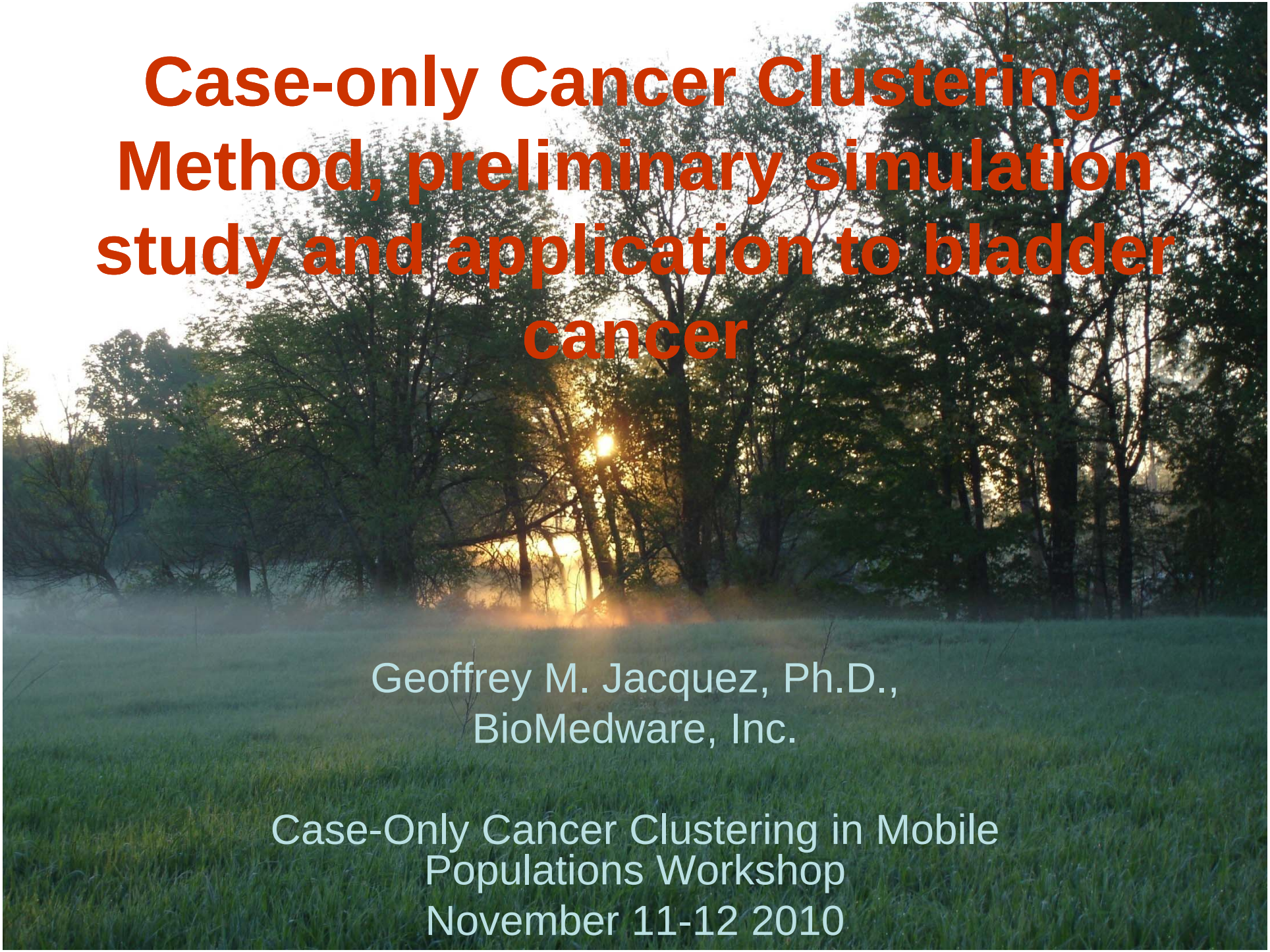


Case-only Cancer Clustering: Method, preliminary simulation study and application to bladder cancer

Geoffrey M. Jacquez, Ph.D.,
BioMedware, Inc.

Case-Only Cancer Clustering in Mobile Populations Workshop
November 11-12 2010

Outline

- Need
- Janus and Vesta statistics for case-only data
- Discussion

This research was funded by the NCI and the NIEHS. The perspectives in this presentation are those of the author and do not necessarily represent the official views of the funding agencies.

Outline

- **Need**
- Janus and Vesta statistics for case-only data
- Discussion

“The effect of [human] mobility could be a time-space lag between causes and effects that makes conventional mapping spurious”

-A. Shaerstrom (2003)

Schaerstrom, A. (2003) ‘*The potential for time geography in medical geography*’, *in* L. Toubiana, C. Viboud, A. Flahault and A.-J. Valleron (eds.), Geography and Health. Paris: Inserm. pp. 195-207.

Need for improved clustering approaches

- Clustering of cases at time of diagnosis or death is often of little scientific or practical interest in terms of enhancing our understanding of health-environment relationships

- Existing methods often
  - Treat individuals as immobile
  - Assume latency between causative exposures and health events (e.g. diagnosis, death) is negligible

- The duration of time from initial etiologic action of a causative exposure to disease detection has been called the empirical induction period (Rothman, 1981, AJE)

- Failure to account for mobility during this period can make it impossible to detect clustering of cases in relation to the spatial distribution of their causative exposures

The Need cont'd

- *A priori* hypotheses concerning the timing of clustering often do not exist

- Health researchers may then wish to investigate whether clustering at any point in time is associated with development of disease

- Inappropriate assumptions about the length of the empirical induction period can result in nondifferential misclassification and bias the results toward the null (Rothman, 1981, AJE)

Temporal orientations

- Without clear intuition regarding the appropriate temporal model, researchers may consider alternative orientations:
  - calendar year
  - years prior to diagnosis/recruitment
  - participants' age
  - ...
- Although this is well known, most analyses use only one temporal orientation
  - Implicit (e.g., Cumulative exposure/clustering of all residences on one map)
  - Multiple interval (e.g. location of residence in calendar years)
  - Life stage (e.g., Exposure/clustering at age of menarche)

Outline

- Need
- **Janus and Vesta statistics for case-only data**
- Discussion

Definition

- Interaction: Nearby cases occur at about the same time

- Causes
  - Contagion/infection
  - Localized exposure in space and time

Background

- Mantel, Knox, Barton & David, Jacquez, Aldstadt, Kulldorff etc.

- No clustering methods for case data that simultaneously account for
  - residential mobility
  - Known risk factors and covariates
  - Empirical Induction Period (EIP)

- $\{x_i, y_i, t_i\}$ is not sufficient for many diseases!

Objectives

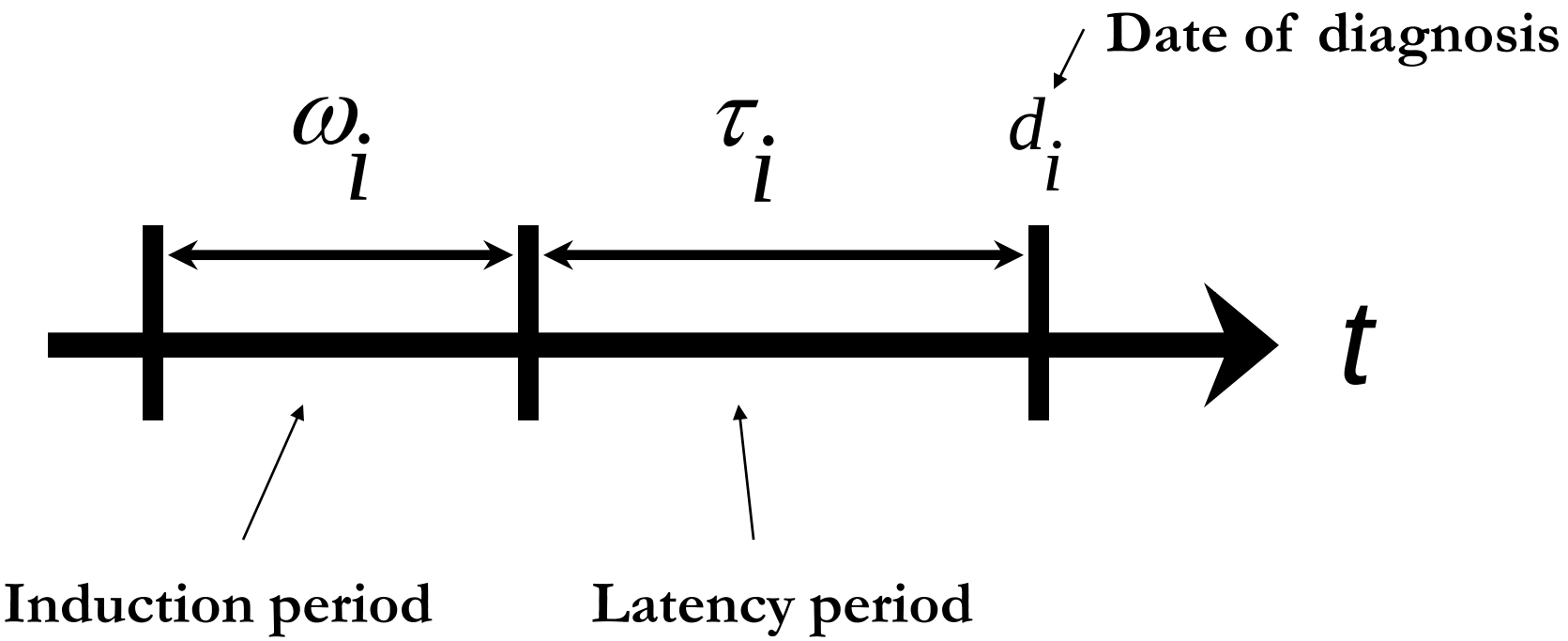
- Develop local and global tests for clustering in **case-only** data that account for
  - Residential mobility
  - EIP
  - Known risk factors and covariates
- Evaluate sensitivity to EIP
- Or use reasonable estimates of EIP

Janus and Vesta Statistics

Jacquez GM, J Meliker and A Kaufmann. 2007. In search of induction and latency periods: space-time interaction accounting for residential mobility, risk factors and covariates. *International Journal of Health Geographics*, **6**:35 http://www.ij-healthgeographics.com/content/6/1/35

Jacquez, GM and J. Meliker. 2009. *Geographic Clustering for Mobile Populations*. Chapter 19 *In* "A Handbook of Spatial Analysis", S. Fotheringham and P. Rogerson (Eds.). Sage Publications.

Model of Empirical Induction Period (EIP)

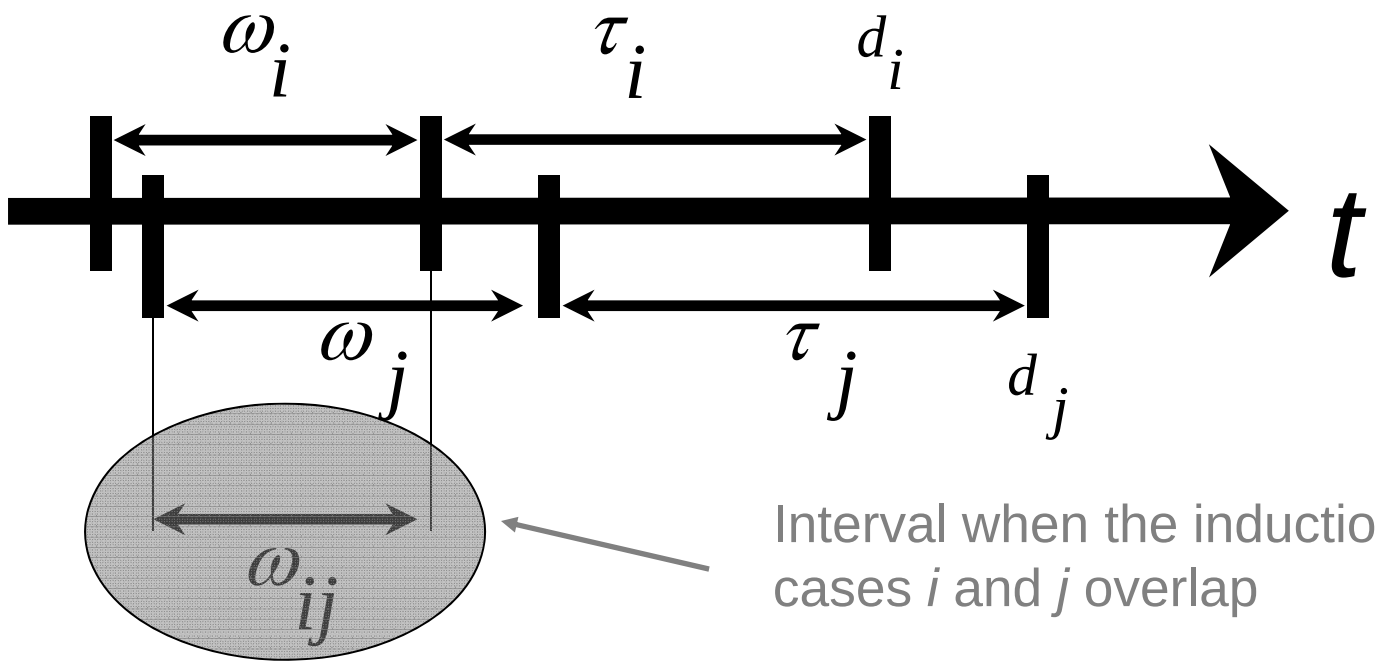


$$EIP = \omega + \tau$$

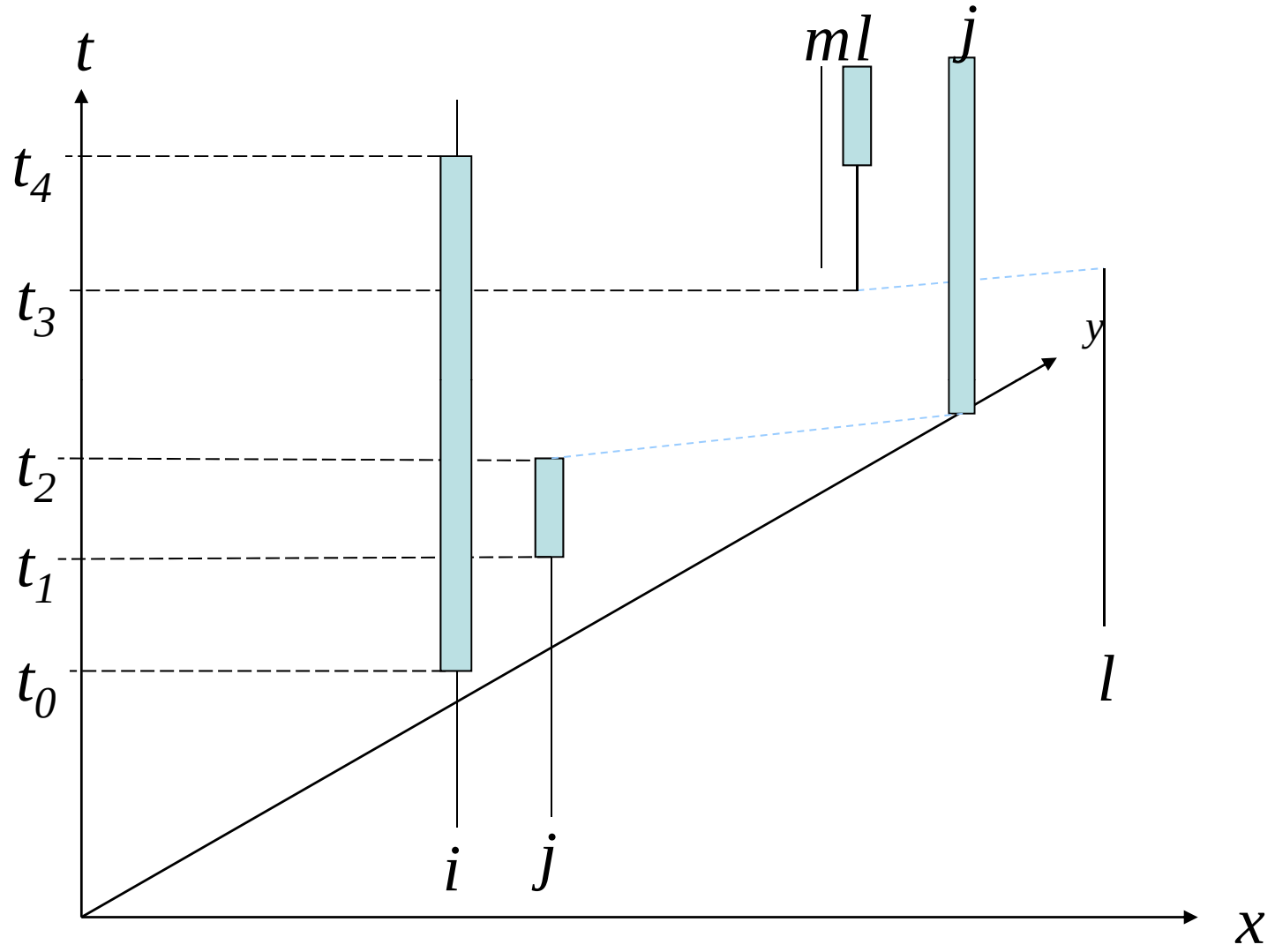
Rothman KJ. Induction and latent periods. Am J Epidemiol 1981;114:253-9

Intersection of Induction Periods

Let $\omega_{ij} = \omega_i \cap \omega_j$



Model of Exposure Trace



The space-time path of places where a person lived during their induction period is the *exposure trace*

Janus Statistic for Local Spatial Clustering of Exposure Traces at Time *t*

$$c_{it} = \begin{cases} 1 \text{ IFF case } i \text{ is in its exposure trace at time } t \ (t \in \omega_i) \\ 0 \text{ otherwise} \end{cases}$$

$$S_{ik\omega t} = c_{it} \sum_{j=1}^{k} c_{jt}$$

This is the count, at time *t*, of the number of *k* NN's of case *i* that were in their induction period at time *t*

The summation is over case *i*'s *k* nearest neighbors

Local Vesta Statistic for Interaction in Exposure Traces

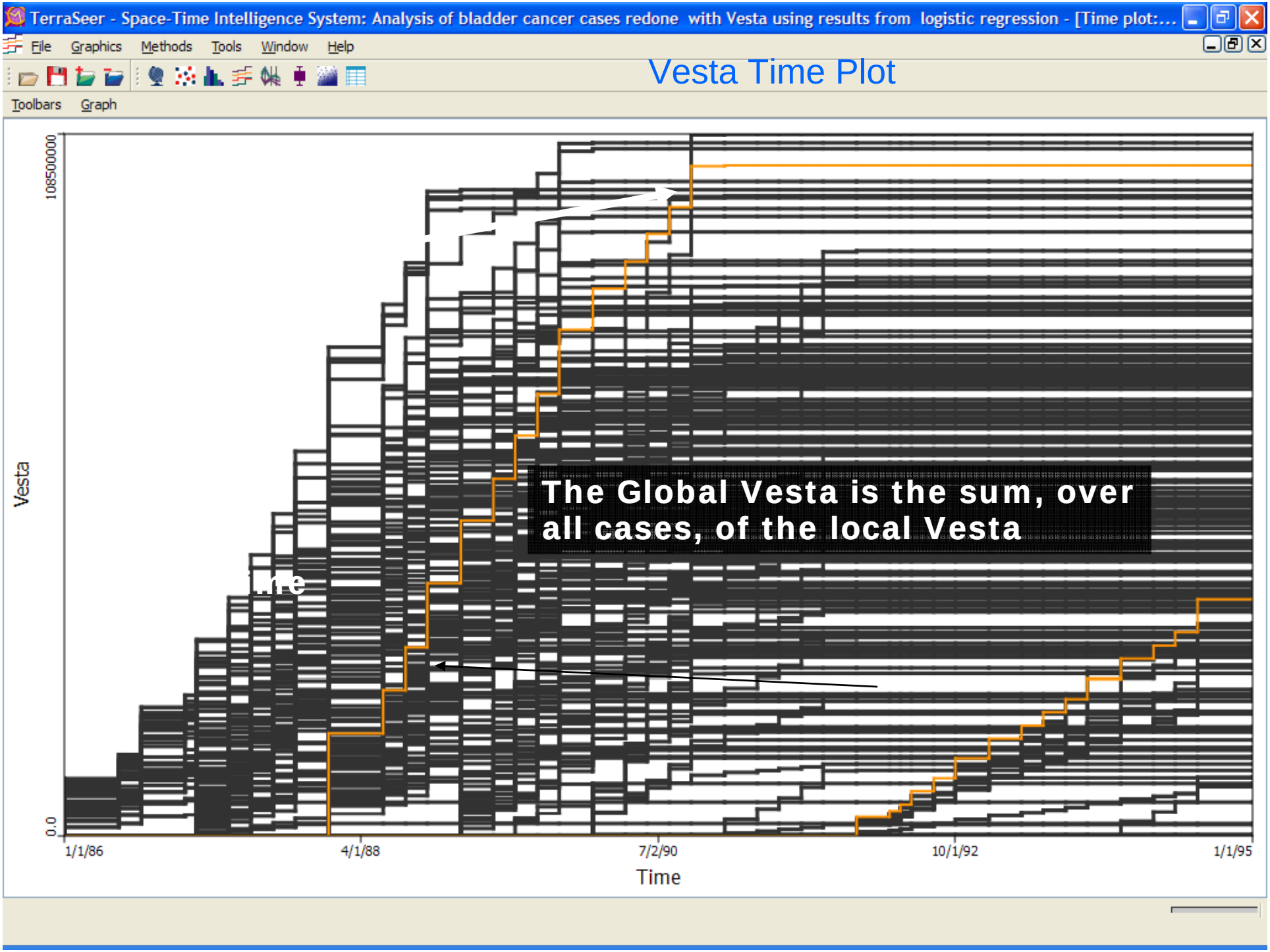
Measures proximity of ET's in space-time

$$V_{ik\omega} = \sum_{\substack{j=1 \\ i \neq j}}^{N} \eta_{ijk\omega}$$

This is the count of the *k*-nearest neighbors of case *i* whose induction periods overlapped with case *i*'s induction period

It quantifies interaction about the exposure trace of case *i*

The summation is over *N*, the number of cases


TerraSeer - Space-Time Intelligence System: Analysis of bladder cancer cases redone with Vesta using results from logistic regression - [Time plot:...
File
Graphics
Methods
Tools
Window
Help
Toolbars
Graph
Vesta Time Plot
108500000
Vesta
0.0
The Global Vesta is the sum, over all cases, of the local Vesta
1/1/86
4/1/88
7/2/90
10/1/92
1/1/95
Time

Sensitivity to EIP

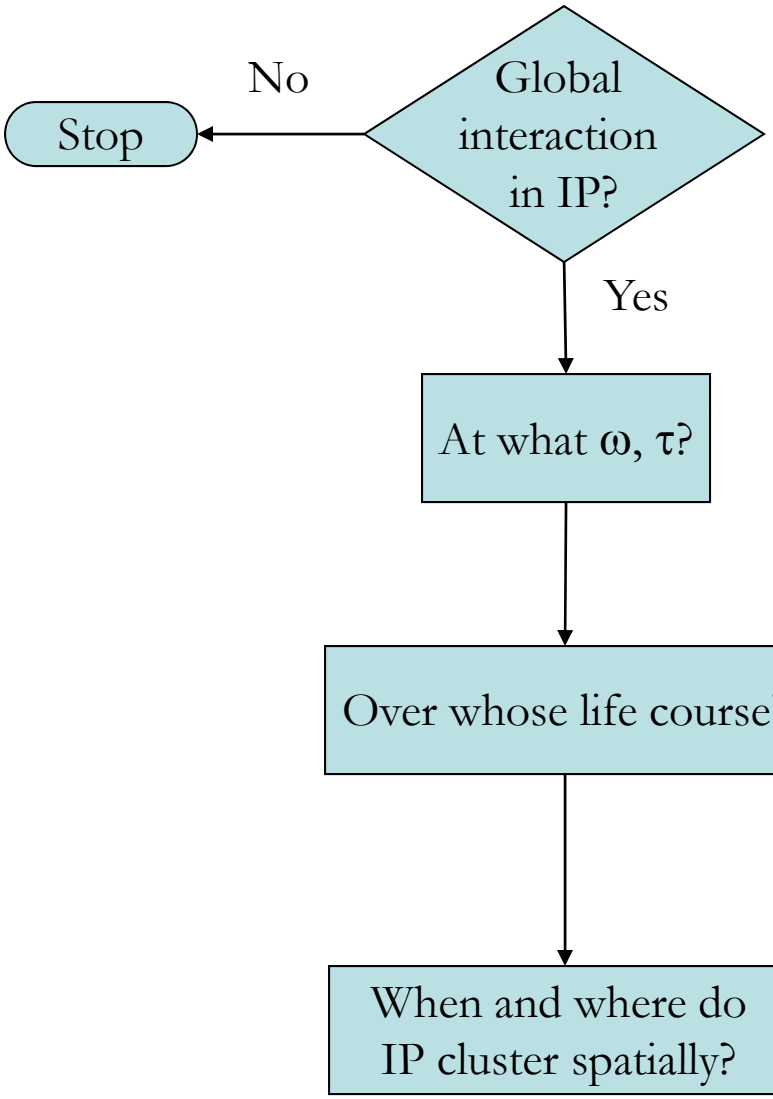
1) Specify EIP parameter space

$$\mathbf{S} = \{\omega_1,\tau_1;\ \omega_2,\tau_2;\ \ \ldots\ \ ;\ \omega_m,\tau_m\}$$

2) Evaluate the global Vesta statistic and its probability over space **S**

3) Identify those values of EIP that yield significant global Vesta statistics

Diagnostic Process for Clustering Induction Periods



Assess sensitivity of global Vesta to specification of latency and induction periods (IP)

Identify significant induction and latency period(s)

Identify cases and their exposure traces with significant local Vesta

Identify time intervals and places of residence of cases with significant local Janus

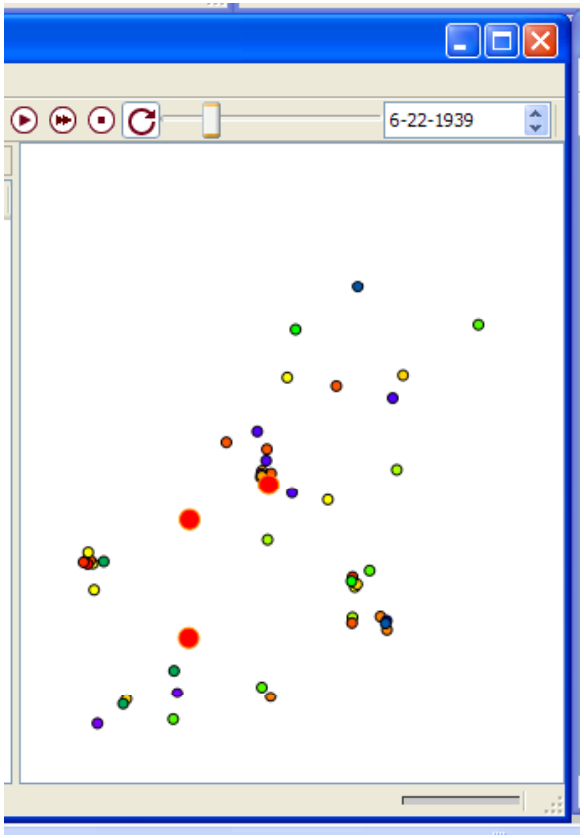
Simulation Study

- Use residential histories and dates of diagnosis for cases from bladder cancer study

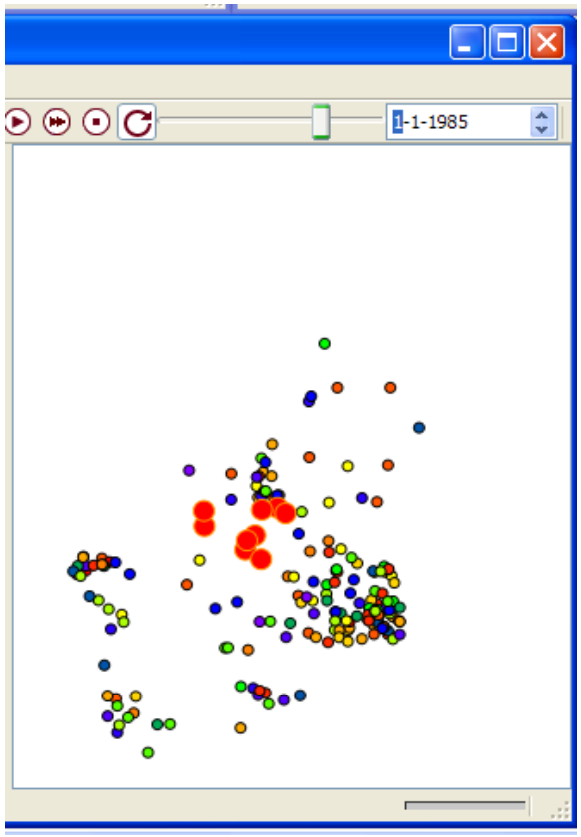
- Simulate
  - No clustering in exposure traces
  - Cluster of size 10 with ω=1, τ=15 years
  - Cluster of size 25 with ω=1, τ=15 years

- Can Janus and Vesta
  - Not declare clustering when there isn't any?
  - Find the clusters when they really are there?

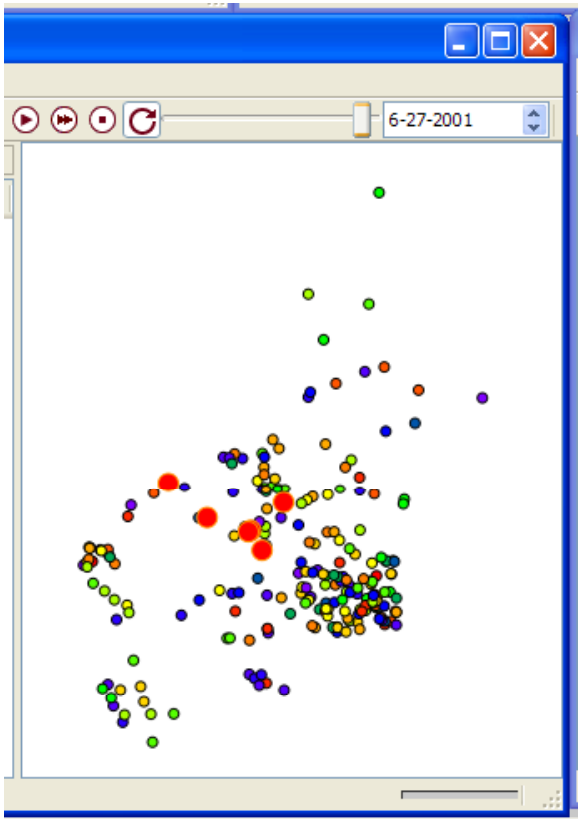
Cluster Evolution Through Time
Cluster of size 10 with ω=1, τ=15 years



1939



1985

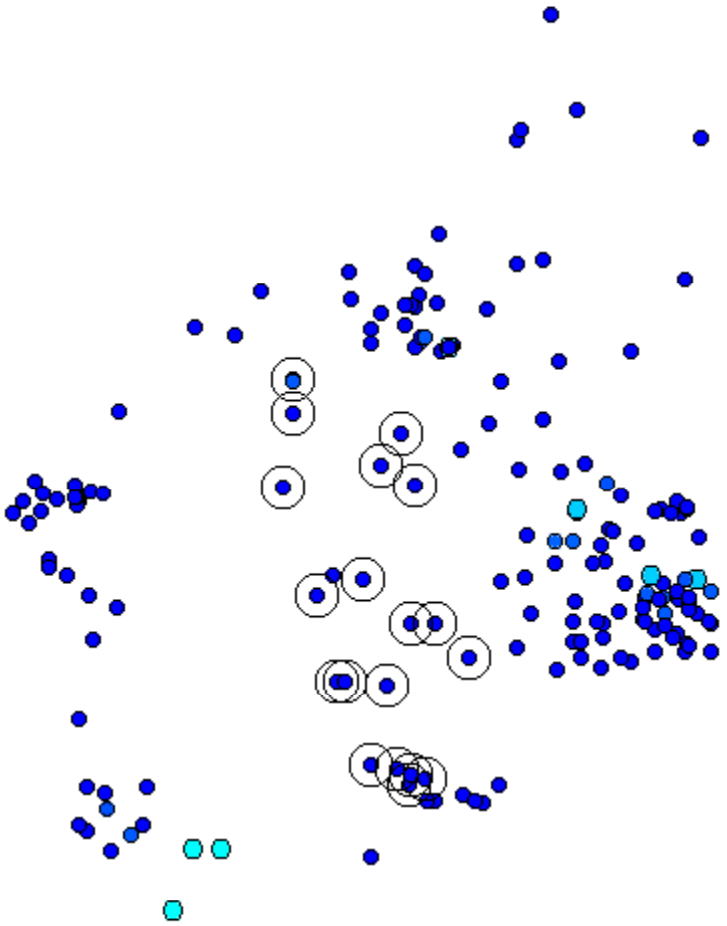


2001

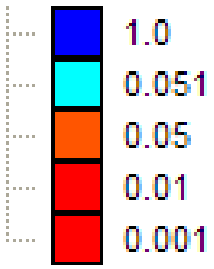
Simulation Study Results

Simulation	Global Vesta Probability	Parameters	Conclusion
No Interaction	P=0.107	NA	No interaction
Cluster size 10	P=0.011	ω=1, τ=15 years	Correct inference
Cluster size 25	P=0.011	ω=1, τ=15 years	Correct inference

Example: Janus finds the cluster of size 25



1 Jan 1980



Comments: Simulation Study

- New approach may be capable of quantifying EIP in real populations.
- Needs more realistic simulations, application to real populations and cancers.
- Cautious optimism.

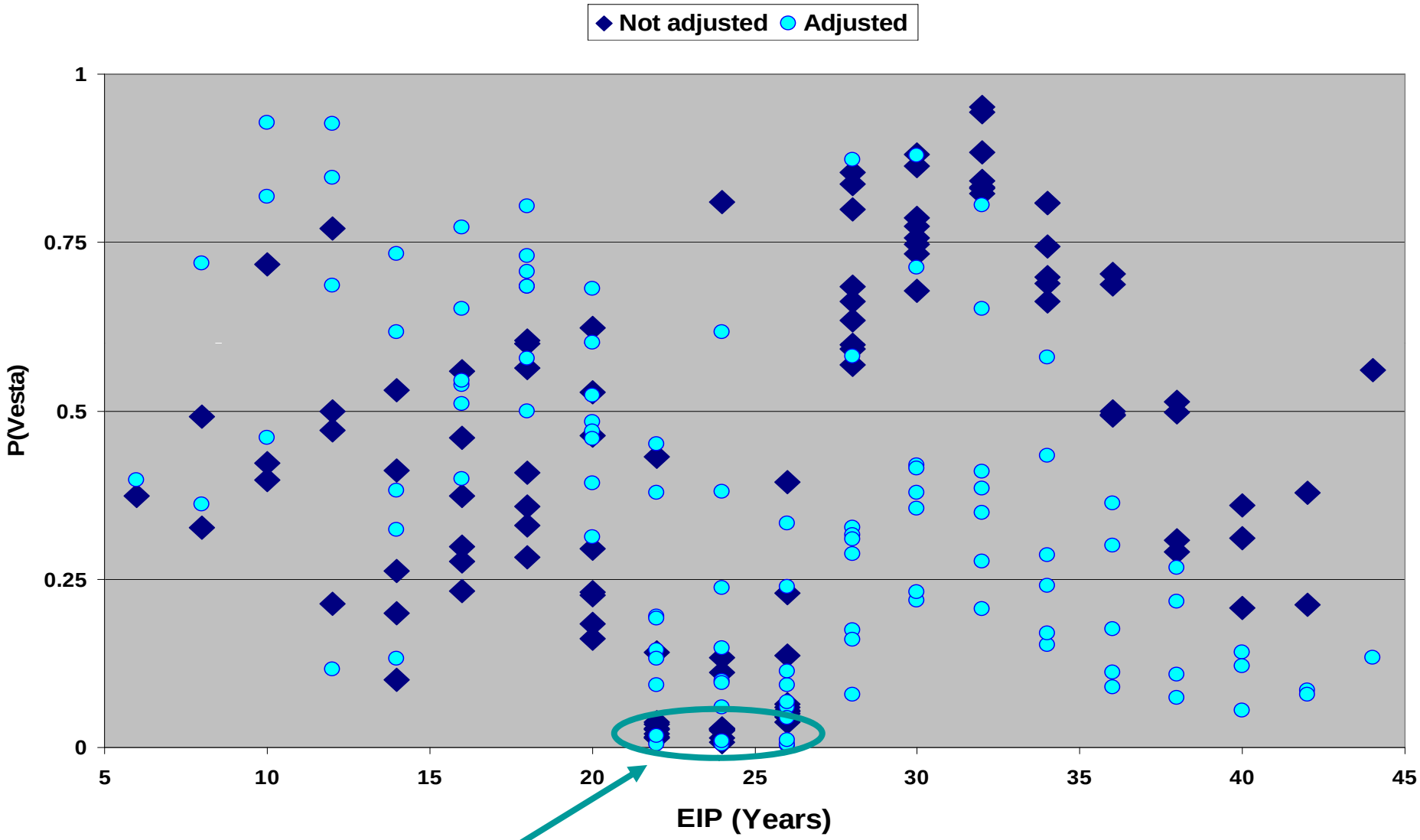
Bladder Cancer Case-Control Study

- Population-based
- Enrollment began in Fall, 2003
- Requirements of participation
  - Reside in 1 of 11 counties (Genesee, Huron, Ingham, Jackson, Lapeer, Livingston, Oakland, Sanilac, Shiawassee, Tuscola, Washtenaw) for previous 5 years
  - No previous cancer (exception of non-melanoma skin cancer)
- Bladder cancer cases (392)
  - Michigan State Cancer Registry
  - Aged 21-80, when diagnosed
  - Diagnosed 2000-2004
- Controls (492)
  - Selected by population-based random digit dialing (RDD) and RDD of age-weighted lists
  - Frequency matched to cases: race, sex, age (±5 yr) (not satisfactorily matched yet)

Bladder Cancer Study Design

- Use case data from bladder cancer study, with actual dates of diagnosis
- Evaluate 110 combinations of induction and latency
  - Induction period 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 years
  - Latency 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25 years
- Repeat after adjustment for smoking, age, gender, race, education
- Did **Not** use full data set – enrollment was in process

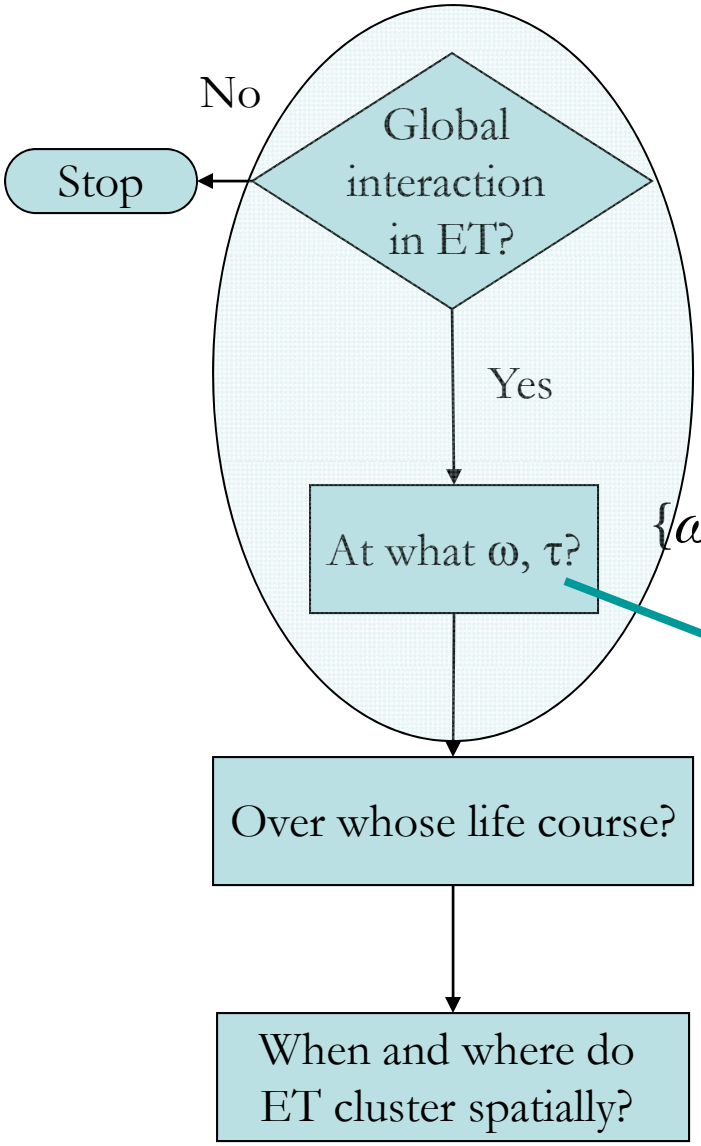
P(Global Vesta) vs EIP, *k*=1



After adjustment Minimum p=0.003

ω=5, τ=19.5 years

Diagnostic Process for Clustering Exposure Traces



$\Delta V_{k\omega}$

Assess sensitivity of global Vesta to specification of EIP

Suggests exposure on local spatial scale, occurring over 5 years, resulting in cancers diagnosed 19 years later

$\{\omega, \tau\} \Rightarrow \min[\, p(\Delta V_{k\omega})]$

Identify significant induction and latency period(s)

k=1, ω=5, τ=19

$\Delta V_{ik\omega}$

Identify cases and residential locations with significant local Vesta

$S_{ik\omega t}$

Identify time intervals and places of residence of cases with significant local Janus

Local Janus Movie, Adjusted, k=1, ω=5, τ=19

Who

Where

Comments: Janus & Vesta

- Technique evaluates clustering in case data accounting for residential mobility, EIP, risk factors and covariates

- This methodology allows researchers to
  - (1) Identify those EIP that maximize interaction;
  - (2) Evaluate clustering over an individual's life course and
  - (3) Localize clustering in exposure traces to specific cases, places & times

- Simulation studies suggest the methods do not declare clustering to exist when it actually does not, even when assessing a range of EIP

- Studies are required to more fully understand the statistical properties of this method, and possible impacts of multiple testing

Comments (continued)

- When apriori knowledge is available regarding EIP one does *not* need the sensitivity analysis to EIP

- May be able to use stage-based models of cancer with new data on tumor progression in planned study of pancreatic cancers in Michigan

NATIONAL CANCER INSTITUTE

NCI Cancer Bulletin
A Trusted Source for Cancer Research News

November 2, 2010 • Volume 7 / Number 21

About the Bulletin | Archive/Search

Pancreatic Cancers May Develop Slowly Over Many Years

Pancreatic cancers may take more than a decade to progress to a lethal stage, according to an analysis of genetic changes in tumors from seven patients. The findings, which appeared October 28 in *Nature*, surprised even the researchers themselves by indicating that there is a long lag time between the first cancer-causing genetic change in a pancreatic lesion and the development of life-threatening metastatic disease. This lag time represents a window of opportunity for detecting the cancer in its early stages, the researchers noted.

On average, the researchers estimated, 11.7 years had passed between the initiating mutation in the tumor cell and the development of the first cancer cell that gave rise to the "parental clone," which contains all of the mutations known to drive pancreatic cancer development. Another 6.8 years passed before at least one subclone had gained the genetic potential to spread. From that point, another 2.7 years, on average, passed before the patients' deaths.

EIP = 21.2 years

"We were all surprised by how slow the natural history of pancreatic cancer seems to be," said lead researcher Dr. Christine Iacobuzio-Donahue. "This disease seems to be so lethal, and the feeling amo many people has been that you can't do anything about pancreatic cancer. But we now know that it takes years for metastases to develop. So we finally have an idea of what we need to do in terms of early detection and when we need to do it."

Future of individual-level cluster studies

- Significant innovation: Allocation of unexplained risk to specific locations, times and small groups of individuals

- New era: From pre-epidemiology to post-epidemiology

Acknowledgments

BioMedware: Jaymie Meliker, Pierre Goovaerts, Gillian AvRuskin, Andrew Kaufmann, Robert Rommel, Yanna Pallicaris

U of Michigan, EHS: Jerome Nriagu, Al Franzblau, Melissa Slotnick, Stacey Fedewa, Lisa Bailey, Nicholas Mank, Danielle Movsky, Aaron Linder, Zorimar Rivera, Luis Rivera

U of Michigan, Epidemiology: David Schottenfeld, Mark Wilson

Michigan Public Health Institute (conducting telephone interviews)

Michigan State Cancer Registry (Glenn Copeland)