**Appendix** **1.**

**Poisson-regression-based CEM clustering algorithm**

The CEM clustering algorithm employed to cluster registry-specific cervical cancer incidence data, which are count data. The method relies on an iterative process to obtain an optimal clustering based on likelihood under a Poisson regression model. The algorithm operates with a prefixed number of clusters *k*. The iterative process is initialized by an initial C-step and an initial M-step:

* Initial C-step: randomly assign the age-specific cervical cancer incidence of each registry *r* to one of the *k* clusters.
* Initial M-step: estimate the initial parameters under the Poisson regression model as defined in Box 1 based on the initial assignment and compute the proportion of registries belonging to each cluster *c*.

Upon initialization, the *m-*th iteration consists of an E-step, C-step, and M-step:

* E-step: compute the probability for the age-specific cervical cancer incidence of each registry to belong to each cluster *c* given the previously obtained parameter estimates :
* C-step: assign each the age-specific cervical cancer incidence of each registry to the cluster with the maximum probability.
* M-step: estimate the parameters under the Poisson regression model as defined under “Poisson regression model” based on the assignment obtained in the C-step and compute the proportion of registries belonging to each cluster *c*.

The iterative process terminates when the model fit no longer improves, which is defined as when the difference between two consecutive log-likelihoods is smaller than the given threshold value (0.001).

As different initial assignments could result in different final assignments, the above iterative process was repeated 100 times with different initial assignments, randomly generated from a multinomial distribution. The 100 different final assignments were compared based on the Bayesian information criterion (BIC), and only the one with the highest BIC was selected.

**Poisson regression model**

A Poisson regression model with cluster-specific effects for intercept, age group id (for 5-year age groups 15-19, …, 75-79) and the square of age group id is fitted to the cervical cancer incidence data (and for the number of cases and women-years in the data) estimated based on the classified cluster obtained:

For notational convenience, we introduced the following auxiliary coefficients for :

To facilitate interpretation of the cluster patterns, we then computed the maximum incidence and the 5-year age group of maximum incidences , as follows:

, and , .