Statistical methods for real time detection of the start and subsequent epidemic states of HIV outbreaks among People Who Inject Drugs: a comparative simulation study

V. Baralou[1], N. Demiris[2], C. Thomadakis[1], I. Gountas[3] and G. Touloumi[1]

[1] Department of Hygiene, Epidemiology & Medical Statistics, Medical School, National and Kapodistrian University of Athens, Athens, Greece
[2] Department of Statistics, Athens University of Economics and Business, Athens, Greece
[3] Medical School, University of Cyprus, Nicosia, Cyprus

Background

- HIV outbreaks among people who inject drugs (PWID) are not uncommon; they are related to increased morbidity, mortality and economic burden.
- A reliable automated surveillance system for early detection and control of possible epidemics is of high public health importance.
- Statistical methods have been proposed for the prospective outbreak detection for diseases with seasonality and for epidemics of limited duration.
- HIV outbreaks among PWID follow different patterns with respect to their magnitude, duration and post-epidemic level.
- In addition to the outbreak onset, it is also useful to identify the subsequent epidemic states, but no study has addressed this problem prospectively.

Objective: To compare through simulations the performance of different methods for detecting in real time the start, the non-increasing state (plateau and/or decline) and the end of an outbreak.

Methods

Motivating example

- ❑ A previous HIV outbreak among PWID in Greece (Figure 1).
- ❑ An epidemic curve usually comprises the following five states:
  1. pre-epidemic, 2. increase, 3. plateau, 4. decline, 5. post-epidemic.
- ❑ Data on new weekly HIV diagnoses were provided by The European Surveillance System (TESSy) of the European Centre for Disease Prevention and Control (ECDC).

Simulation study

- ❑ Simulations were roughly based on the TESSy data.
- ❑ We first simulated the epidemic state $Z_t \in \{1,2,3,4,5\}$ for week t.
  - ➢ Weeks $t = 1, \dots, 263$ did not contain outbreak cases and were used as baseline by the methods for outbreak detection.
  - ➢ The starting week s of the outbreak was randomly selected among weeks $t = 264, \dots, 315$ that spanned 1 year ($Z_t = 1, t < s$).
  - ➢ For the remaining epidemic states ($Z_t, t \geq s$), their durations were identical with those of the motivating example for all settings.
- ❑ The weekly counts Y_t for week t were then simulated from a Negative Binomial distribution with overdispersion $r_{t|Z_t}$ and mean $\mu_{t|Z_t}$ that depended on the state of the current week t as follows:

$$\mu_{t|Z_t=1,3,5} = \mu_{t-1}$$
$$\mu_{t|Z_t=2,4} = f(\mu_{t-1}),$$

where f is the growth or decline function as defined in each scenario.

- ❑ Six simulation scenarios were set up assuming:
  - ➢ Different number of epidemic states (with or without plateau)
  - ➢ Different growth and decline shapes (constant, exponential)
  - ➢ Different epidemic progression over time (abrupt, gradual).

Statistical methods to identify each state

- ❑ A two-state hidden Markov model (HMM)
- ❑ A new method based on the prediction interval (PI) (Figure 2)
- ❑ A novel approach that combines the former two by assigning weights to HMM's transition probabilities at each time point based on the PI limits (HMM-PI)
- ❑ Classic methods for outbreak detection
  - ➢ Regression-based methods: OutbreakP, Farrington algorithm
  - ➢ Control charts: cumulative sum (PCUSUM) and exponentially weighted moving average (PEWMA) chart for Poisson distribution

Performance evaluation

- ❑ Methods were applied prospectively.
- ❑ Performance was assessed with:
  - ➢ Timeliness: difference between the start of the epidemic and the first alarm signal after its onset
  - ➢ Sensitivity: true positive rate
  - ➢ Specificity: true negative rate

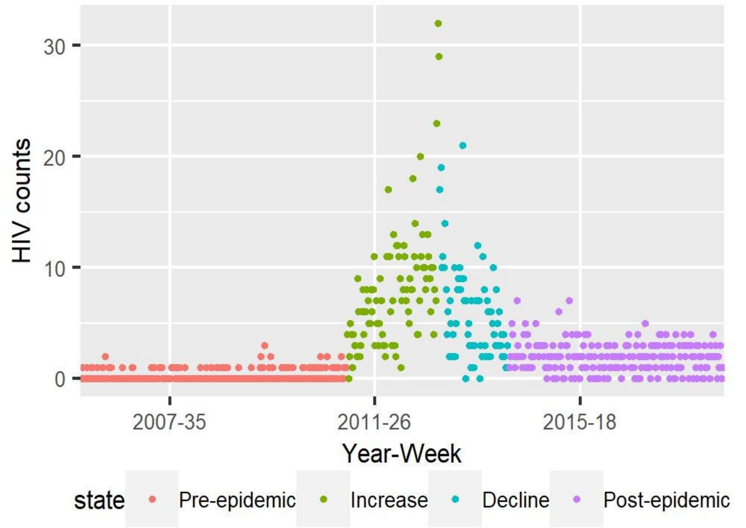

Figure 1. Weekly HIV cases among PWID in Greece segmented in four epidemic states

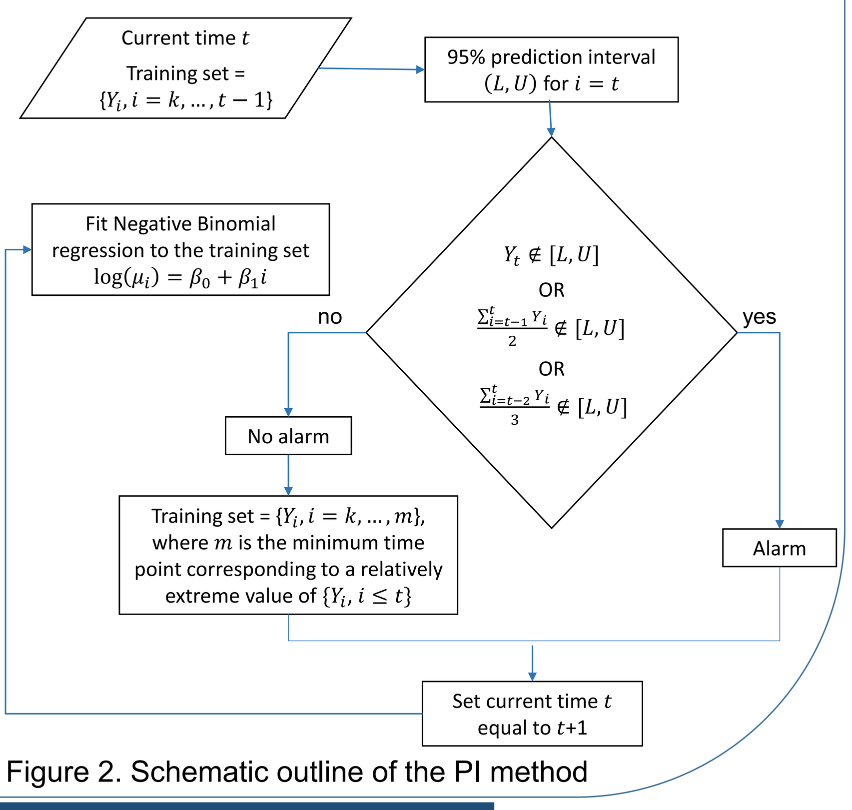

Figure 2. Schematic outline of the PI method

Results

Simulation results

- ❑ Start of the outbreak:
  - ➢ Specificity: excellent by all methods in all scenarios.
  - ➢ Sensitivity (Figure 3):
    - ➢ Regression-based methods (i.e., Farrington, OutbreakP, PI) and HMM had the worst performance.
    - ➢ HMM-PI improved the sensitivity of both HMM and PI.
  - ➢ Timeliness (Figure 4):
    - ➢ HMM had by far the worst performance among all methods.
    - ➢ OutbreakP and PCUSUM were the most timely methods.
    - ➢ HMM-PI performed better than HMM but similarly with PI.

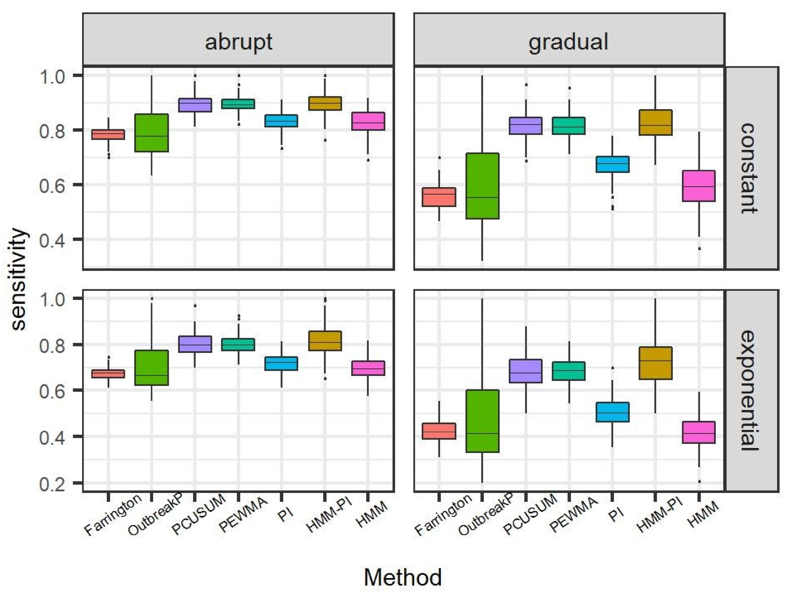

Figure 3. Sensitivity of methods for outbreak detection under different epidemic shapes and time progress

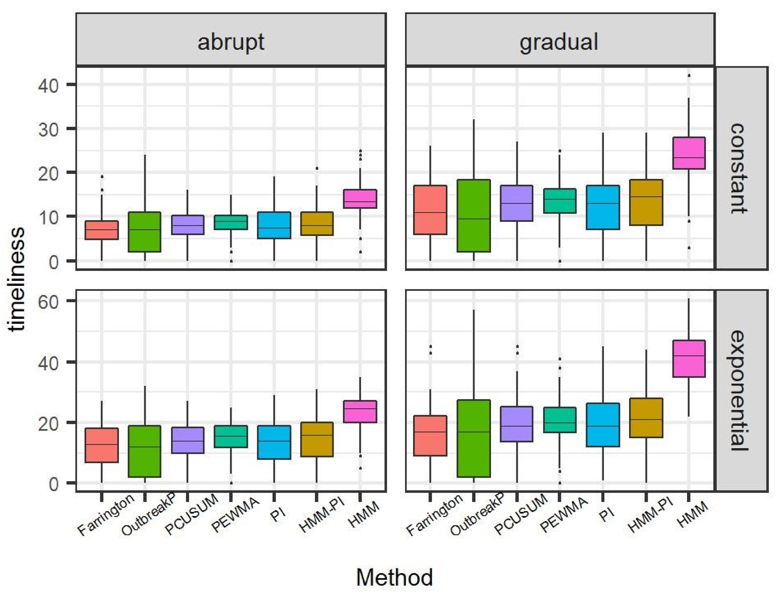

Figure 4. Timeliness of methods for outbreak detection under different epidemic shapes and time progress

- ❑ Non-increasing state (Figure 5, 6, 7):
  - ➢ PI gave the best balance between all metrics in all settings.
  - ➢ HMM gave the highest values of specificity but had by far the worst performance in terms of sensitivity and timeliness.
  - ➢ Although HMM-PI resulted in at least similar sensitivity and timeliness with PI, the latter performed similarly or better in terms of specificity.
- ❑ End of the outbreak:
  - ➢ No method reached a satisfying balance between all metrics.
  - ➢ PI performed better in most scenarios giving high specificity and moderate timeliness.
  - ➢ Although HMM-PI and HMM performed similarly in terms of sensitivity and timeliness, HMM-PI had by far better specificity in all scenarios but did not reach satisfying values (<50%) when assuming gradual epidemic progress.

Real data application

- ❑ Similarly with the scenario for abrupt and constant growth, all methods for outbreak detection gave excellent values for all metrics with HMM-PI ranking first (Table) and moderate specificity only for OutbreakP.

Method	Sensitivity	Specificity	Timeliness
Farrington	96.6%	100%	0
OutbreakP	98.9%	58%	0
PCUSUM	100%	92%	0
PEWMA	100%	90%	0
PI	97.7%	100%	0
HMM	89.7%	98%	0
HMM-PI	100%	98%	0

Table. Methods' performance when applied on real data to detect the start of the outbreak

- ❑ All methods for identifying the non-increasing state gave excellent values of specificity (i.e., 100%). HMM-PI outperformed both HMM and PI in terms of sensitivity and timeliness.
- ❑ The best-performing method for detecting the post-epidemic state was PI as it had excellent sensitivity (100%) and no delay in signaling an alarm. Similarly with the scenarios assuming a gradual and exponential decline, its specificity was lower (50%). Although HMM and HMM-PI gave good values of timeliness, they both failed in terms of specificity and sensitivity (<50%).

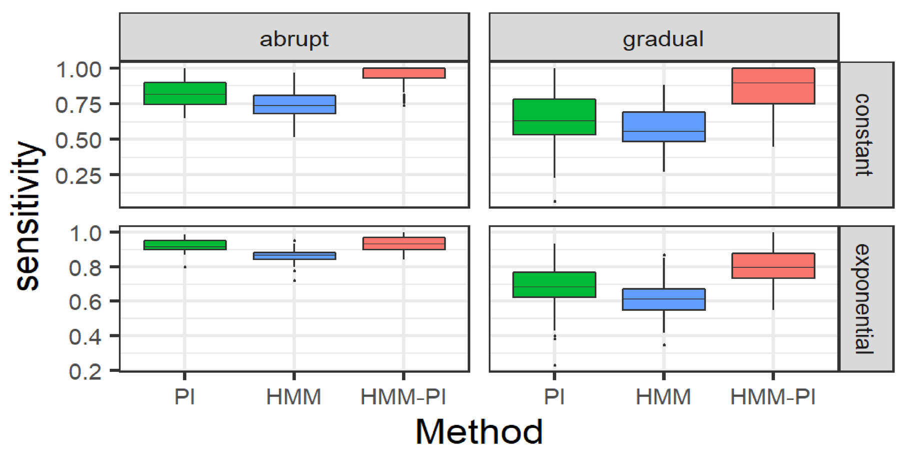

Figure 5. Sensitivity of methods for detecting the non-increasing state in scenarios without plateau

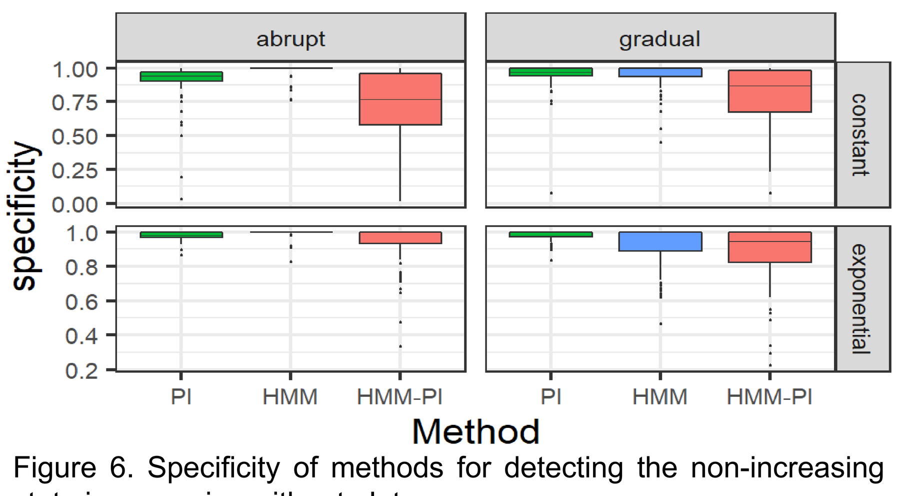

Figure 6. Specificity of methods for detecting the non-increasing state in scenarios without plateau

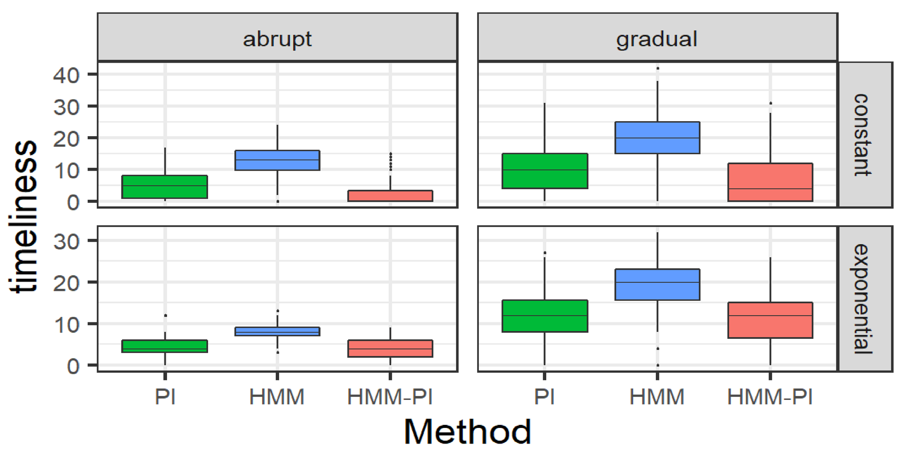

Figure 7. Timeliness of methods for detecting the non-increasing state in scenarios without plateau

Conclusions

- None of the investigated methods is panacea for detecting all epidemic states.
- Our HMM-PI proposal seems quite promising for detecting the outbreak onset since it achieved a performance similar or superior to well-established statistical methods.
- PI should be used for identifying the subsequent epidemic states.
- Although HMM performed poorly in real time, it can provide useful insights if applied retrospectively.

The present work was co-funded by the European Union and Greek national funds through the Operational Program "Competitiveness, Entrepreneurship and Innovation" under the call "RESEARCH-CREATE-INNOVATE" (project code: T1EΔK-03812).

European Union European Regional Development Fund
EPAnEK 2014-2020 OPERATIONAL PROGRAMME COMPETITIVENESS • ENTREPRENEURSHIP • INNOVATION
With the co-financing of Greece and European Union
ΕΣΠΑ 2014-2020 Partnership Agreement