A Method to Evaluate Estimates Produced by the Capture-recapture Model

Guoping Rong, He Zhang, Dong Shao, Ouwu

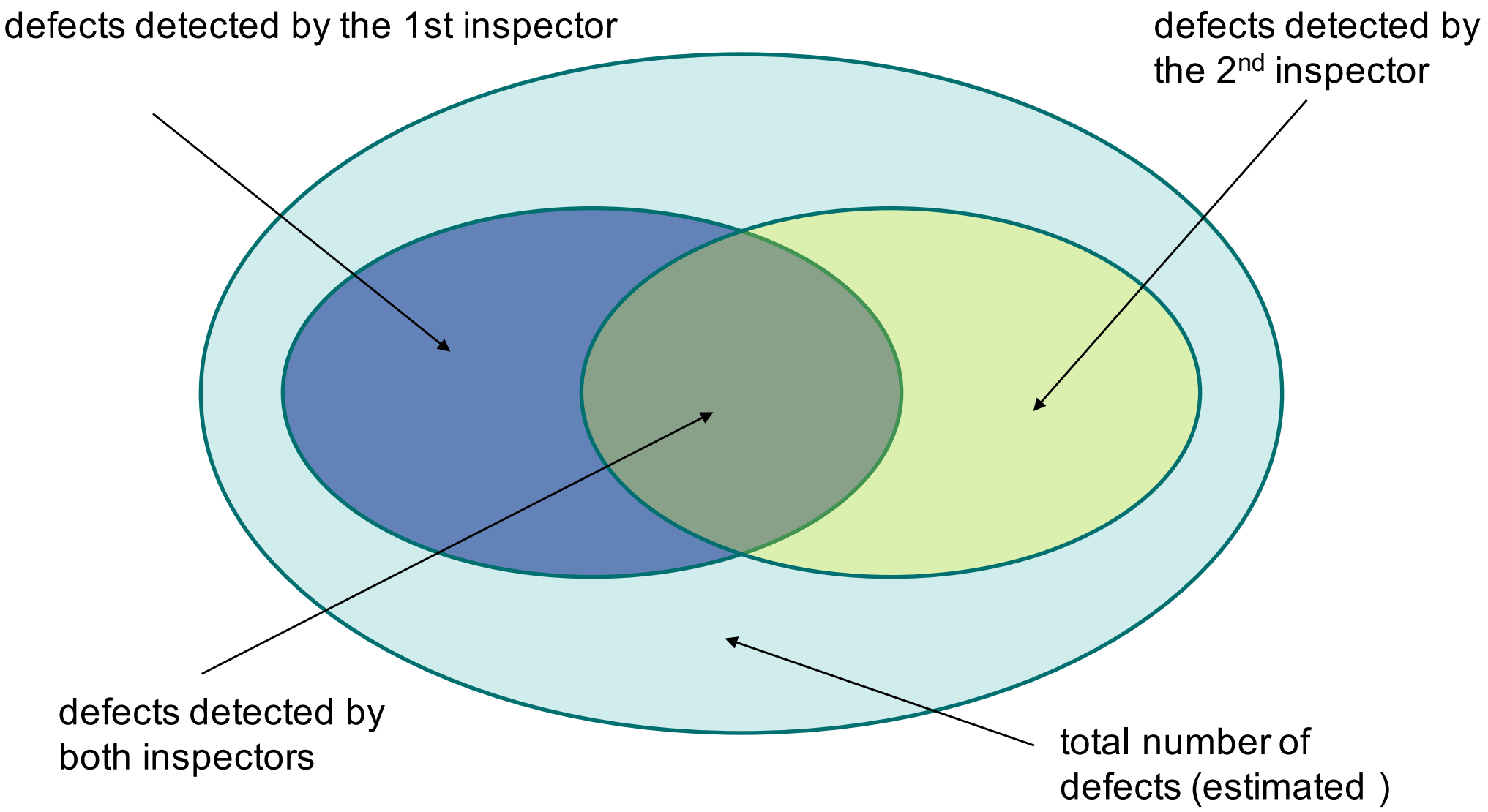
ronggp@software.nju.edu.cn

EMSE Group@Software Institute, Nanjing University

Nov. 9th, 2015

Contents

The Capture-recapture(CR) Method



Motivation

- Estimates created by CR could be very extreme
- Most studies focused on improvements to the CR method
- Few studies on evaluation of the estimation quality

  --May I trust a certain CR estimation?

  --Are There any clues to support decision?

What Is Good Estimation

- ***Accuracy*** is typically measured as the Relative Error (RE). Usually R.E. can be calculated as follow:

RE = (^N −N) / N *, where* ***^N*** *denotes the estimated number of defects and* ***N*** *is the actual number of defects. Therefore, A positive RE means overestimating the total (and residual) defects and vice versa.*

RE may correlate to...

- Predicted Detection Rate(PDR)

PDR = D / N where D denotes the number of unique defects detected by the whole inspection team and ^N is the predicted number of the defects produced by the capture-recapture method.

- Overlap Rate(OR)

OR = M/D where M denotes the number of unique defects detected by more than one inspector and D is the number of unique defects detected by the whole inspection team.

The Data Source

- A controlled experiment[8]

57 student inspectors were required to read a software requirement specification document with 30 seeded defects.

[8] Qi Shan, Guoping Rong, He Zhang and Dong Shao, An Empirical Evaluation of Capture-recapture Estimators in Software Inspection, Adelaide, ASWEC 2015.

Results (RE VS. PDR)

Team size	Correlation	Significance
3	-0.4466814	0.00*
4	-0.7540374	0.00*
10	-0.8796201	0.00*
20	-0.9499149	0.00*
30	-0.9644953	0.00*
40	-0.9405431	0.00*
50	-0.9308268	0.00*

* the value approximates zero.

RE decreased with increased PDR in general. With increased team size(i.e., more inspectors involved), the phenomenon is more and more significant. Besides, smaller inspection teams tended to underestimate the residual defects.

There is a reliable correlation between RE and PDR.

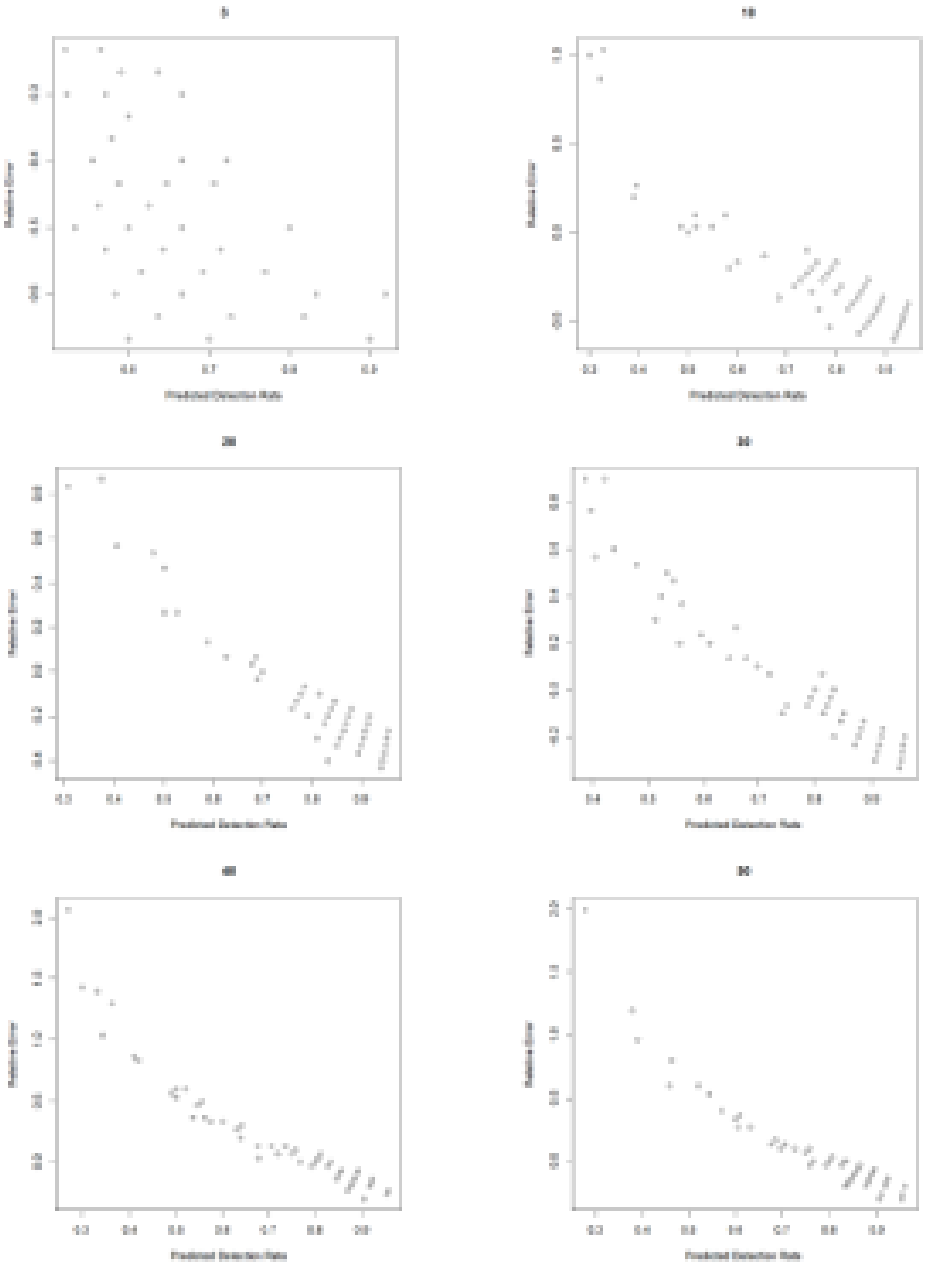


Fig. 1: Correlation between *PDR* and *RE* for Estimator M_h-JK with Various Team Size

Results (RE VS. OR)

Team size	Correlation	Significance
3	-0.3066081	0.0067
4	-0.6347219	0.00*
10	-0.6515608	0.00*
20	-0.6181362	0.00*
30	-0.6713394	0.00*
40	-0.5364338	0.00*
50	-0.1552372	0.00*

* the value approximates zero.

In general, the RE also decreased with increased OR, which to a certain degree indicates the fact that with larger portion of duplicated defects, it is more likely to underestimate the residual defects. However, this phenomenon is mitigated with more inspectors involved.

Although the significance is close to zero, the corresponding correlation value could not suggest reliable correlation existing between OR and RE.

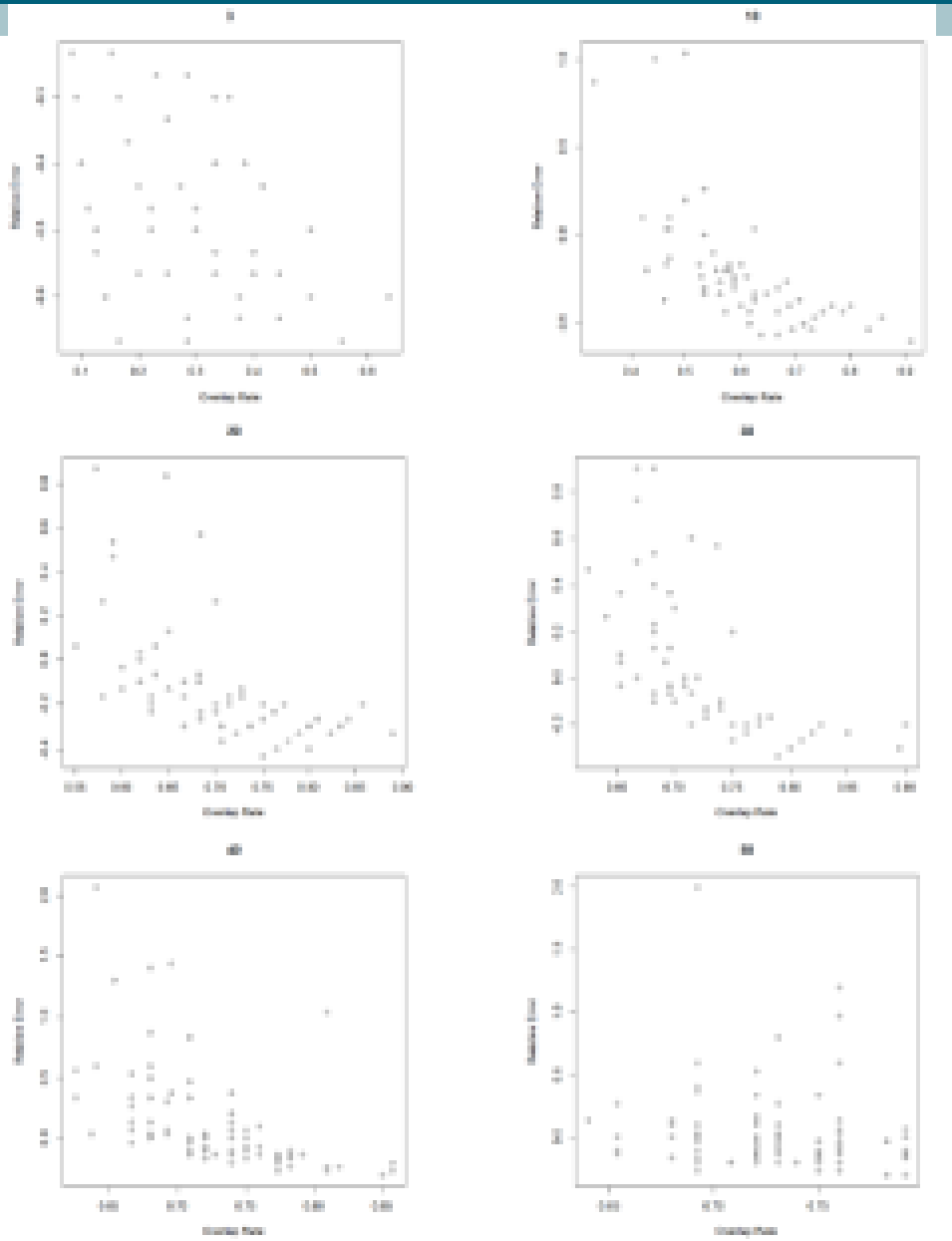


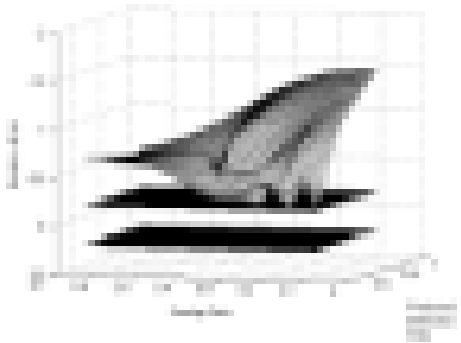
Fig. 2: Correlation between *OR* and *RE* for Estimator M_h-JK

南京大学 软件学院
NANJING UNIVERSITY · SOFTWARE INSTITUTE

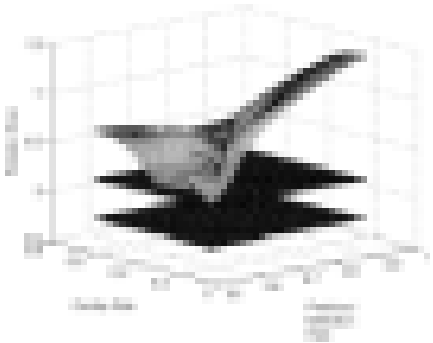
Results (Orthogonal Evaluation)

- By combining two types of correlation mentioned above, we could illustrate the relation between PDR, OR and RE in 3 dimensional diagram.
  - for each team size, we could draw the 3 dimensional diagram with
    - left view(to indicate the correlation between OR and RE),
    - right view(to indicate the correlation between PDR and RE)
    - central view(to indicate both).
  - With horizontal planes to indicate ±20% RE

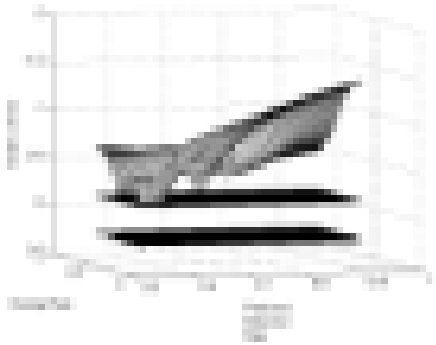
3D diagram for orthogonal evaluation



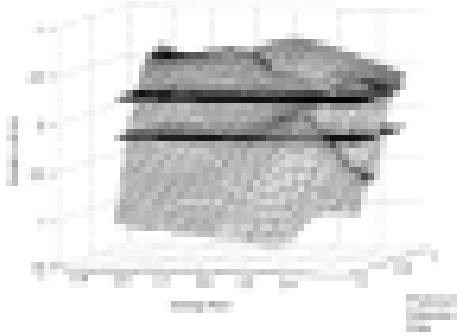
(a) left view for team size 3



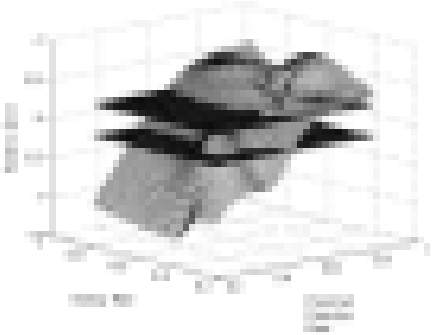
(b) central view for team size 3



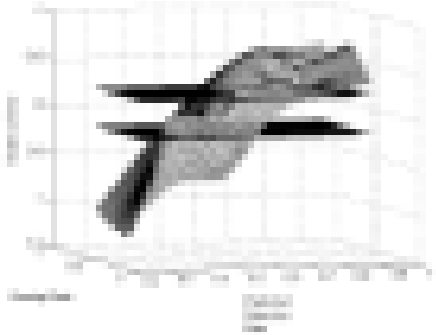
(c) right view for team size 3



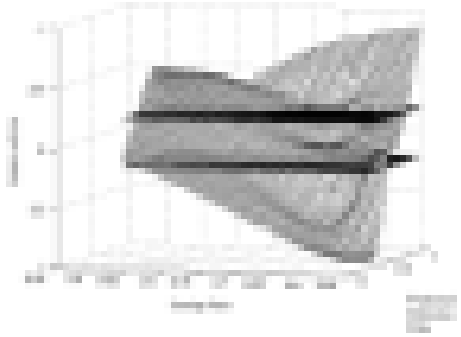
(d) left view for team size 10



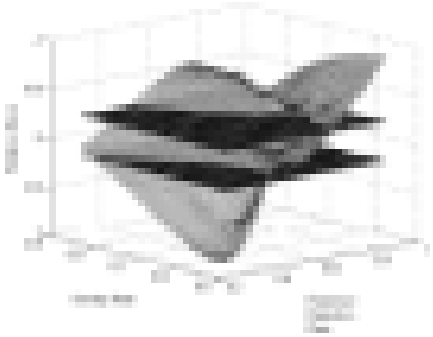
(e) central view for team size 10



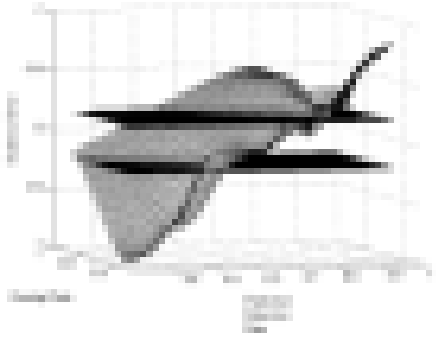
(f) right view for team size 10



(g) left view for team size 20



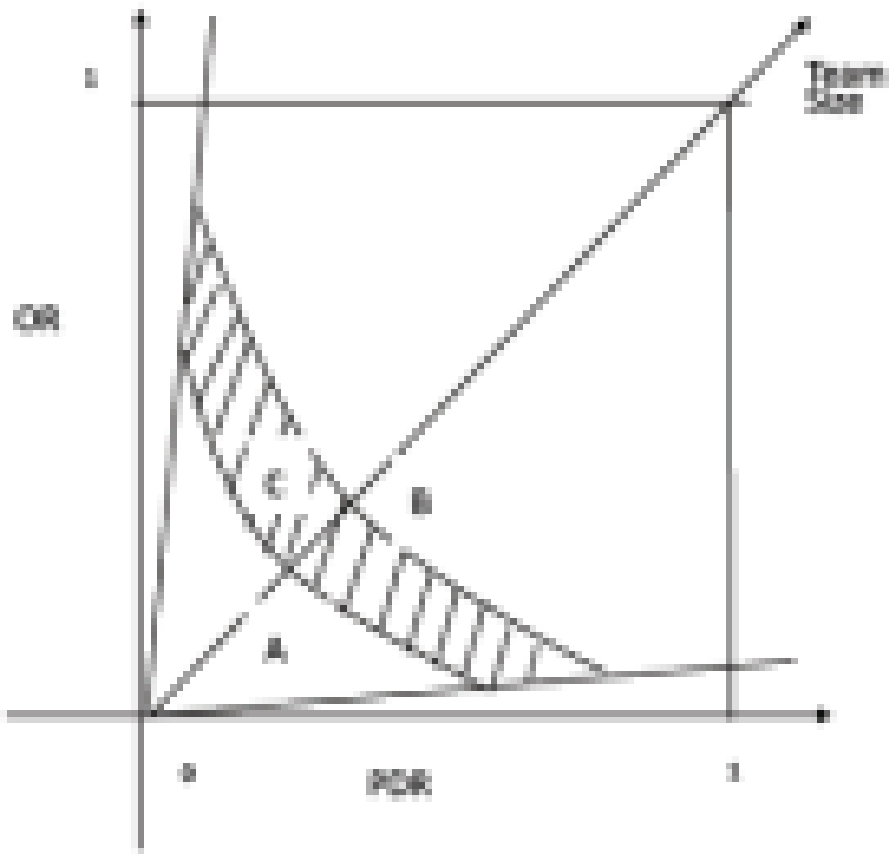
(h) central view for team size 20



(i) right view for team size 20

南京大学软件学院
NANJING UNIVERSITY · SOFTWARE INSTITUTE

How to evaluate CR estimates



- We could then summarize the phenomenon in this Figure.
  - Area A indicates overestimating of the residual defects after a certain inspection where both PDR and OR are relatively small.
  - Area B indicates the opposite phenomenon.
  - The estimates locating in the area C (the shadow area) usually imply relatively accurate estimation.

Discussion

- We still lack reliable benchmark on precise ranges of PDR and OR with various team sizes.
- Secondly, we only evaluated the estimates produced by estimator Mh-JK
- Thirdly, student inspectors and requirement documents are two contextual factors

Conclusions and Future work

- we proposed an orthogonal evaluation method to evaluate the estimates produced by capture-recapture method.

- Future research work regarding the topic in this paper may focus on solving the limitations above.


南京大学软件学院
NANJING UNIVERSITY · SOFTWARE INSTITUTE
Thank You !