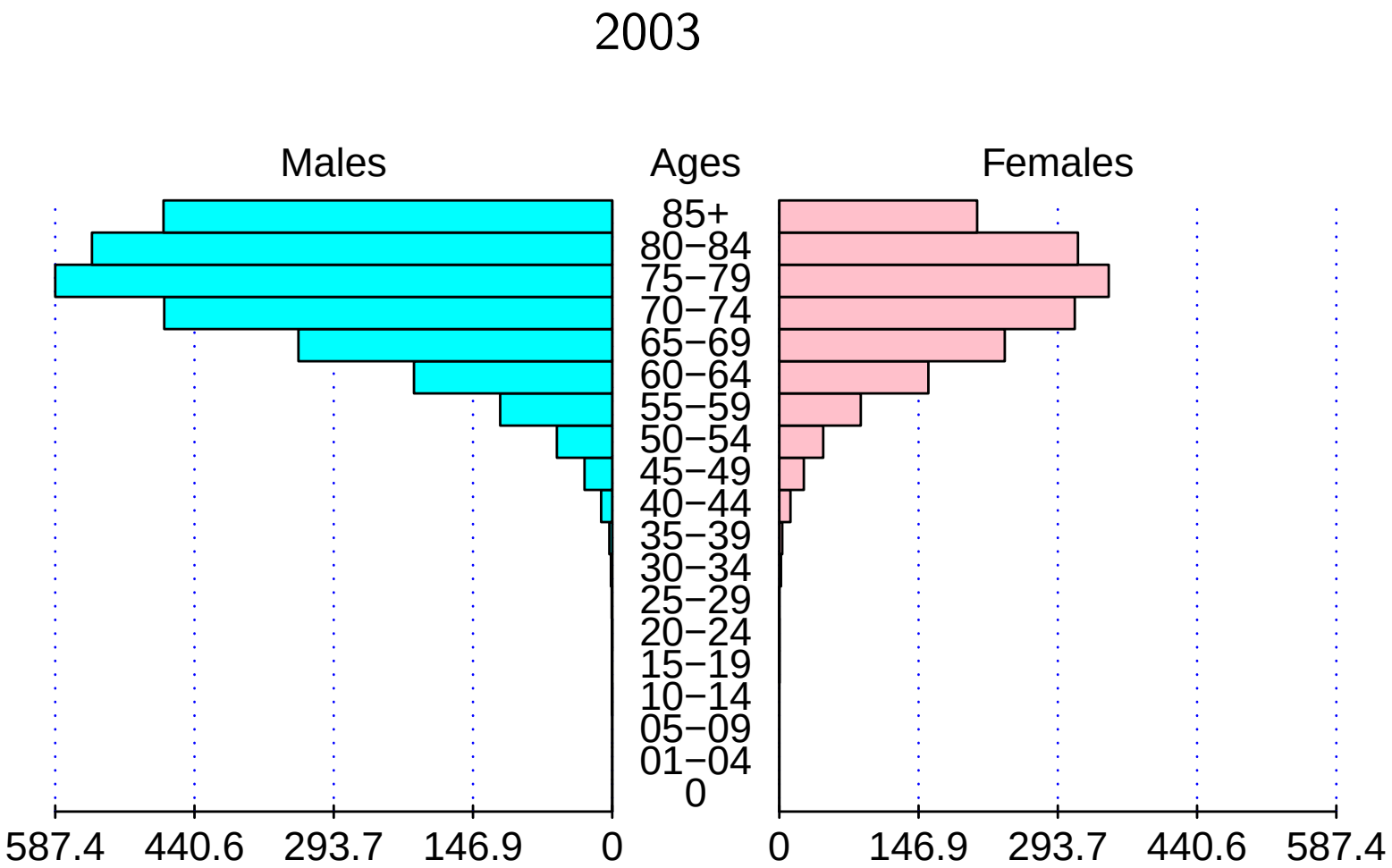
Incidence


2001
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
549.4
412
274.7
137.3
0
0
137.3
274.7
412
549.4

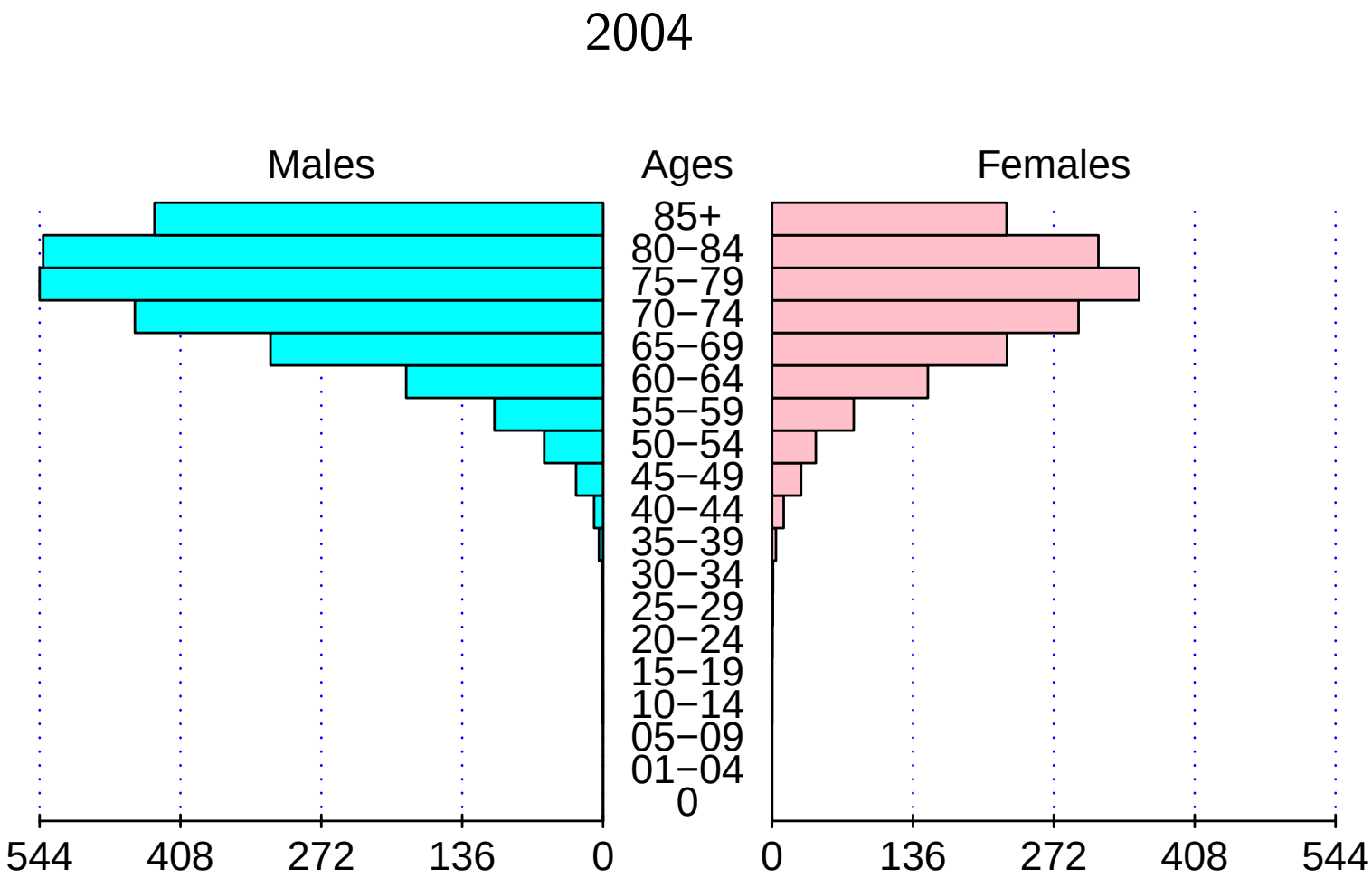
Incidence


2002
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
572.4
429.3
286.2
143.1
0
0
143.1
286.2
429.3
572.4

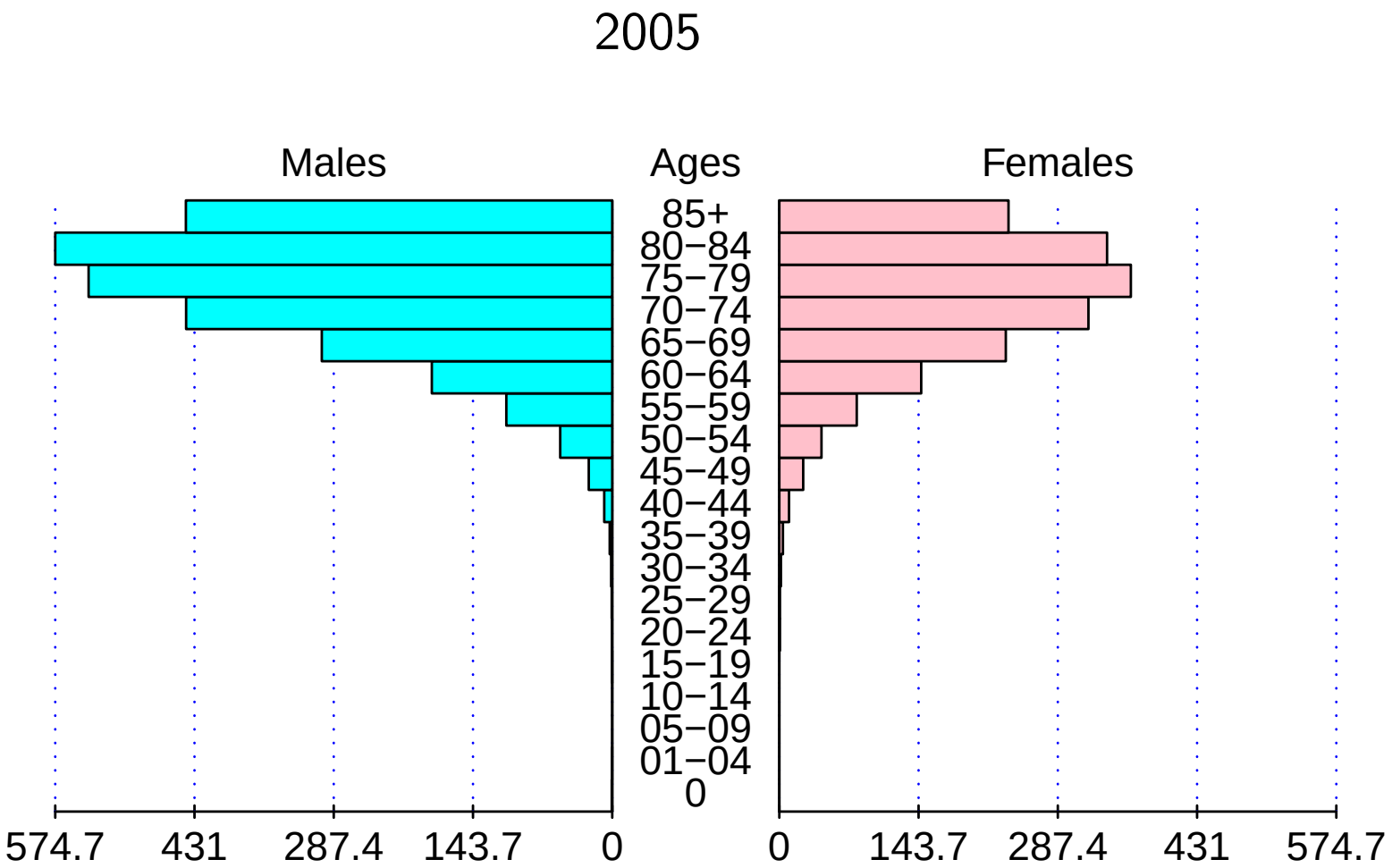
Incidence


2003
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
587.4
440.6
293.7
146.9
0
0
146.9
293.7
440.6
587.4

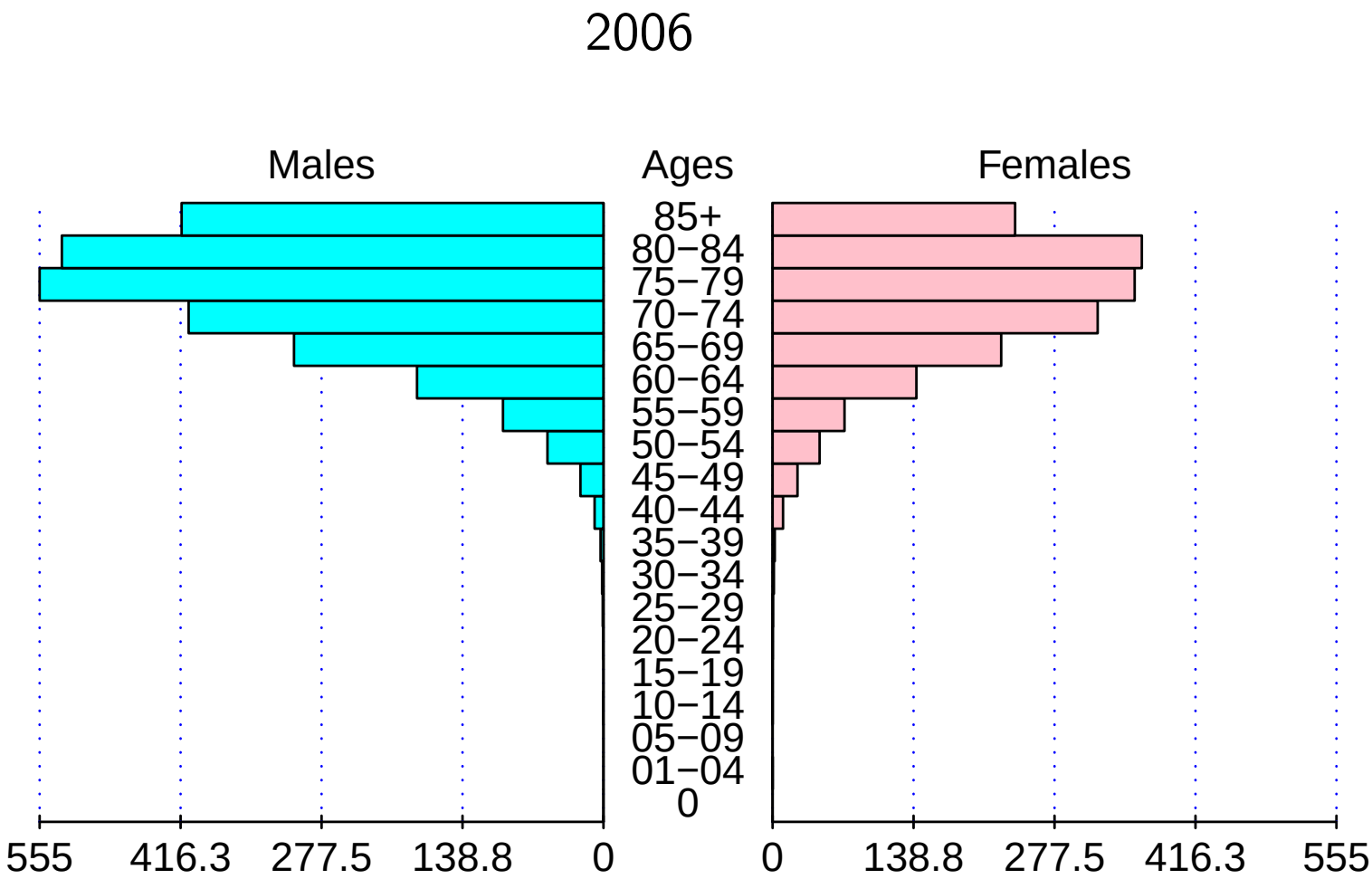
Incidence


2004
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
544
408
272
136
0
0
136
272
408
544

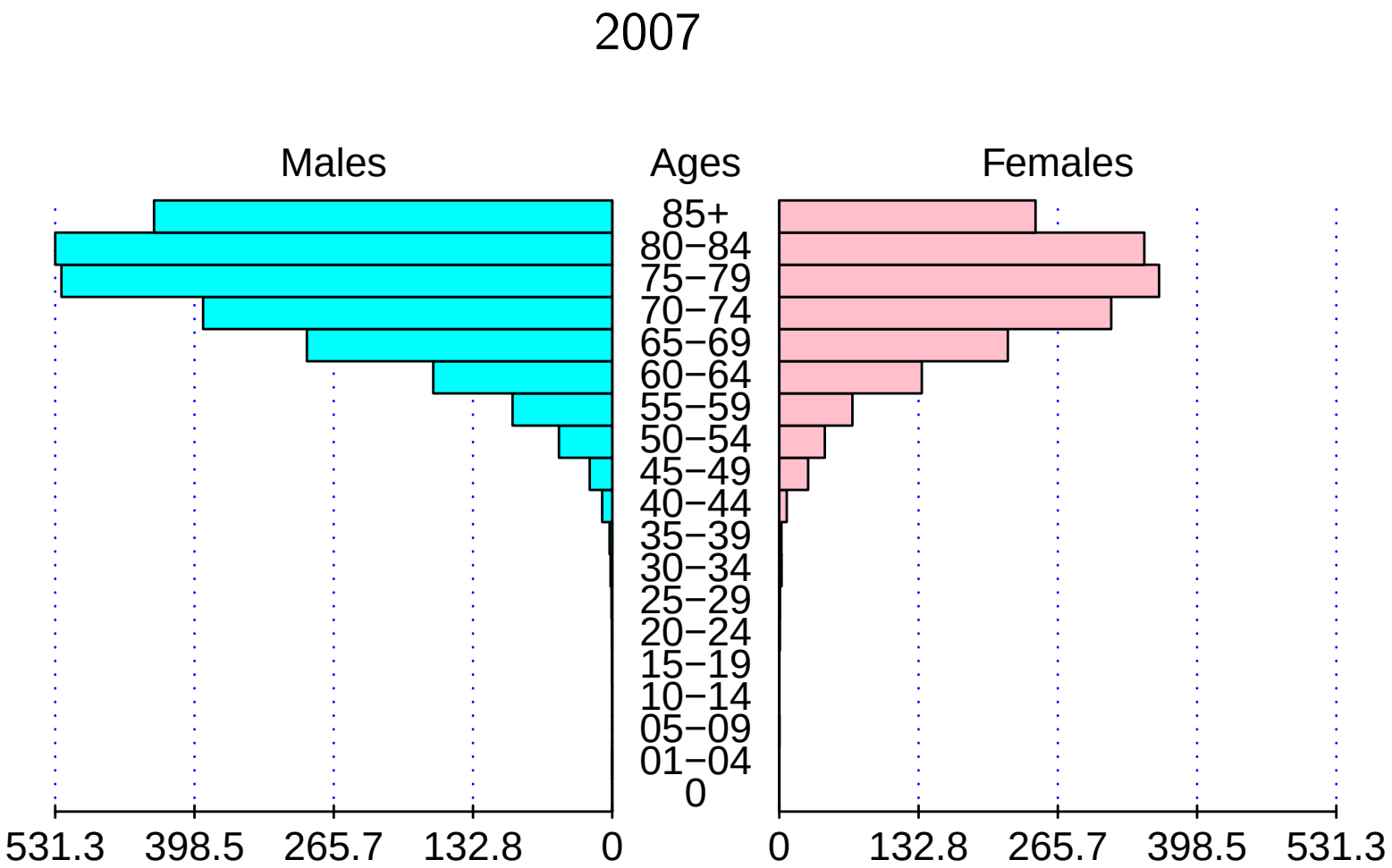
Incidence


2005
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
574.7
431
287.4
143.7
0
0
143.7
287.4
431
574.7

Incidence


2006
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
555
416.3
277.5
138.8
0
0
138.8
277.5
416.3
555

Incidence


2007
Males
Ages
Females
85+
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
05−09
01−04
0
531.3
398.5
265.7
132.8
0
0
132.8
265.7
398.5
531.3

Prevalence

Prevalence represents the fraction of **existing cases** in a population:

- the ratio $(d/(d+f))$ between the number of diseased animals (d) and the total number of animals at risk $(d+f)$
- the probability that a randomly-choosen animal is diseased

Point prevalence is the proportion of infected individuals in a defined population at a given time point.

Period prevalence is the proportion of infected individuals in a defined population **found over** a specified time period.

Misclassification

Case definition based on different diagnostic methods, qualitative or quantitative tests.
Quantification of **misclassification**:

	Diseased	Not diseased
Test +	TP	FP
Test -	FN	TN

- **Sensitivity**: the probability that a truly diseased animal will be classified as diseased. $SE = TP/(TP + FN)$
- **Specificity**: the probability that a truly non-diseased animal will be classified as non-diseased. $SP = TN/(FP + TN)$

Hepatitis B virus test

Andrew got a tattoo. Two months later he was refused as a blood donor. The phlebotomist explained that he had to wait a year to make sure he didn't get hepatitis B from the tattoo. That got him worried, so he ordered a home test kit for hepatitis B virus (HBV) from a website. The website said that the sensitivity of the test was 0.99 and the specificity was 0.995. Hepatitis B is rare among those who are not intravenous drug users – about 2 cases per 100,000 people. Studies suggest that getting a tattoo from an operator who follows accepted hygenie standards does not greatly increase the risk. Let's assume that Andrew believed that his risk was about 3 in 100,000.

If Andrew expect 10 million people as population at risk, then about 300 would have HBV, and the rest would not. As we know HBV test has 99% sensitivity, which means that it will catch 99% of the HBV cases (297 of the 300 cases) and miss the rest. The test has 99.5% specificity, which means that 99.5% of the noninfected people will test negative, but 0.5% of them will be false positives.

Hepatitis B virus test

	HBV +	HBV −	$\sum$
Test +	297	49,998	50,295
Test −	3	9,949,702	9,949,705
$\sum$	300	9,999,700	10,000,000

Suppose Andrew tests negative. There are 9,949,705 people like him – negative. Of these only 3 have HBV, so there are 3 chances in 9,949,705 (about 1 in 3.3 million) that a person who tests negative actually is infected.

On the other hand, suppose Andrew tests positive. There are 50,295 people like him – positive. Out of this group, only 297 really do have HBV (about 1 of 170). That means that even if Andrew tests positive, there is still only about 0.6% chance that he is actually infected.

Another example: `http://yudkowsky.net/rational/bayes`

Prevalence

- Apparent (P_A): the probability that a randomly-choosen unit of observation will test positive

$$\hat{P}_A = x/n$$

- Bayesian estimation:
  - $n/N \leq 0.1$: $x \sim binomial(n, P_A)$
  - $n/N > 0.1$: $x \sim hypergeometric(N, n, P_A)$
- Diagnostic misclassification:
  - Sensitivity: $p(+|Infected) \neq 100\%$
  - Specificity: $p(-|Not\ infected) \neq 100\%$

Prevalence

- Apparent (P_A): the probability that a randomly-choosen unit of observation will test positive

$$\hat{P}_A = x/n$$

- Bayesian estimation:
  - $n/N \leq 0.1$: $x \sim binomial(n, P_A)$
  - $n/N > 0.1$: $x \sim hypergeometric(N, n, P_A)$
- Diagnostic misclassification:
  - Sensitivity: $p(+|Infected) \neq 100\%$
  - Specificity: $p(-|Not\ infected) \neq 100\%$

Prevalence

- Apparent (P_A): the probability that a randomly-choosen unit of observation will test positive

$$\hat{P}_A = x/n$$

- Bayesian estimation:
  - $n/N \leq 0.1$: $x \sim binomial(n, P_A)$
  - $n/N > 0.1$: $x \sim hypergeometric(N, n, P_A)$
- Diagnostic misclassification:
  - Sensitivity: $p(+|Infected) \neq 100\%$
  - Specificity: $p(-|Not\ infected) \neq 100\%$
  - Rogan-Gladen estimator

$$P_T = \frac{P_A + Sp - 1}{Se + Sp - 1}$$

  - Bayesian binomial

$$x|P_A, Se, Sp \sim binom(n, Se \cdot P_T + (1 - Sp)(1 - P_T))$$

Prevalence

- Apparent (P_A): the probability that a randomly-choosen unit of observation will test positive

$$\hat{P}_A = x/n$$

- Bayesian estimation:
  - $n/N \leq 0.1$: $x \sim binomial(n, P_A)$
  - $n/N > 0.1$: $x \sim hypergeometric(N, n, P_A)$
- Diagnostic misclassification:
  - Sensitivity: $p(+|Infected) \neq 100\%$
  - Specificity: $p(-|Not\ infected) \neq 100\%$
  - Rogan-Gladen estimator:

$$\hat{P}_T = (\hat{P}_A + Sp - 1)/(Se + Sp - 1)$$

  - Bayesian binomial:

$$x|P_A, Se, Sp \sim binomial(n, P_T Se + (1 - P_T)(1 - Sp))$$

Prevalence

- Apparent (P_A): the probability that a randomly-choosen unit of observation will test positive

$$\hat{P}_A = x/n$$

- Bayesian estimation:
  - $n/N \leq 0.1$: $x \sim binomial(n, P_A)$
  - $n/N > 0.1$: $x \sim hypergeometric(N, n, P_A)$
- Diagnostic misclassification:
  - Sensitivity: $p(+|Infected) \neq 100\%$
  - Specificity: $p(-|Not\ infected) \neq 100\%$
  - Rogan-Gladen estimator:

$$\hat{P}_T = (\hat{P}_A + Sp - 1)/(Se + Sp - 1)$$

  - Bayesian binomial:

$$x|P_A, Se, Sp \sim binomial(n, P_T Se + (1 - P_T)(1 - Sp))$$

Prevalence

- Apparent (P_A): the probability that a randomly-choosen unit of observation will test positive

$$\hat{P}_A = x/n$$

- Bayesian estimation:
  - $n/N \leq 0.1$: $x \sim binomial(n, P_A)$
  - $n/N > 0.1$: $x \sim hypergeometric(N, n, P_A)$
- Diagnostic misclassification:
  - Sensitivity: $p(+|Infected) \neq 100\%$
  - Specificity: $p(-|Not\ infected) \neq 100\%$
  - Rogan-Gladen estimator:

$$\hat{P}_T = (\hat{P}_A + Sp - 1)/(Se + Sp - 1)$$

  - Bayesian binomial:

$$x|P_A, Se, Sp \sim binomial(n, P_T Se + (1 - P_T)(1 - Sp))$$

Prevalence

- Apparent (P_A): the probability that a randomly-choosen unit of observation will test positive

$$\hat{P}_A = x/n$$

- Bayesian estimation:
  - $n/N \leq 0.1$: $x \sim binomial(n, P_A)$
  - $n/N > 0.1$: $x \sim hypergeometric(N, n, P_A)$
- Diagnostic misclassification:
  - Sensitivity: $p(+|Infected) \neq 100\%$
  - Specificity: $p(-|Not\ infected) \neq 100\%$
  - Rogan-Gladen estimator:

$$\hat{P}_T = (\hat{P}_A + Sp - 1)/(Se + Sp - 1)$$

  - Bayesian binomial:

$$x|P_A, Se, Sp \sim binomial(n, P_T Se + (1 - P_T)(1 - Sp))$$

Prevalence

- $x = 2,\ n = 60,\ Se = 0.3,\ Sp = 0.96,\ N = 675$
- Rogan-Gladen estimator: $\hat{P}_T = (\hat{P}_A + Sp - 1)/(Se + Sp - 1) = (2/60 + 0.96 - 1)/(0.3 + 0.96 - 1) = -0.026$
- Bayesian approach accounts uncertainity of P_T, Se, Sp:
  - 95% certain that $Se < 0.5$ and $Sp > 0.94$
  - Beta prior distributions
  - Posterior distributions
  - $P_T = 0.02$, 95% credible interval $0 - 0.456$
  - $Se = 0.29$, 95% credible interval $0.11 - 0.52$
  - $Sp = 0.96$, 95% credible interval $0.94 - 0.98$
  - 97.1% certain $P_T < 0.456$
  - 52% certain population is infected

Prevalence

- $x = 2,\ n = 60,\ Se = 0.3,\ Sp = 0.96,\ N = 675$
- Rogan-Gladen estimator: $\hat{P}_T = (\hat{P}_A + Sp - 1)/(Se + Sp - 1) = (2/60 + 0.96 - 1)/(0.3 + 0.96 - 1) = -0.026$
- Bayesian approach accounts uncertainity of P_T, Se, Sp:
  - 95% certain that $Se < 0.5$ and $Sp > 0.94$
  - Beta prior distributions
  - Posterior distributions
    - $P_T = 0.02$, 95% credible interval 0 – 0.056
    - $Se = 0.29$, 95% credible interval 0.11 – 0.52
    - $Sp = 0.96$, 95% credible interval 0.94 – 0.98
    - 97.1% certain $P_T < 0.056$
    - 95% certain population is infected

Prevalence

- $x = 2$, $n = 60$, $Se = 0.3$, $Sp = 0.96$, $N = 675$
- Rogan-Gladen estimator: $\hat{P}_T = (\hat{P}_A + Sp - 1)/(Se + Sp - 1) = (2/60 + 0.96 - 1)/(0.3 + 0.96 - 1) = -0.026$
- Bayesian approach accounts uncertainity of P_T, Se, Sp:
  - 95% certain that $Se < 0.5$ and $Sp > 0.94$
  - Beta prior distributions
  - Posterior distributions
    - $\hat{P}_T = 0.02$, 95% credible interval $0 - 0.456$
    - $\hat{S}e = 0.29$, 95% credible interval $0.11 - 0.52$
    - $\hat{S}e = 0.96$, 95% credible interval $0.94 - 0.98$
    - 97.5% certain $P_T < 0.456$
    - 58% certain population is infected

Prevalence

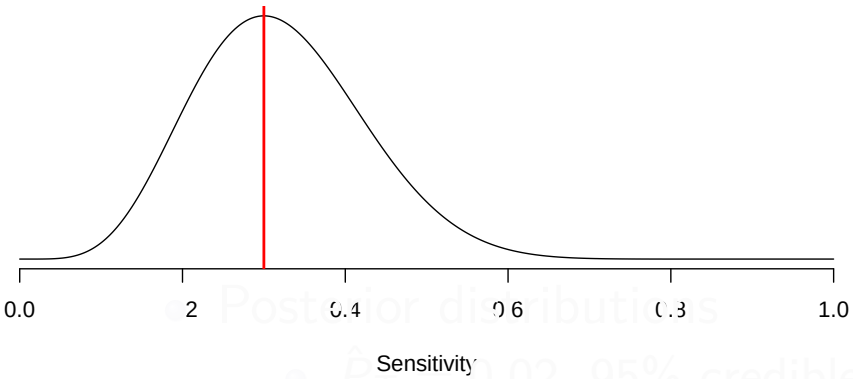
- $x = 2$, $n = 60$, $Se = 0.3$, $Sp = 0.96$, $N = 675$
- Rogan-Gladen estimator: $\hat{P}_T = (\hat{P}_A + Sp - 1)/(Se + Sp - 1) = (2/60 + 0.96 - 1)/(0.3 + 0.96 - 1) = -0.026$
- Bayesian approach accounts uncertainity of P_T, Se, Sp:
  - 95% certain that $Se < 0.5$ and $Sp > 0.94$
  - Beta prior distributions
  - Posterior distributions
    - $\hat{P}_T = 0.02$, 95% credible interval $0 - 0.456$
    - $\hat{S}e = 0.29$, 95% credible interval $0.11 - 0.52$
    - $\hat{S}e = 0.96$, 95% credible interval $0.94 - 0.98$
    - 97.5% certain $P_T < 0.456$
    - 58% certain population is infected

Prevalence

- $x = 2$, $n = 60$, $Se = 0.3$, $Sp = 0.96$, $N = 675$
- Rogan-Gladen estimator: $\hat{P}_T = (\hat{P}_A + Sp - 1)/(Se + Sp - 1) = (2/60 + 0.96 - 1)/(0.3 + 0.96 - 1) = -0.026$
- Bayesian approach accounts uncertainity of P_T, Se, Sp:
  - 95% certain that $Se < 0.5$ and $Sp > 0.94$
  - Beta prior distributions
  - Posterior distributions
    - $\hat{P}_T = 0.02$, 95% credible interval $0 - 0.456$
    - $\hat{S}e = 0.29$, 95% credible interval $0.11 - 0.52$
    - $\hat{S}e = 0.96$, 95% credible interval $0.94 - 0.98$
    - 97.5% certain $P_T < 0.456$
    - 58% certain population is infected

Prevalence

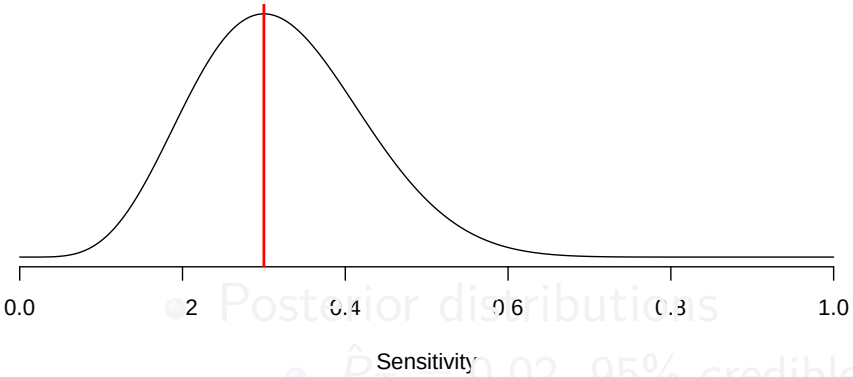
- $x = 2$, $n = 60$, $Se = 0.3$, $Sp = 0.96$, $N = 675$
- Rogan-Gladen estimator: $\hat{P}_T = (\hat{P}_A + Sp - 1)/(Se + Sp - 1) = (2/60 + 0.96 - 1)/(0.3 + 0.96 - 1) = -0.026$
- Bayesian approach accounts uncertainity of P_T, Se, Sp:
  - 95% certain that $Se < 0.5$ and $Sp > 0.94$
  - Beta prior distributions

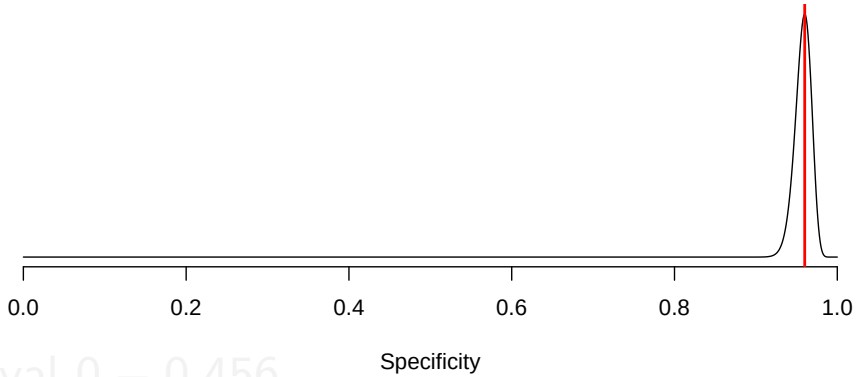


- Posterior distributions
  - $\hat{P}_T = 0.02$, 95% credible interval $0 - 0.456$
  - $\hat{Se} = 0.29$, 95% credible interval $0.11 - 0.52$
  - $\hat{Se} = 0.96$, 95% credible interval $0.94 - 0.98$
  - 97.5% certain $P_T < 0.456$
  - 58% certain population is infected

Prevalence

- $x = 2$, $n = 60$, $Se = 0.3$, $Sp = 0.96$, $N = 675$
- Rogan-Gladen estimator: $\hat{P}_T = (\hat{P}_A + Sp - 1)/(Se + Sp - 1) = (2/60 + 0.96 - 1)/(0.3 + 0.96 - 1) = -0.026$
- Bayesian approach accounts uncertainity of P_T, Se, Sp:
  - 95% certain that $Se < 0.5$ and $Sp > 0.94$
  - Beta prior distributions

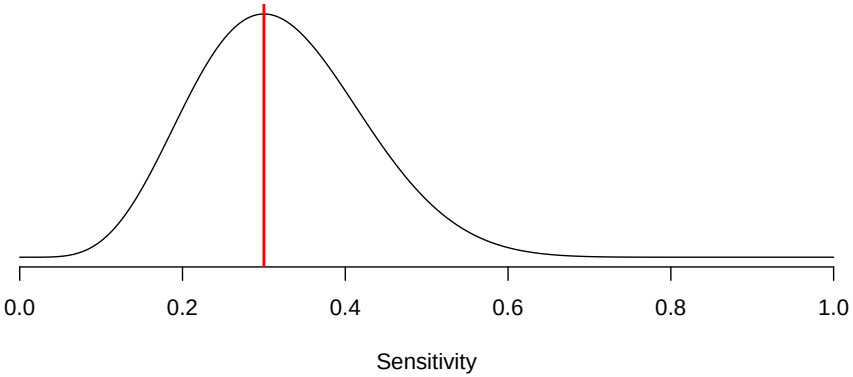


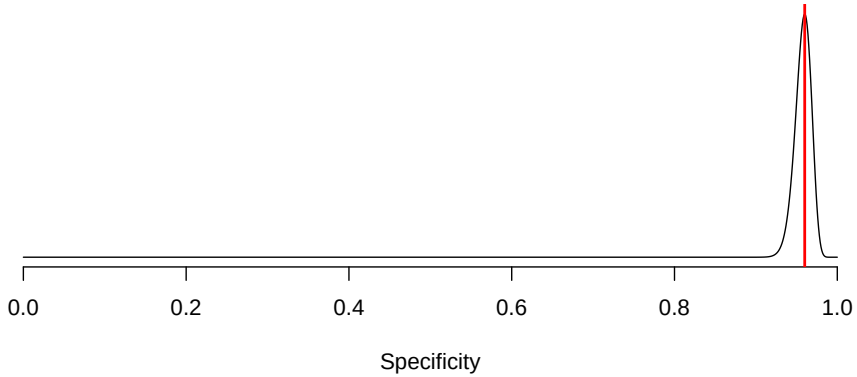


- Posterior distributions
  - $\hat{P}_T = 0.02$, 95% credible interval $0 - 0.456$
  - $\hat{S}e = 0.29$, 95% credible interval $0.11 - 0.52$
  - $\hat{S}e = 0.96$, 95% credible interval $0.94 - 0.98$
  - 97.5% certain $P_T < 0.456$
  - 58% certain population is infected

Prevalence

- $x = 2$, $n = 60$, $Se = 0.3$, $Sp = 0.96$, $N = 675$
- Rogan-Gladen estimator: $\hat{P}_T = (\hat{P}_A + Sp - 1)/(Se + Sp - 1) = (2/60 + 0.96 - 1)/(0.3 + 0.96 - 1) = -0.026$
- Bayesian approach accounts uncertainity of P_T, Se, Sp:
  - 95% certain that $Se < 0.5$ and $Sp > 0.94$
  - Beta prior distributions

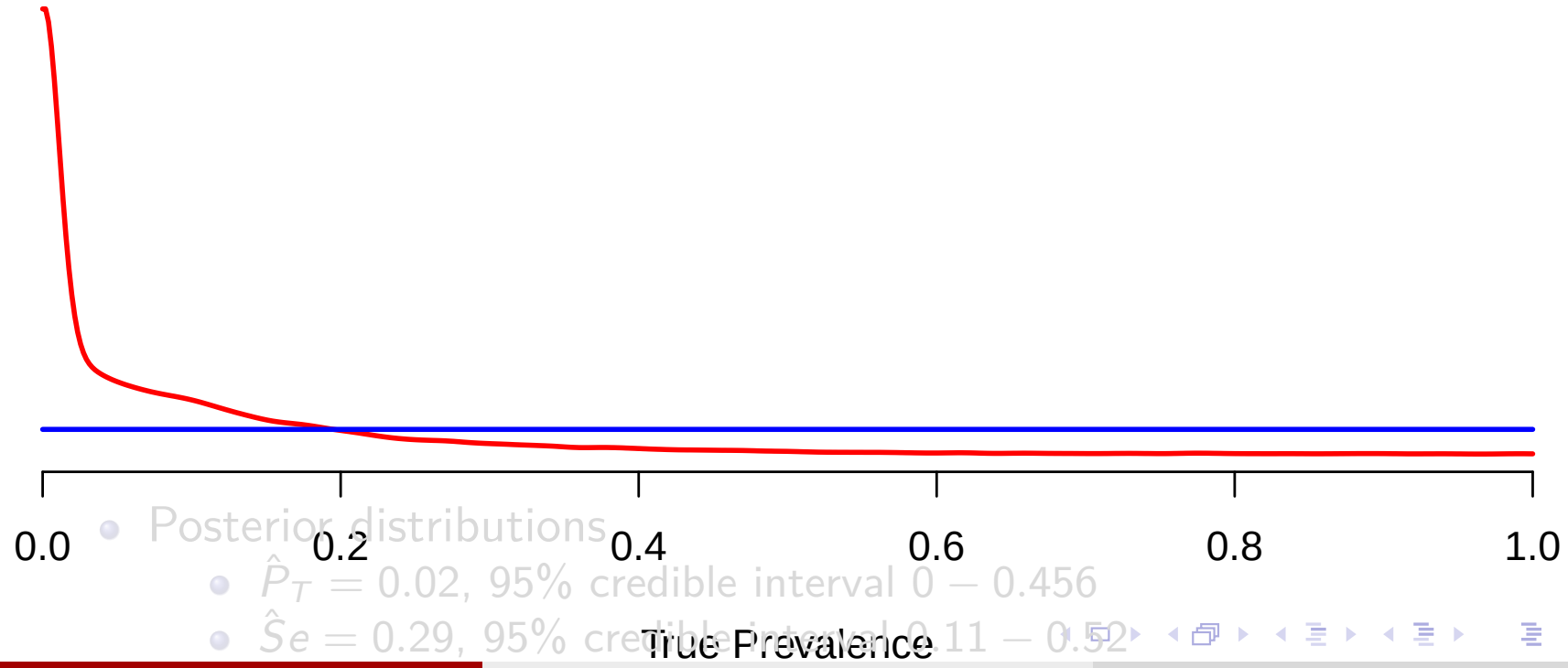




- Posterior distributions
  - $\hat{P}_T = 0.02$, 95% credible interval $0 - 0.456$
  - $\hat{S}e = 0.29$, 95% credible interval $0.11 - 0.52$
  - $\hat{S}p = 0.96$, 95% credible interval $0.94 - 0.98$

Prevalence

- $x = 2$, $n = 60$, $Se = 0.3$, $Sp = 0.96$, $N = 675$
- Rogan-Gladen estimator: $\hat{P}_T = (\hat{P}_A + Sp - 1)/(Se + Sp - 1) = (2/60 + 0.96 - 1)/(0.3 + 0.96 - 1) = -0.026$
- Bayesian approach accounts uncertainity of P_T, Se, Sp:
  - 95% certain that $Se < 0.5$ and $Sp > 0.94$
  - Beta prior distributions

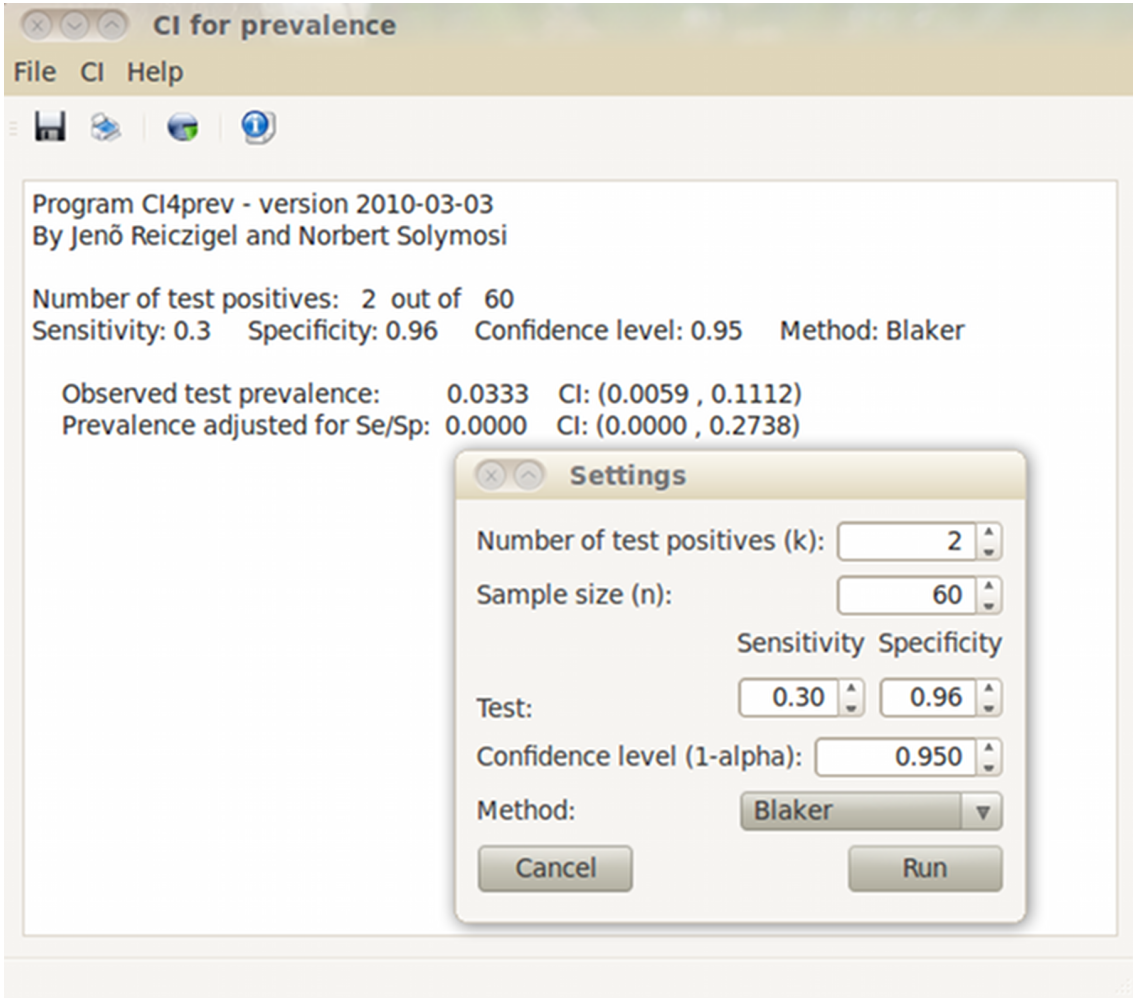


- Posterior distributions
  - $\hat{P}_T = 0.02$, 95% credible interval $0 - 0.456$
  - $\hat{S}e = 0.29$, 95% credible interval $0.11 - 0.52$

Prevalence

- $x = 2$, $n = 60$, $Se = 0.3$, $Sp = 0.96$, $N = 675$
- Rogan-Gladen estimator: $\hat{P}_T = (\hat{P}_A + Sp - 1)/(Se + Sp - 1) = (2/60 + 0.96 - 1)/(0.3 + 0.96 - 1) = -0.026$
- Bayesian approach accounts uncertainity of P_T, Se, Sp:
  - 95% certain that $Se < 0.5$ and $Sp > 0.94$
  - Beta prior distributions
  - Posterior distributions
    - $\hat{P}_T = 0.02$, 95% credible interval $0 - 0.456$
    - $\hat{S}e = 0.29$, 95% credible interval $0.11 - 0.52$
    - $\hat{S}e = 0.96$, 95% credible interval $0.94 - 0.98$
    - 97.5% certain $P_T < 0.456$
    - 58% certain population is infected

Prevalence – tools



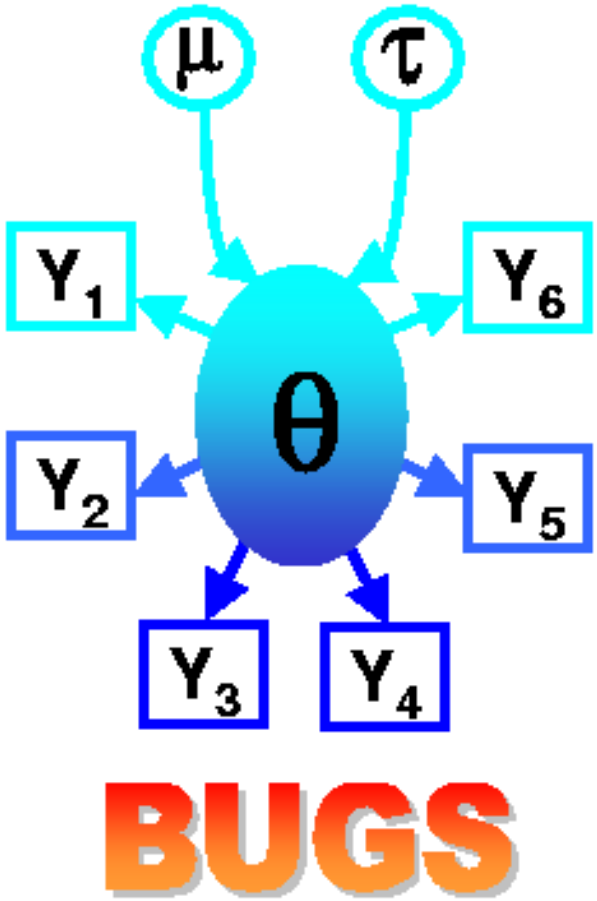
http://www.univet.hu/users/jreiczig/prevalence-with-se-sp.html

http://www.ausvet.com.au/content.php?page=software

Prevalence – tools

Thomas Bayes (1702 – 1761)

$$p(\theta|x) = \frac{p(x|\theta)}{p(x)} \times p(\theta)$$



http://www.mrc-bsu.cam.ac.uk/bugs/

http://www.epi.ucdavis.edu/diagnostictests/software.html#PrevalenceEstimation

Monitoring and Surveillance Systems (MOSS)

Monitoring: Systematic, ongoing or repeated, measurement, collection, collation, analysis, interpretation and timely dissemination information of animal health related data **without** an associated pre-defined plan of (control) action

Surveillance: Systematic, ongoing or repeated, measurement, collection, collation, analysis, interpretation and timely dissemination of animal health related data, essential for describing hazard occurrence and for the planning, implementation, and evaluation of risk mitigation (control) measures

Sampling

Census: if all animals in a population are investigated.

If a survey is designed well, then a reasonably accurate and acceptable estimate of a variable can be made by examining some of the animals in the relevant population; that is, a **sample**.

The **target population** is the total population about which information is required.

The **study population** is the population from which a sample is drawn.

The study population consists of **elementary units**, which cannot be divided further.

A collection of elementary units, grouped according to a common characteristic, is a **stratum**.

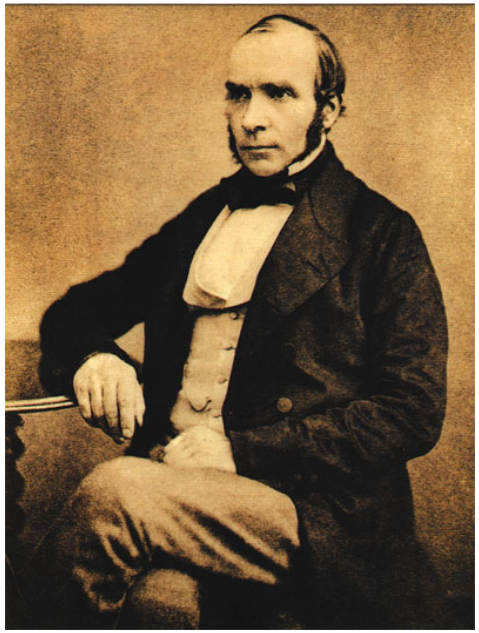
Sampling

Main types of sampling:

- **non-probability** sampling in which the choice of the sample is left to the investigator;
- **probability** sampling in which the selection of the sample is made using a deliberate, unbiased process, so that each sampling unit in a group has an equal probability of being selected; this is the basis of random sampling. Types:
  - simple random sampling
  - systematic sampling
  - stratified sampling
  - clustered sample (one, two and multi stage)

Sample size calculation:

- Presence, prevalence study: misclassification
- Survey Toolbox: `http://www.ausvet.com.au/content.php?page=softwarest`
- WinEpiscope: `http://www.clive.ed.ac.uk/winepiscope/`



John Snow
(1813-1858)

The objectives of spatial epidemiology:

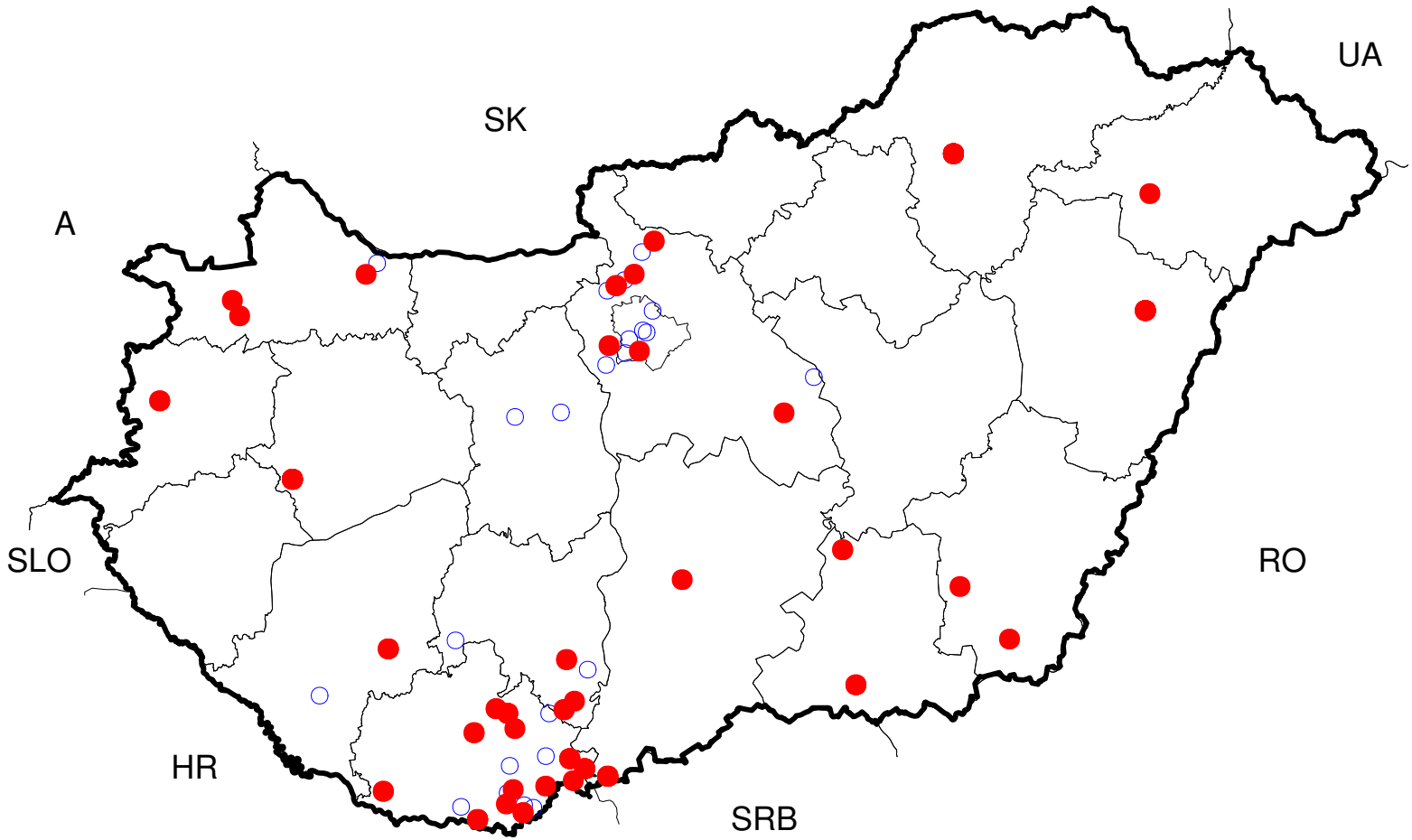
- disease mapping
- description of spatial patterns
- explanation or prediction of disease risk

Fundamental to these objectives is the need for data which, in addition to the classical data attribute information describing the characteristics of the entity studied, require the availability of georeferenced feature data, be they points or areas:

- Attribute data:
  - Tables, text files
  - Databases: relational and hierarchical
- Georeferenced feature:
  - Vector graphical maps (points, lines, polygons)
  - Raster, pixel graphical maps
  - Databases: relational and hierarchical

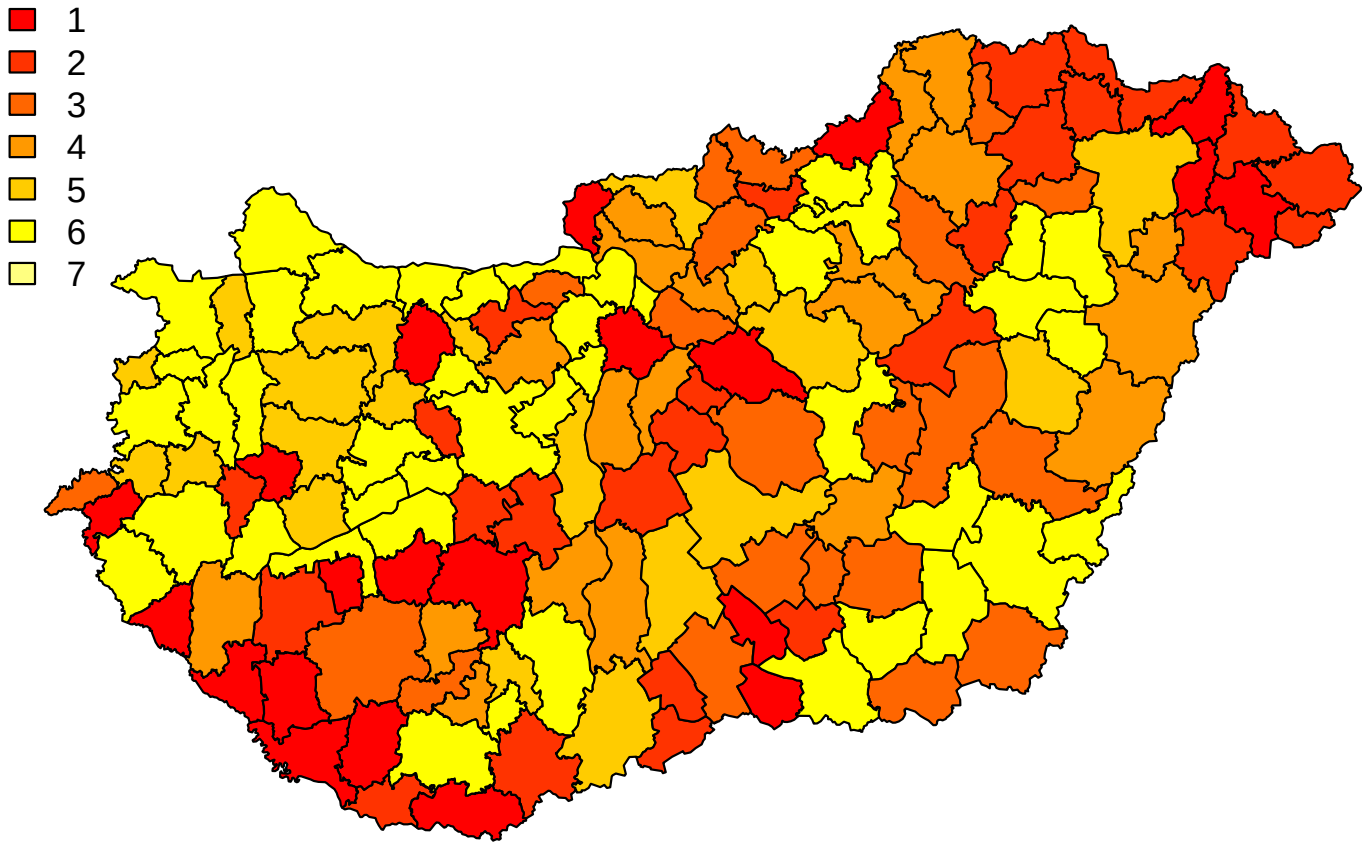
- Mapping
  - Pattern analysis
  - Ecological analysis

- Point
  - Choropleth
  - Isopleth



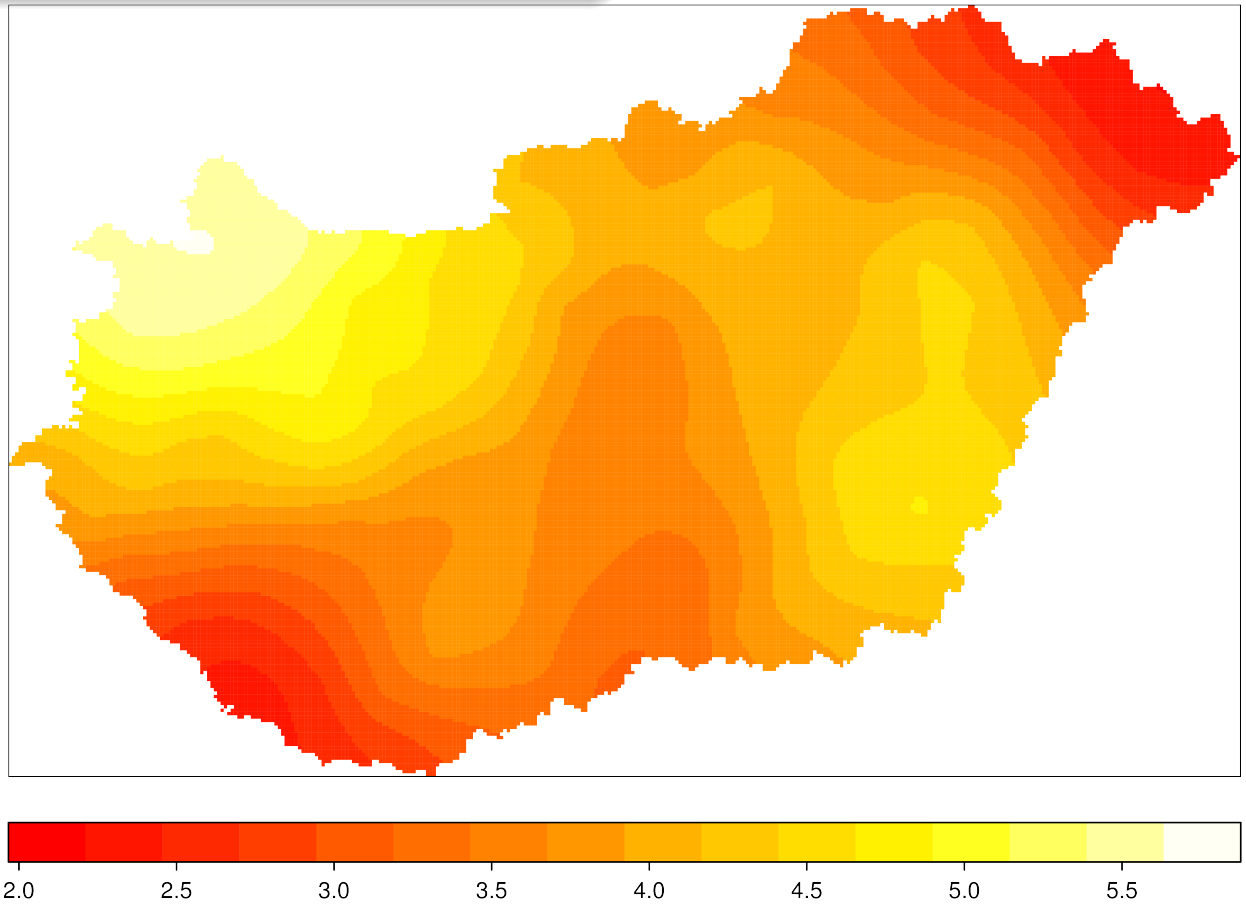
- Mapping
- Pattern analysis
- Ecological analysis

- Point
- Choropleth
- Isopleth



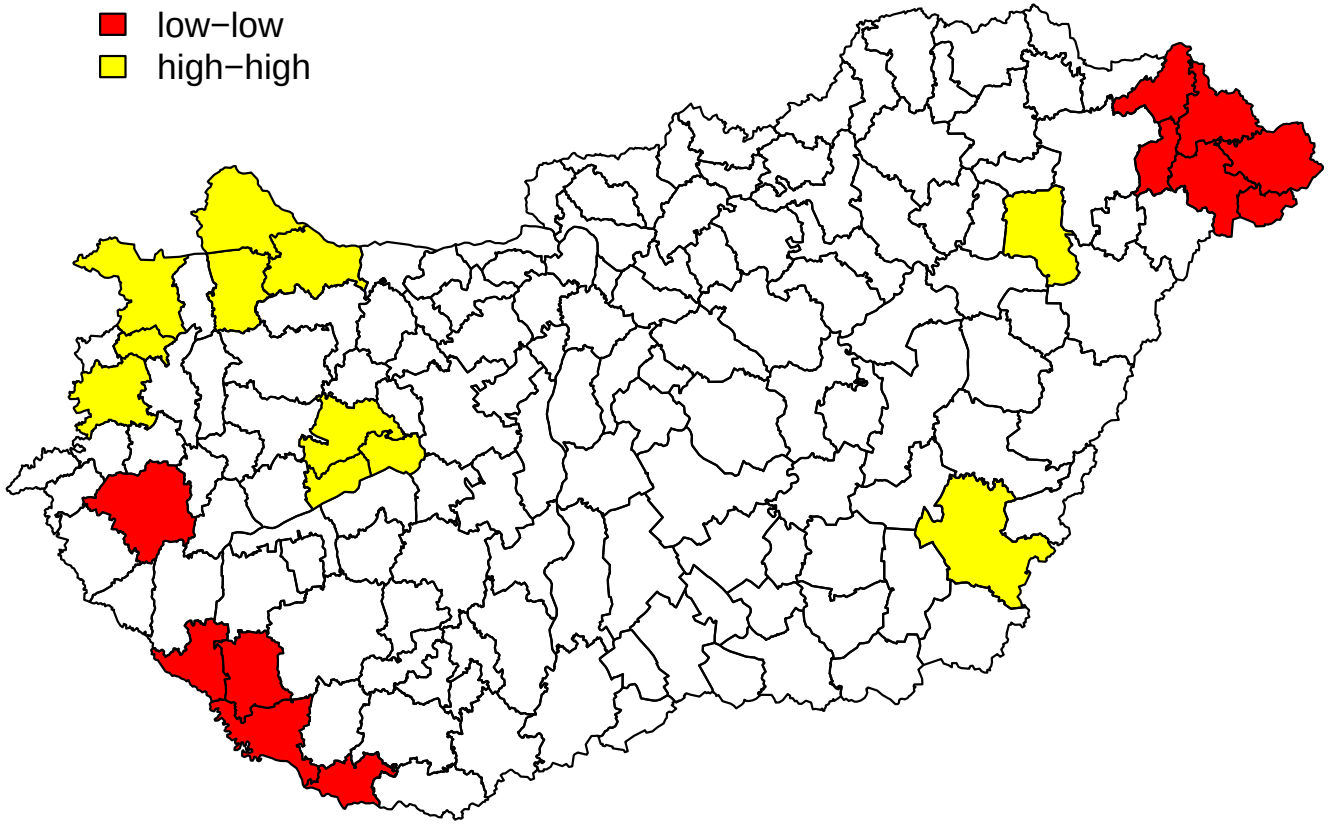
- Mapping
- Pattern analysis
- Ecological analysis

- Point
- Choropleth
- Isopleth



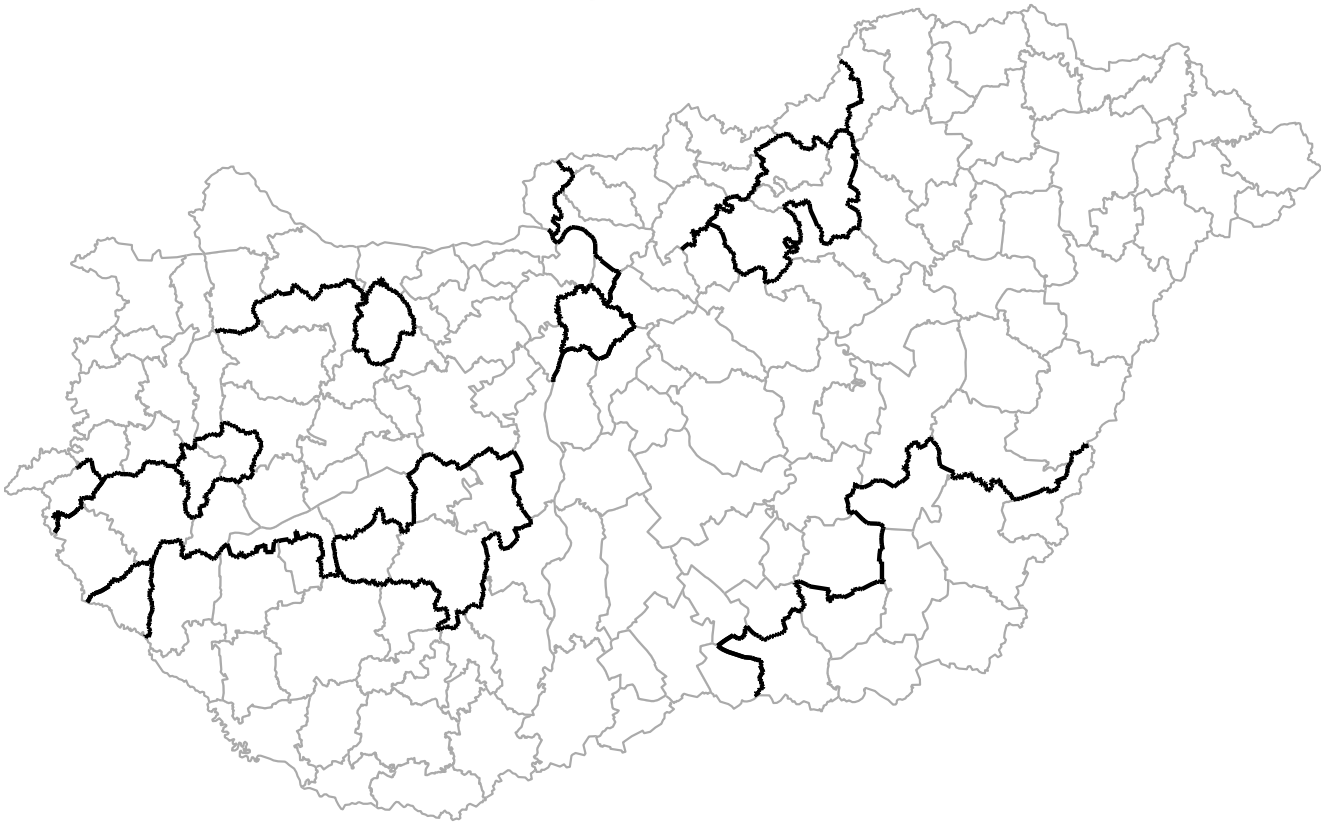
- Mapping
- Pattern analysis
- Ecological analysis

- Cluster
- Barrier



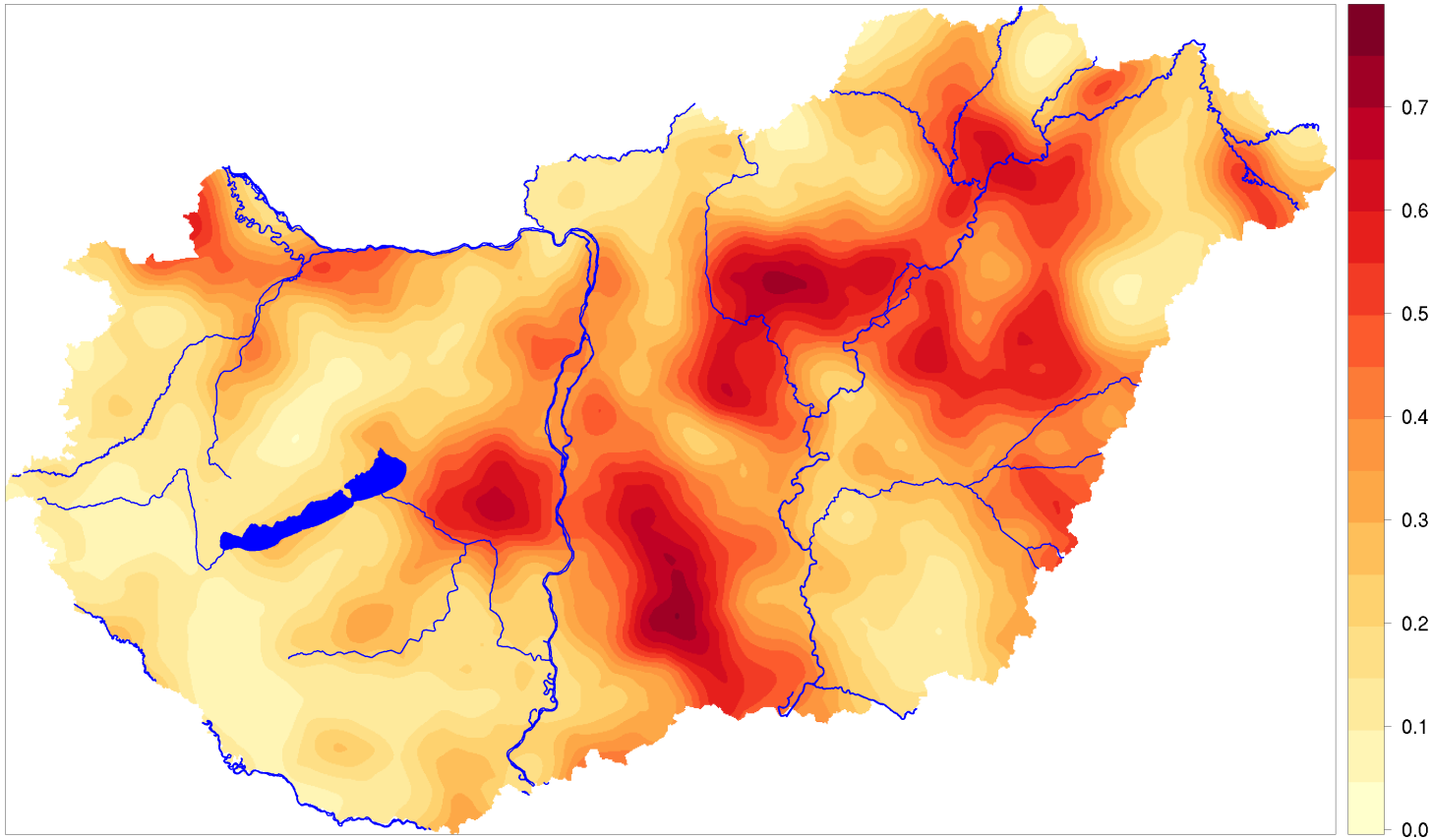
- Mapping
- Pattern analysis
- Ecological analysis

- Cluster
- Barrier



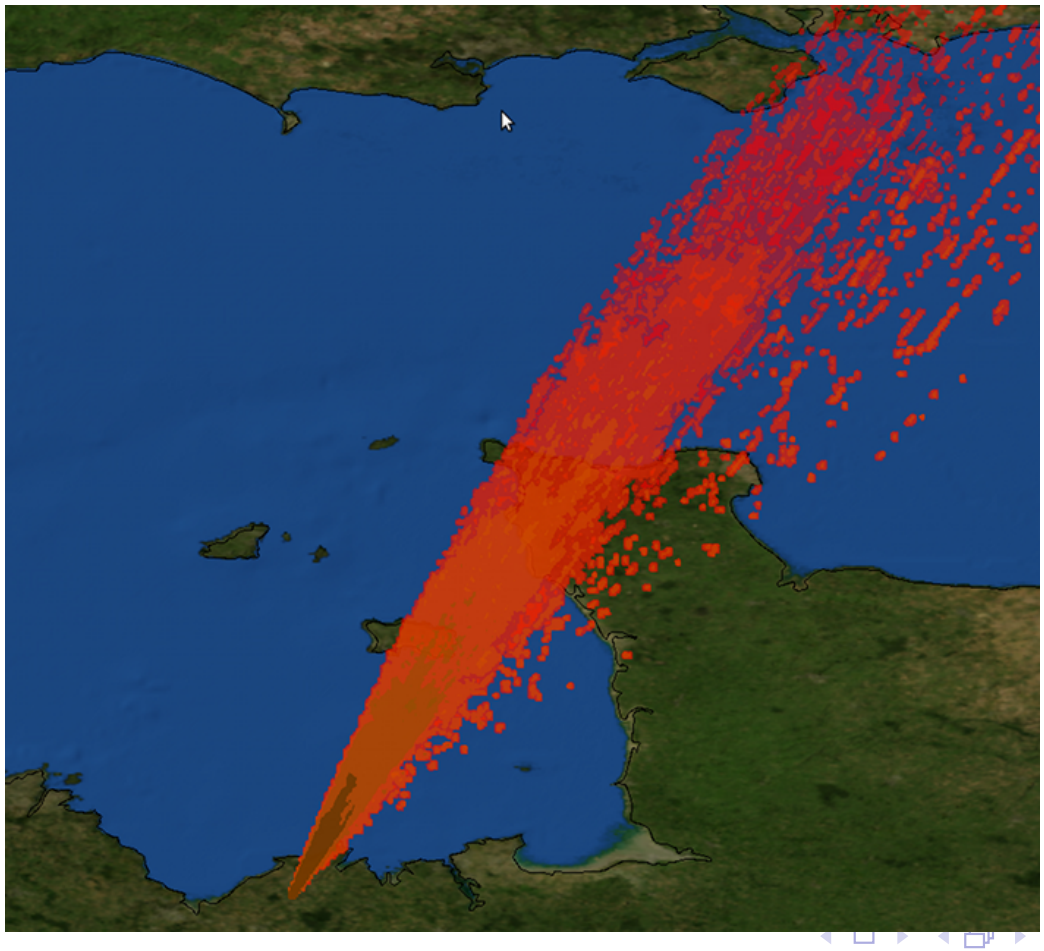
- Mapping
- Pattern analysis
- Ecological analysis

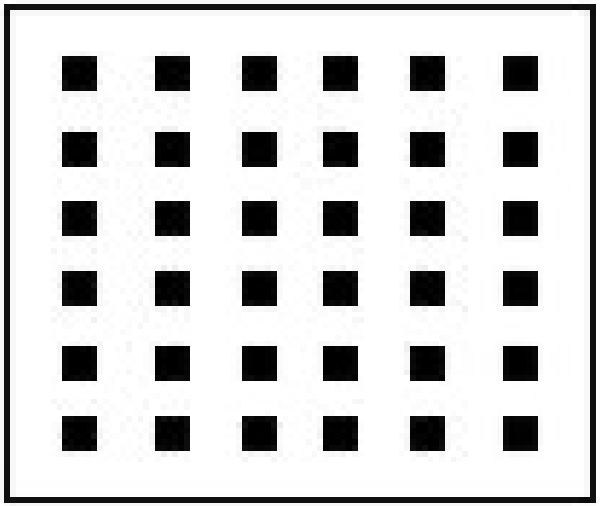
- $log(\theta_i) = p + x_i\beta + \mu_i + v_i$
- Spreading



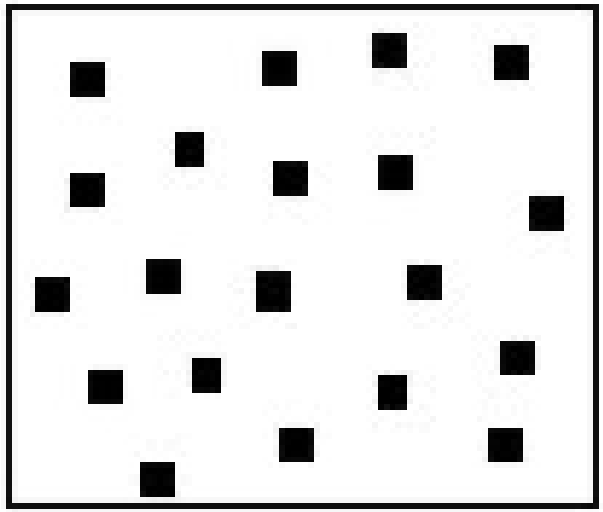
- Mapping
- Pattern analysis
- Ecological analysis

- $log(\theta_i) = p + x_i\beta + \mu_i + v_i$
- Spreading

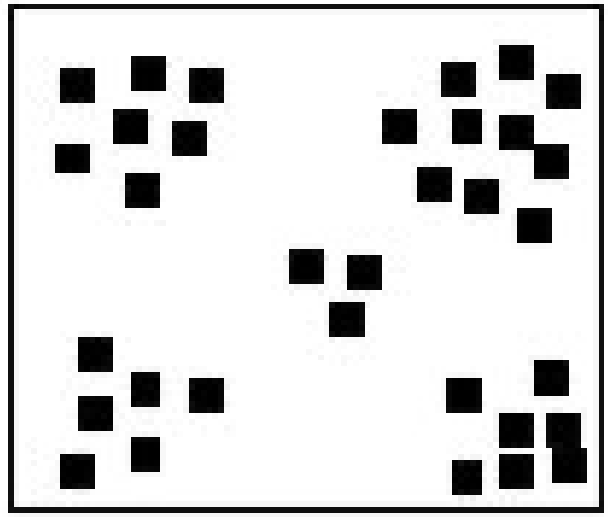




Regular



Random



Clustered

"cluster" is an unusual aggregation, real or perceived, of health events that are grouped together in time and space (CDC July 27, 1990 / 39(RR-11);1-16)

Methods

Global clustering methods:

- Aggregated data
  - Geary's c
  - Moran's I
- Point data
- Space-time

Methods

Global clustering methods:

- Aggregated data
  - Geary's c
  - Moran's I
- Point data
  - Cuzick-Edwards' k-nearest neighbour test
  - Ripley's K-function
  - Rogerson's cumulative sum (CUSUM) method
- Space-time

Methods

Global clustering methods:

Moran's *I*

$$I = n\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)\left(\sum\sum_{i\neq j} w_{ij}\right)}$$

Spatial distribution	Geary's c	Moran's I
Clustered	$0 \leq c < 1$	$I > 0$
Random	$c = 1$	$I = 0$
Uniform	$1 < c < 3$	$I < 0$

Methods

Global clustering methods:

- Aggregated data
  - Geary's *c*
  - Moran's *I*
- Point data
  - Cuzick-Edwards' *k*-nearest neighbour test
  - Ripley's K-function
  - Rogerson's cumulative sum (CUSUM) method
- Space-time

Methods

Global clustering methods:

- Aggregated data
  - Geary's *c*
  - Moran's *I*
- Point data
  - Cuzick-Edwards' *k*-nearest neighbour test
  - Ripley's K-function
  - Rogerson's cumulative sum (CUSUM) method
- Space-time
  - Barton's test
  - Ederer-Myers-Mantel (EMM) test
  - Jacquez's k-nearest neighbours test
  - Knox test
  - Mantel's test
  - Space-time k-function

Methods

Global clustering methods:

Cuzick-Edwards' k-nearest neighbour test

For each case, the test counts how many of the k-nearest neighbours are also cases, such that if there are n_1 cases, and $m_i(k)$ represents the number of cases among the k nearest neighbours of case i so that $0 \leq m_i(k) \leq k$, for $i = 1, \ldots, n_1$, a test statistic T_k can be calculated as:

$$T_k = \sum_{i=1}^{n_1} m_i(k)$$

Thus, when cases are clustered, the nearest neighbour to a case tends to be another case and T_k will be large. However, when all cases have controls as their nearest neighbours T_k will be zero. The observed value of T_k can be compared with the distribution of values computed using Monte Carlo randomization of the dataset.

Methods

Global clustering methods:

- Aggregated data
  - Geary's *c*
  - Moran's *I*
- Point data
  - Cuzick-Edwards' *k*-nearest neighbour test
  - Ripley's K-function
  - Rogerson's cumulative sum (CUSUM) method
- Space-time
  - Barton's test
  - Ederer-Myers-Mantel (EMM) test
  - Jacquez's k nearest neighbours test
  - Knox test
  - Mantel's test
  - Space-time k-function

Methods

Global clustering methods:

- Aggregated data
  - Geary's *c*
  - Moran's *I*
- Point data
  - Cuzick-Edwards' *k*-nearest neighbour test
  - Ripley's K-function
  - Rogerson's cumulative sum (CUSUM) method
- Space-time
  - Barton's test
  - Ederer-Myers-Mantel (EMM) test
  - Jacquez's k nearest neighbours test
  - Knox test
  - Mantel's test
  - Space-time k-function

Methods

Local clustering methods:

- Aggregated data
  - Getis and Ord's local $Gi(d)$ statistic
  - Local Moran test
- Point data
  - Openshaw's Geographical Analysis Machine (GAM)
  - Turnbull's Cluster Evaluation Permutation Procedure (CEPP)
  - Besag and Newell's method
  - Kulldorff's spatial scan statistic
- Space-time

Methods

Local clustering methods:

- Aggregated data
  - Getis and Ord's local $Gi(d)$ statistic
  - Local Moran test
- Point data
  - Openshaw's Geographical Analysis Machine (GAM)
  - Turnbull's Cluster Evaluation Permutation Procedure (CEPP)
  - Besag and Newell's method
  - Kulldorff's spatial scan statistic
- Space-time

Methods

Local Moran test

The local Moran test detects local spatial autocorrelation in aggregated data by decomposing Moran's *I* statistic into contributions for each area within a study region. Termed Local Indicators of Spatial Association (LISA), it's statistic for each area is calculated as:

$$I_i = Z_i \sum_{j, j \neq i}^{n} w_{ij} Z_j$$

where Z_i and Z_j are the observed values in standardized form, and w_{ij} is a spatial weights.

Methods

Local clustering methods:

- Aggregated data
  - Getis and Ord's local $Gi(d)$ statistic
  - Local Moran test
- Point data
  - Openshaw's Geographical Analysis Machine (GAM)
  - Turnbull's Cluster Evaluation Permutation Procedure (CEPP)
  - Besag and Newell's method
  - Kulldorff's spatial scan statistic
- Space-time
  - Kulldorff's space-time scan statistic

Methods

Kulldorff's spatial scan statistic

Gradually scanning a window across time and/or space, noting the number of observed and expected observations inside the window at each location. The window with the maximum likelihood is the most likely cluster, that is, the cluster least likely to be due to chance. Scan statistics use a different probability model depending on the nature of the data. A Bernoulli, discrete Poisson or space-time permutation model is used for count data. The standard purely spatial scan statistic imposes a circular window on the map. The window is in turn centered on each of several possible grid points positioned throughout the study region. For each grid point, the radius of the window varies continuously in size from zero to some upper limit specified by the user. In this way, the circular window is flexible both in location and size. In total, the method creates an infinite number of distinct geographical circles with different sets of neighboring data locations within them. Each circle is a possible candidate cluster.

Methods

Local clustering methods:

- Aggregated data
  - Getis and Ord's local $Gi(d)$ statistic
  - Local Moran test
- Point data
  - Openshaw's Geographical Analysis Machine (GAM)
  - Turnbull's Cluster Evaluation Permutation Procedure (CEPP)
  - Besag and Newell's method
  - Kulldorff's spatial scan statistic
- Space-time
  - Kulldorff's space-time scan statistic

Methods

Local clustering methods:

- Aggregated data
  - Getis and Ord's local $Gi(d)$ statistic
  - Local Moran test
- Point data
  - Openshaw's Geographical Analysis Machine (GAM)
  - Turnbull's Cluster Evaluation Permutation Procedure (CEPP)
  - Besag and Newell's method
  - Kulldorff's spatial scan statistic
- Space-time
  - Kulldorff's space-time scan statistic

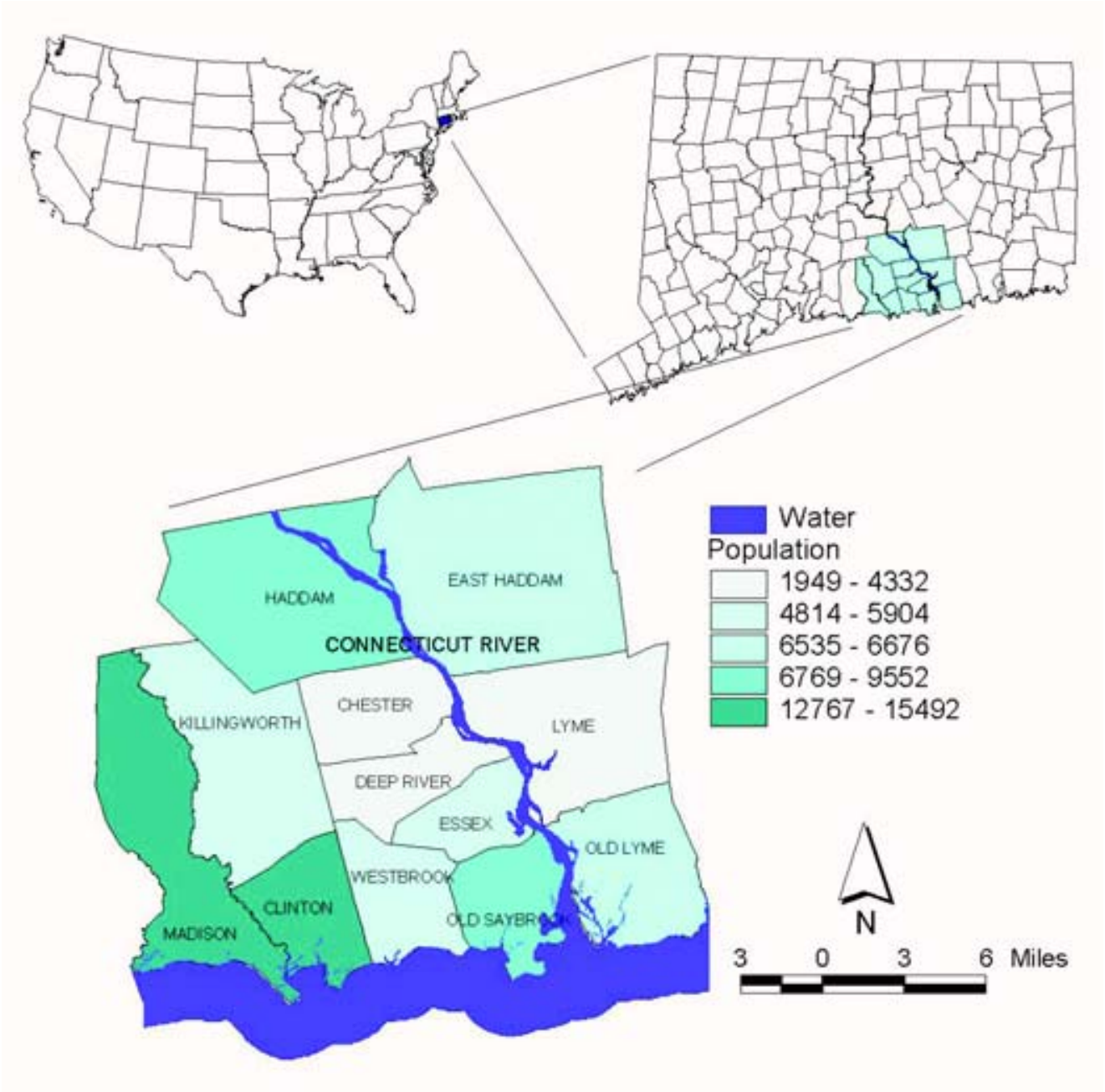
Focused clustering methods

Tools

- Geographic Information Systems (GIS)
  - Visualization (e.g. Google Earth)
  - Geoprocessing tools
  - Data management procedures
  - Quantum GIS (`http://www.qgis.org/`)

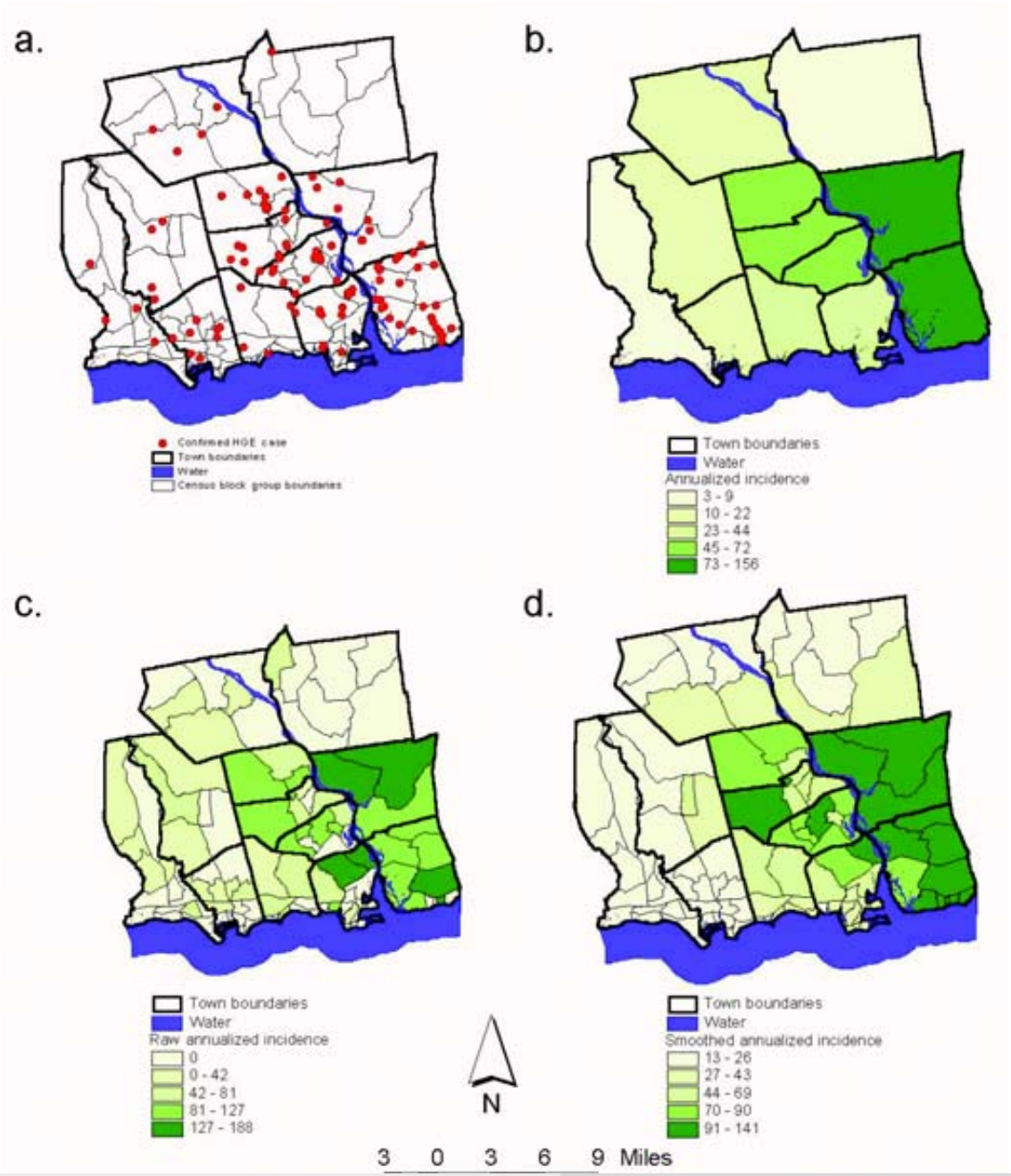
- Spatial analysis, modelling
  - R with packages e.g. `spdep`, `splancs`, `Dcluster` (`http://www.r-project.org/`)
  - SaTScan (`http://www.satscan.org/`)
  - ClusterSeer (`http://www.biomedware.com/?module=Page&sID=clusterseer`)

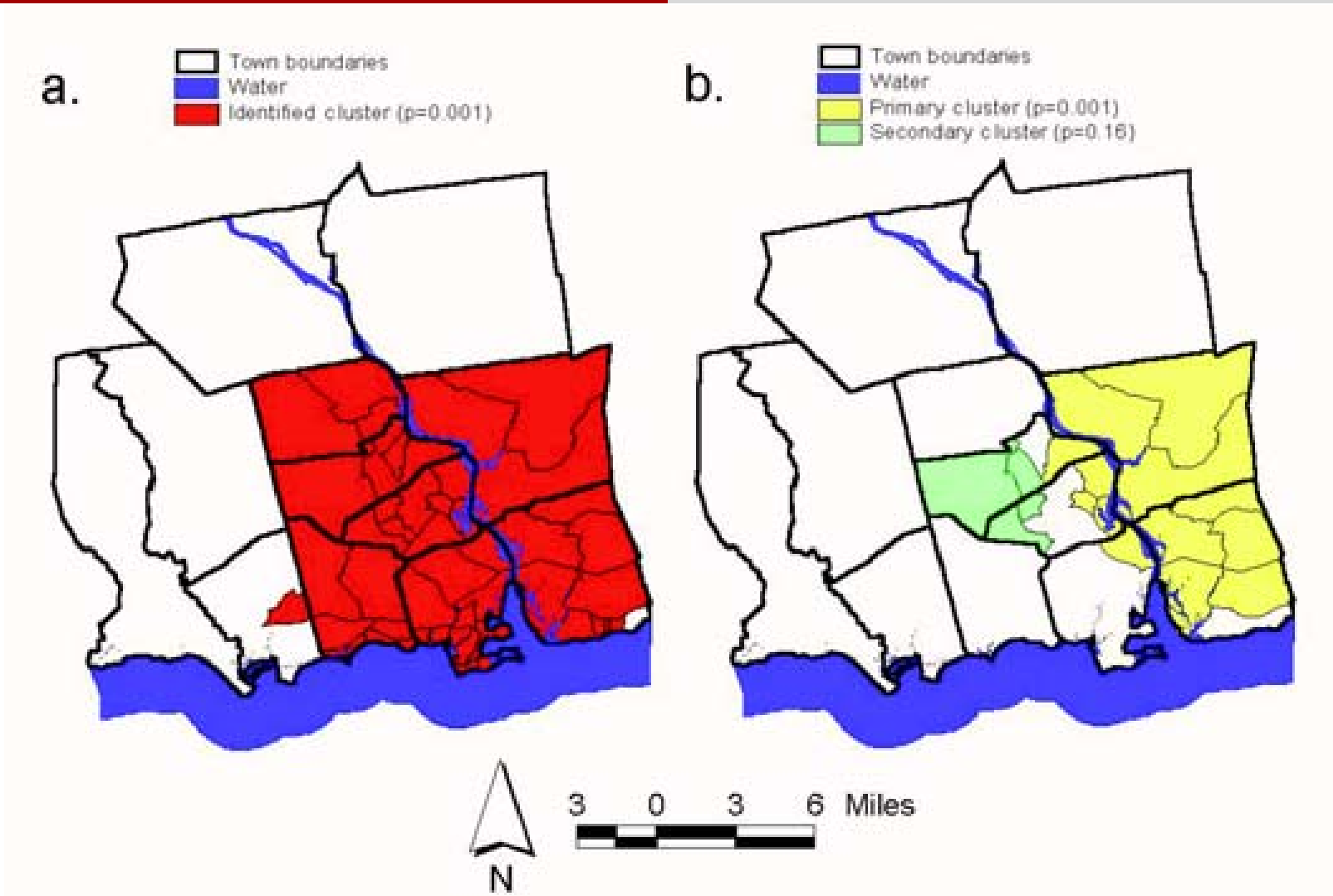
- Human granulocytic ehrlichiosis (HGE)
- The symptoms of HGE may include a sudden high fever, headache, muscle aches (myalgia), chills, and a general feeling of weakness and fatigue (malaise) within a week or so after initial infection.
- The agent of HGE is most closely related to *Ehrlichia phagocytophila*, which infects sheep and cattle, and *E. equi*, which causes disease in horses.
- HGE is transmitted to humans by the tick vector, *Ixodes scapularis*
- surveillance system for HGE was established in 1997 in a 12-town area around Lyme, Connecticut, USA
- During the 4 years of surveillance (1997–2000), the average annual incidence of confirmed cases of HGE in the 12-town area was 42 cases per 100,000 persons.
- Cluster analysis (Kulldorff spatial scan statistic) was performed with the default maximum spatial cluster size of $\leq 50\%$ of the population and again with a smaller maximum cluster size of $\leq 25\%$ to look for possible subclusters.


HADDAM
EAST HADDAM
CONNECTICUT RIVER
KILLINGWORTH
CHESTER
LYME
DEEP RIVER
ESSEX
OLD LYME
WESTBROOK
CLINTON
MADISON
OLD SAYBROOK
Water
Population
1949 - 4332
4814 - 5904
6535 - 6676
6769 - 9552
12767 - 15492
N
3 0 3 6 Miles

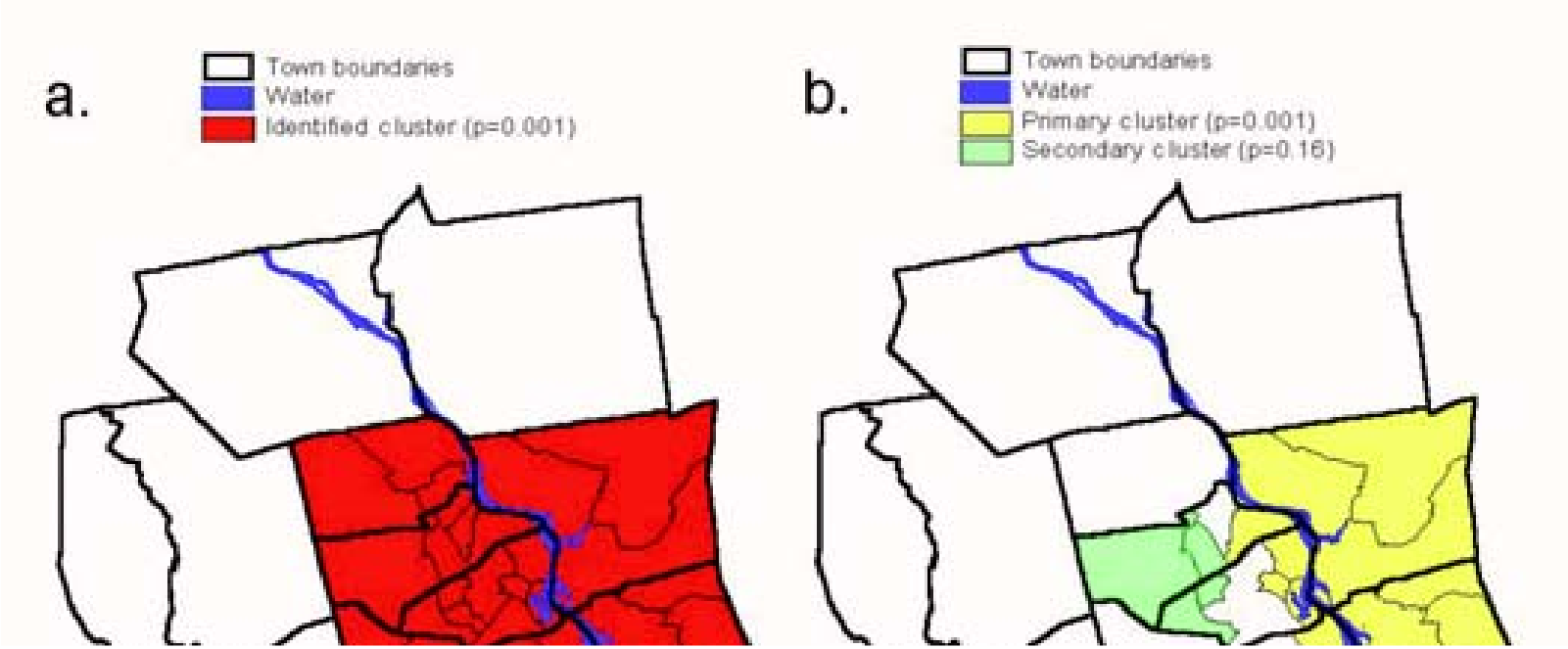
a. Confirmed human granulocytic ehrlichiosis (HGE) cases identified through active and passive surveillance systems, 1997–2000;
b. Raw annualized incidence of confirmed HGE cases by town, 1997– 2000*;
c. Raw annualized incidence of confirmed HGE cases by census block group*;
d. Smoothed annualized incidence of confirmed HGE cases by census block group.

*per 100,000 persons





a. Single identified cluster of human granulocytic ehrlichiosis (HGE) cases within the 12-town area (maximum cluster size 50% total population), relative risk (RR)=1.8, p=0.001; **b.** Two identified clusters of HGE cases within the 12-town area (maximum cluster size 25% total population): primary cluster: RR=2.6, p=0.001, secondary cluster: RR=2.6, p=0.16.

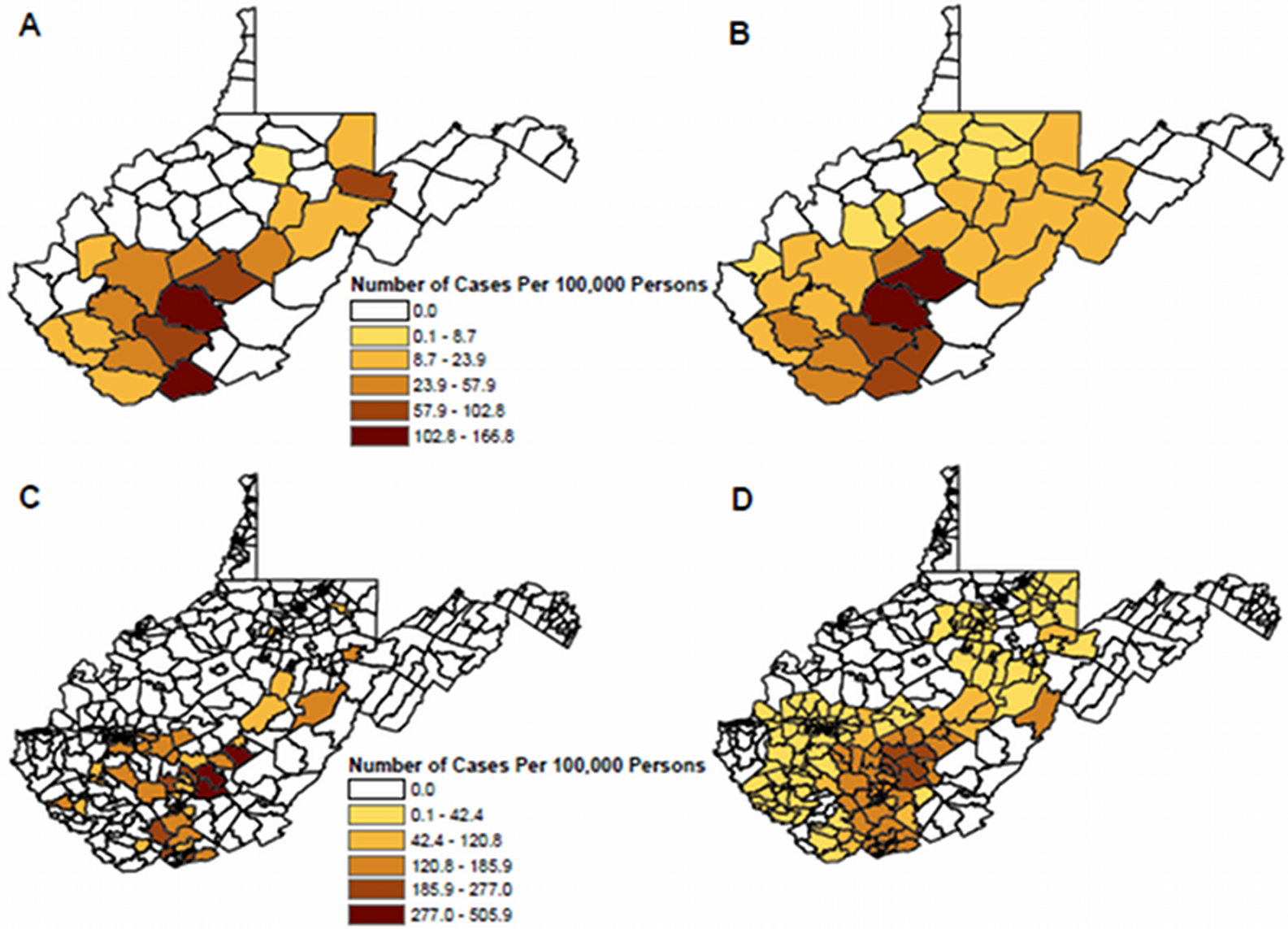


Reference

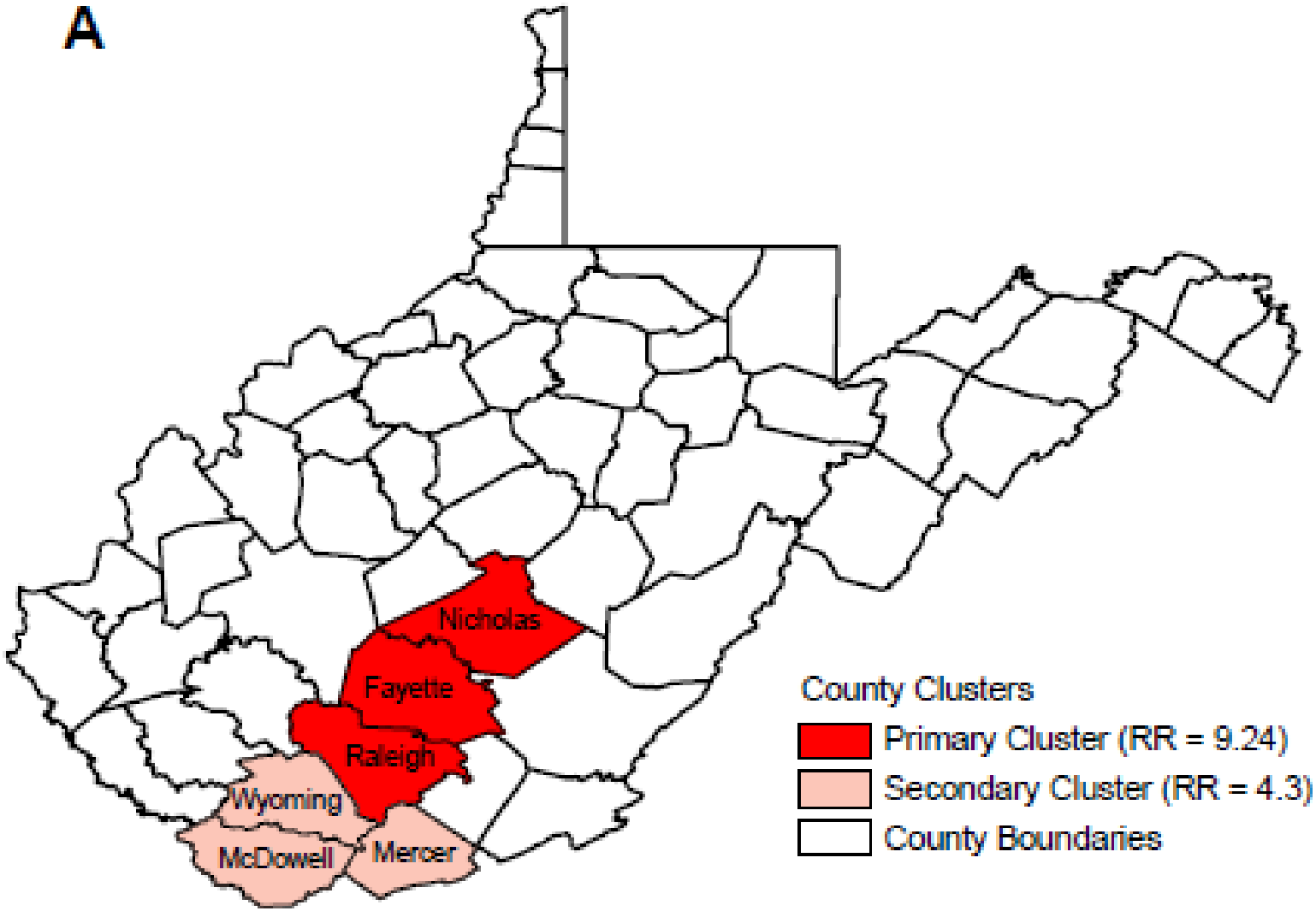
Emma K. Chaput, James I. Meek, and Robert Heimer (2002) Spatial Analysis of Human Granulocytic Ehrlichiosis near Lyme, Connecticut. Emerging Infectious Diseases Vol. 8, No. 9, 943-948

a. Single identified cluster of human granulocytic ehrlichiosis (HGE) cases within the 12-town area (maximum cluster size 50% total population), relative risk (RR)=1.8, p=0.001; **b.** Two identified clusters of HGE cases within the 12-town area (maximum cluster size 25% total population): primary cluster: RR=2.6, p=0.001, secondary

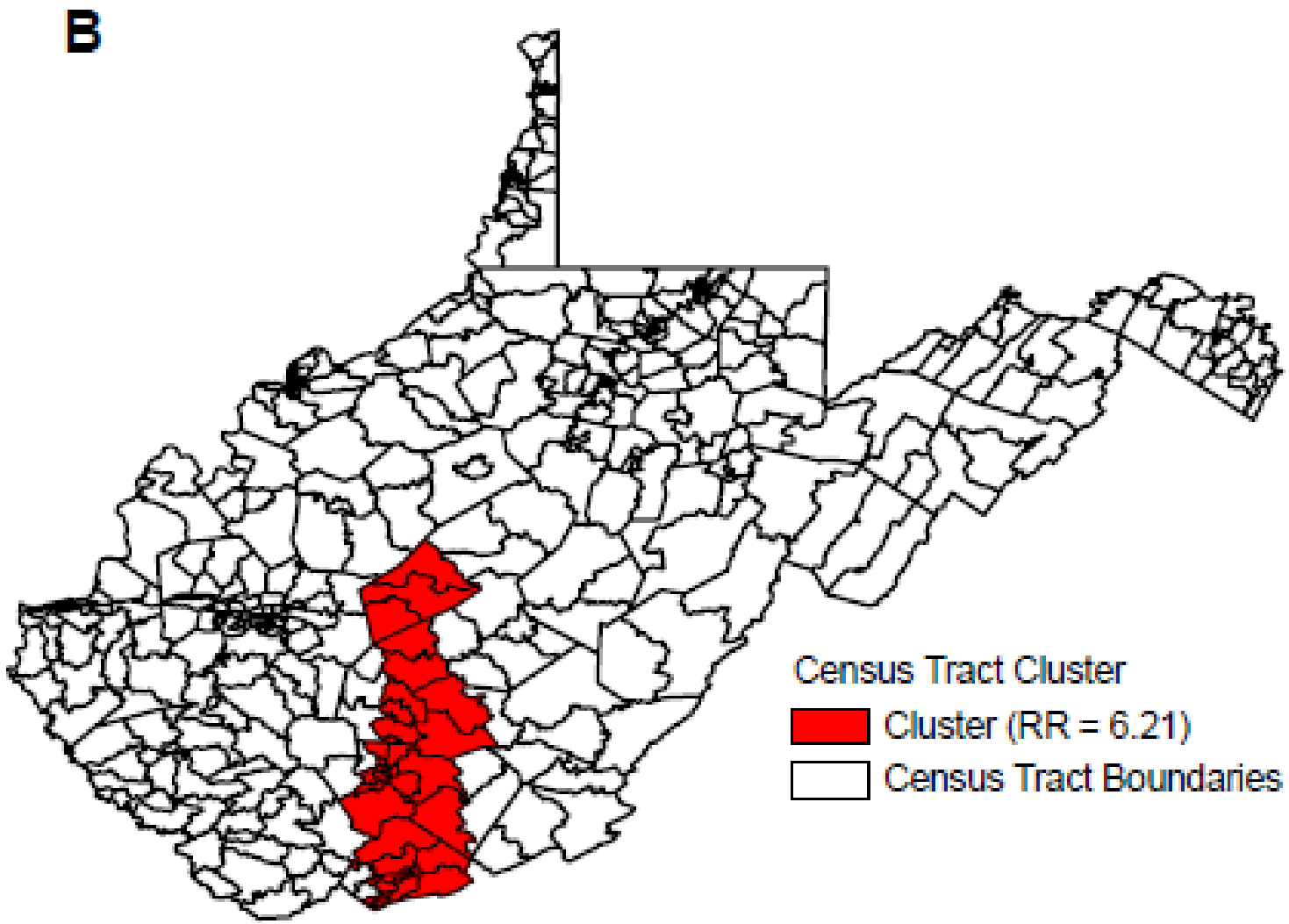
- La Crosse virus is a major cause of pediatric encephalitis in the United States (genus *Orthobunyavirus*, family *Bunyaviridae*)
- The virus is transmitted to humans through the bite of infective mosquitoes, the primary vector being the eastern tree-hole mosquito, *Aedes triseriatus*, though two invasive species, the Asian tiger mosquito, *Ae. albopictus*, and the Asian bush mosquito, *Ae. japonicus*, have been incriminated as possible secondary or bridge vectors and both species are known to feed on humans.
- Cases reported to the West Virginia Department of Health from 2003 to 2007, 81 were 15 years or younger, of which 68 had data available on the location of their primary residence.
- Statistically significant global clustering was detected at the county and census tracts with global Moran's I values of 0.4986 ($p = 0.0001$) and 0.2935 ($p = 0.0001$), respectively.
- Similarly, by Kulldorff's Spatial Scan Statistic statistically significant local clusters ($p < 0.05$) of high-risk were detected at both the county and the census tract levels



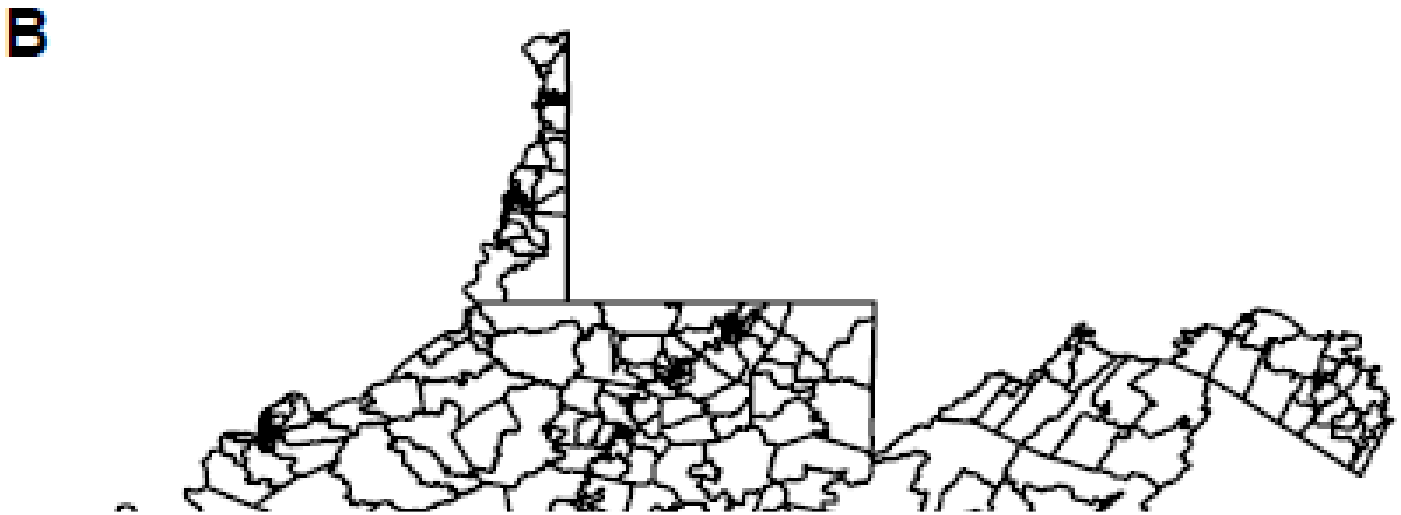
The unsmoothed and smoothed cumulative incidence of La Crosse virus infections at the county and census tract levels in children 15 years and younger. The distribution of unsmoothed risk of La Crosse virus infections at the county (A) and the census tract levels (C) for West Virginia. The distribution of spatial empirical Bayesian smoothed risk for La Crosse virus infections in West Virginia at the county (B) and the census tract levels (D).



Spatial clustering of La Crosse virus infection risk at the county and census tract levels in children 15 years and younger. These maps show the significant high-risk clusters for La Crosse virus infection in West Virginia at the county (A) and at the census tract levels (B) detected by Kulldorff's Spatial Scan Statistic. RR = relative risk.



Spatial clustering of La Crosse virus infection risk at the county and census tract levels in children 15 years and younger. These maps show the significant high-risk clusters for La Crosse virus infection in West Virginia at the county (A) and at the census tract levels (B) detected by Kulldorff's Spatial Scan Statistic. RR = relative risk.



Reference

Andrew D Haddow, Danae Bixler, Agricola Odoi (2011) The spatial epidemiology and clinical features of reported cases of La Crosse Virus infection in West Virginia from 2003 to 2007. BMC Infectious Diseases, 11:29

Spatial clustering of La Crosse virus infection risk at the county and census tract levels in children 15 years and younger. These maps show the significant high-risk clusters for La Crosse virus infection in West Virginia at the county (A) and at the census tract levels (B) detected by Kulldorff's Spatial Scan Statistic. RR = relative risk.

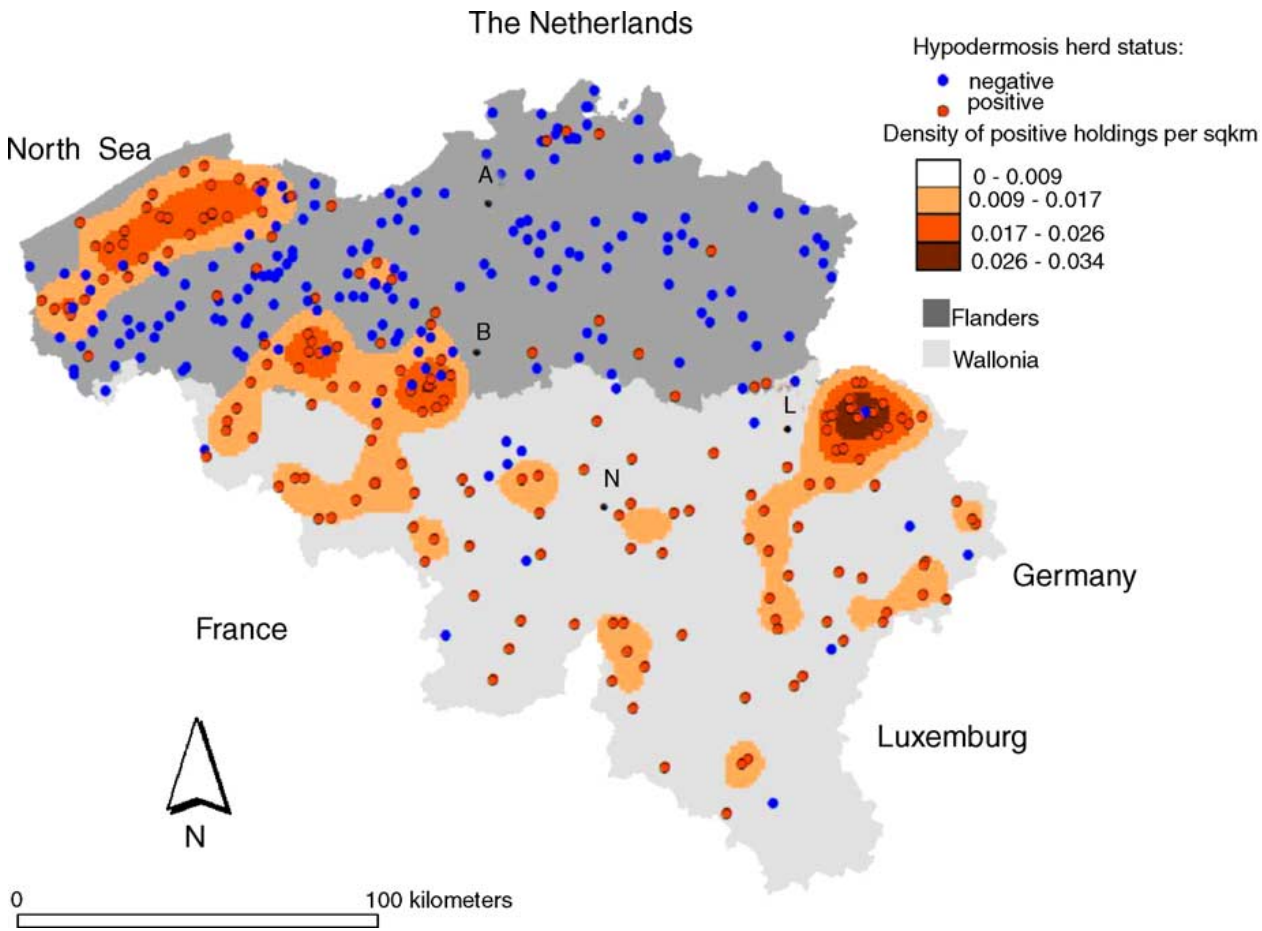
- Warble flies (*Hypoderma* spp.) are common parasites of cattle in the Northern Hemisphere. Infestations of cattle with the larvae of this fly cause serious damage to hides and occasional deaths (due to anaphylactic shock or toxaemia or damage to the central nervous system or oesophagus).
- Survey was carried out in 390 selected herds of all types (dairy, mixed and beef) from December 1997 to March 1998, which were included in a national infectious bovine rhinotracheitis and paratuberculosis (Johne's-disease) survey. All animals >24 months old were blood sampled and an ELISA was used on pooled serum samples (10 animals per pool).
- An ELISA to detect antibodies against *Hypoderma bovis* and *Hypoderma lineatum* infections was developed to screen cattle; the ELISA uses individual sera, pooled sera of up to 10 animals or milk
- The herd seroprevalence was 48.7% (95% confidence interval: 43.6–53.8); positive herds were mainly in the south of the country and along the North Sea coast.

The most likely cluster based on the spatial scan statistic represented negative herds (relative risk (RR) 0.29, P = 0.001) in the provinces of Antwerp and Limburg. There are three significant secondary clusters:

East of Belgium (province of Liége, Herve, German border) with RR 1.88 (P = 0.001) compared to the surrounding area, and 45 cases (23.96 expected) out of a population of 49.

Province of Hainaut (west of Wallonia), RR = 1.84 (P = 0.001), 35 cases (19.1 expected) out of 39.

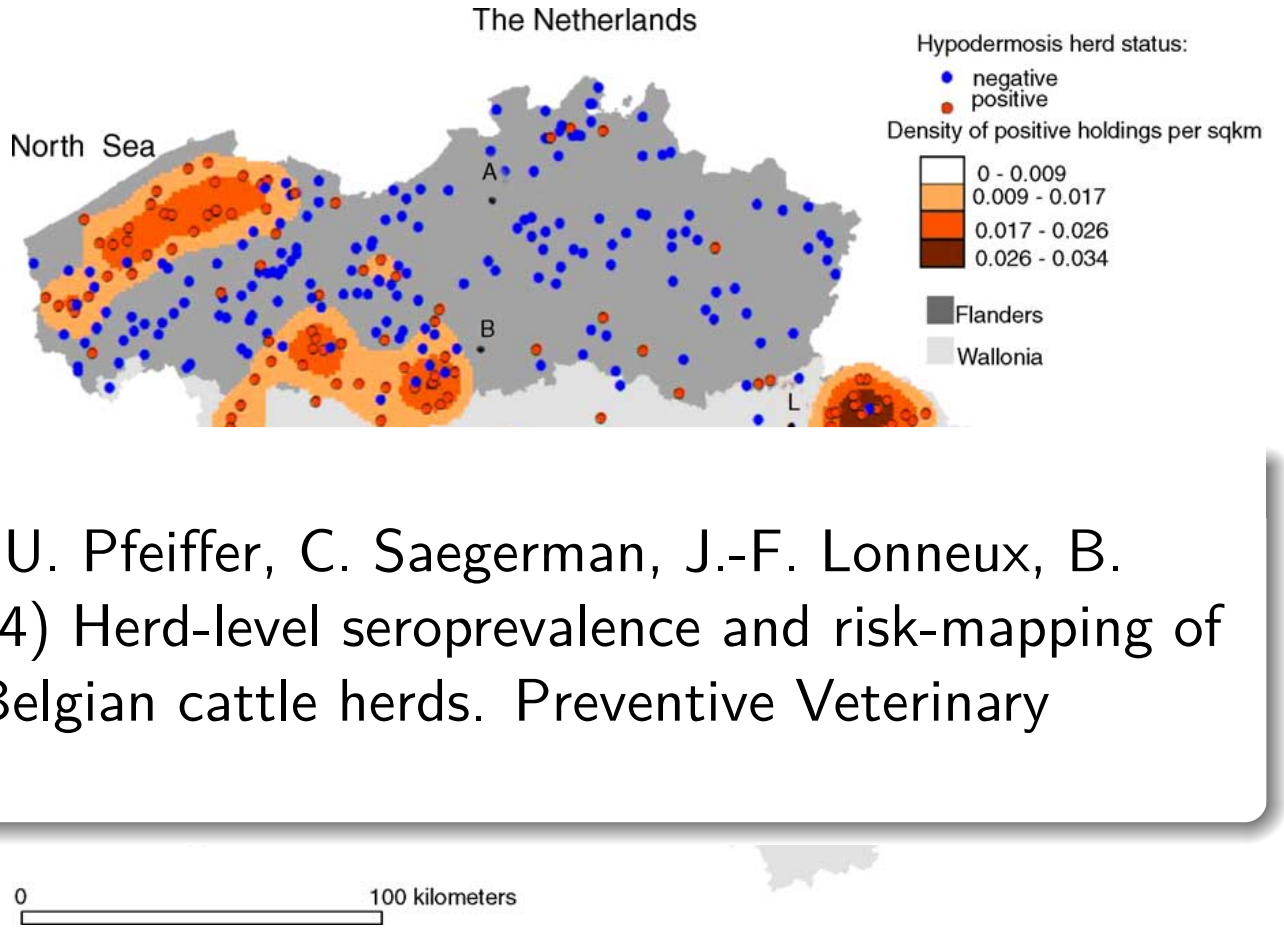
Province of East-Flanders (along the North Sea), RR = 0.20 (P = 0.006), three cases (15.2 expected) out of 31.



Serological hypodermosis herd infection status and kernel estimate of positive density of infected holdings: A, Antwerp; B, Brussels; L, Liége; N, Namur.

The most likely cluster based on the spatial scan statistic represented negative herds (relative risk (RR) 0.29, P = 0.001) in the provinces of Antwerp and Limburg. There are three significant secondary clusters:

East of Belgium (province of Liége, Herve, German border) with RR 1.88 (P = 0.001) compared to the surrounding area, and 45 cases (23.96 expected) out of a population of 49.

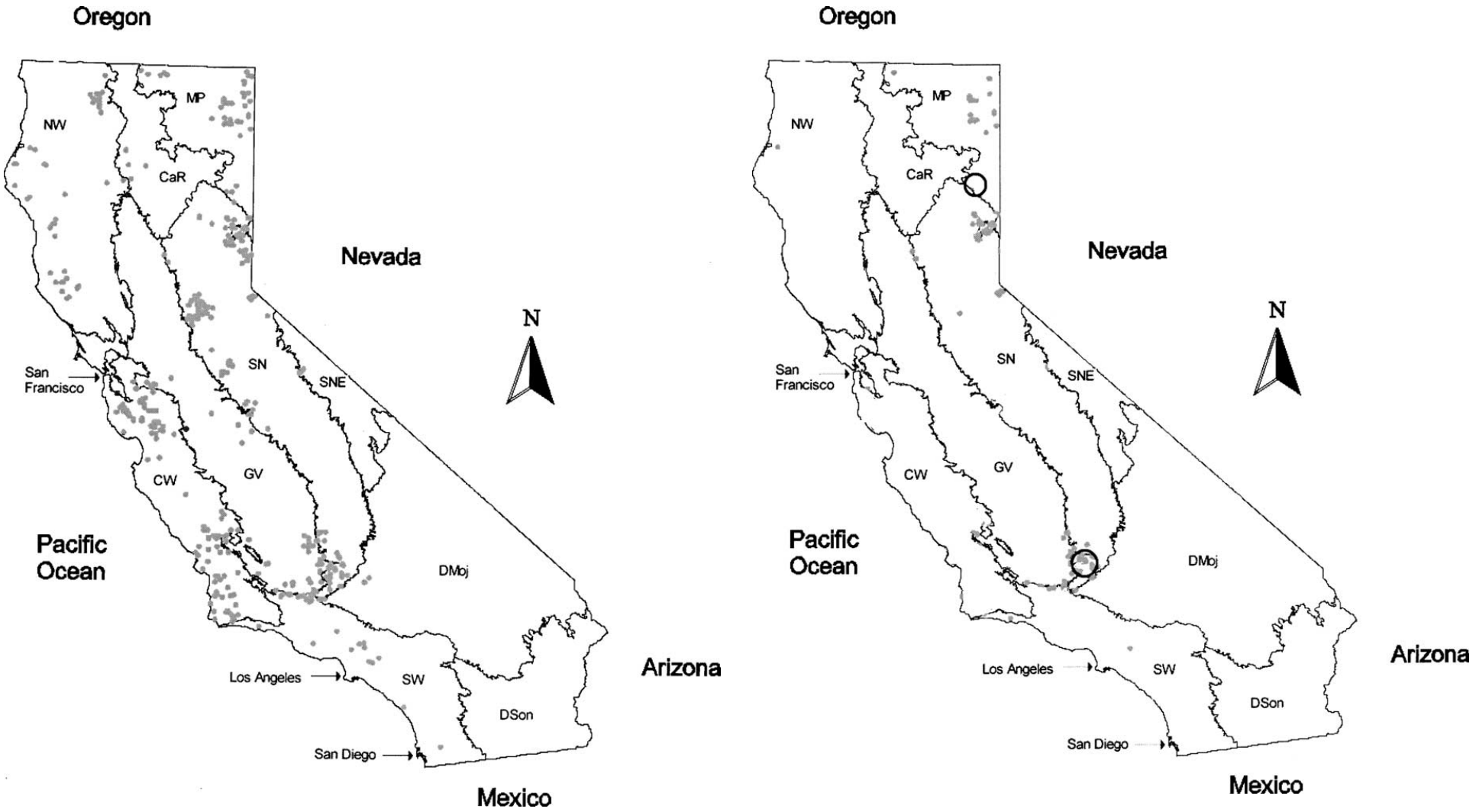


Serological hypodermosis herd infection status and kernel estimate of positive density of infected holdings: A, Antwerp; B, Brussels; L, Liége; N, Namur.

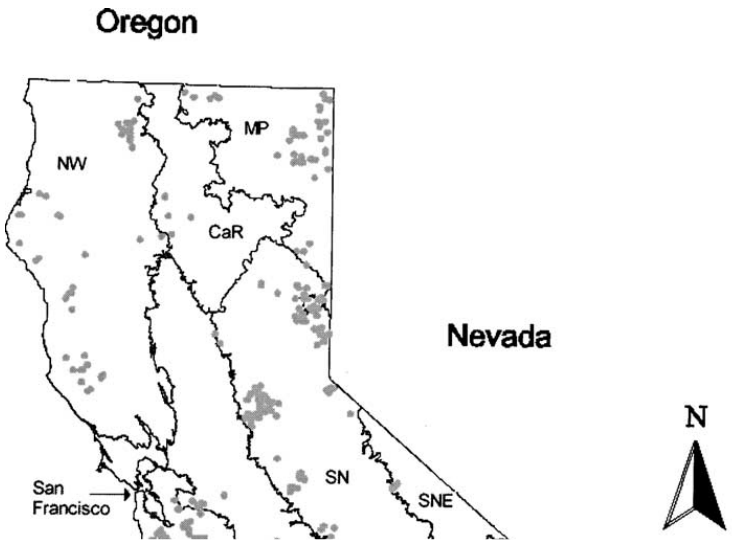
Reference

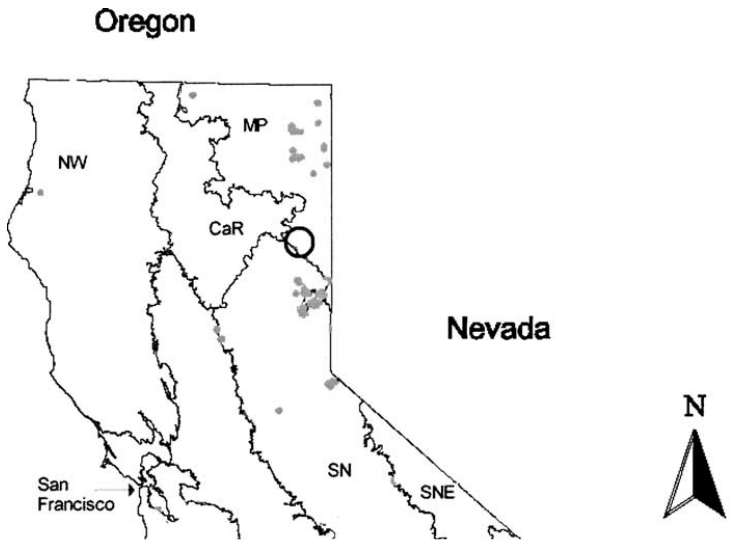
D. Haine, F. Boelaert, D.U. Pfeiffer, C. Saegerman, J.-F. Lonneux, B. Losson, K. Mintiens (2004) Herd-level seroprevalence and risk-mapping of bovine hypodermosis in Belgian cattle herds. Preventive Veterinary Medicine 65, 93–104

- Zoonotic transmission of sylvatic plague caused by *Yersinia pestis* occurs in California, USA. Coyotes are ubiquitous throughout California and can become infected by the agent. Carnivores are involved in the plague cycle as potential carriers of infective fleas to other rodent populations. Canids appear to be highly resistant to infection with *Y. pestis*
- The geographic distribution of 863 coyotes tested was examined between 1994-1998.
- It was 11.7% of specimens positive to *Y. pestis*
- *Y. pestis* was more prevalent in eastern portions of the state
- Cuzick-Edward's test as global tested
- Spatial scan statistic for localization



Location of coyotes sampled (left) and testing positive to *Y. pestis* (right), and location of clusters.





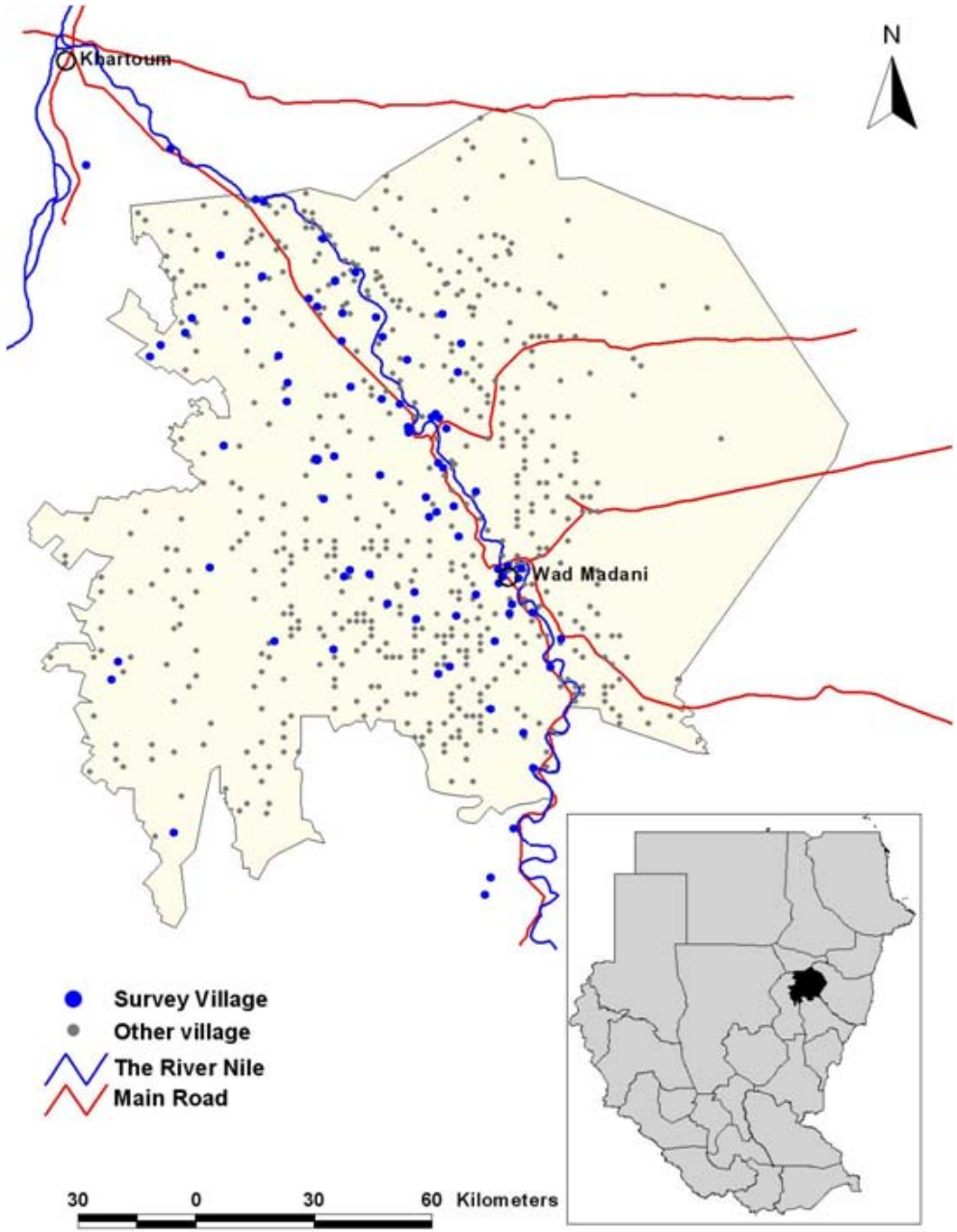
Reference

B.R Hoar, B.B Chomel, D.L Rolfe, C.C Chang, C.L Fritz, B.N Sacks, T.E Carpenter (2003) Spatial analysis of *Yersinia pestis* and *Bartonella vinsonii* subsp. *berkhoffii* seroprevalence in California coyotes (Canis latrans). Preventive Veterinary Medicine 56, 299–311

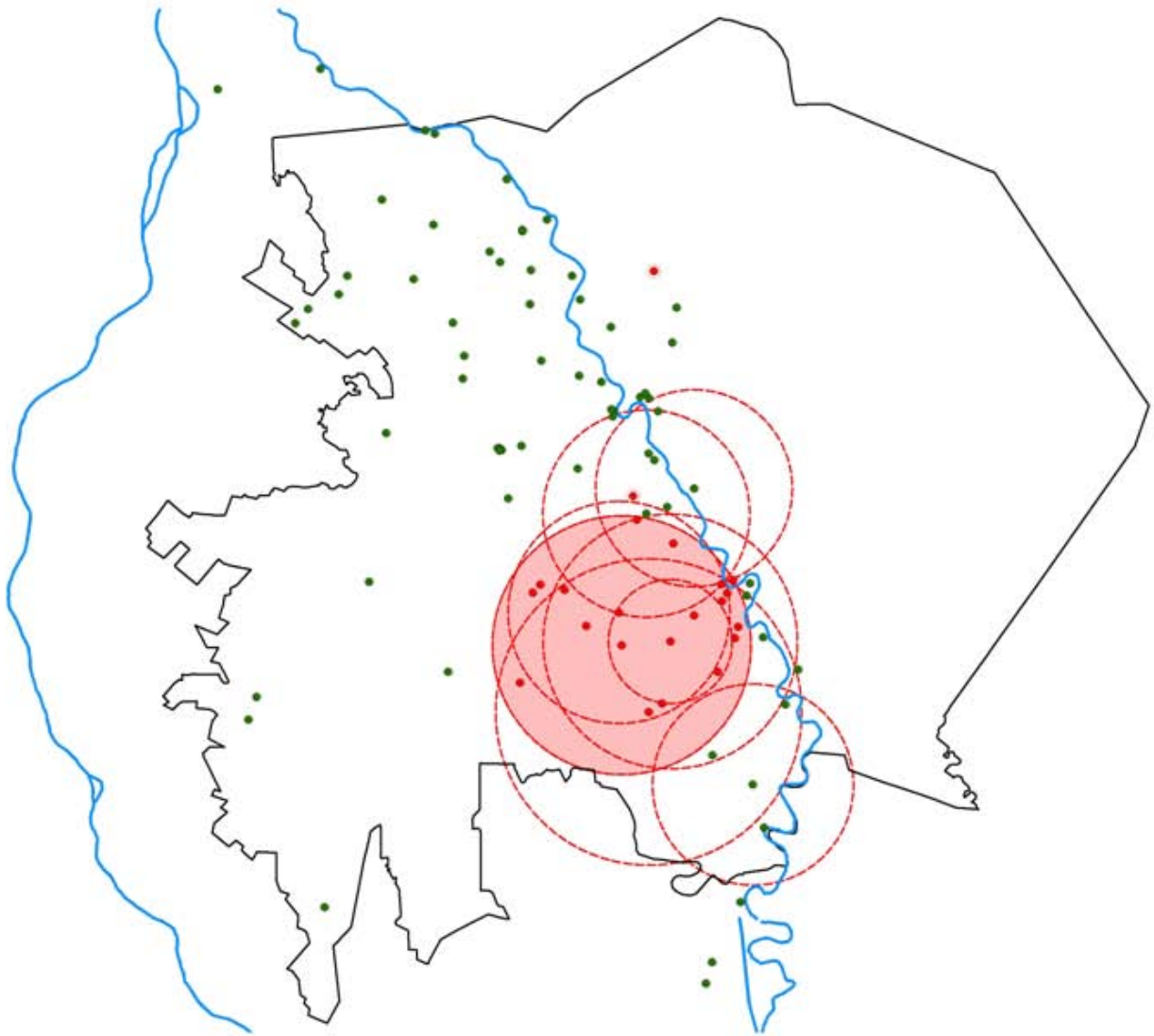
Location of coyotes sampled (left) and testing positive to *Y. pestis* (right), and location of clusters.

- Data from surveys undertaken in January each year from 1999-2009 in 88 villages in the Gezira state were assembled. During each survey, about a 100 children between the ages two to ten years were sampled to examine the presence of *P. falciparum* parasites.
- Spatial-only and space-time clustering, Kulldorff scan statistic
- Over the study period, 96,022 malaria slide examinations were undertaken and the *P. falciparum* prevalence was 8.6% in 1999 and by 2009 this had reduced to 1.6%.
- The cluster analysis showed the presence of one significant spatial-only cluster in each survey year and one significant space-time cluster over the whole study period. The primary spatial-only clusters in 10/11 years were either contained within or overlapped with the primary space-time cluster.

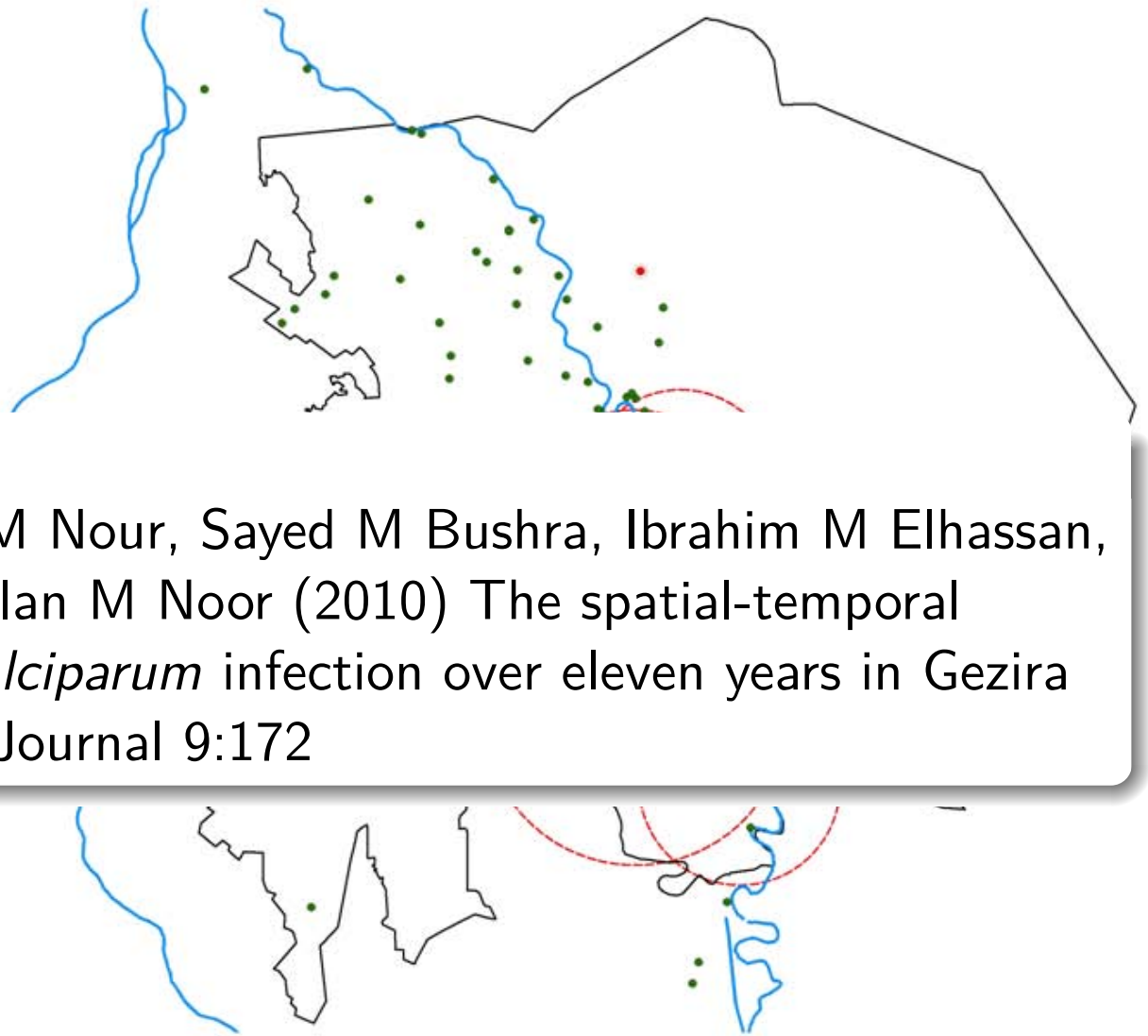
Map of Gezira state showing the location of the state capital (Wad Madani) in relation to the national capital (Khartoum), the distribution of settlements in Gezira and the location of the distribution of 88 survey locations where the *P. falciparum* prevalence surveys were undertaken from 1999-2009. Inset is the state map of the Sudan showing the location of Gezira state.



Location of the space-time primary cluster (Kulldorff statistic was significant at $P < 0.01$, shaded) of P. falciparum prevalence in Gezira state from 1999 to 2009. Shown also are the spatial only primary clusters for each year (circles with broken red boundaries). Except for the spatial-only cluster in 2007 (northwest of Gezira, on the western side of the Blue Nile), all other spatial-only clusters either.



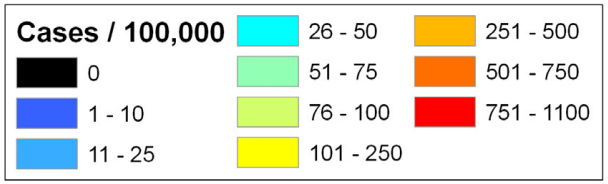
Location of the space-time primary cluster (Kulldorff statistic was significant at $P < 0.01$, shaded) of P. falciparum prevalence in Gezira state from 1999 to 2009. Shown cluster in 2007 (northwest of Gezira, on the western side of the Blue Nile), all other spatial-only clusters either.



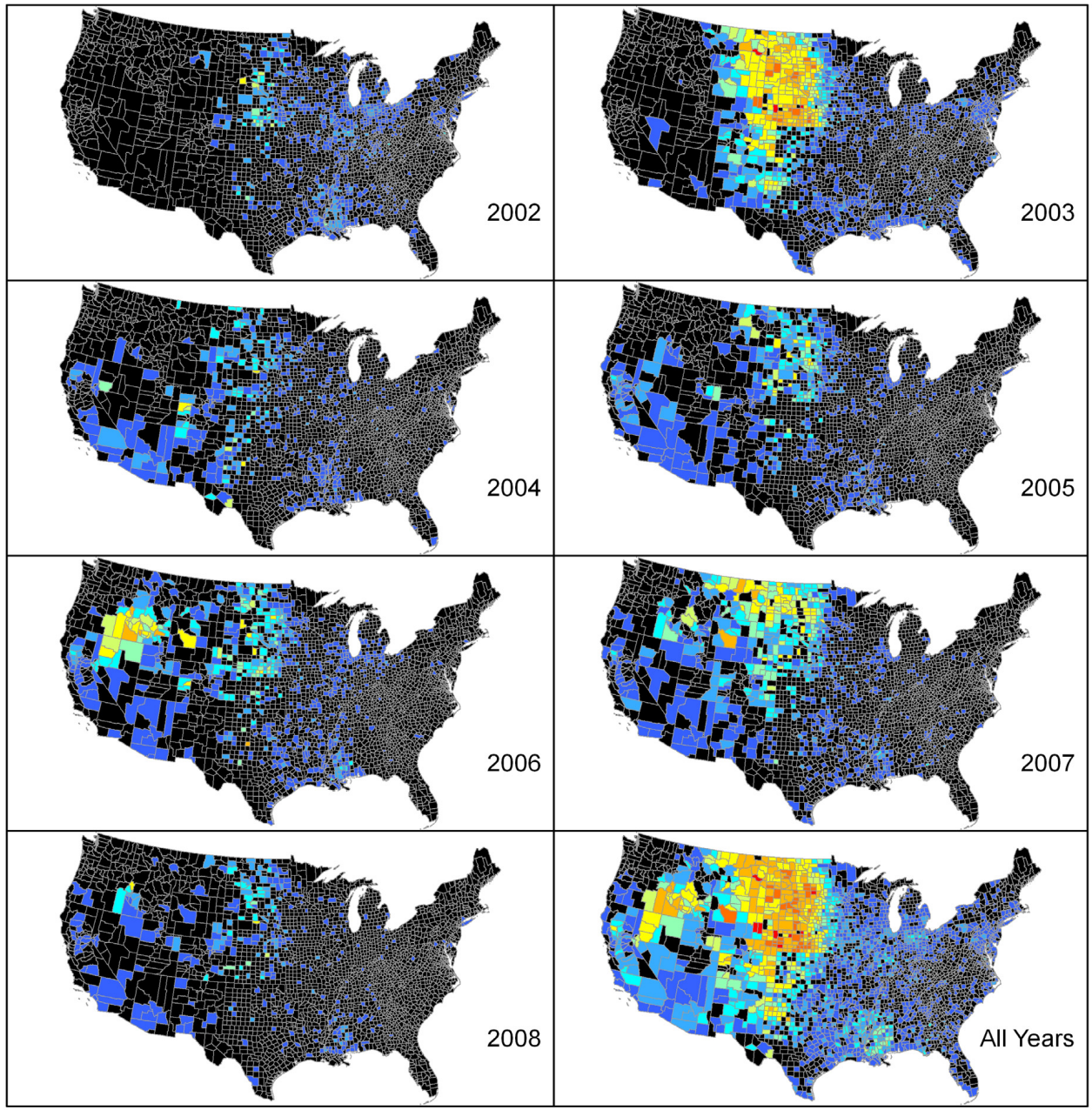
Reference

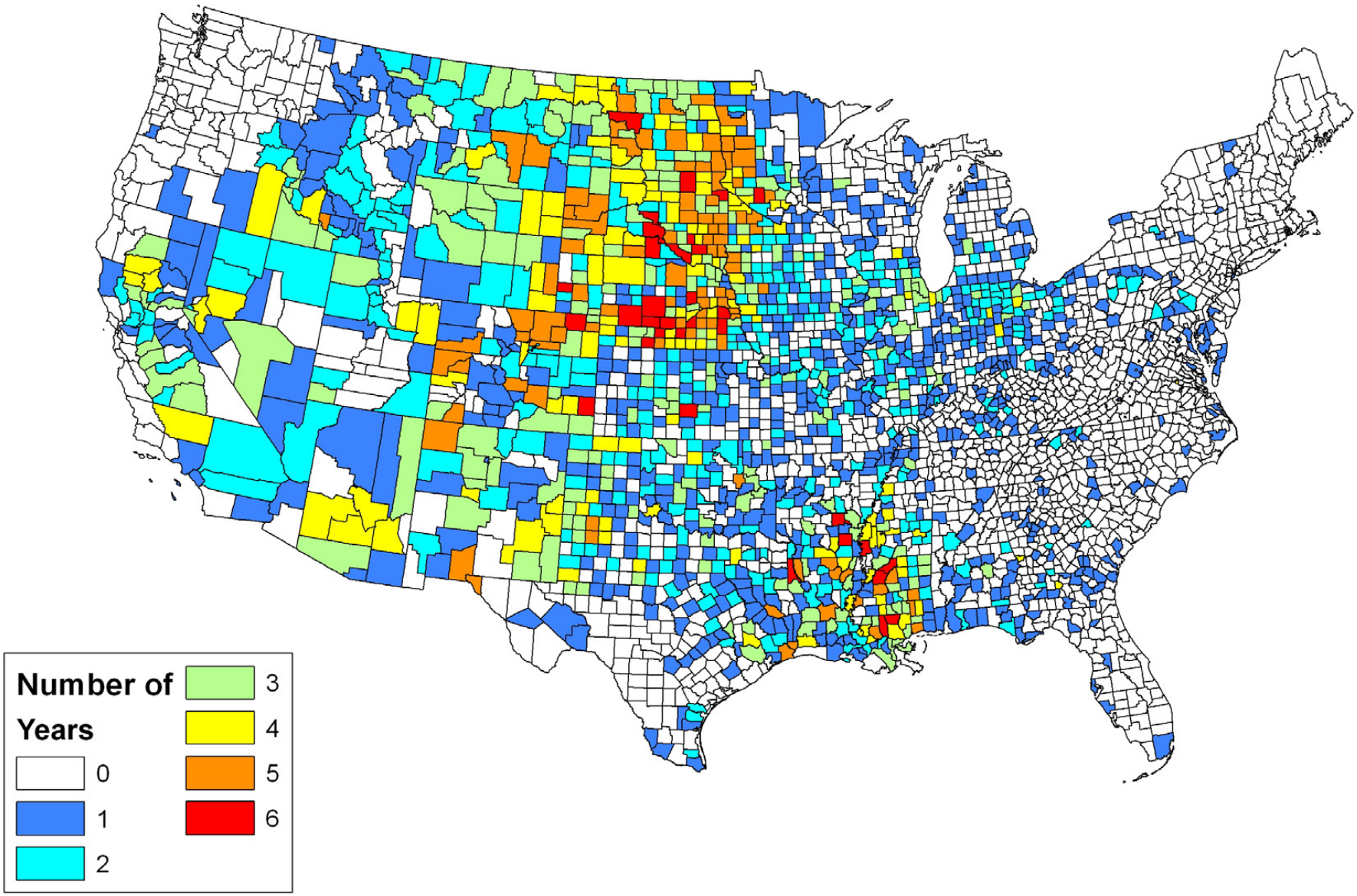
Samia E Mirghani, Bakri YM Nour, Sayed M Bushra, Ibrahim M Elhassan, Robert W Snow and Abdisalan M Noor (2010) The spatial-temporal clustering of *Plasmodium falciparum* infection over eleven years in Gezira State, The Sudan. Malaria Journal 9:172

- West Nile virus (WNV) is a vector-borne illness that can severely affect human health. After introduction on the East Coast in 1999, the virus quickly spread and became established across the continental United States.
- To quantify spatially and temporally variations, Kulldorff's spatial scan statistic and Anselin's Local Moran's I statistic was used, uncovering spatial clustering of human WNV incidence at the county level in the continental United States from 2002–2008.
- The spatial scan and Local Moran's I statistics revealed several consistent, important clusters or hot-spots with significant year-to-year variation.
- In 2002, before the pathogen had spread throughout the country, there were significant regional clusters in the upper Midwest and in Louisiana and Mississippi.
- The largest and most consistent area of clustering throughout the study period was in the Northern Great Plains region including large portions of Nebraska, South Dakota, and North Dakota, and significant sections of Colorado, Wyoming, and Montana.
- In 2006, a very strong cluster centered in southwest Idaho was prominent.

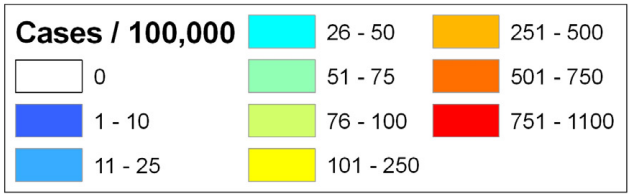


WNV human incidence by year. Each panel shows the number of human WNV cases per 100,000 people for a single year or for all of the years combined.

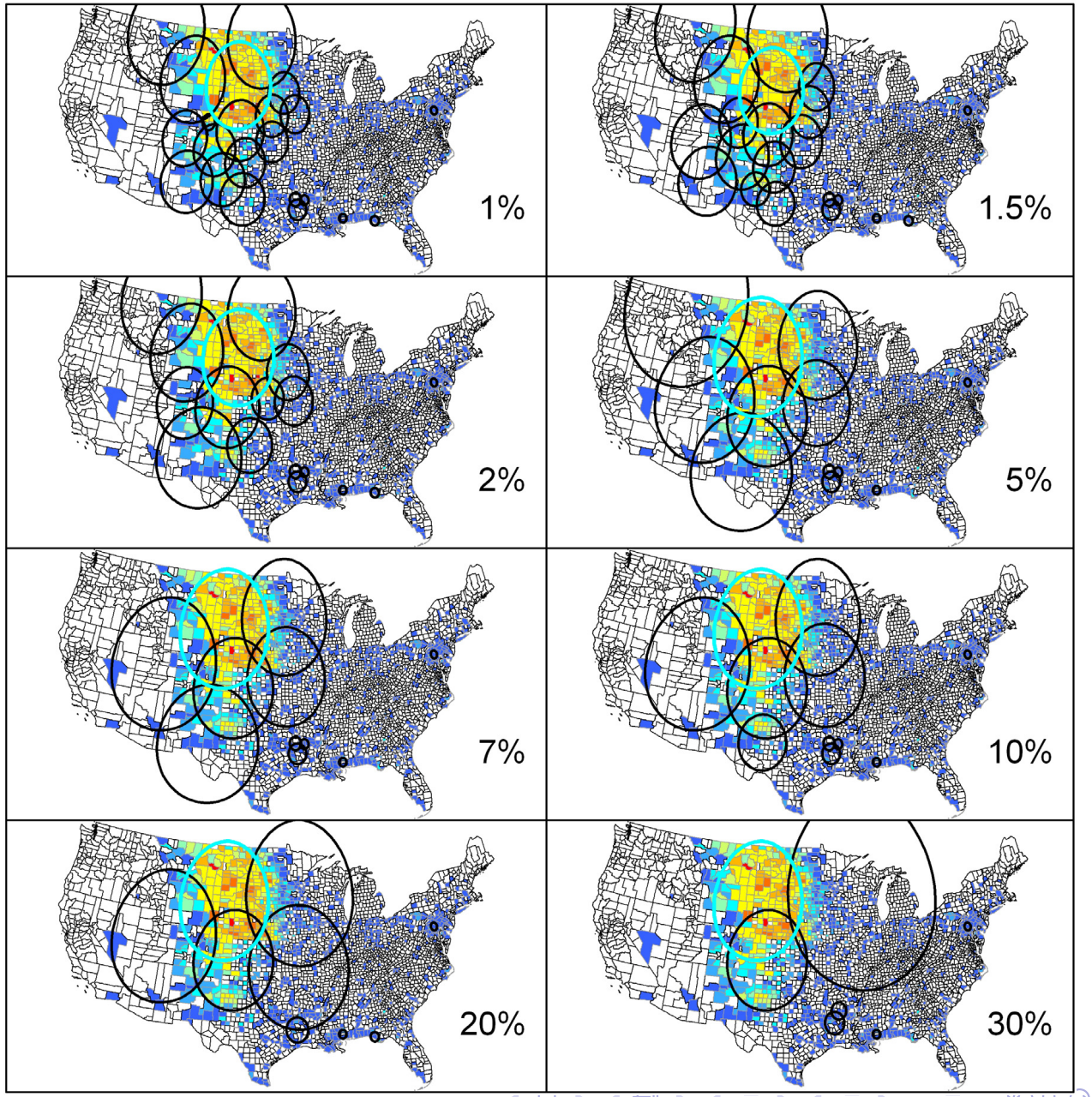


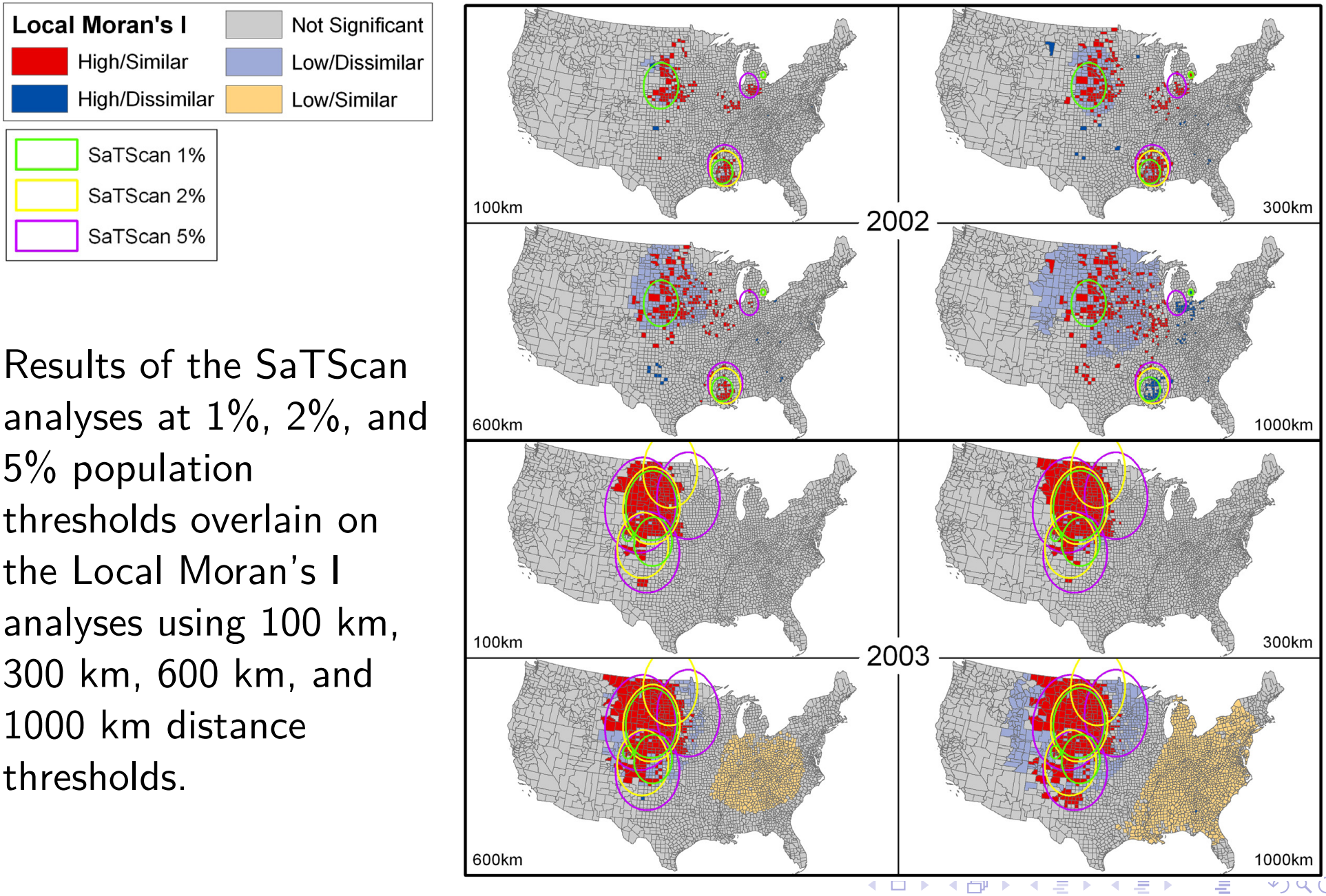


Counties of the US with regards to how many years each county had higher than expected rates of human WNV incidence.

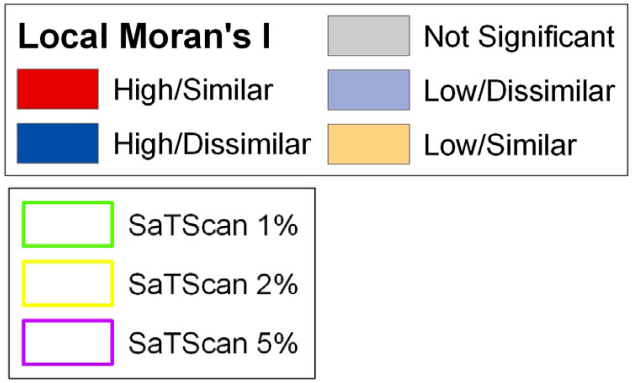


SaTScan results for 2003 at various population thresholds. Results of running the spatial scan statistic with varying population limits for human WNV incidence in the continental United States for 2003 are shown. Blue indicates areas with low rates of human WNV incidence and red represents areas with high rates of human WNV incidence.

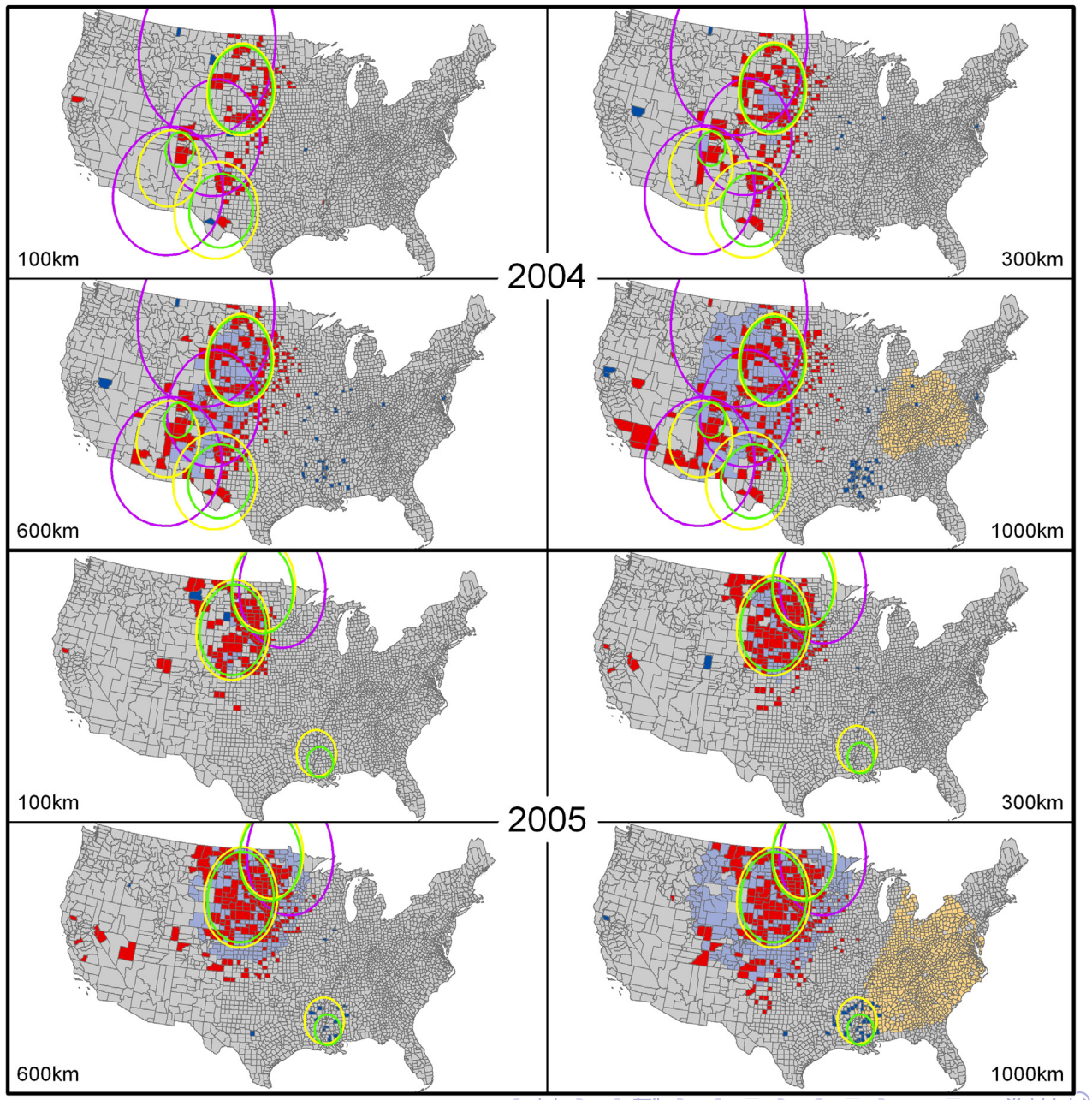


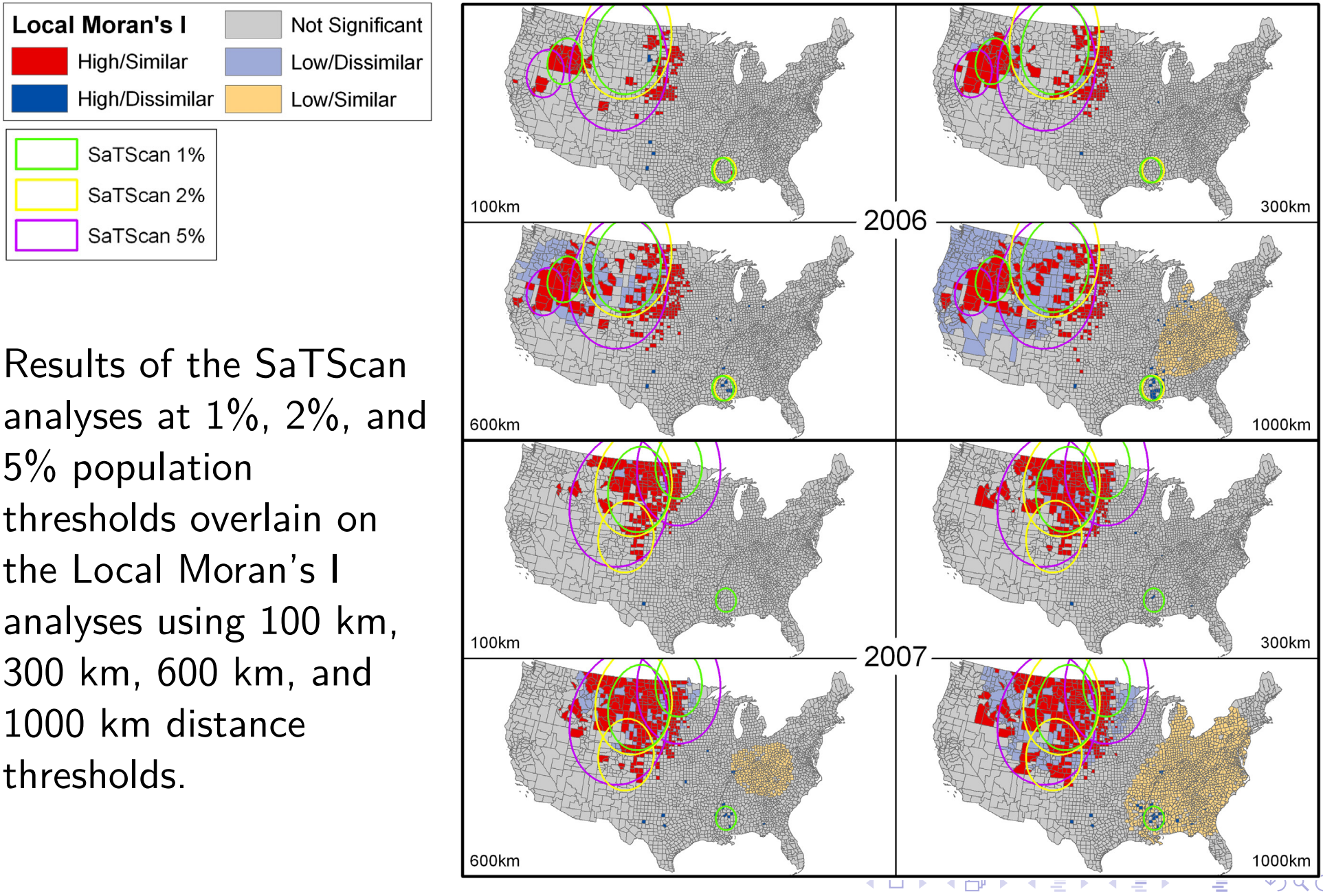


Results of the SaTScan analyses at 1%, 2%, and 5% population thresholds overlain on the Local Moran's I analyses using 100 km, 300 km, 600 km, and 1000 km distance thresholds.

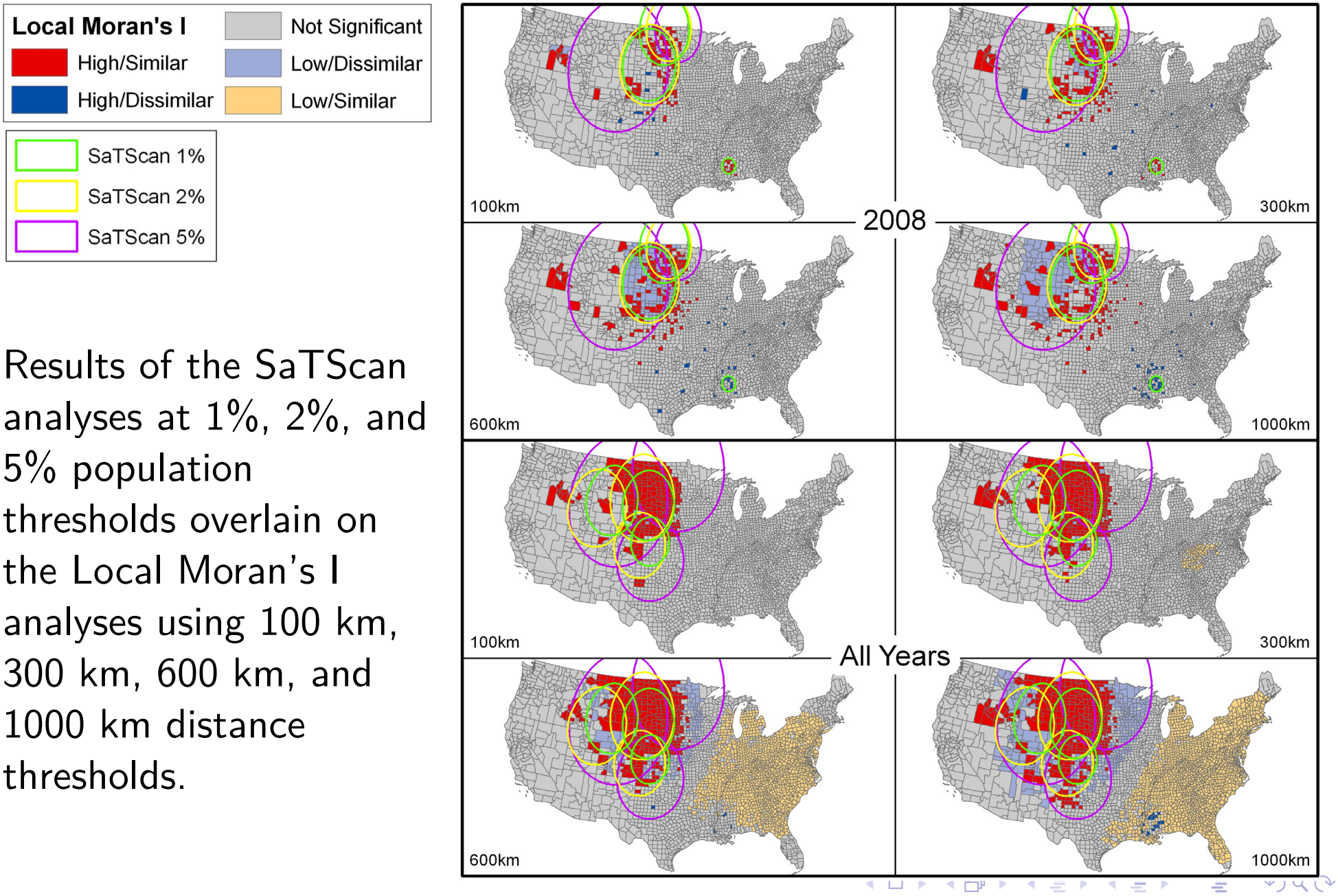


Results of the SaTScan analyses at 1%, 2%, and 5% population thresholds overlain on the Local Moran's I analyses using 100 km, 300 km, 600 km, and 1000 km distance thresholds.

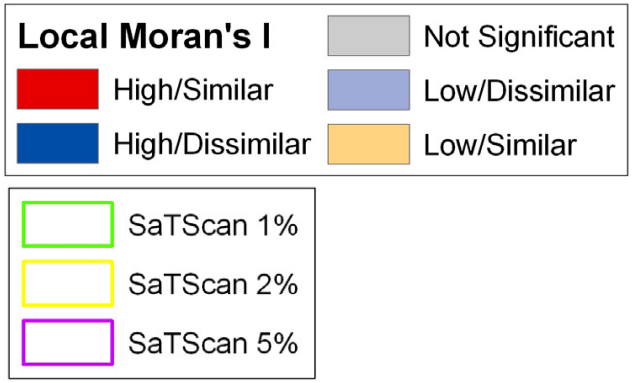


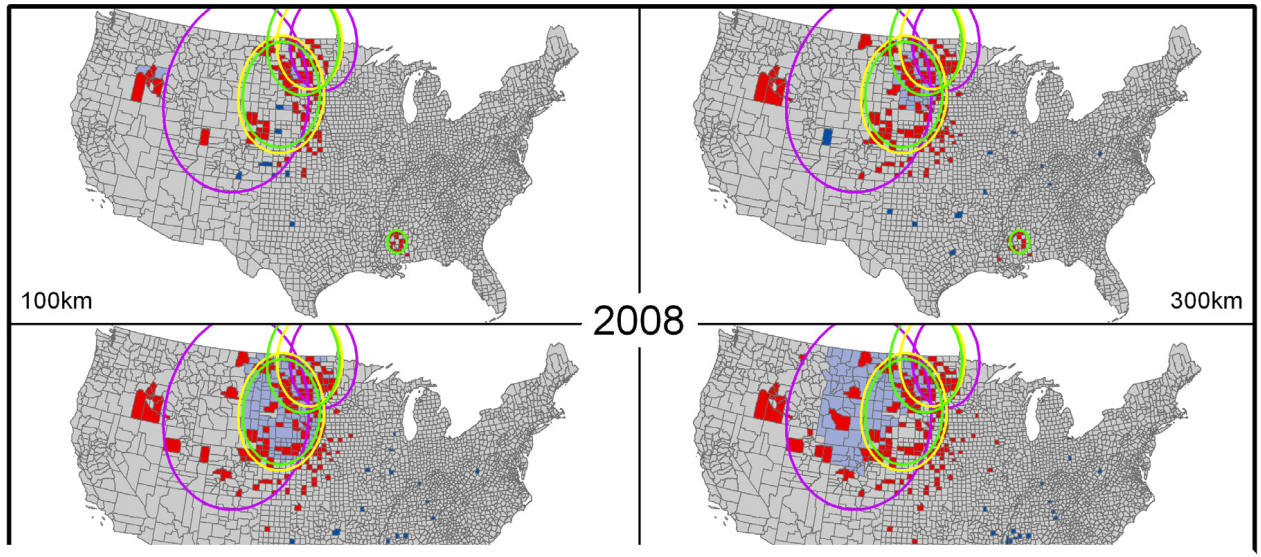


Results of the SaTScan analyses at 1%, 2%, and 5% population thresholds overlain on the Local Moran's I analyses using 100 km, 300 km, 600 km, and 1000 km distance thresholds.



Results of the SaTScan analyses at 1%, 2%, and 5% population thresholds overlain on the Local Moran's I analyses using 100 km, 300 km, 600 km, and 1000 km distance thresholds.



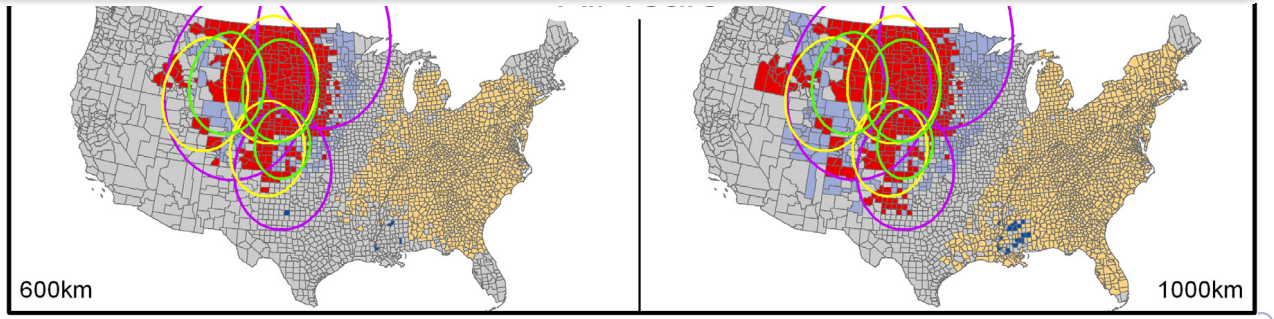


Results of the SaTScan

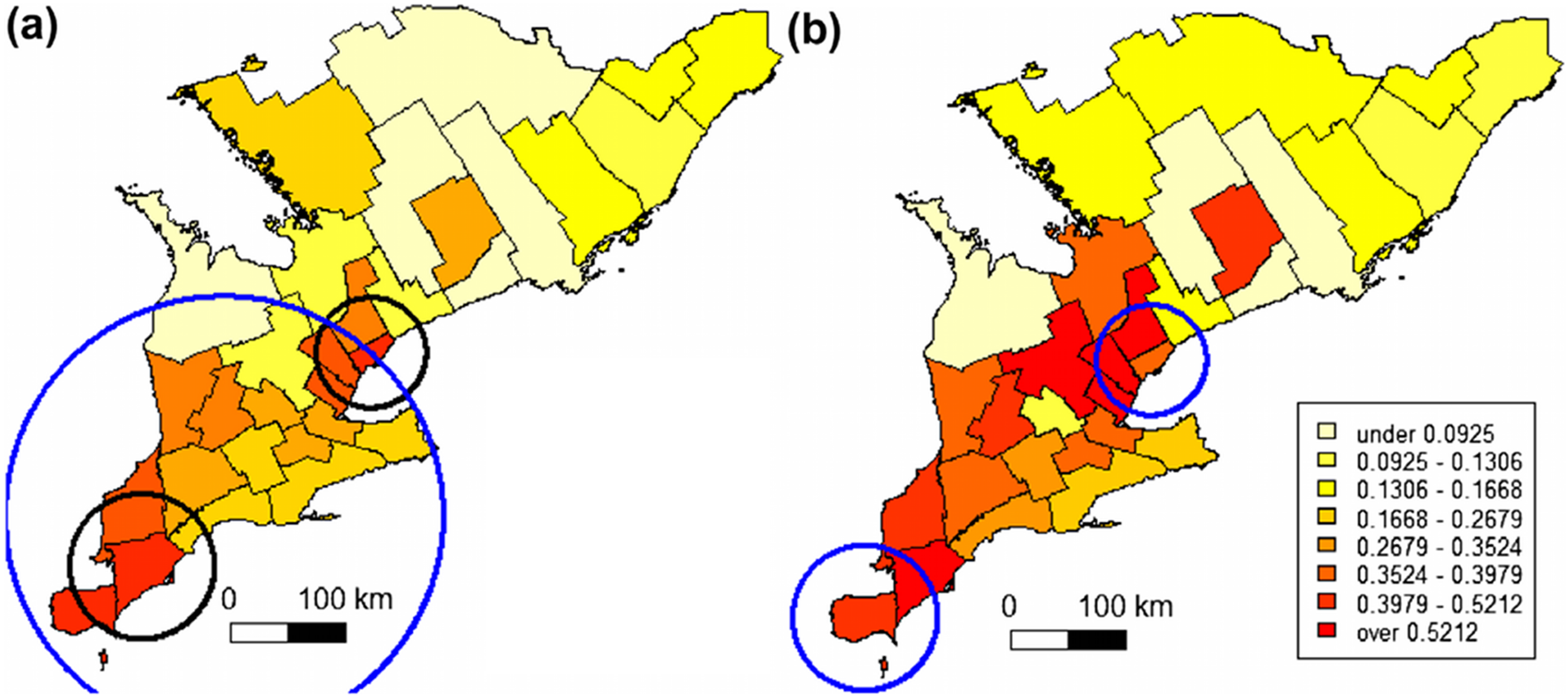
Reference

Ramanathan Sugumaran, Scott R Larson and John P DeGroote (2009) Spatio-temporal cluster analysis of county-based human West Nile virus incidence in the continental United States. International Journal of Health Geographics, 8:43

300 km, 600 km, and 1000 km distance thresholds.



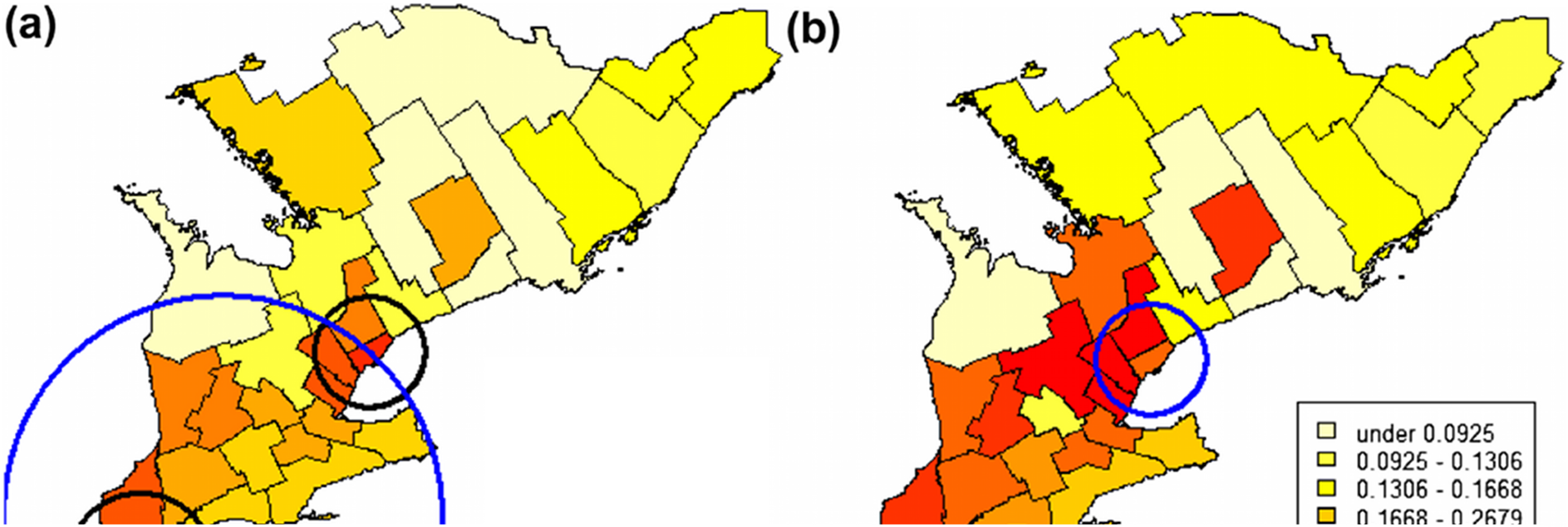
- West Nile virus infections among dead birds sampled from the 30 public health units of southern Ontario in 2005
- Overall a total of 272 out of 1017 dead birds were screened for WNV antibodies and found positive.
- Lindsay et al. (2003): sensitivity 83.9% and specificity 93.6%
- **Stone et al. (2005): sensitivity 82.1% and specificity 100%**
- Ontario Ministry of Health and Long-Term Care (MOHLTC, 2007): sensitivity 85% and specificity 95%
- Rogan and Gladen estimator



Choropleth maps of observed (a) and true (b) West Nile virus dead bird mortality in southern Ontario 2005 with disease clusters identified by the circular spatial scan statistic.

In (a) the large circle indicates the location of a single cluster in the observed mortalities and is based on a maximum cluster population size of 50%. The two smaller circles correspond to clusters identified with a maximum cluster population 30% or 40%.

In (b) the two circles indicate cluster location in the true mortalities as with a maximum cluster population size of 50%, 40% or 30%.



Reference

Olaf Berke, Lance Waller (2010) On the effect of diagnostic misclassification bias on the observed spatial pattern in regional count data – A case study using West Nile virus mortality data from Ontario, 2005. Spatial and Spatio-temporal Epidemiology 1, 117–122

In (a) the large circle indicates the location of a single cluster in the observed mortalities and is based on a maximum cluster population size of 50%. The two smaller circles correspond to clusters identified with a maximum cluster population 30% or 40%.

In (b) the two circles indicate cluster location in the true mortalities as with a maximum cluster population size of 50%, 40% or 30%.