An Empirical Study of the Effect of File Editing Patterns on Software Quality

Feng Zhang, Foutse Khomh, Ying Zou

and Ahmed E. Hassan

Queen's
UNIVERSITY

Do developers follow some file editing patterns?

File Editing

Concurrent?

Extended?

Interrupted?

What's the impact on software quality?

File Editing

Consistent?

Expected?

Interpreted?

Examples

1. Example on concurrent editing
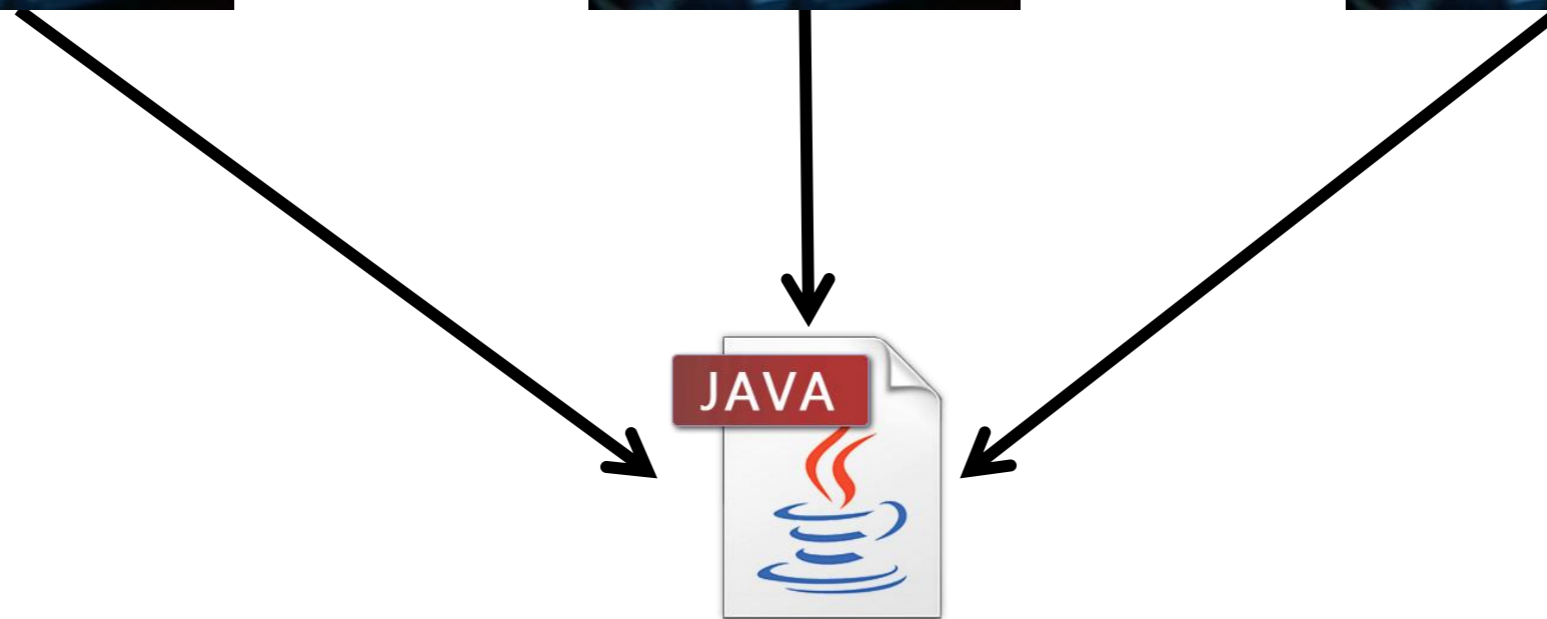
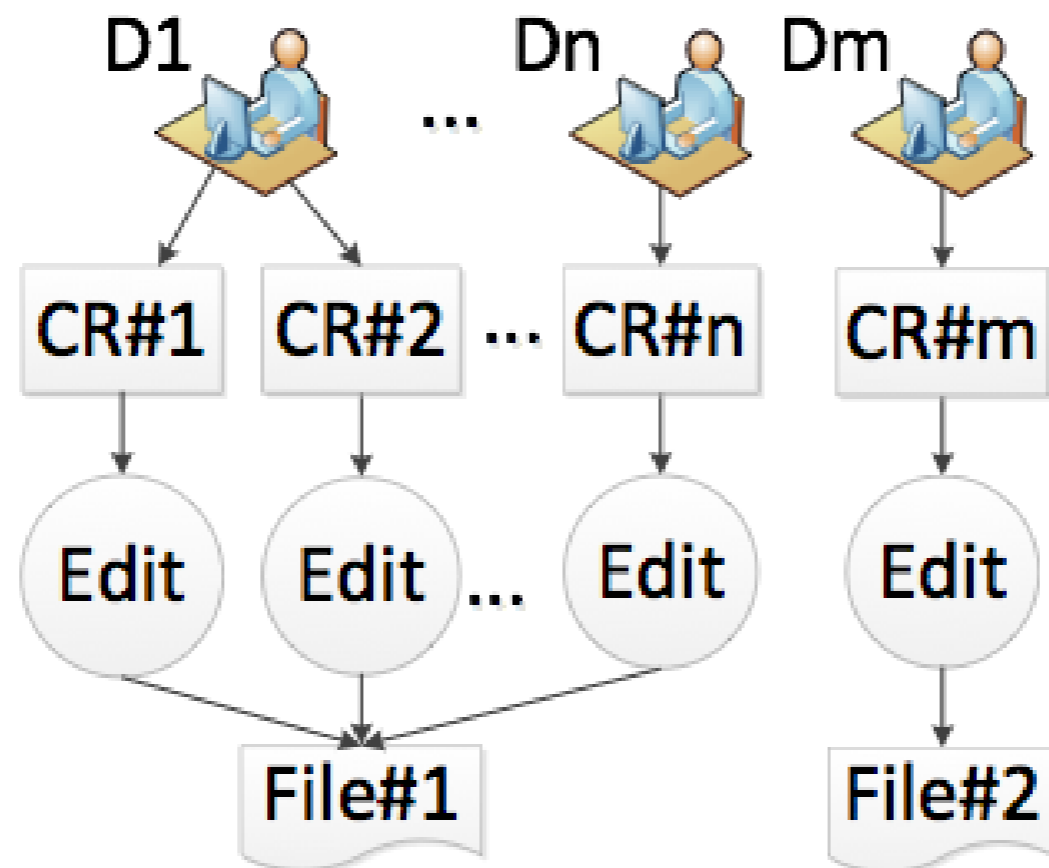Frank   Steffen Pingel   David Green
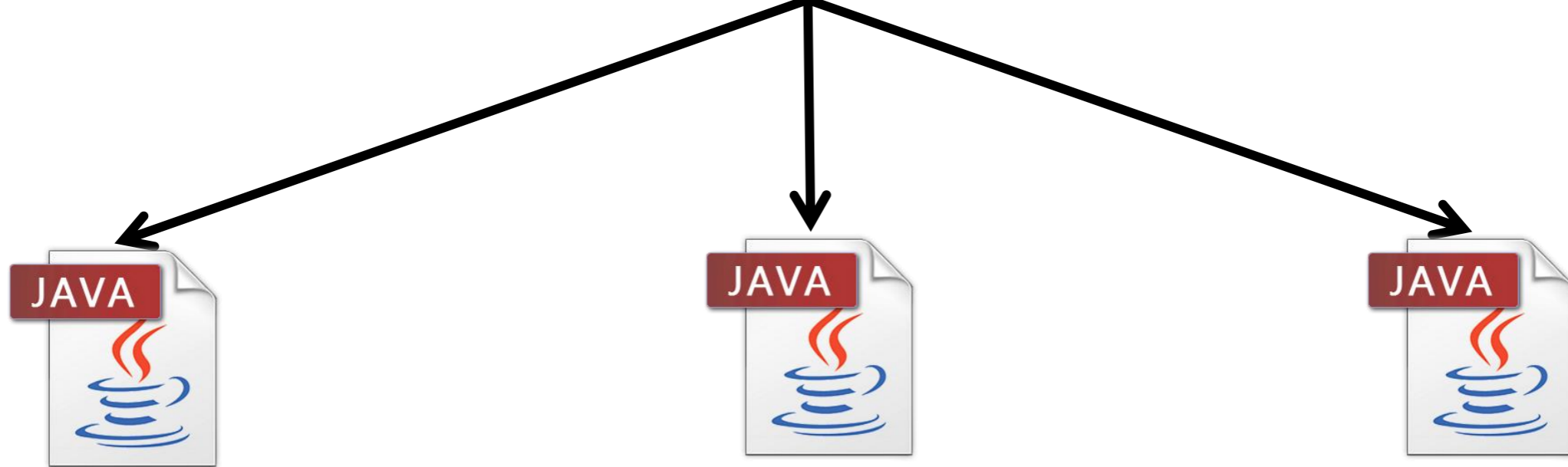

BugzillaTaskEditorPage.java

1. Concurrent Editing

Several developers edit the same file concurrently.
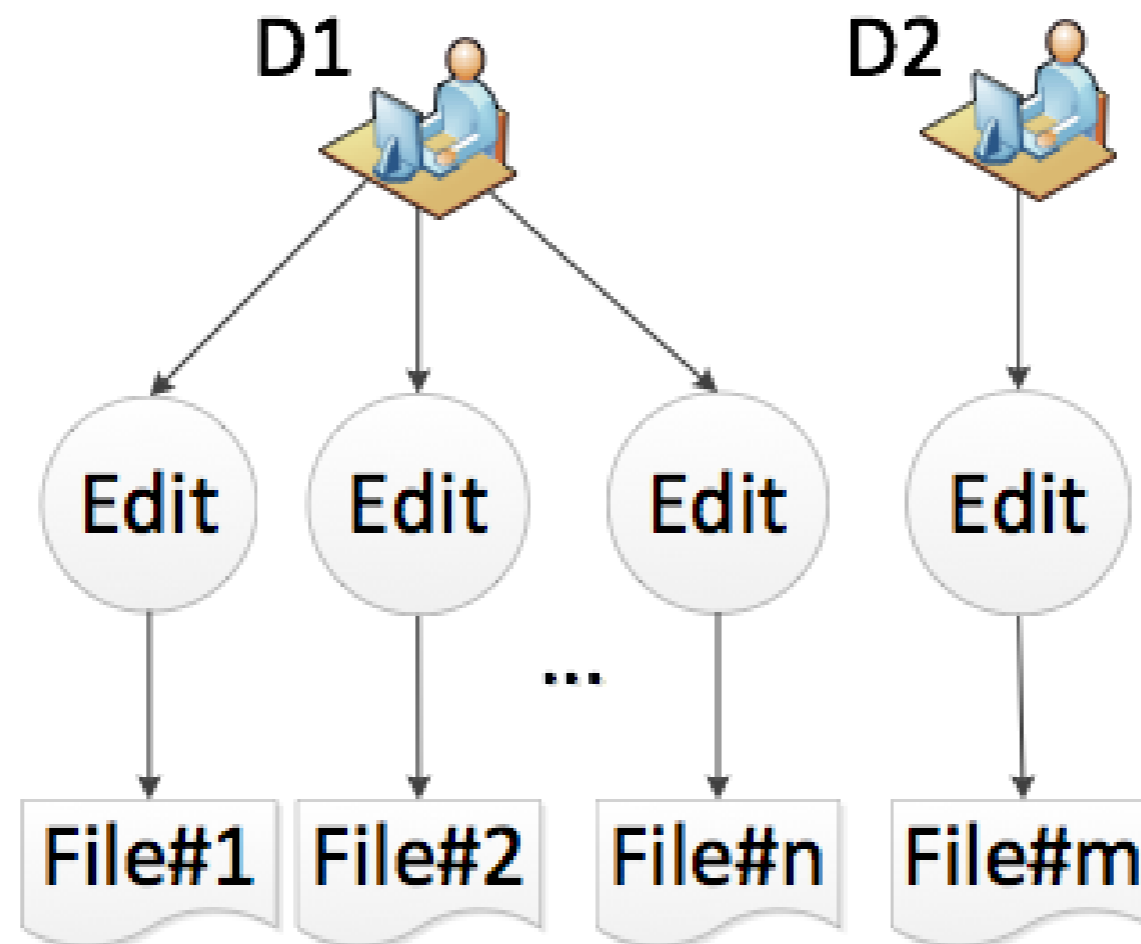
2. Example on parallel editing

Frank

PlanningPerspective
Factory.java

AbstractTaskEditor
Page.java

TasksUiPlugin.java

2. Parallel Editing

Multiple files are edited in parallel by the same developer.

3. Example on extended editing

4.5 hours

1.25 hours

3.6 times

DiscoveryViewer.java

Mylyn Project

3. Extended Editing

Developers spend longer time editing a file.
(threshold: third quantile)

4. Example on interrupted editing

338 hours

TaskCompareDialog.java

164 times

2 hours

Mylyn Project

4. Interrupted Editing

There is a longer idle period during the editing of a file.
(threshold: third quantile)

File Editing Patterns Summary

1. Concurrent Editing

2. Parallel Editing

3. Extended Editing

4. Interrupted Editing

How to evaluate the effect of file editing patterns?

File Editing Patterns

Split Date
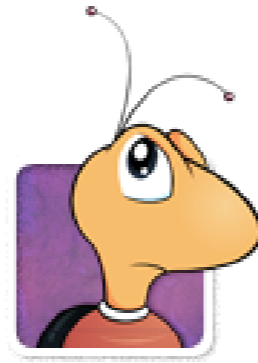
Time

Density of Future Bug

Subject Systems

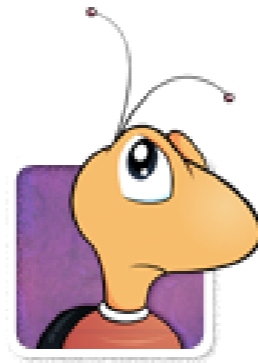|				
Mylyn	2,722 bugs	⟺	3,883 logs	
Eclipse Platform	606 bugs	⟺	793 logs	
PDE	524 bugs	⟺	638 logs	

15

Description of Data

119 developers

2,140 files

5,070 CVS logs

98 bugs

Split Date: 2011-01-01

Data Processing

1. Recovering File Edit History

Mylyn LOG ← bug Id → CVS LOG

2. Recovering File Change History

CVS LOG → density of bug
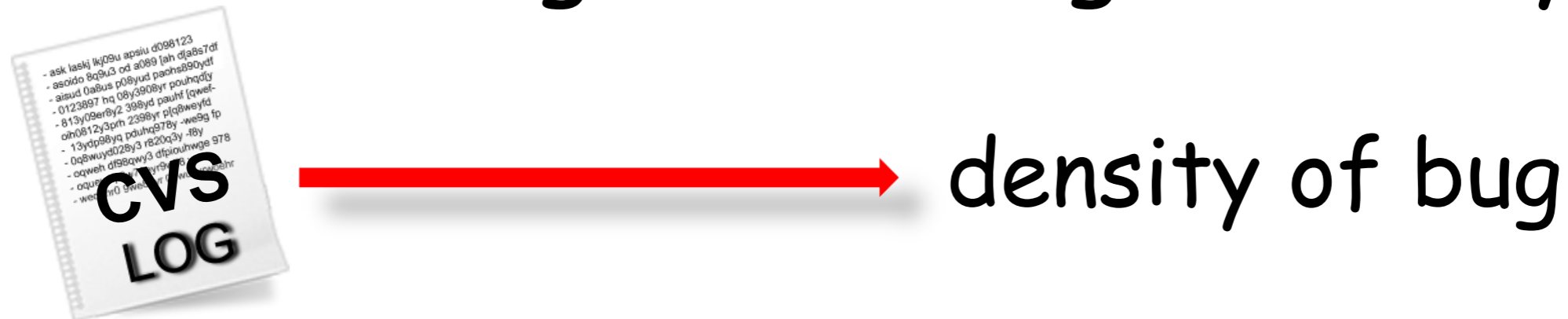
3. Identifying File Editing Patterns
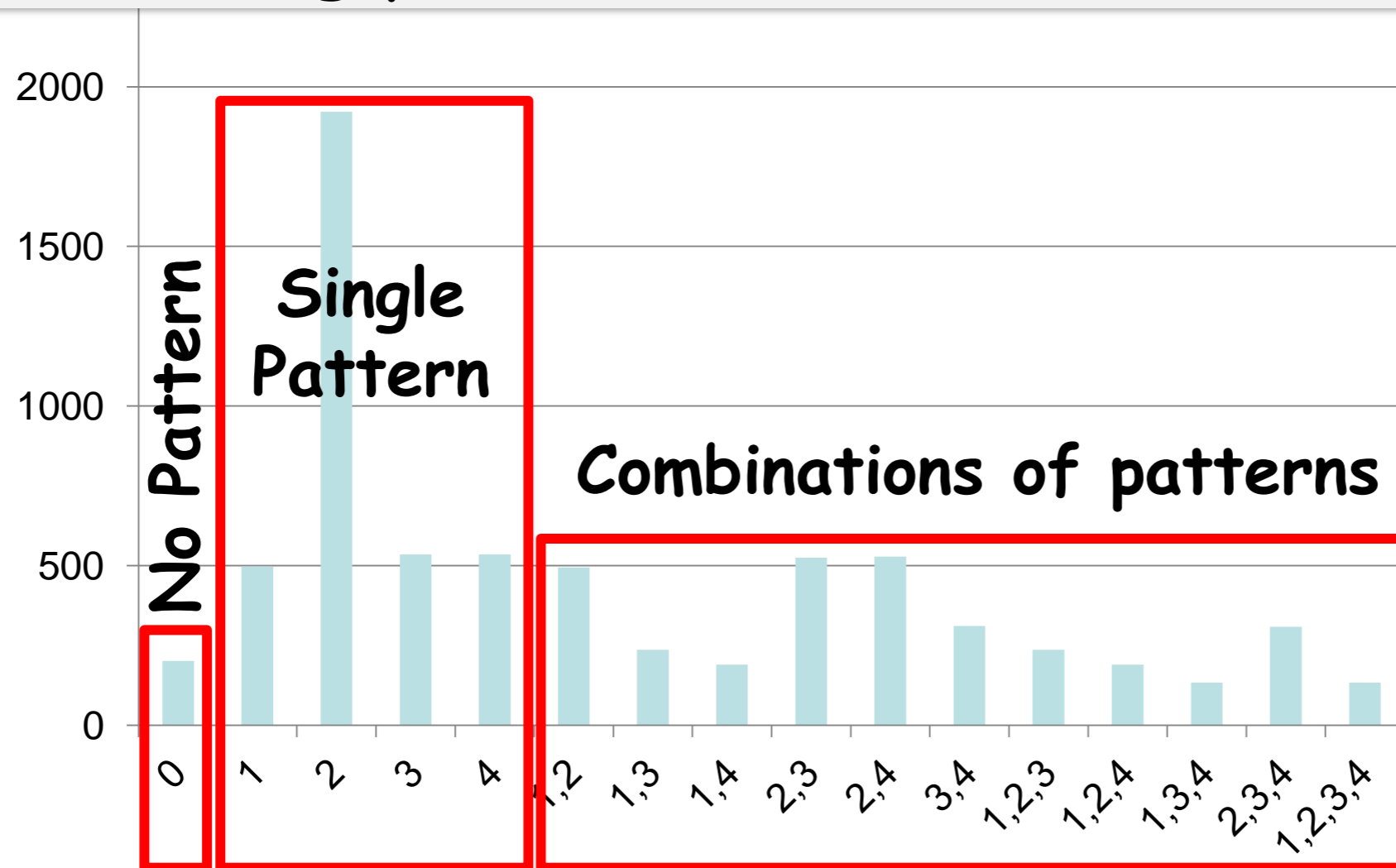
17

Research Questions

RQ1:   Are there different file editing patterns?

RQ2:   Do file editing patterns lead to more bugs?

RQ3:   Do interactions among file editing patterns lead to more bugs?

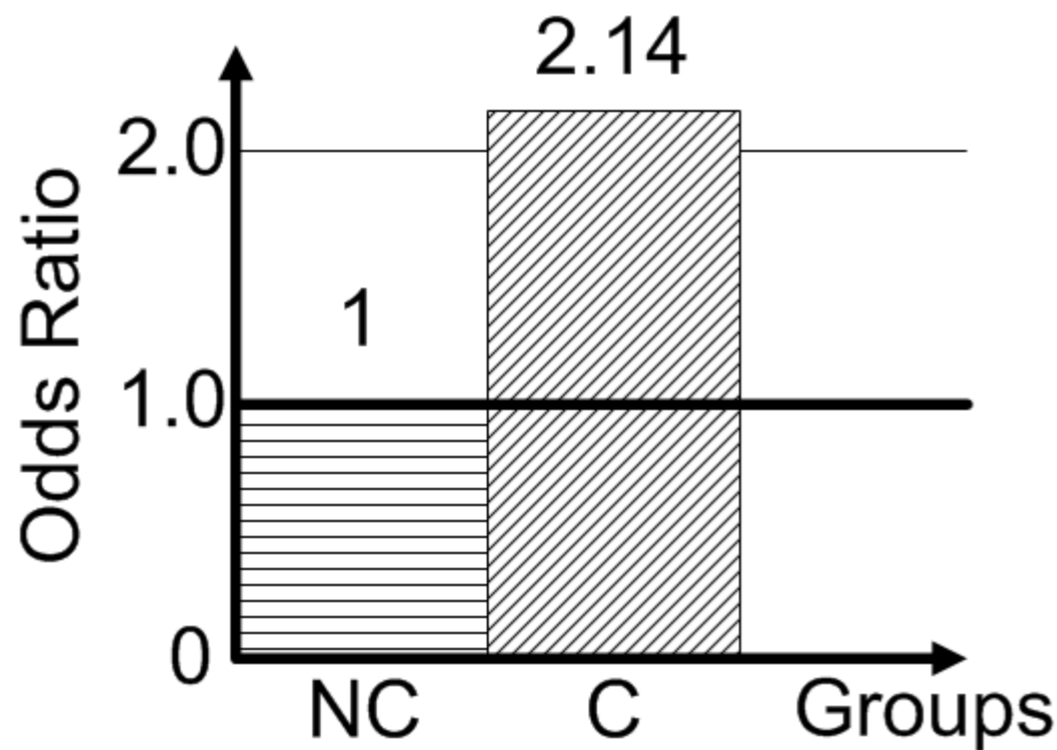# RQ1: Are there different file editing patterns?

File editing patterns do exist

1. Concurrent

2. Parallel

3. Extended

4. Interrupted

Occurrences of file editing patterns and their interactions

19

RQ2: Do concurrent editing pattern lead to more bugs?

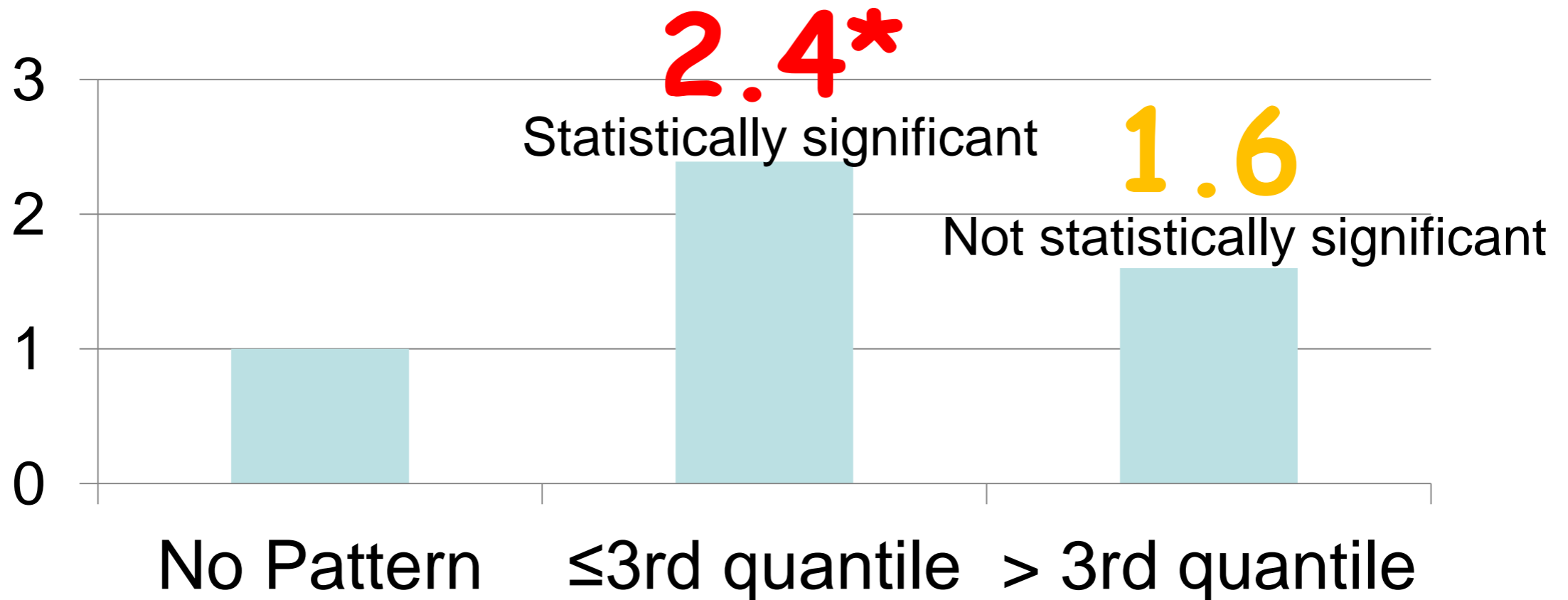2.1 times more likely to experience a future bug if concurrent editing pattern happens

Odds Ratio of two groups:
NC: no concurrent editing
C: concurrent editing
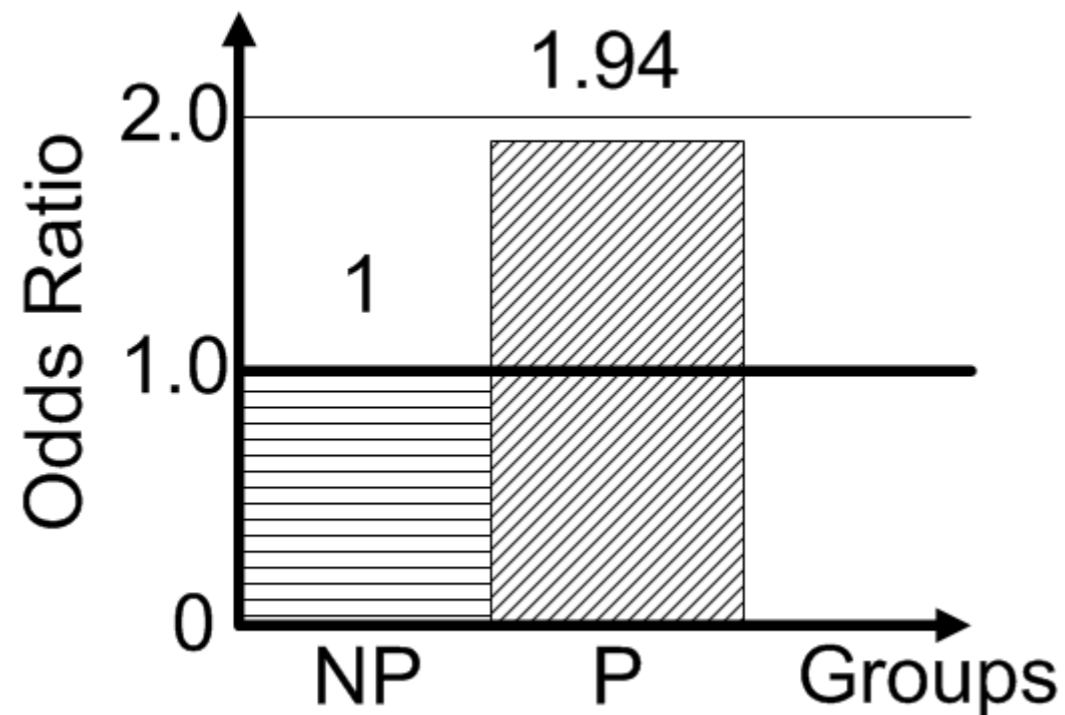
RQ2: Level of involvement in concurrent editing pattern

Definition of Level

average number of developers involved in a concurrent editing of a file.

2.4*
Statistically significant

1.6
Not statistically significant

No Pattern	≤3rd quantile	> 3rd quantile

RQ2: Do parallel editing patterns lead to more bugs?

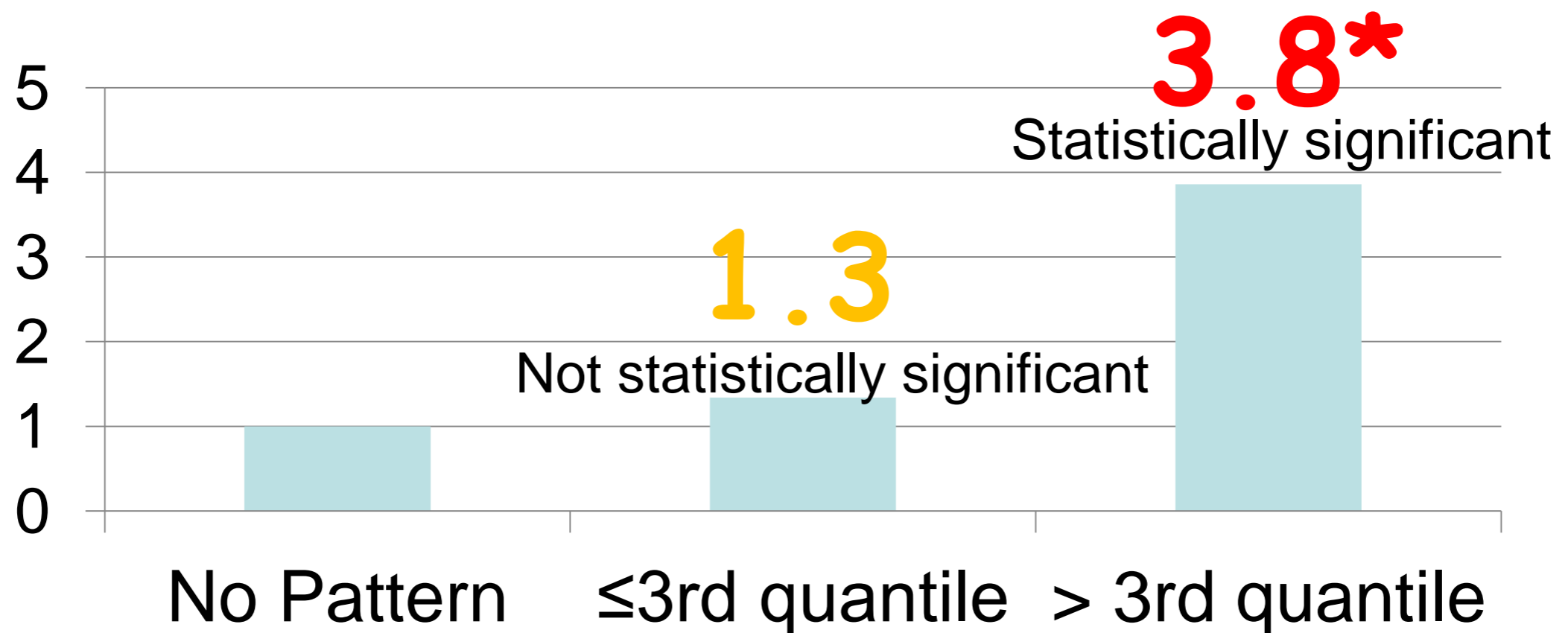1.9 times more likely to experience a future bug if parallel editing pattern happens

Odds Ratio of two groups:
NP: no parallel editing
P: parallel editing
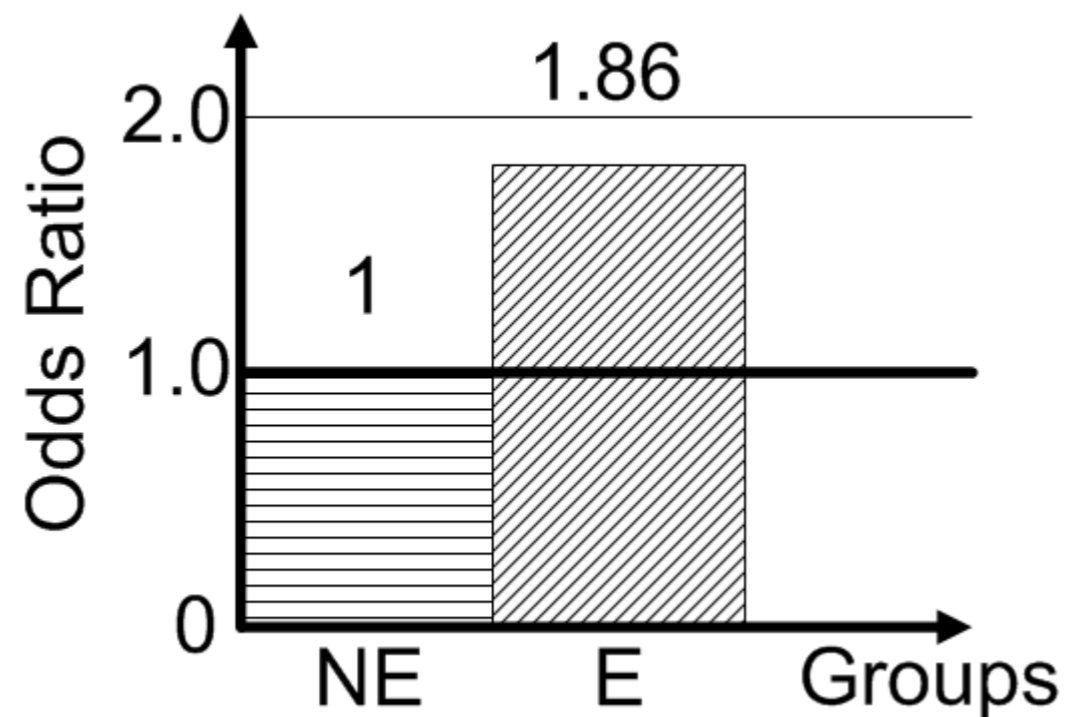
RQ2: Level of involvement in parallel editing pattern

Definition of Level

average number of files edited in parallel with a file.

3.8*
Statistically significant

1.3
Not statistically significant

5
4
3
2
1
0

No Pattern   ≤3rd quantile   > 3rd quantile

RQ2: Do extended editing patterns lead to more bugs?

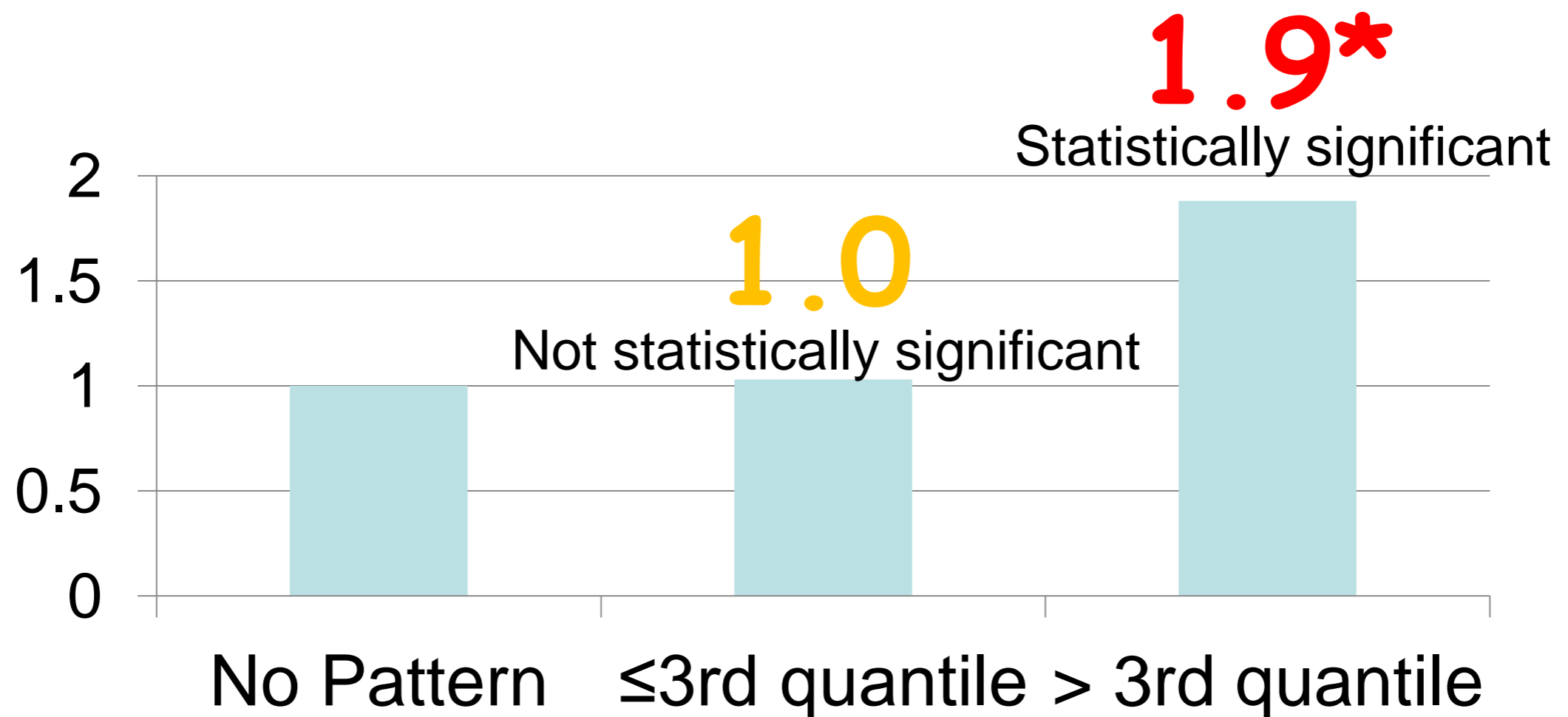1.9 times more likely to experience a future bug if extended editing pattern happens

Odds Ratio of two groups:
NE: no extended editing
E: extended editing
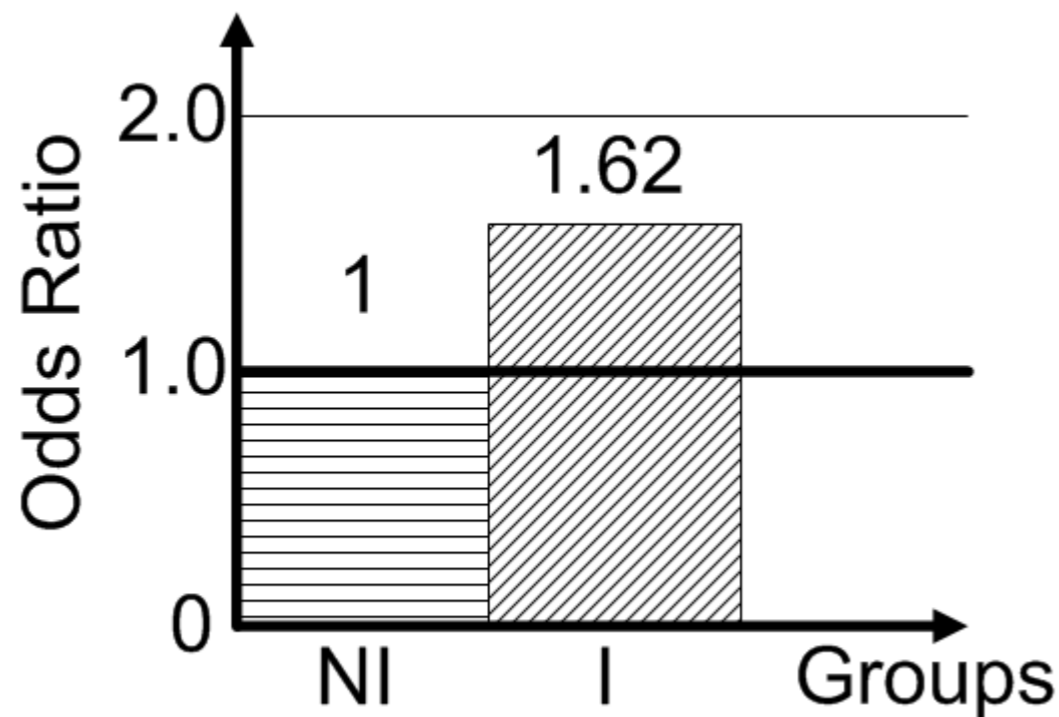
RQ2: Level of involvement in extended pattern

Definition of Level
average editing time of a file.

1.9*
Statistically significant

1.0
Not statistically significant

2	
1.5	
1	
0.5	
0	

No Pattern   ≤3rd quantile > 3rd quantile

RQ2: Do interrupted editing patterns lead to more bugs?

1.6 times more likely to experience a future bug if interrupted editing pattern happens
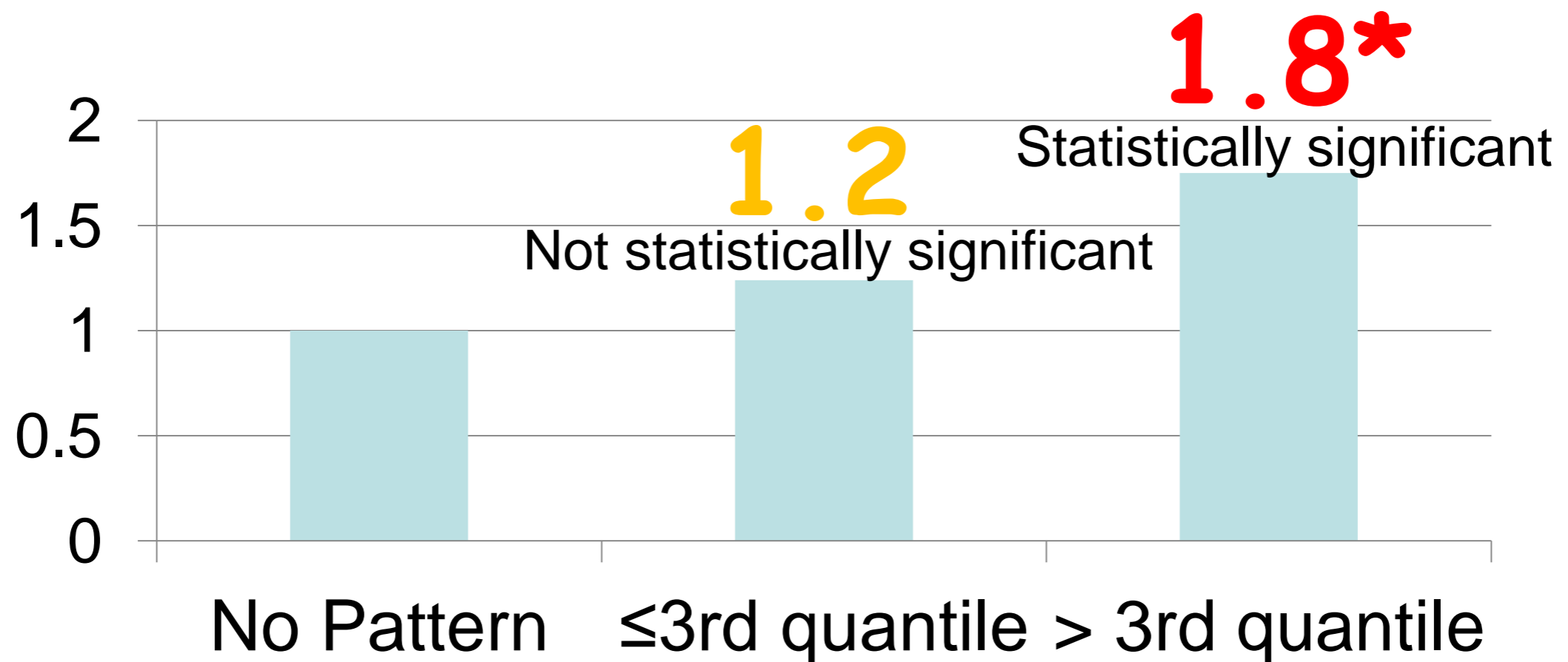
Odds Ratio of two groups:
NI: no interrupted editing
I: interrupted editing

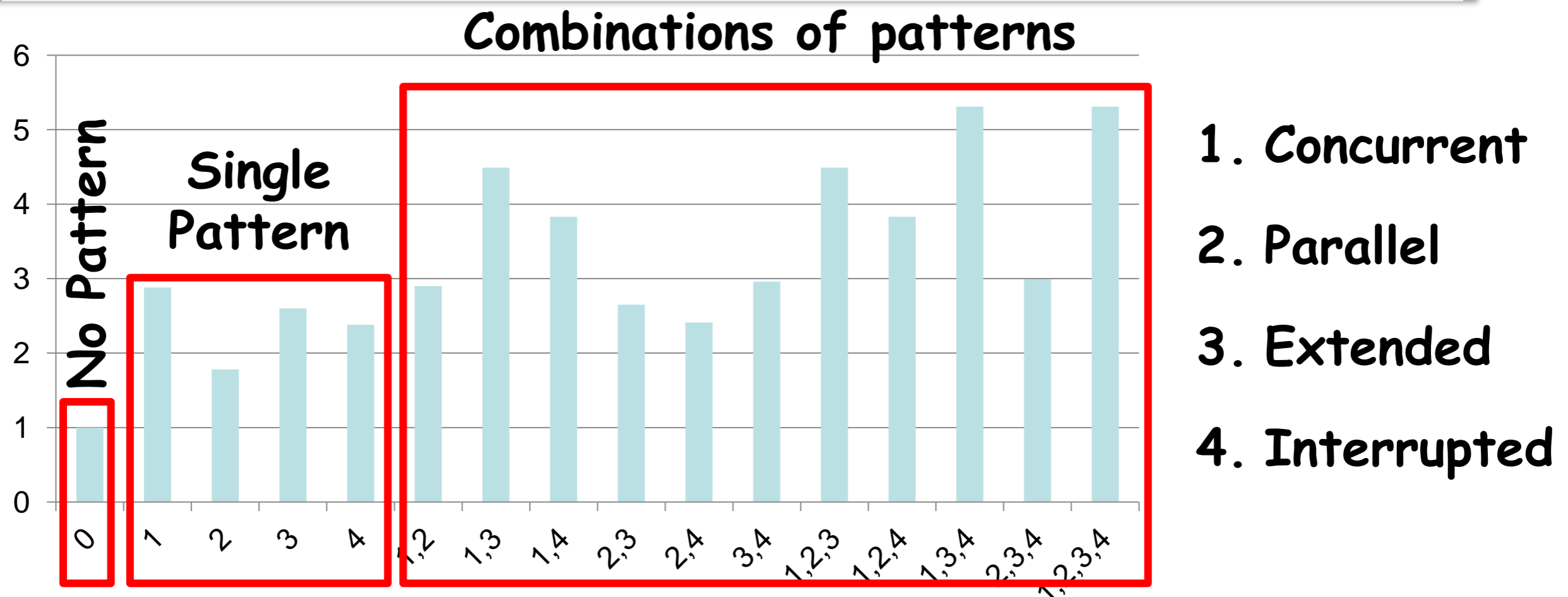# RQ2: Level of involvement in interrupted pattern

Definition of Level
  average interruption time of a file.

RQ3: Do interactions among file editing patterns lead to more bugs?

Combinations of patterns are more risky than single pattern

Combinations of patterns

1. Concurrent

2. Parallel

3. Extended

4. Interrupted

Odds ratio of 16 groups.

28

Conclusion

Odds Ratio **2.1** Concurrent

Odds Ratio **1.9** Parallel

Extended Interrupted

Odds Ratio **1.9**

Odds Ratio **1.6**

Single v.s. Interactions

Average Bug Density:

as high as 1.7 times