University of Science and Technology Beijing

Shrink: A Breast Cancer Risk Assessment Model Based on Medical Social Network

Rui Wang

wangrui@ustb.edu.cn

The 2nd China-Europe Forum on Smart Healthy Aging and Privacy Protection (Share2018)
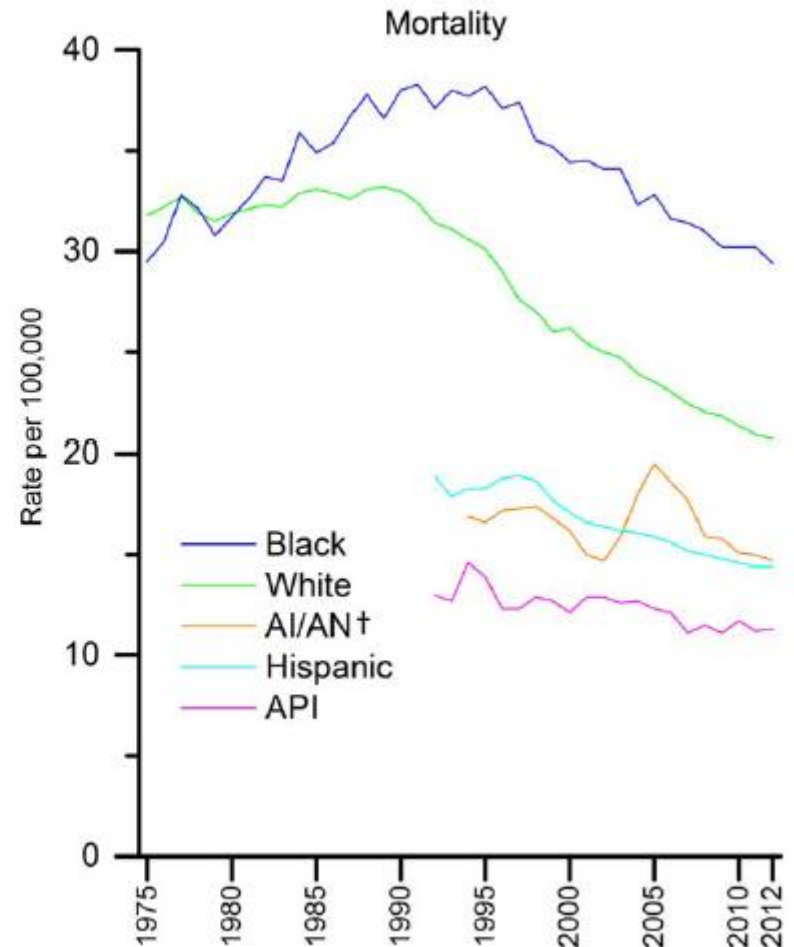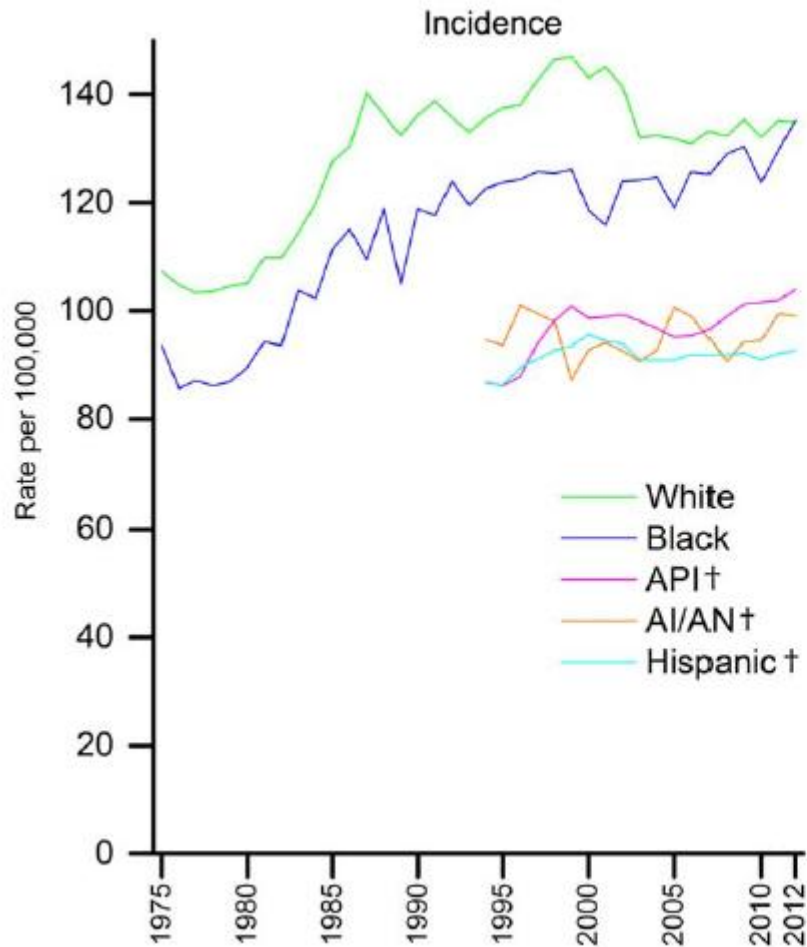
Outline

- ✓ Research background
- ✓ Shrink model
- ✓ Experiment

Outline

- ✓ Research background
- ✓ Shrink model
- ✓ Experiment

Research background

-The serious incidence of breast cancer

Breast cancer incidence and mortality

Research background

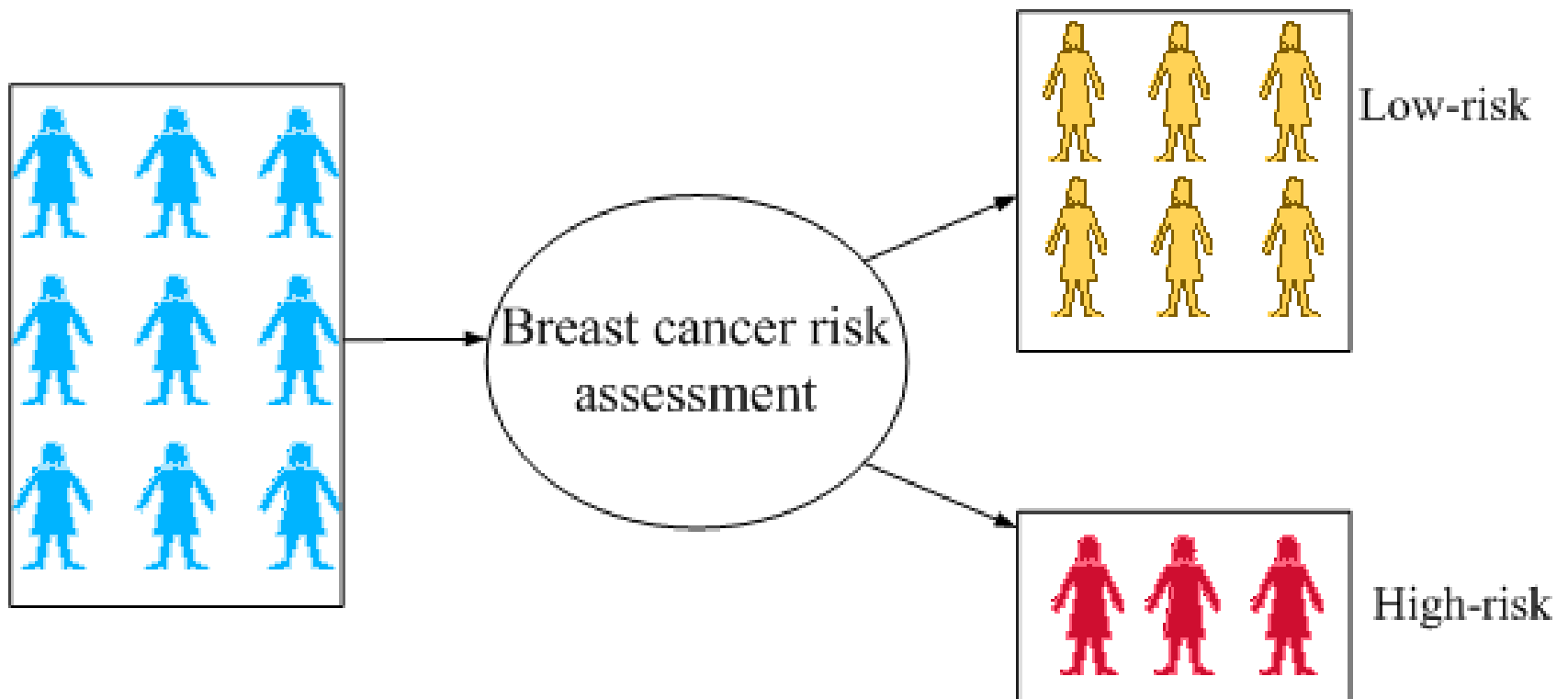-The difficulty in implementing early detection

Community hospital; backward

Populous country; Cost

Research background

-breast cancer risk assessment model

Research background

-Existing risk assessment models

- https://www.cancer.gov/bcrisktool/

Outline

- ✓ Research background
- ✓ Shrink model
- ✓ Experiment

Shrink

-Basic idea

similar habits and customs

probability of suffering the same kind of disease

High

Shrink

-The purpose

Medical social network

High-risk group

Low-risk group
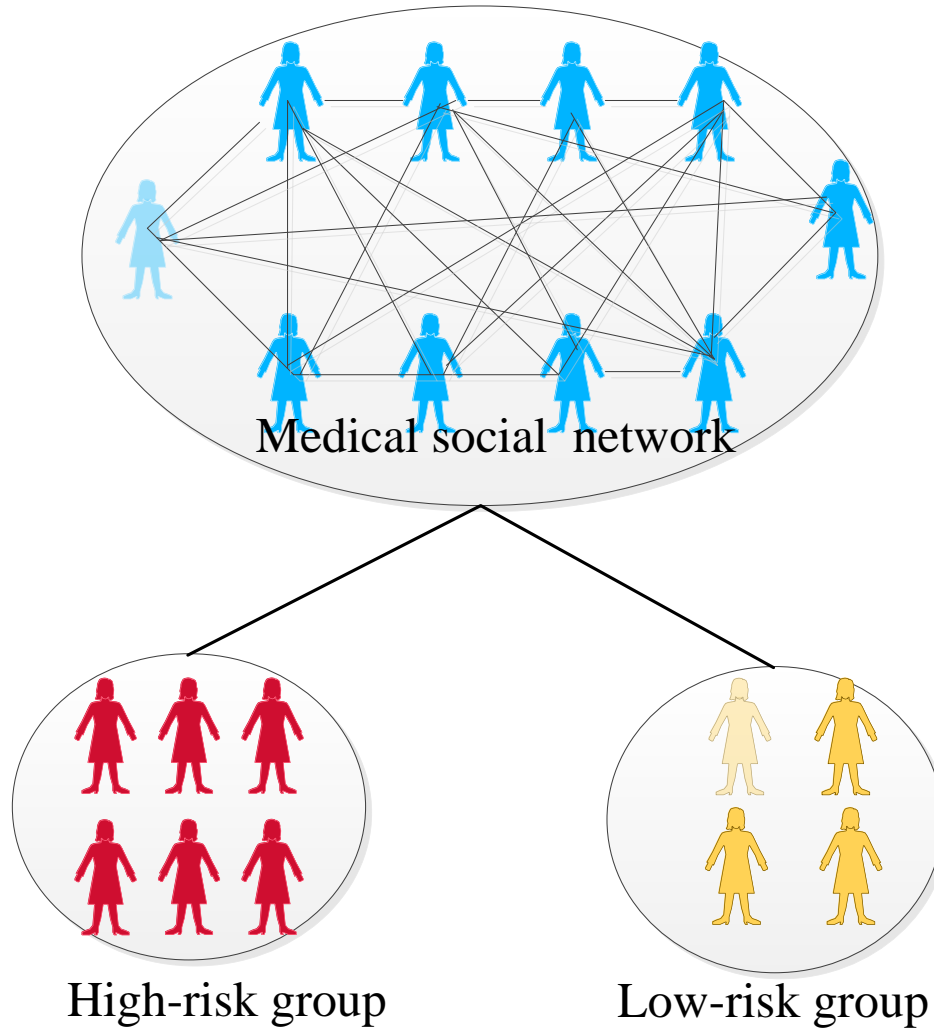
Shrink

-The module

Shrink

GA

GB
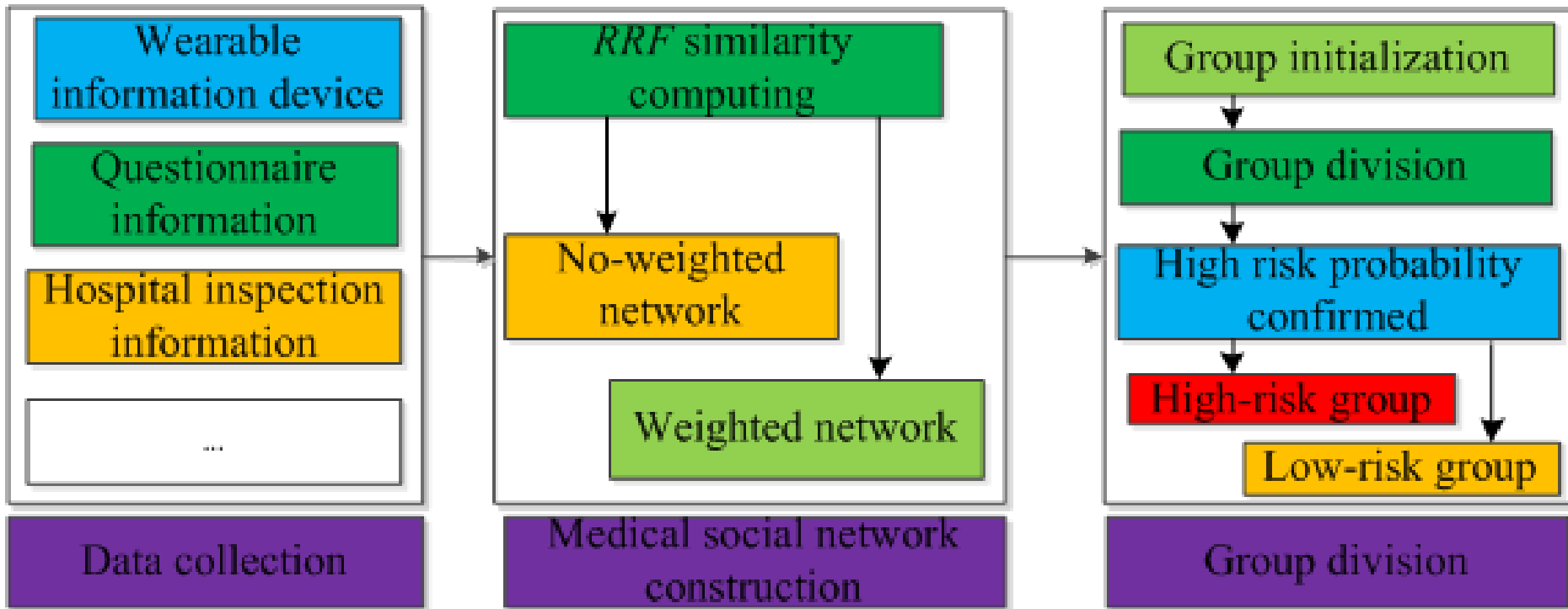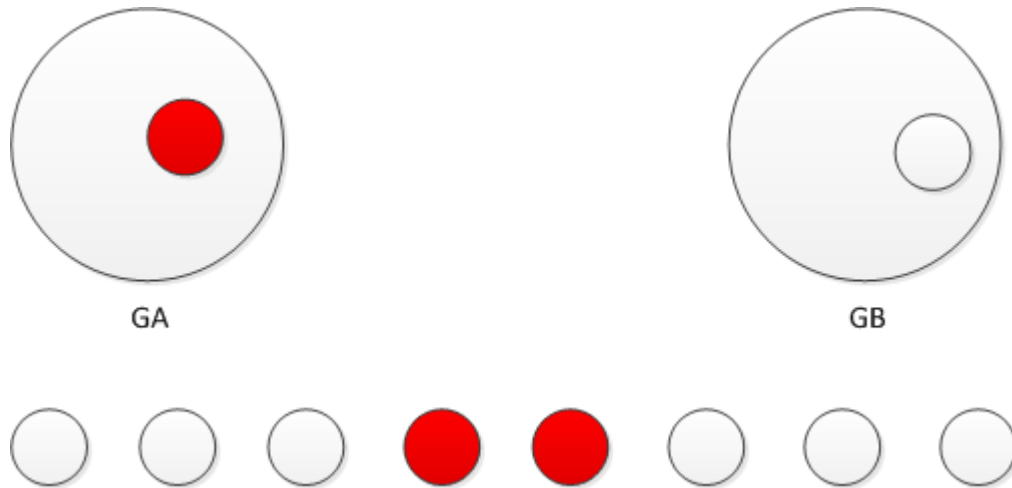
$GA= \{x \mid$ only one x is selected from P according to medical knowledge$\}$,

$GB= \{y \mid \quad s_{mf_i}(x, y)$ is minimum $, x \in GA, y \in N\}$ (2)

Group division

GA

GB
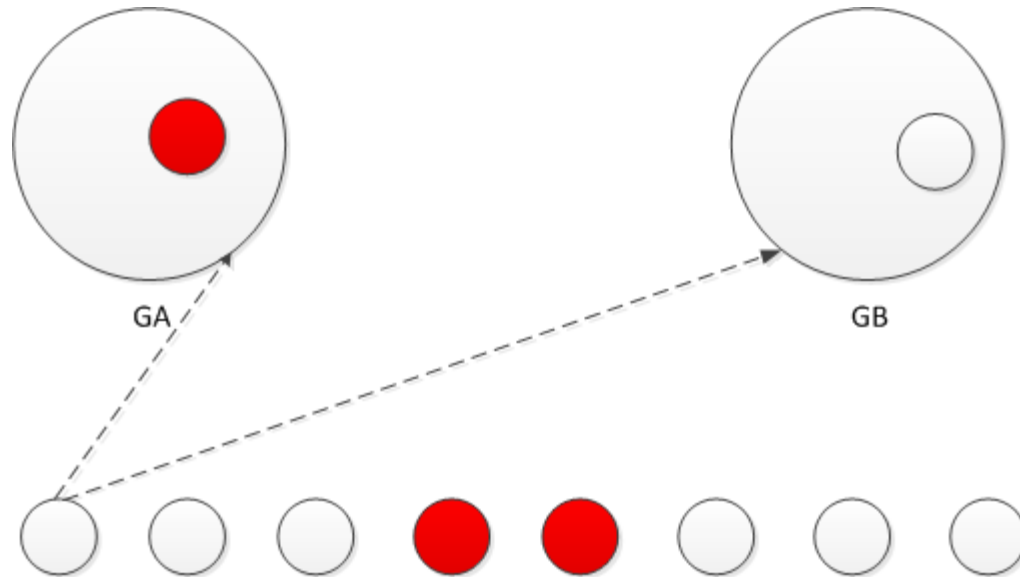
*The change of **modularity** is computed.*

Shrink

- Modularity (Q) is the fraction of the edges that fall within the given groups minus the expected such fraction if edges were distributed at random.

- Q=
$$\sum_{i=1}^{c}(e_{ii} - a_i^2)$$

- E_{ij} means that the number of edges that connect group i and j.
- a_i means that the number of edges that connect group i.

$$\Delta Q = \left[\frac{\Sigma\ in + 2K_{i,in}}{2m} - \left(\frac{\Sigma\ tot + K_i}{2m}\right)^2\right] - \left[\frac{\Sigma\ in}{2m} - \left(\frac{\Sigma\ tot}{2m}\right)^2 - \left(\frac{K_i}{2m}\right)^2\right]$$

Shrink

- If $\Delta Q_{GA} > \Delta Q_{GB}$, the node joins in *GA* group, else it joins in *GB* group.

- Lastly, each node will join in *GA* or *GB*.
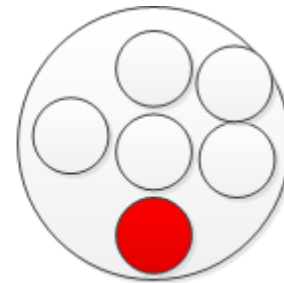
GA   GB

Outline

- ✓ Research background
- ✓ Shrink model
- ✓ Experiment

Experiment

—Data

- The survey data is collected through questionnaire. (1) demographic characteristics (2) female physiological and reproductive factors: (3) medical and family history: (4) dietary habits: frequency of the intake of various types of food; (5) lifestyle habits: (6) breast-cancer-related knowledge:

- More than 120,000 people accepts breast cancer epidemiological survey. In the end, 103,679 hale people and 301 breast cancer patients are confirmed as test cases.

Experiment
—Setting

- Test 1: Firstly, the test will prove the assessment value of Shrink with different *RRF* in order to discover suitable *RRF* to be used by Shrink.

- *RRF4*= {*family history of breast cancer, personal history of benign breast disease, current life satisfactory degree, miscarriages times*},

- *RRF5*=RRF4+ {*age at first birth*}

- *RRF6*=RRF5+ {*BMI*}

- *RRF7*=RRF6+ {*age at menarche*}

- *RRF8*=RRF7+ {*age*}

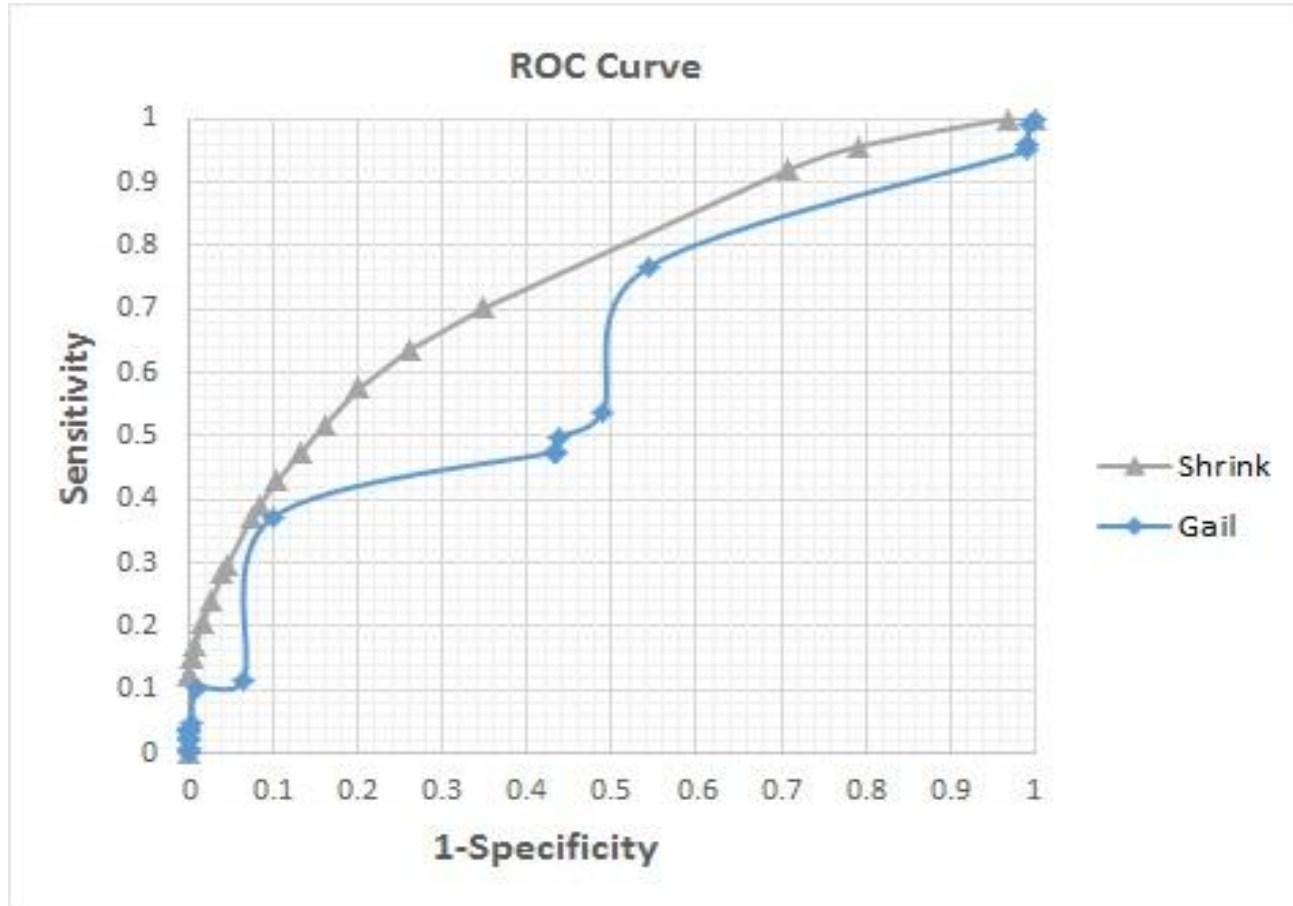- *RRF9*=RRF8+ {*diabetes*}

- *RRF10*=RRF9+ {*urban or rural*}

Experiment —Setting

- Test 2: The test2 will compare the assessment value between Gail and Shrink. The RRF will be selected according to the result of test 1.

Experiment

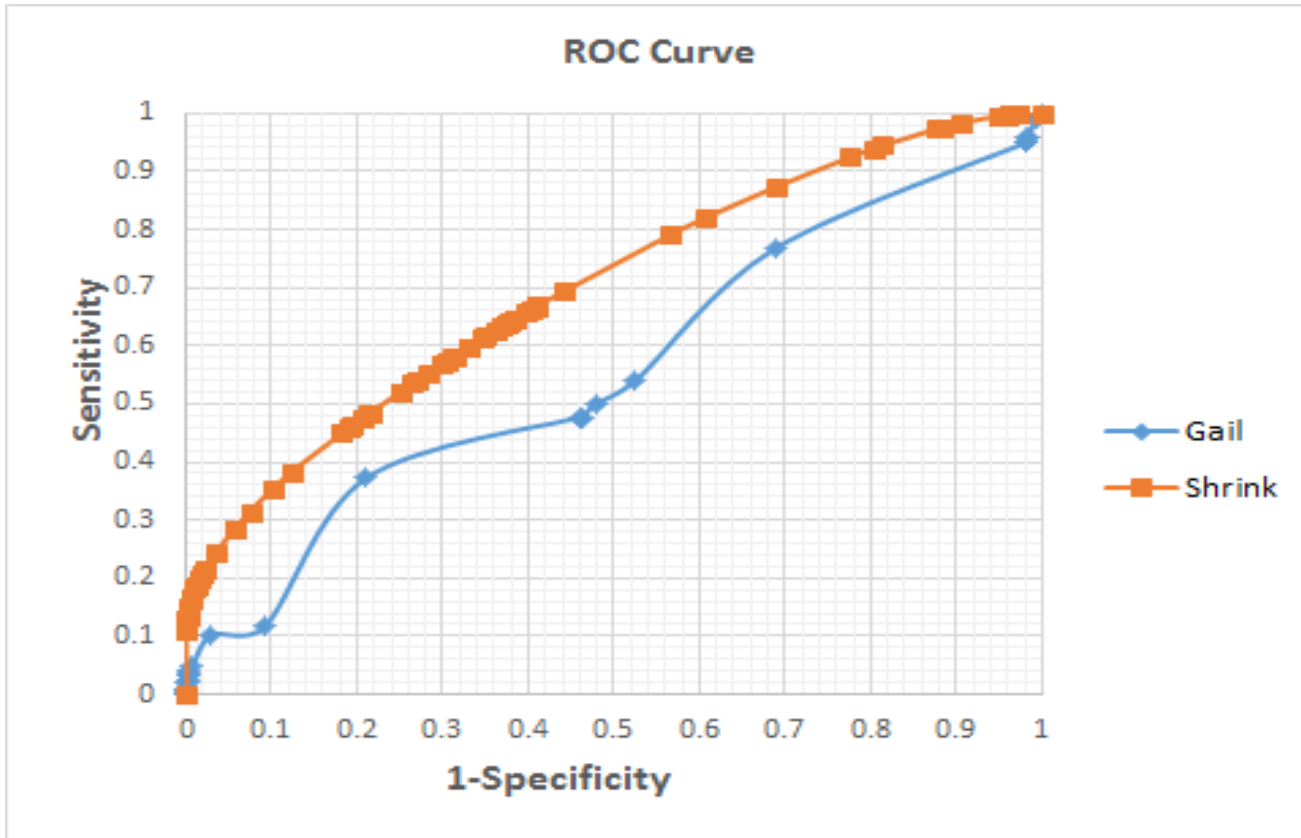- *ROC Curve*

Experiment
—Test 2 result

The ROC Curve of Shrink and Gail based on 5,000 data

Experiment
—Test 2 result

The ROC Curve of Shrink and Gail based on 103,679 data

Conclusion

- We puts forward a new model named Shrink that depends on epidemiological factors, which is more adaptive to the populous country like China than those models. The model utilizes the similarities among epidemiological factors to construct a breast cancer high-risk group, the members of which have high similarity with breast cancer patients.

- The model is meaningful for the prevention and control of breast cancer. And it can be generalized to other countries and regions.

THANK YOU!