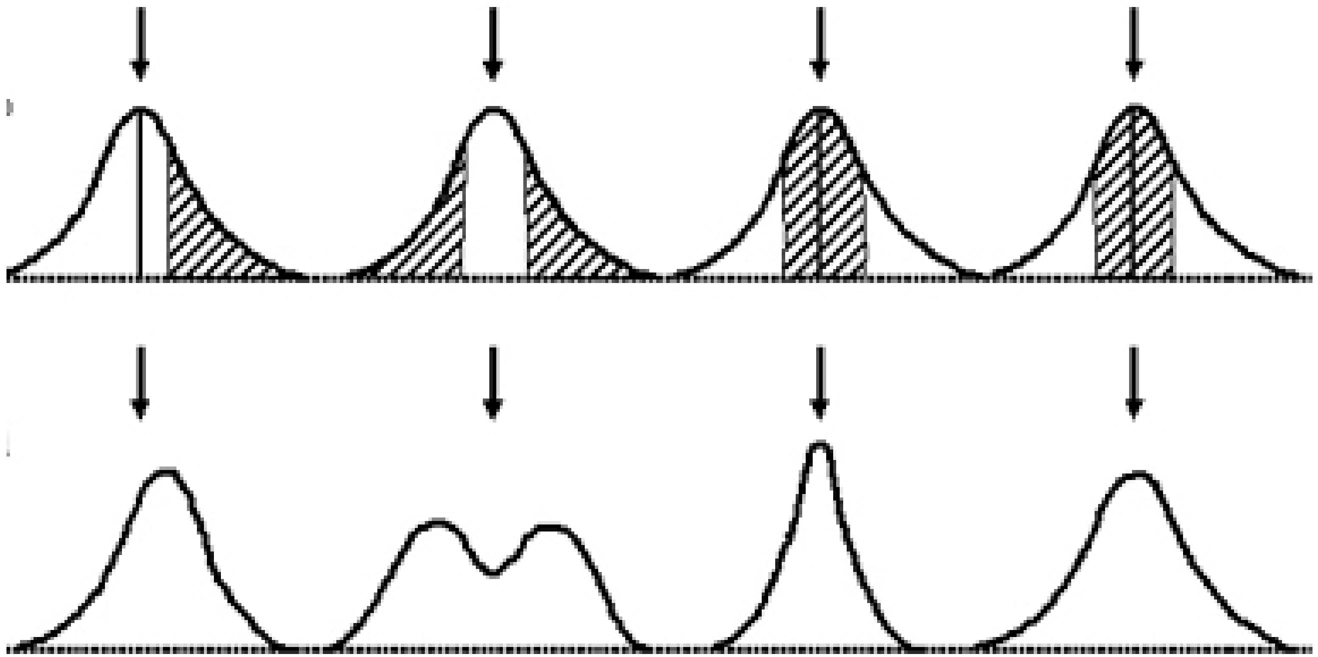
Imperfect Detection and Measuring Selection

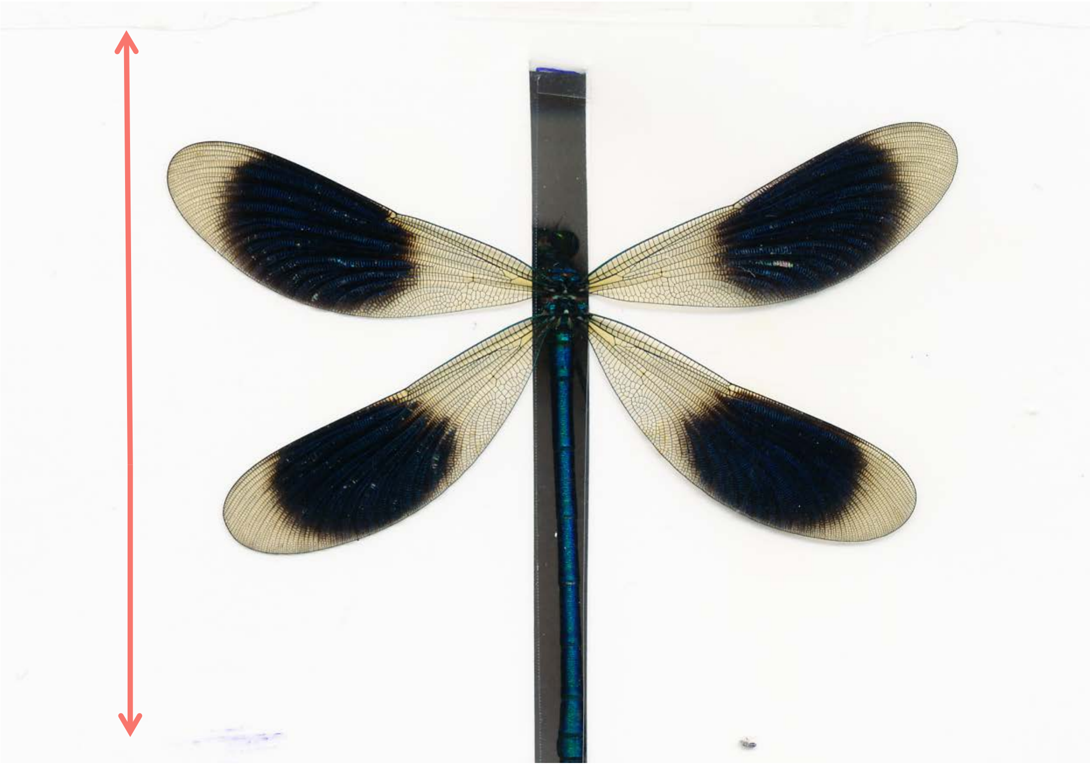


John Waller
PhD EXEB 2013-2017
BLAM 2014

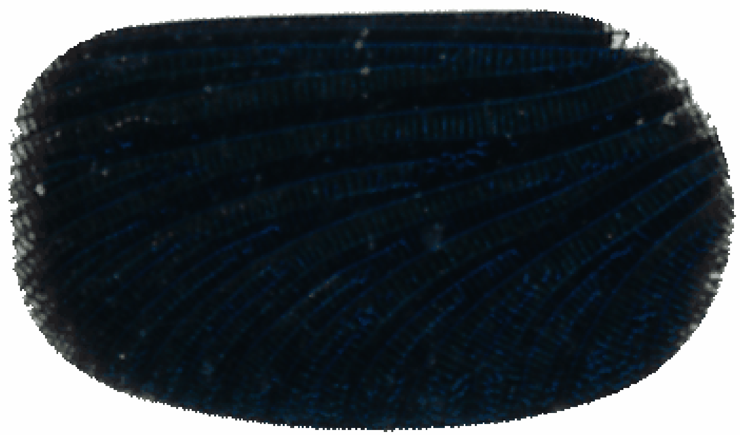
Measuring Selection

- "Natural selection acts on phenotypes"

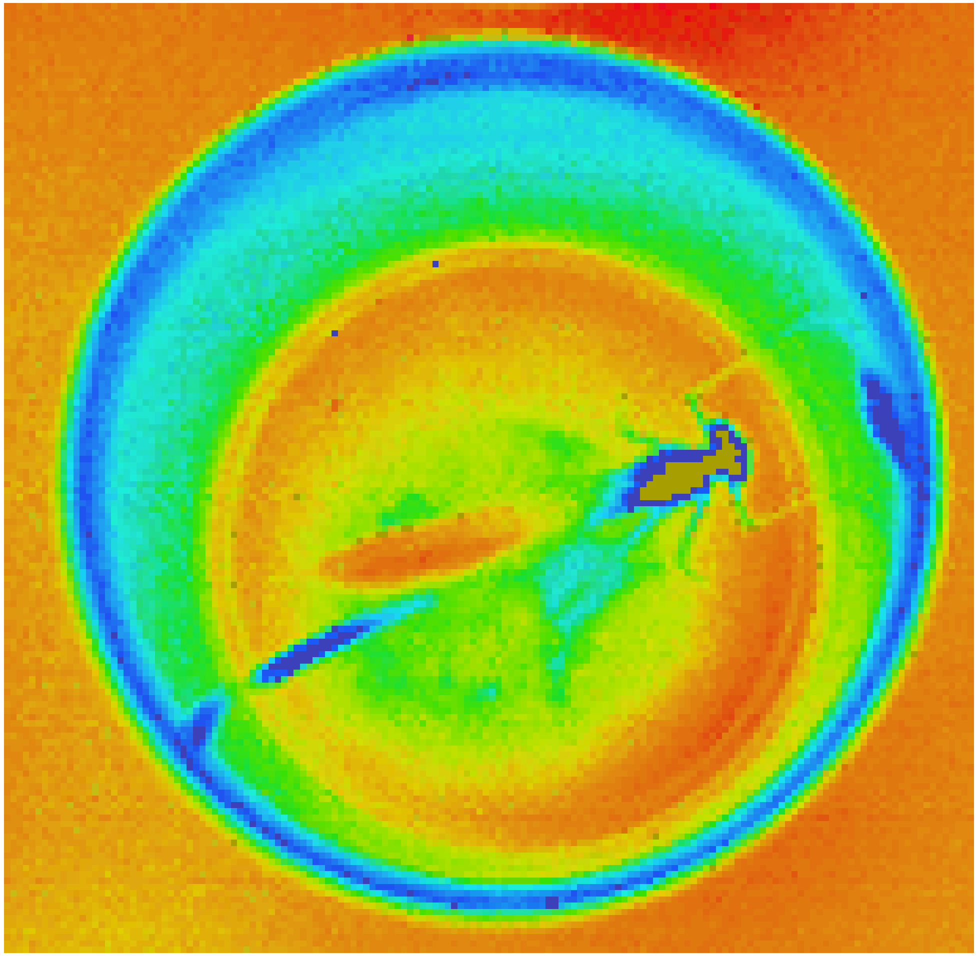
  \- Lande & Arnold 1983



Pigment Variation



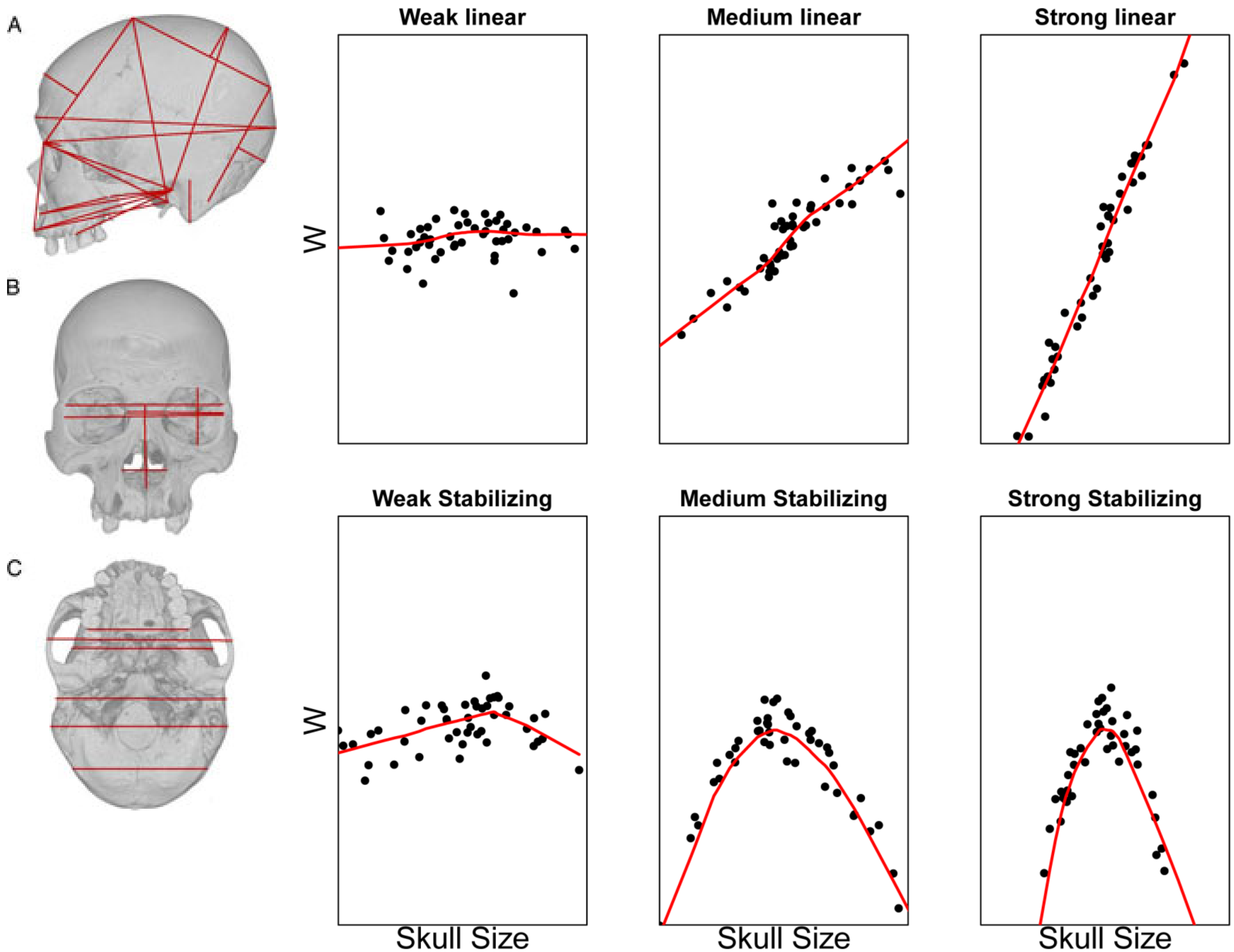
Thermal Plasticity



Flight Performance

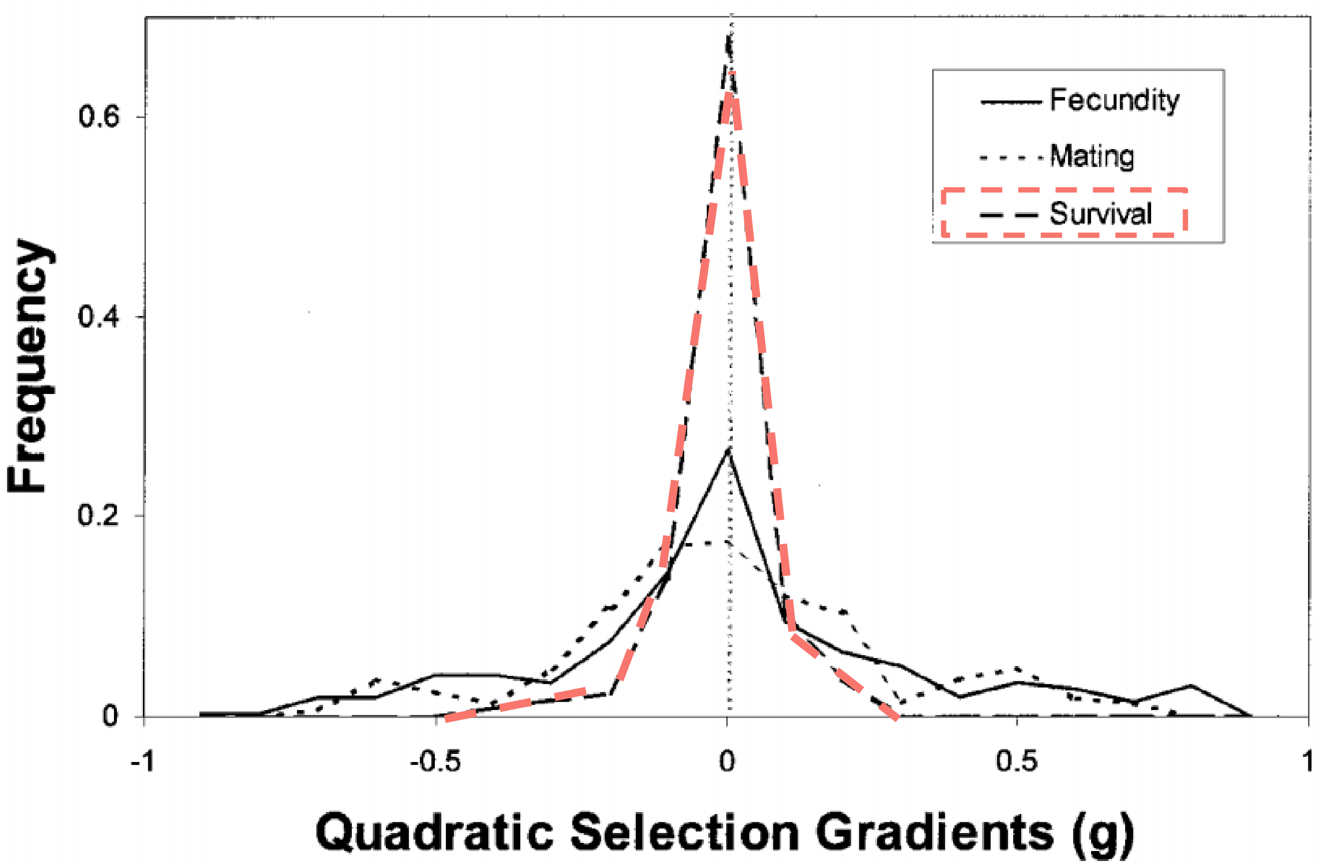
Other Behaviours

Measuring Selection

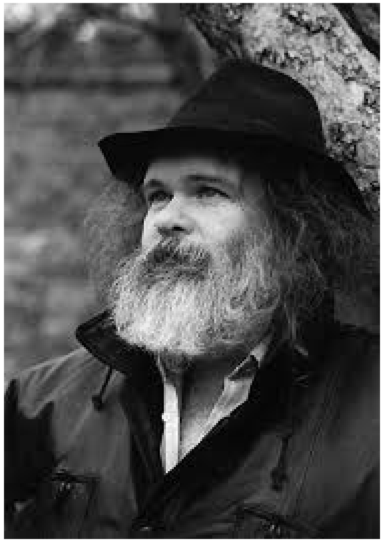


Measuring Selection

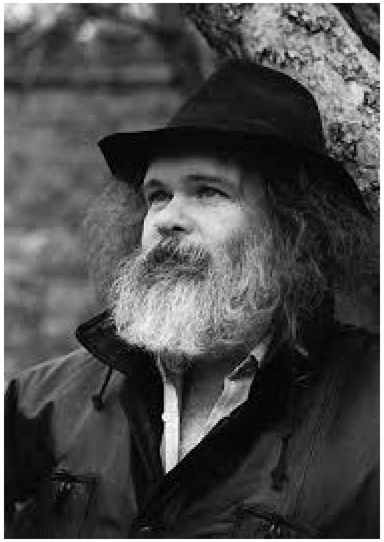
Measuring total fitness (W) is difficult.



Perfect **Detection**

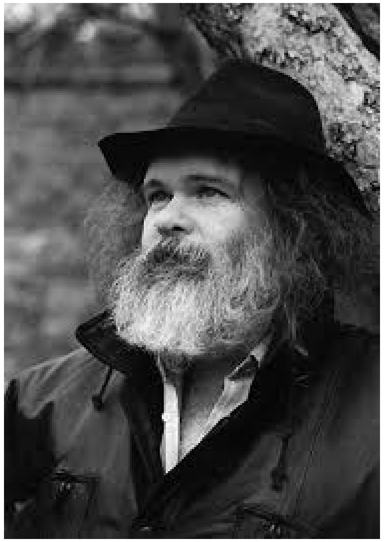


Salomon Schulman



Salomon Schulman

Jessica Abbott

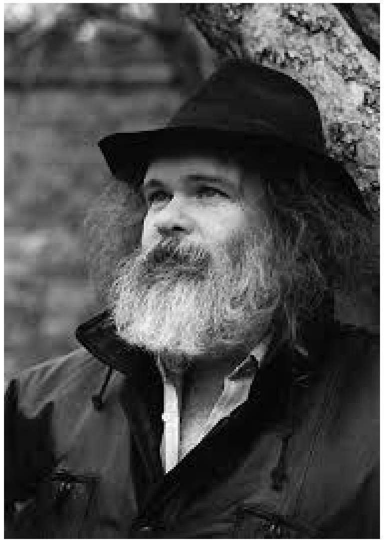


Salomon Schulman

Jessica Abbott

MTWTFSS

1010000



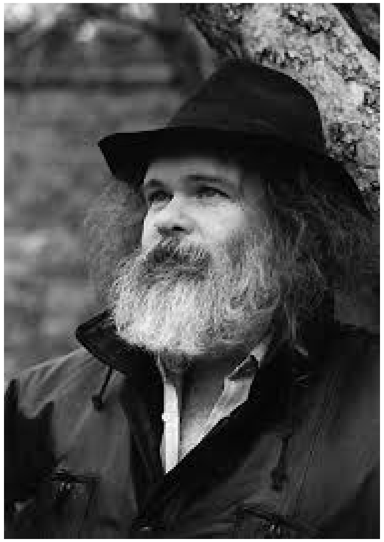
Salomon Schulman

MTWTFSS

1010000

Jessica Abbott

1111110



Salomon Schulman

MTWTFSS

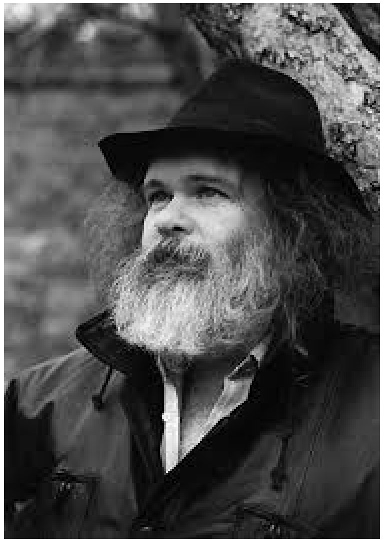
1010000

Lifespan: 3 days

Jessica Abbott

1111110

Lifespan: 6 days



Salomon Schulman

Jessica Abbott

MTWTFSS

$\underbrace{1010}_{\text{But he could still be alive at age 4!}}000$

$\underbrace{111111}_{\text{Likely dead at age 6}}0$

Two statistical methods to measure **survival** selection:

Lande-Arnold (LA)

- **Ignores** imperfect detection
- Uses minimum lifespan
- Used by evolutionary ecologists

MARK

- **Accounts** for it
- Estimates re-capture probability
- Used by wildlife biologists

LA is linear regression

MARK is fancier

Does method matter?

Does method matter?

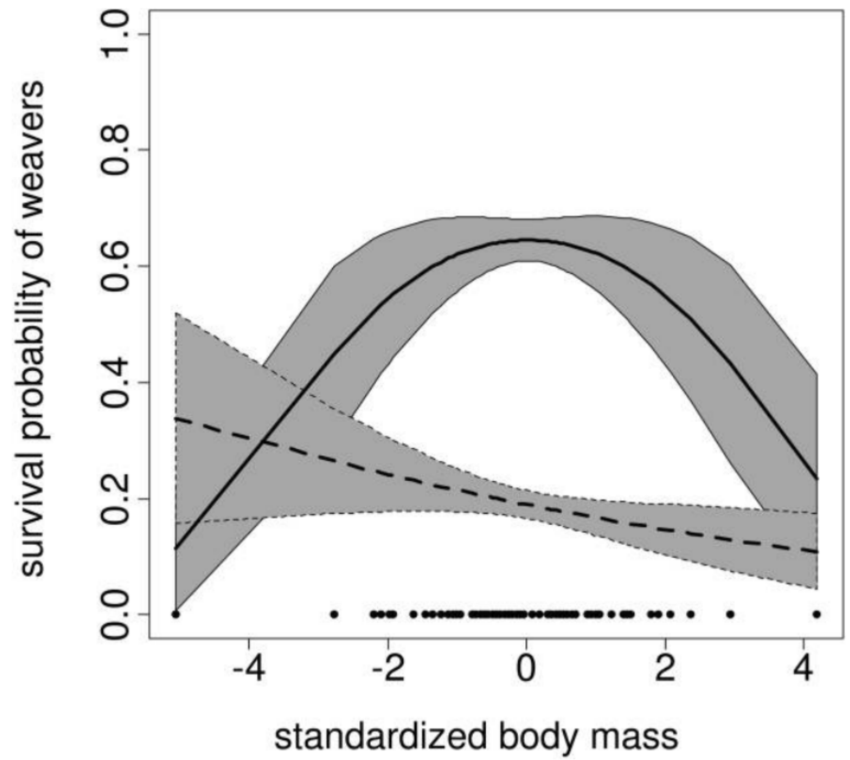


Figure 1: Relationship between survival and body mass in sociable weavers obtained by a mark-recapture analysis (*solid line*) and a naive analysis assuming perfect detection (*dashed line*). Filled circles represent body mass values, and shaded areas represent 95% confidence intervals.

Does method matter?

Mark

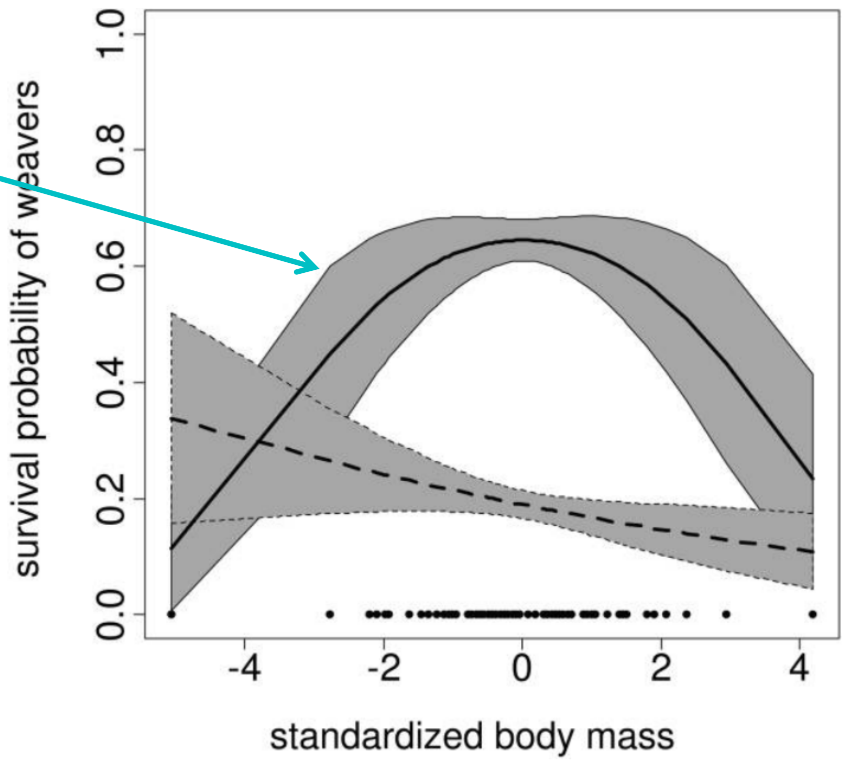


Figure 1: Relationship between survival and body mass in sociable weavers obtained by a mark-recapture analysis (*solid line*) and a naive analysis assuming perfect detection (*dashed line*). Filled circles represent body mass values, and shaded areas represent 95% confidence intervals.

Does method matter?

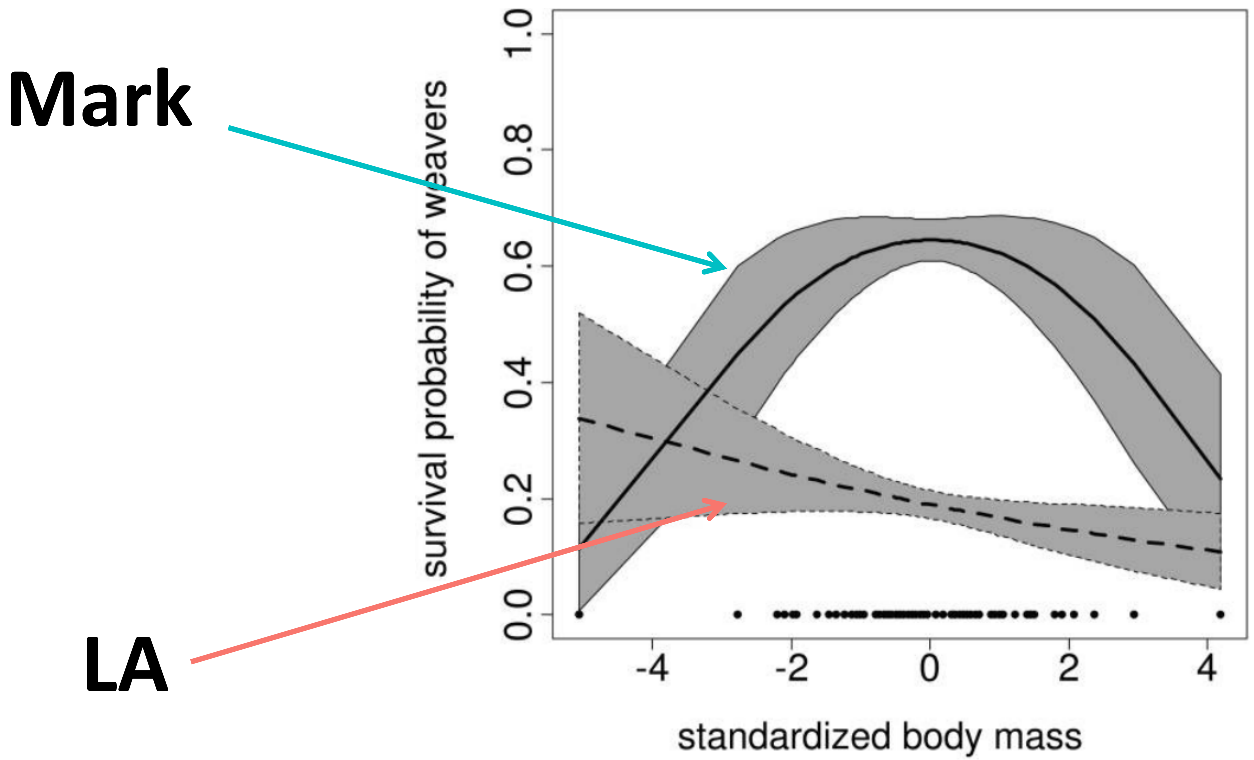
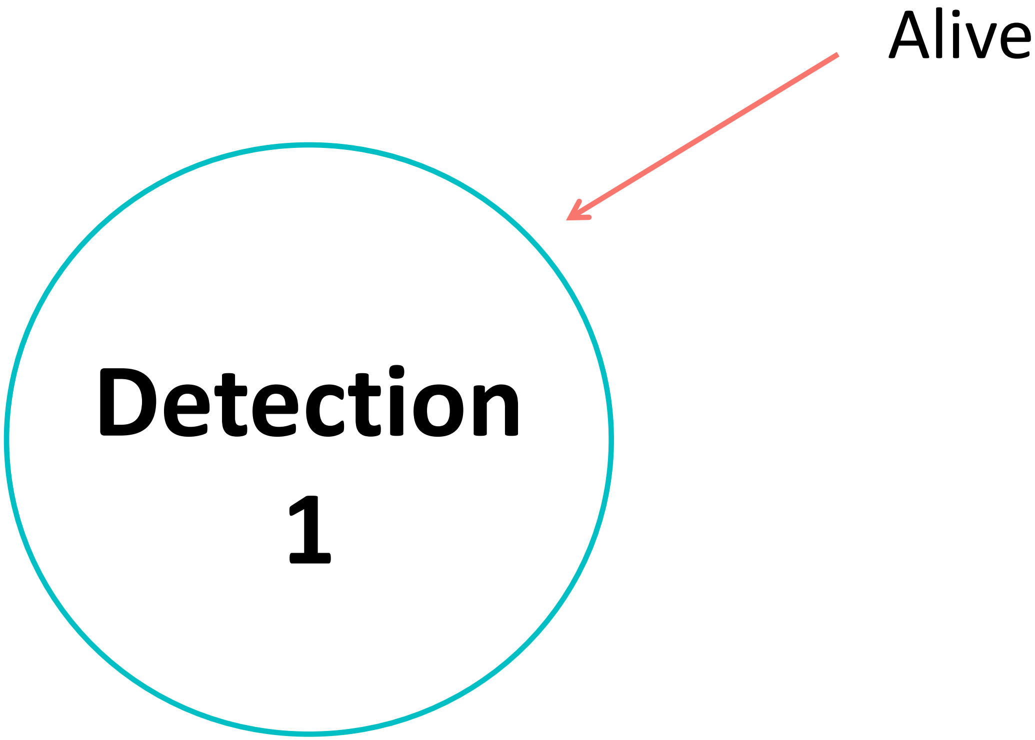
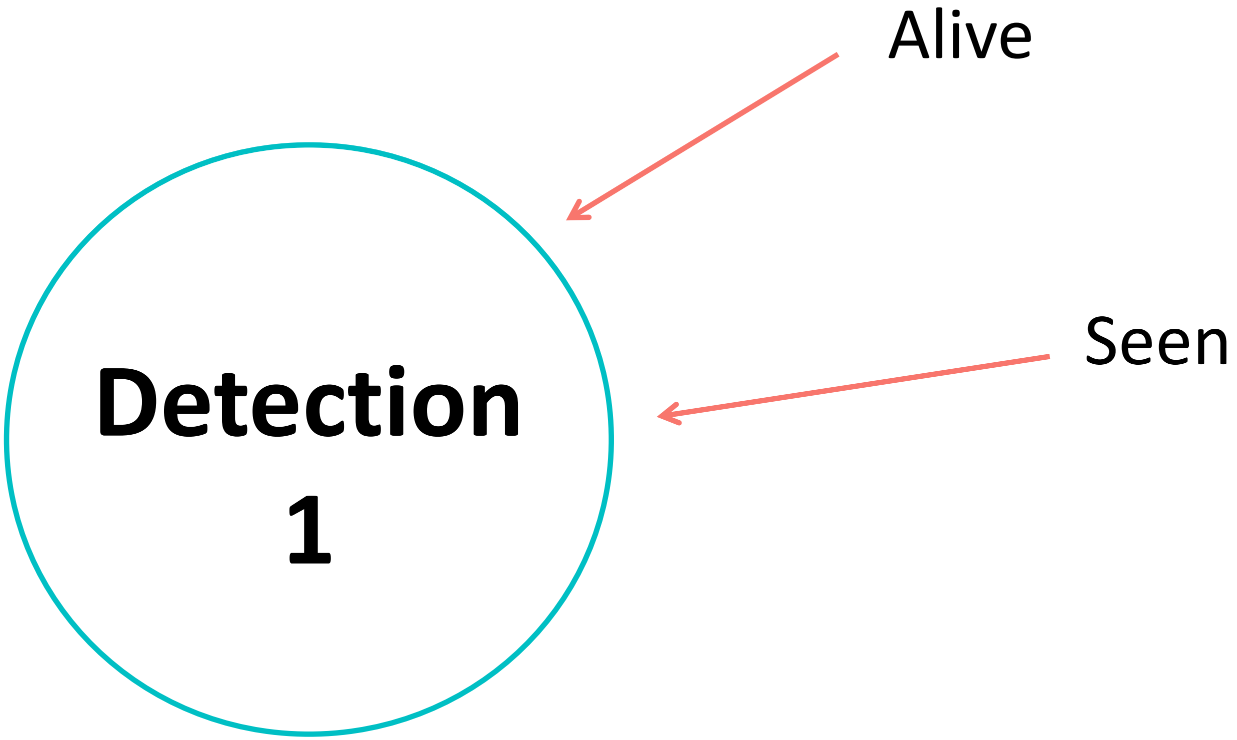
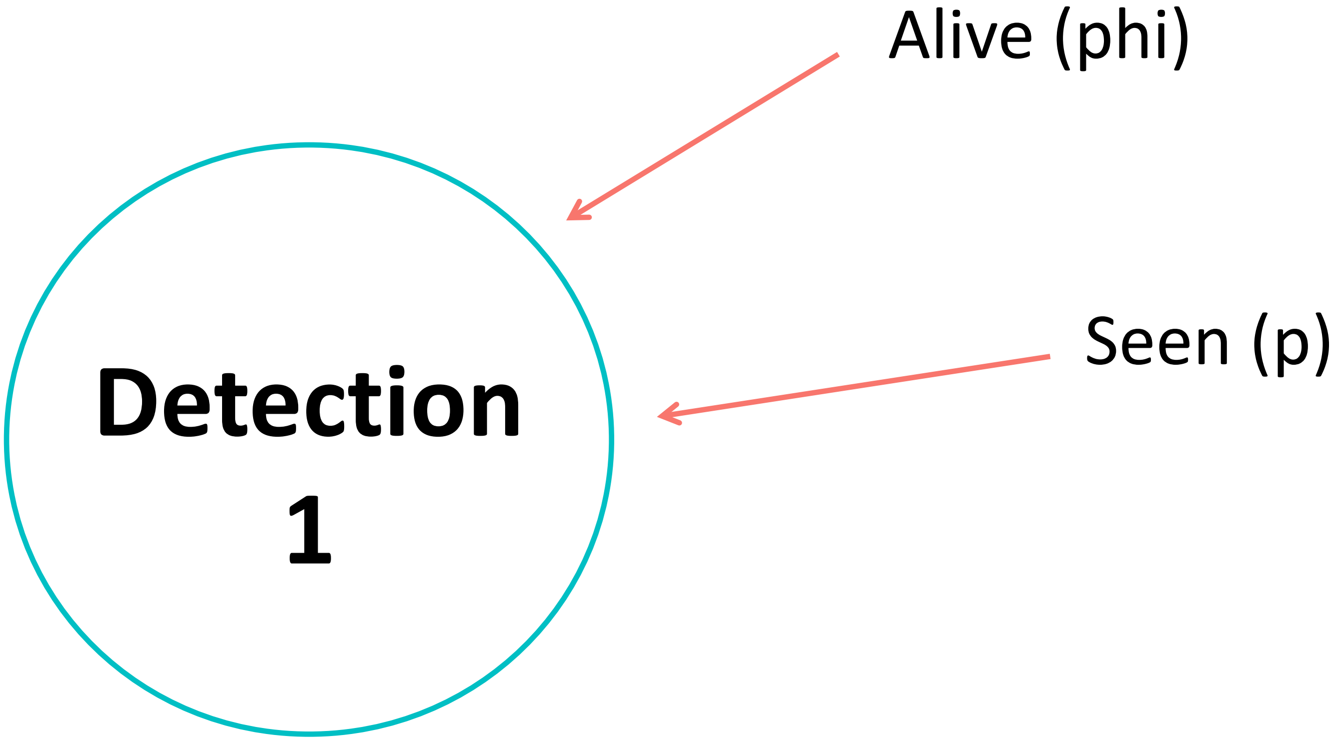


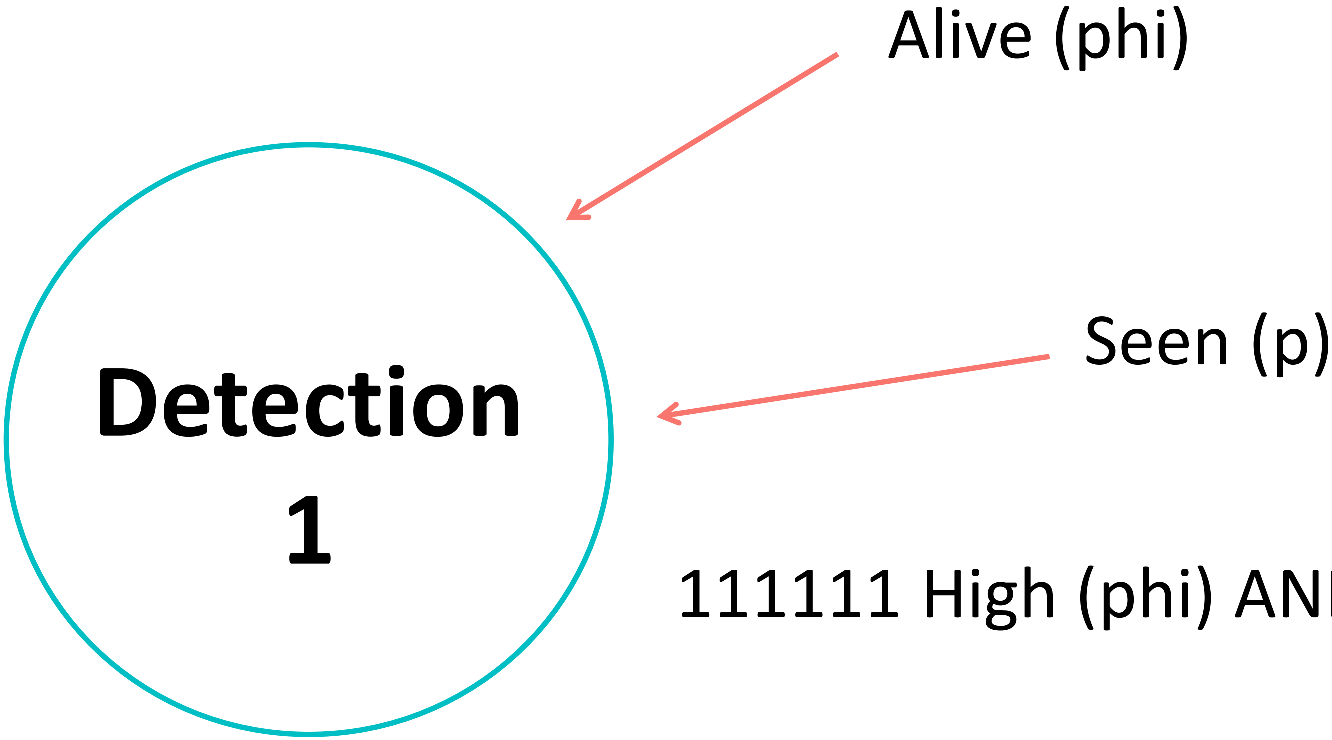
Figure 1: Relationship between survival and body mass in sociable weavers obtained by a mark-recapture analysis (*solid line*) and a naive analysis assuming perfect detection (*dashed line*). Filled circles represent body mass values, and shaded areas represent 95% confidence intervals.

Detection 1


Alive
Detection 1


Alive
Seen
Detection
1


Alive (phi)
Seen (p)
Detection
1



111111 High (phi) AND High (p)

Alive (phi)

Detection 1

Seen (p)

111111 High (phi) AND High (p)
101010 High (phi) AND Med (p)

Alive (phi)

Seen (p)

Detection 1

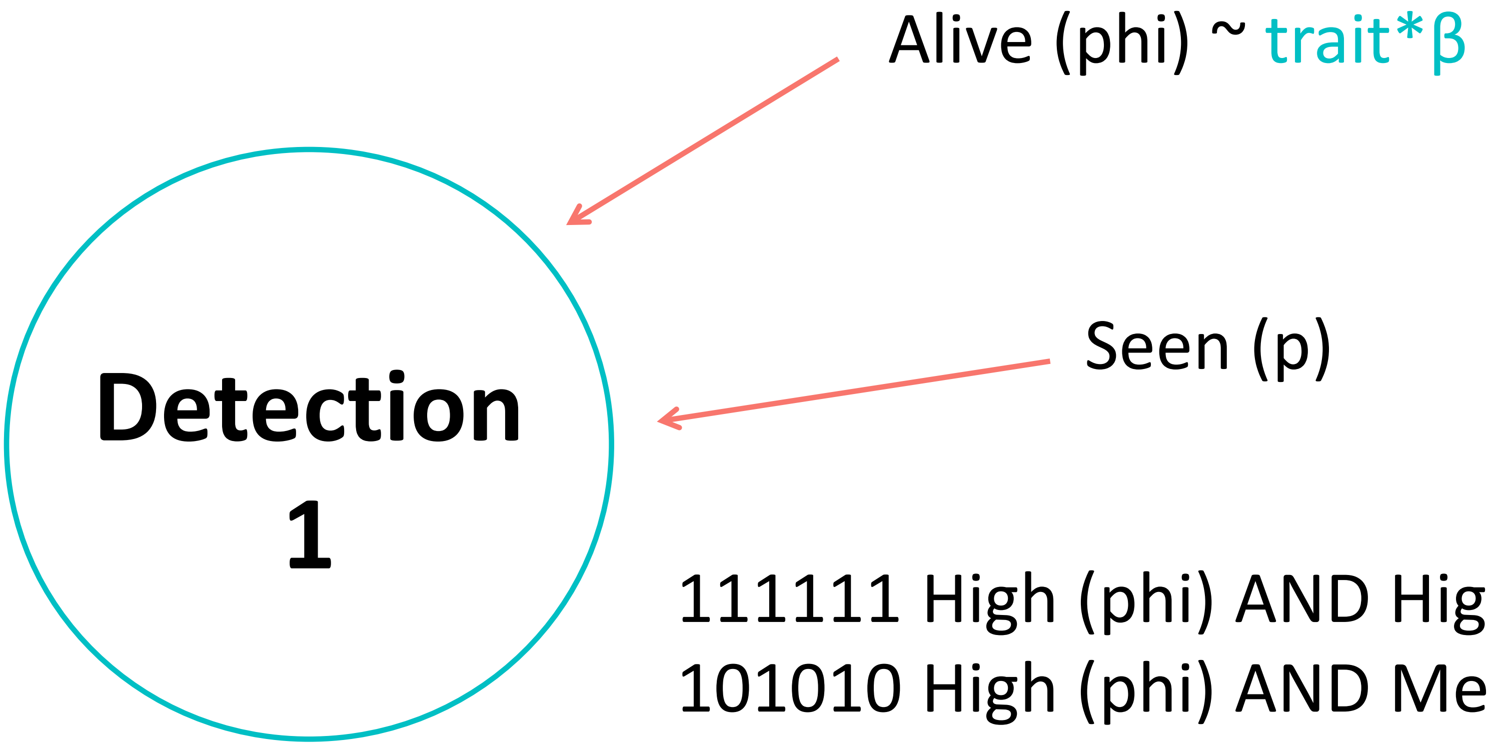
111111 High (phi) AND High (p)
101010 High (phi) AND Med (p)
100000 Low (phi) OR Low (p)

Alive (phi) ~ **trait**

Seen (p)

Detection 1

111111 High (phi) AND High (p)
101010 High (phi) AND Med (p)
100000 Low (phi) OR Low (p)



111111 High (phi) AND High (p)
101010 High (phi) AND Med (p)
100000 Low (phi) OR Low (p)

Alive (phi) ~ $trait*\beta + trait^2*\gamma$

Detection
1

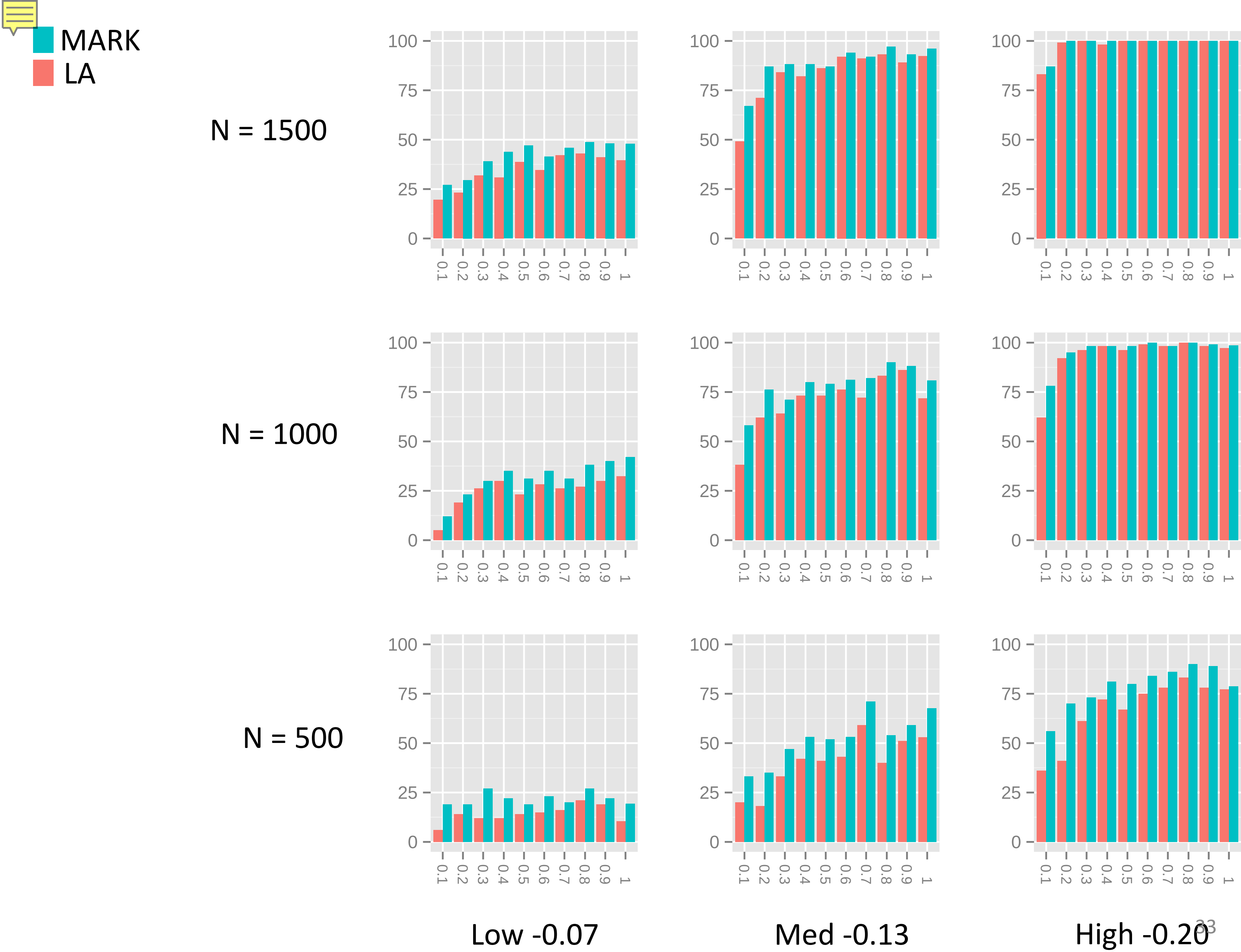
Seen (p)

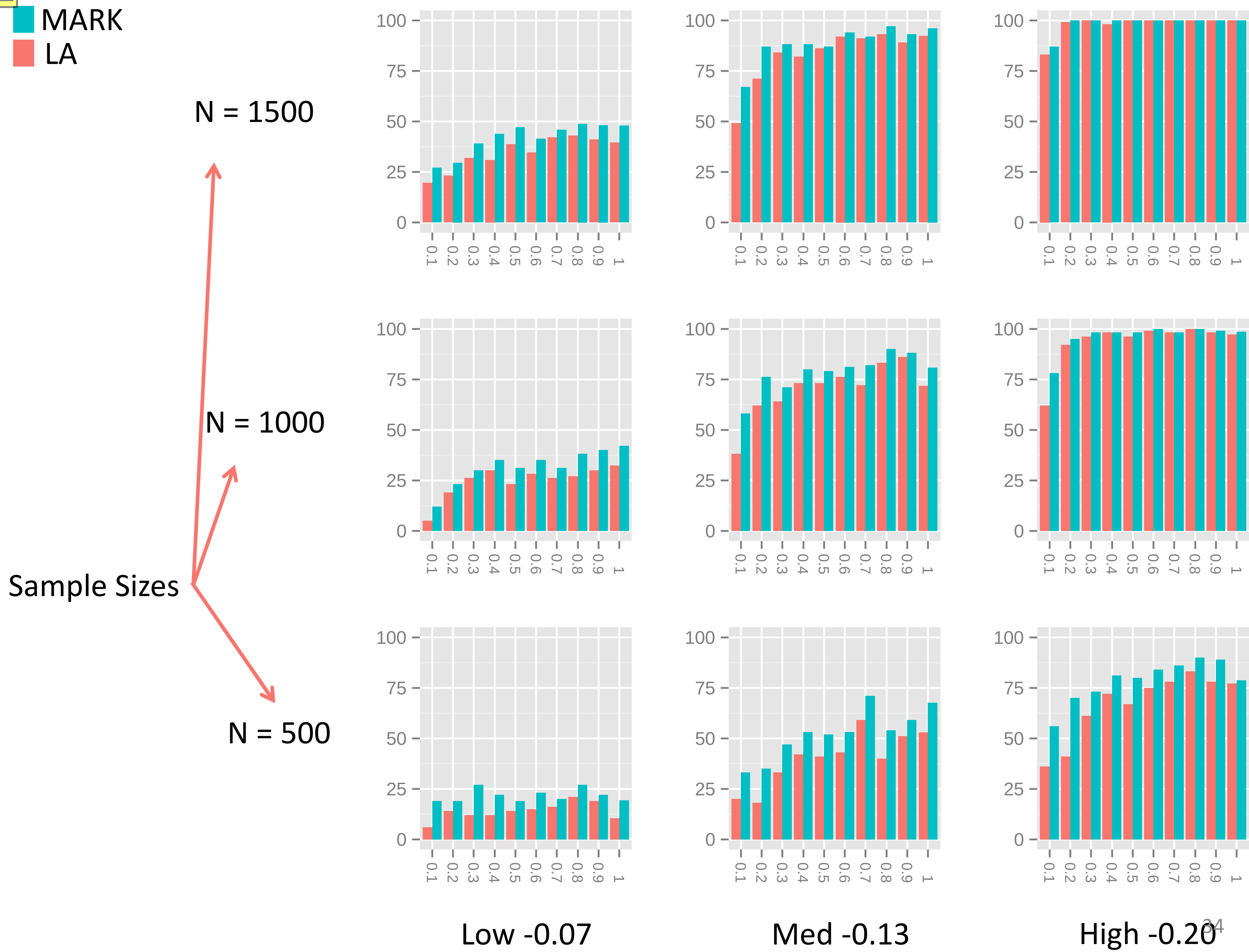
111111 High (phi) AND High (p)

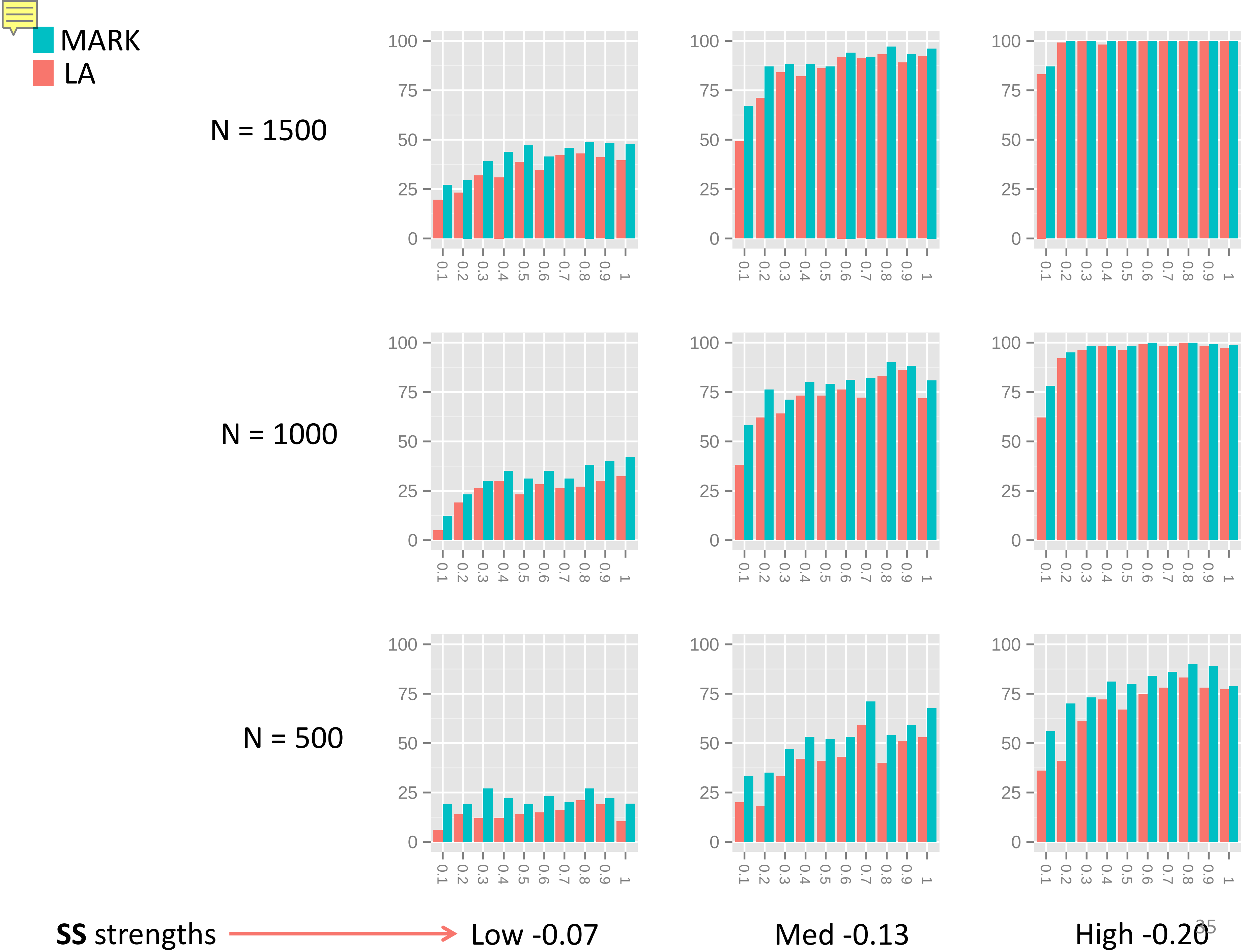
101010 High (phi) AND Med (p)

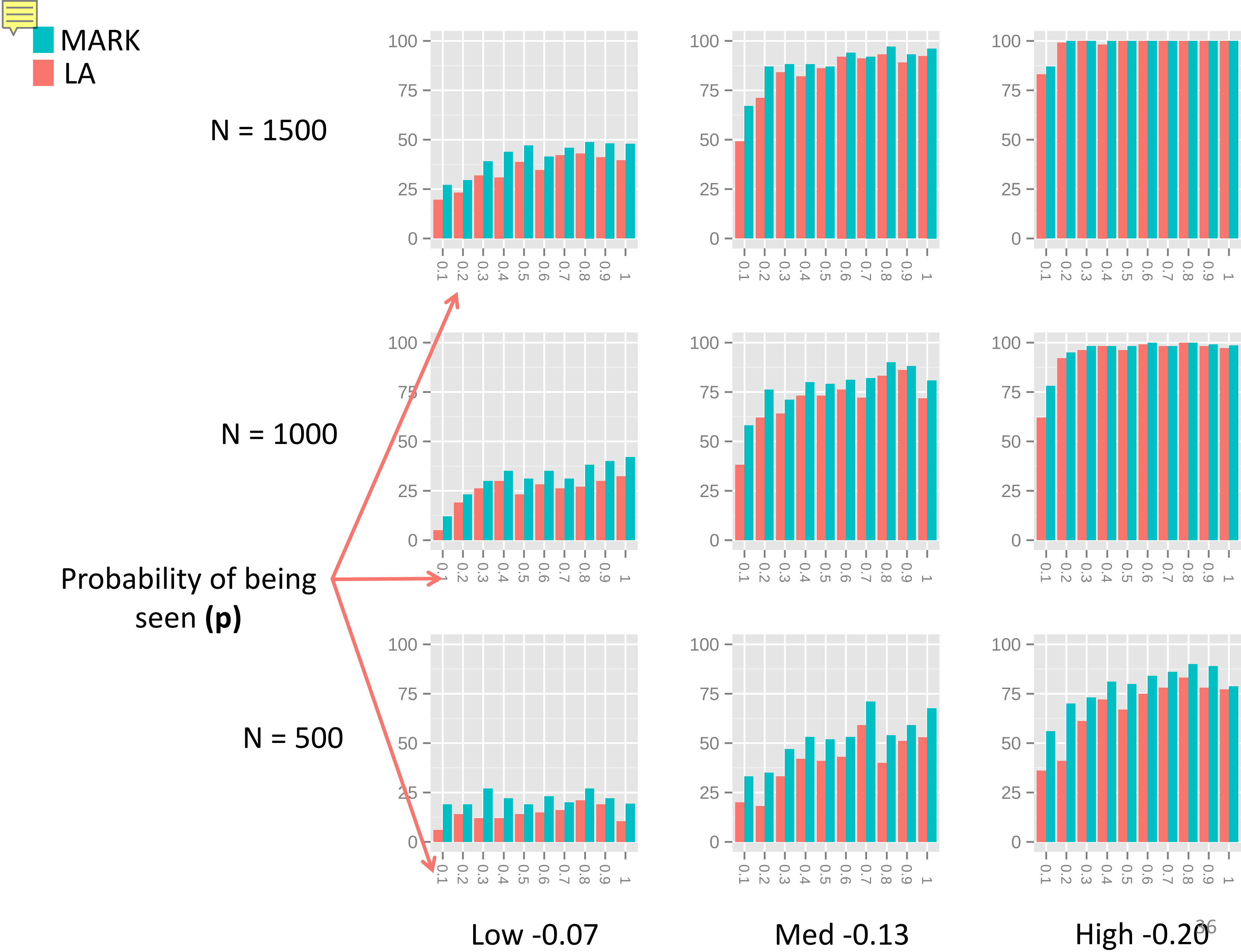
100000 Low (phi) OR Low (p)

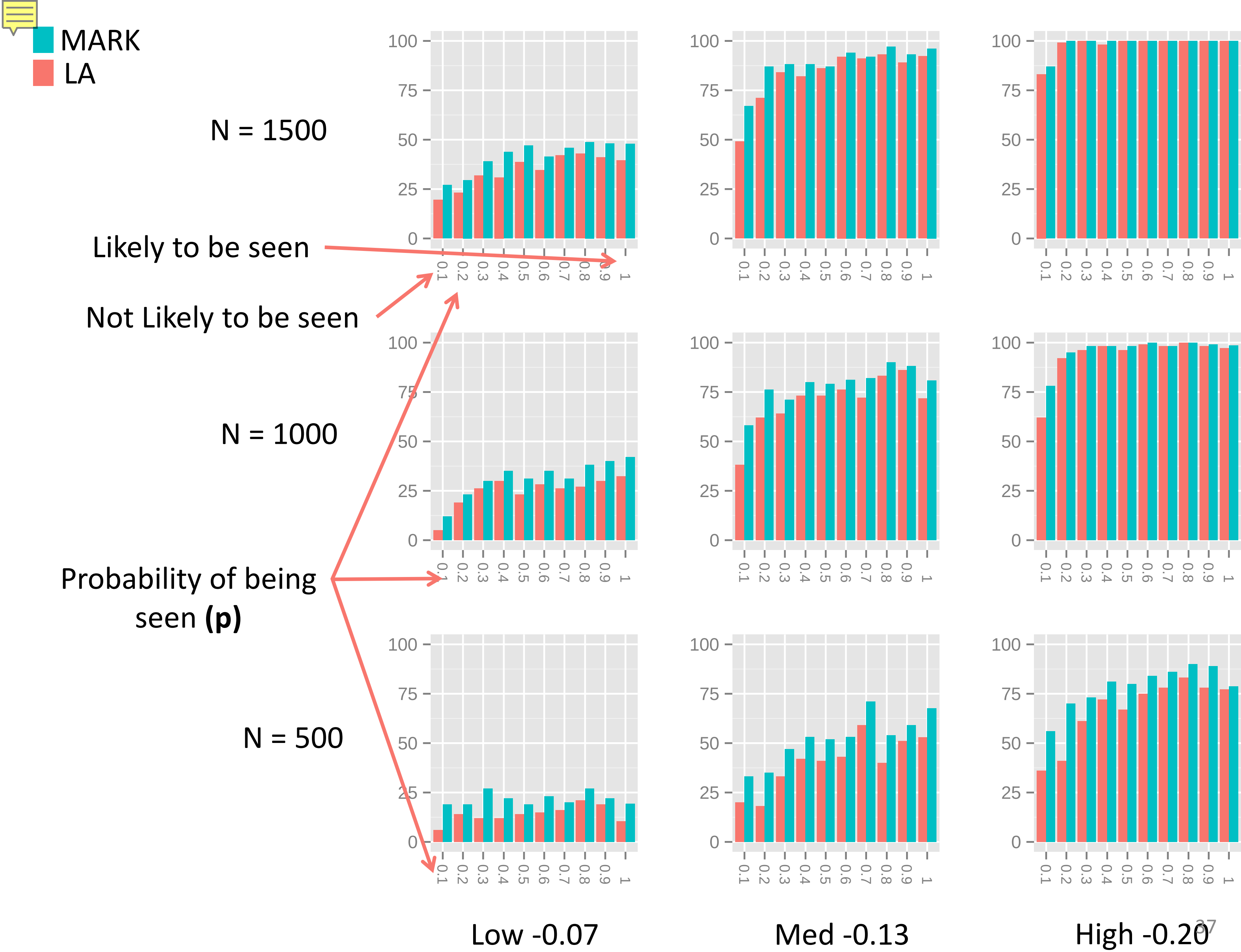
Computer is running....

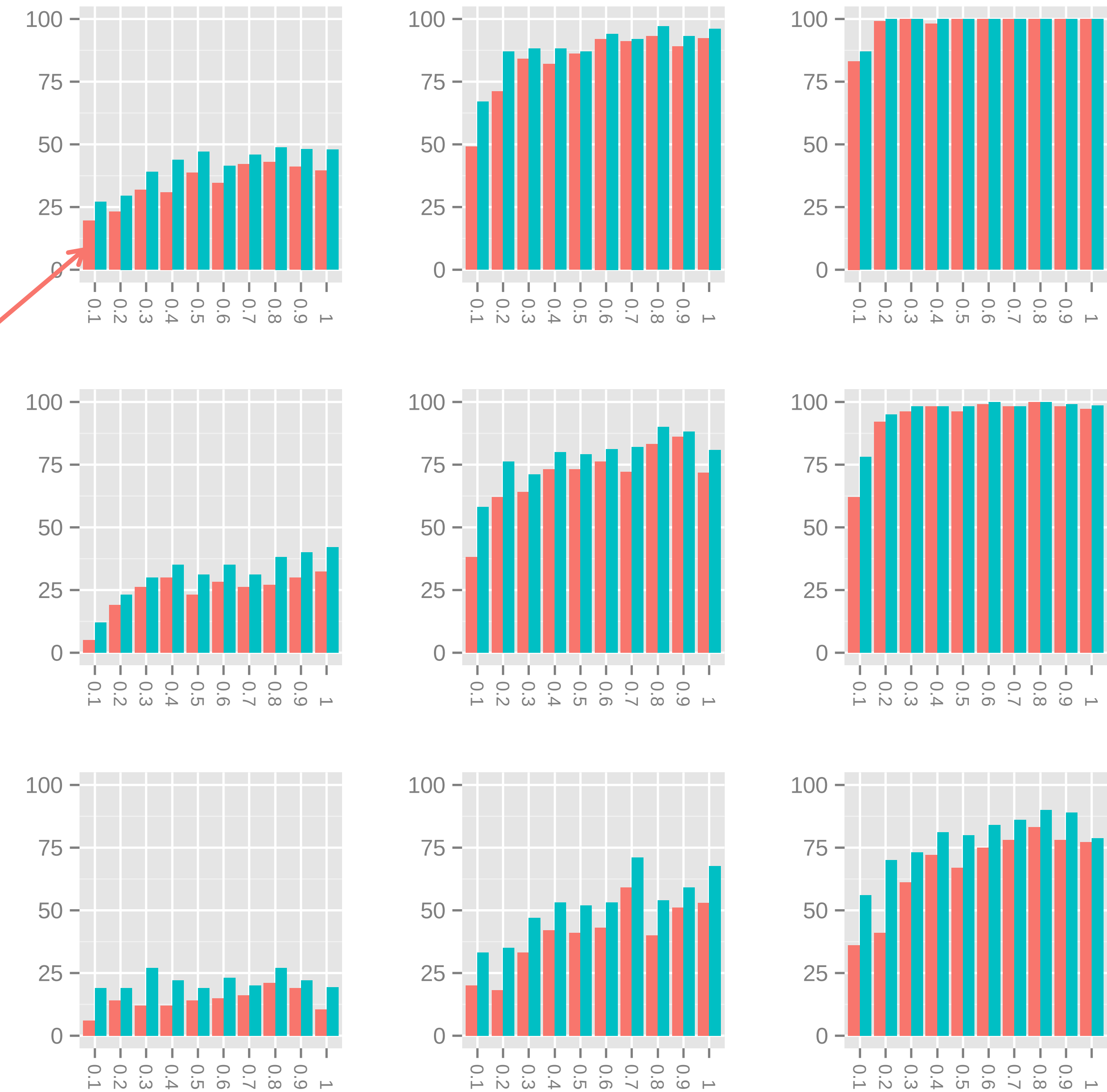

MARK
LA
N = 1500
N = 1000
N = 500
100
75
50
25
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Low -0.07
Med -0.13
High -0.20


MARK
LA
N = 1500
N = 1000
N = 500
Sample Sizes
Low -0.07
Med -0.13
High -0.20

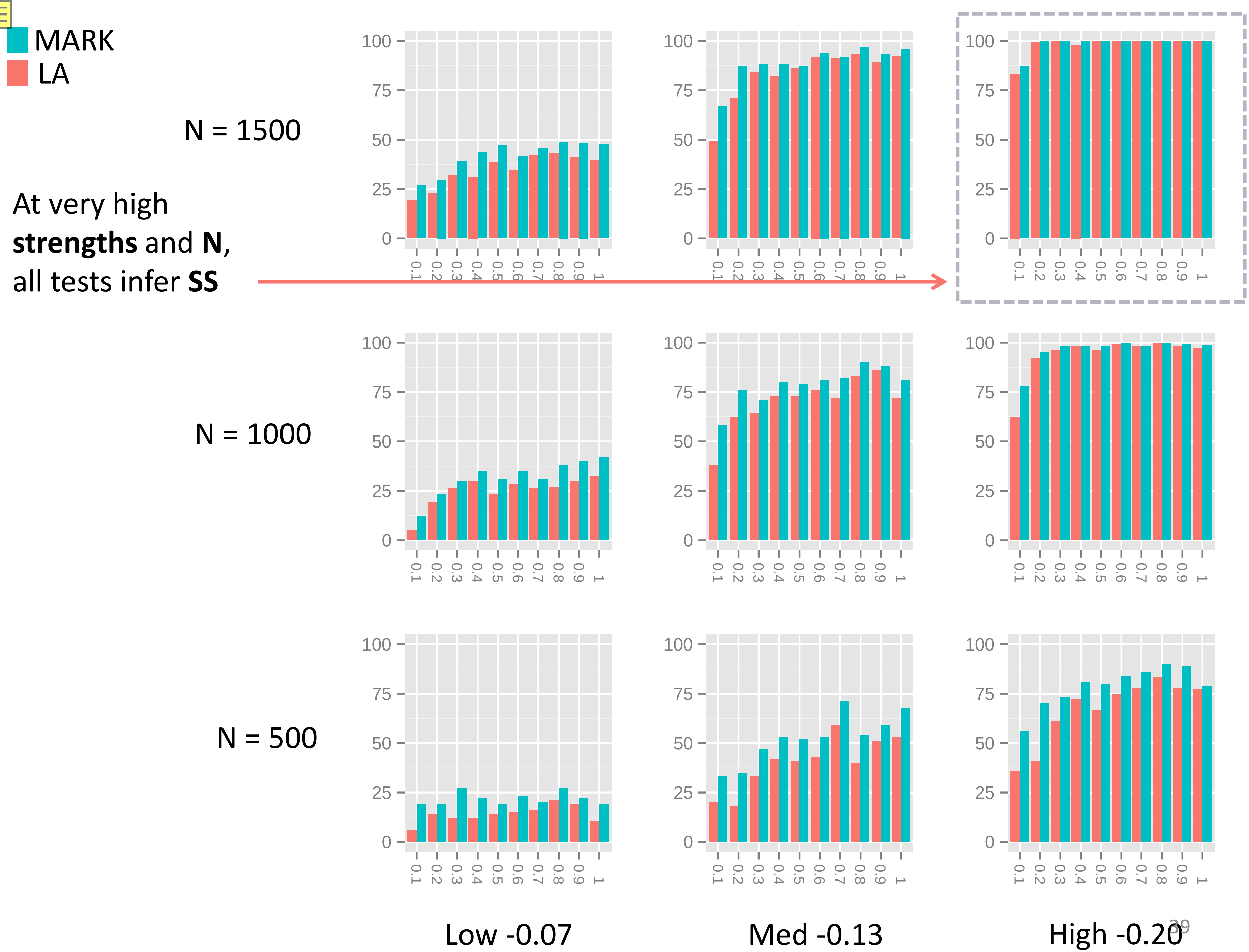

MARK
LA
N = 1500
N = 1000
N = 500
100
75
50
25
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
SS strengths
Low -0.07
Med -0.13
High -0.20


MARK
LA
N = 1500
N = 1000
N = 500
Probability of being seen (p)
Low -0.07
Med -0.13
High -0.20


MARK
LA
N = 1500
N = 1000
N = 500
Likely to be seen
Not Likely to be seen
Probability of being seen **(p)**
Low -0.07
Med -0.13
High -0.20



Each bar tells us the **percentage** of times **SS** was inferred out of 1000 simulations



MARK
LA
N = 1500
N = 1000
N = 500
At very high **strengths** and **N**, all tests infer **SS**
Low -0.07
Med -0.13
High -0.20

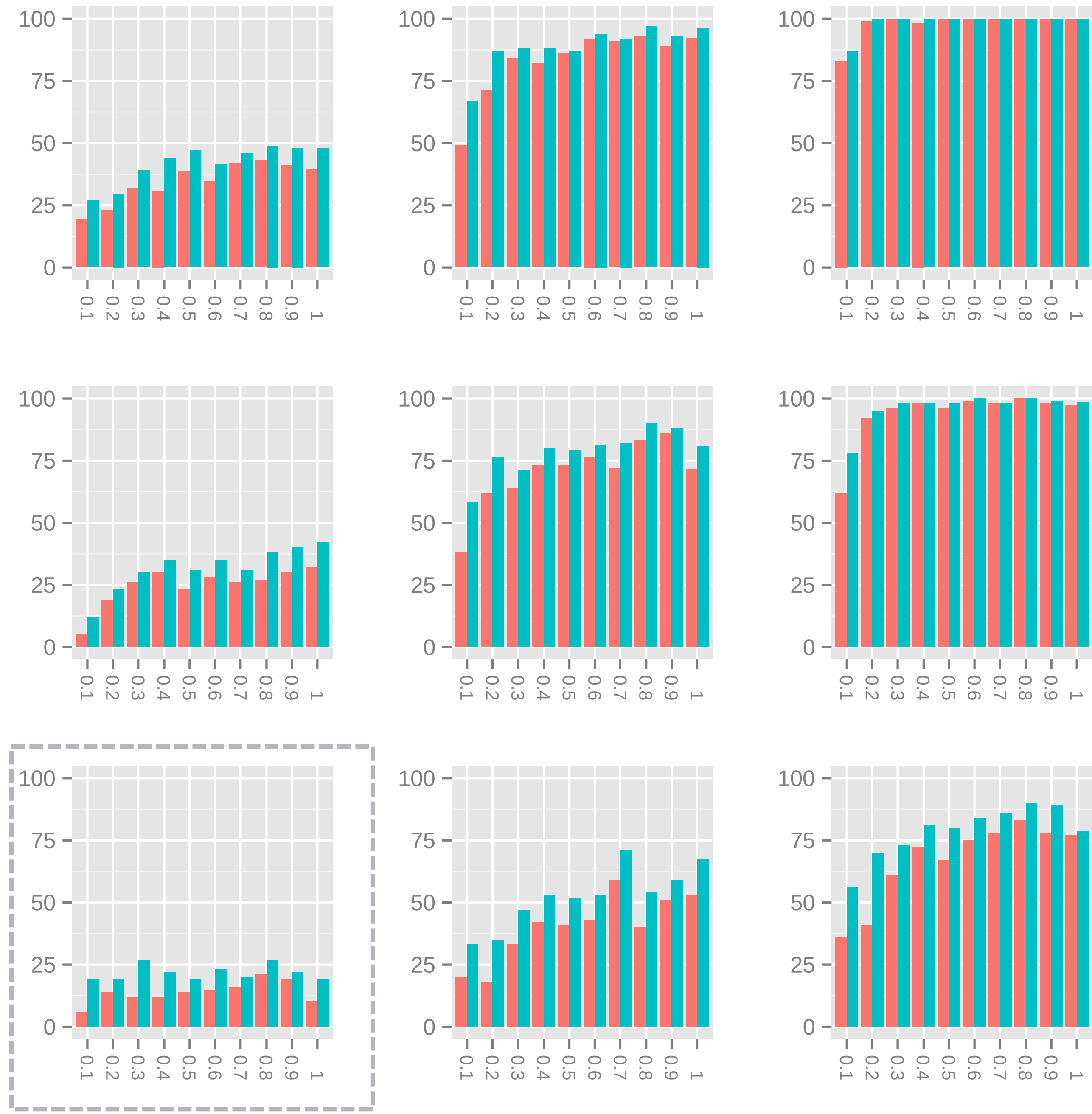
MARK
LA

N = 1500

At very low **strengths** and **N** no tests very good at infering **SS**

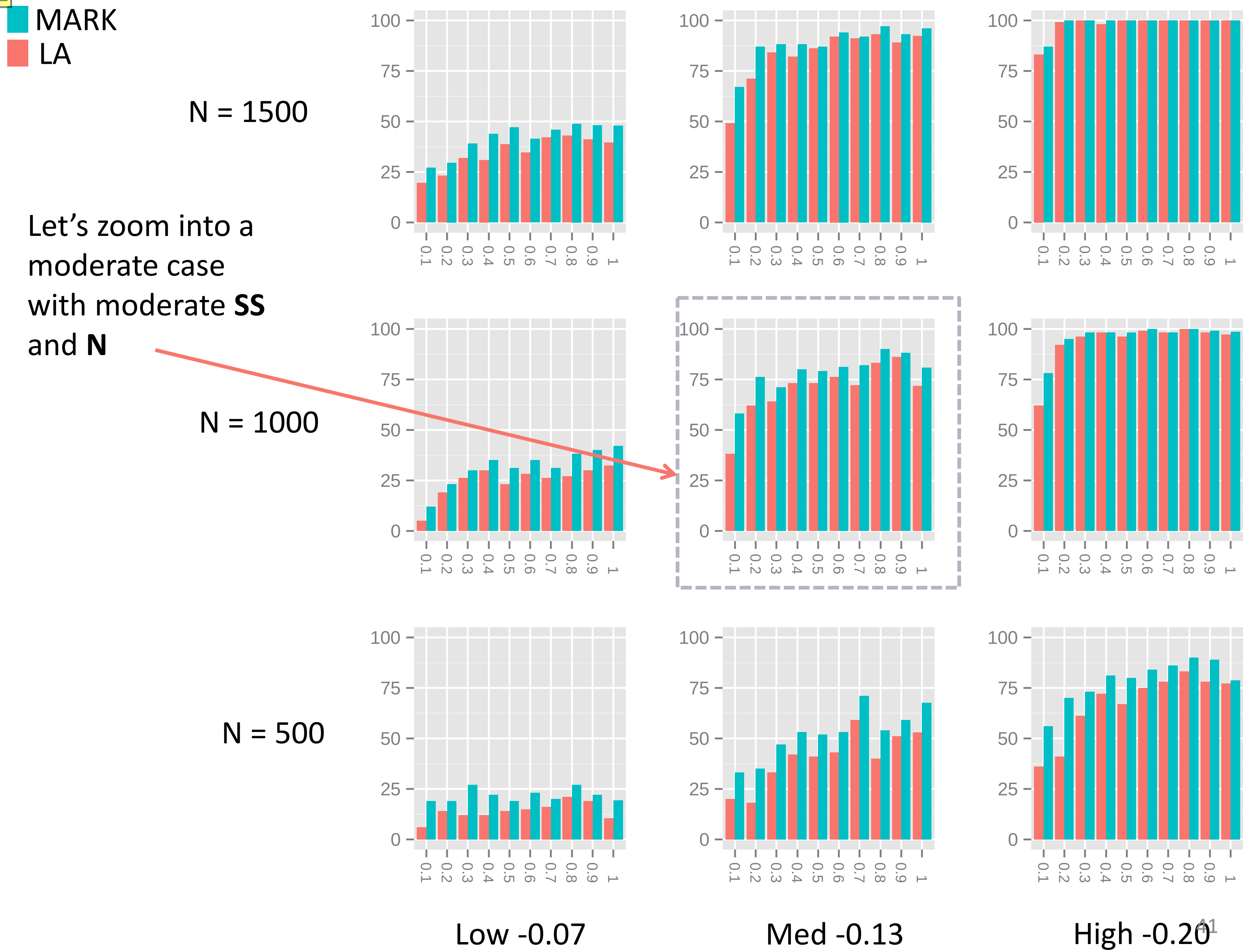
N = 1000

N = 500

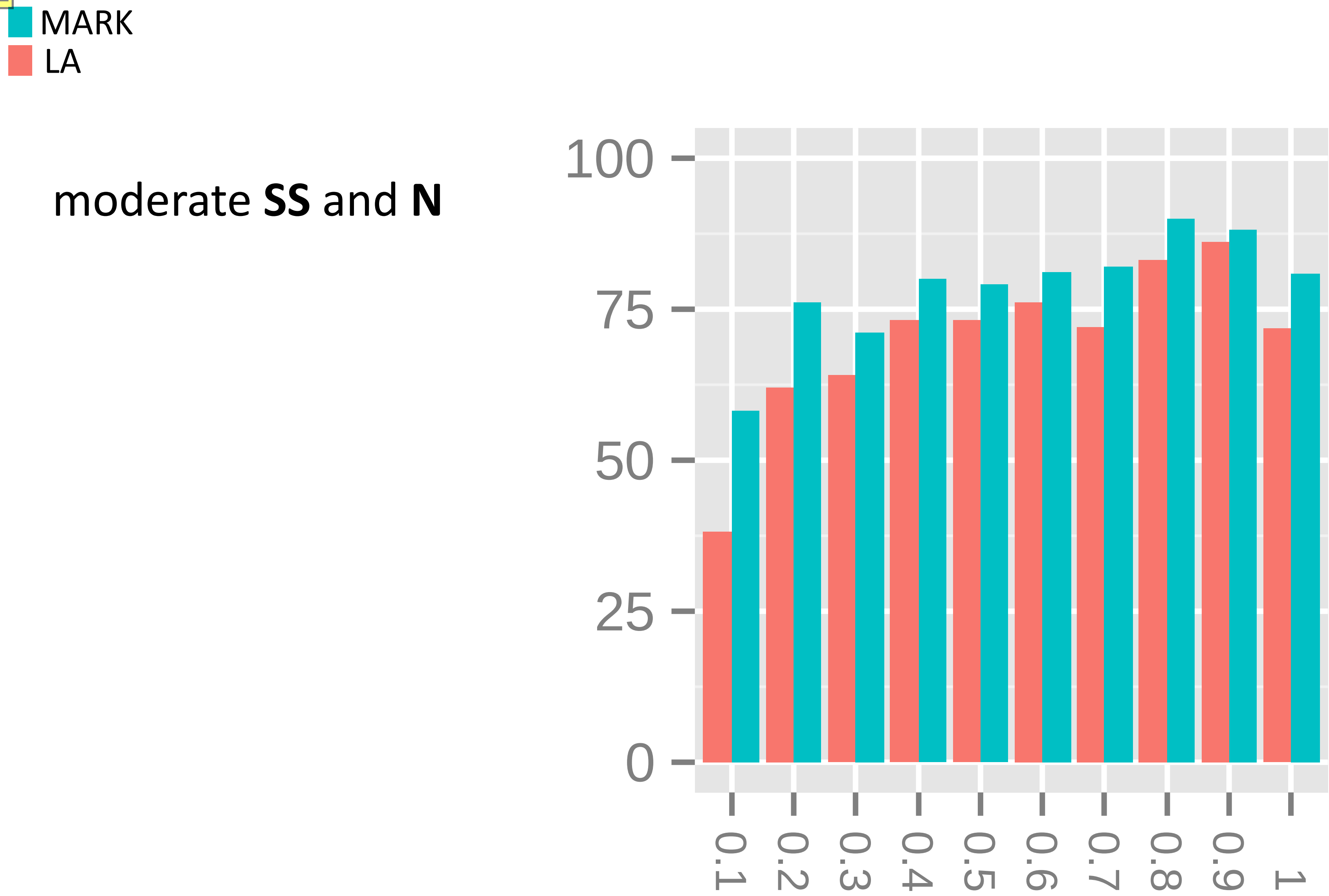


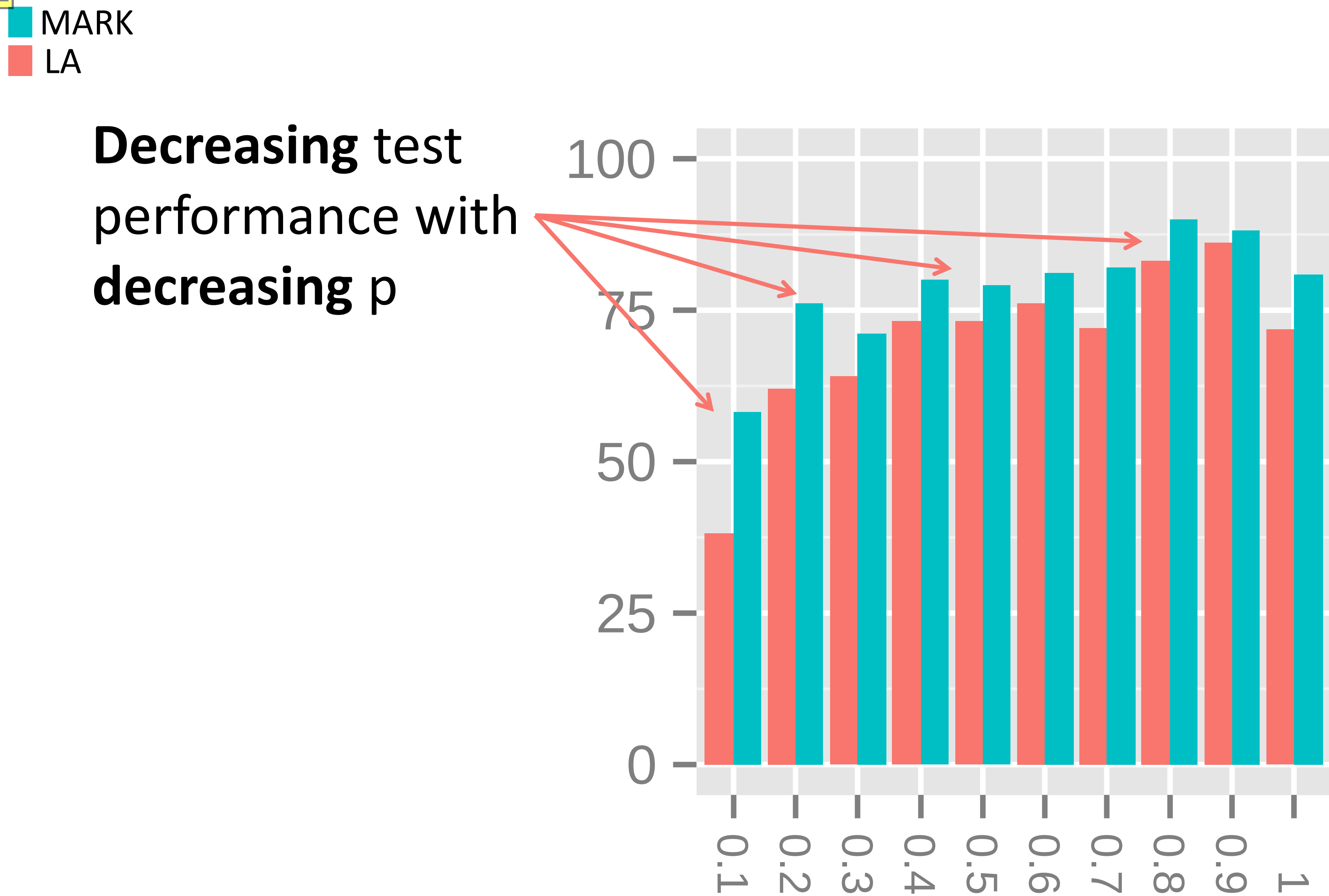
Low -0.07 Med -0.13 High -0.20

Let’s zoom into a moderate case with moderate **SS** and **N**

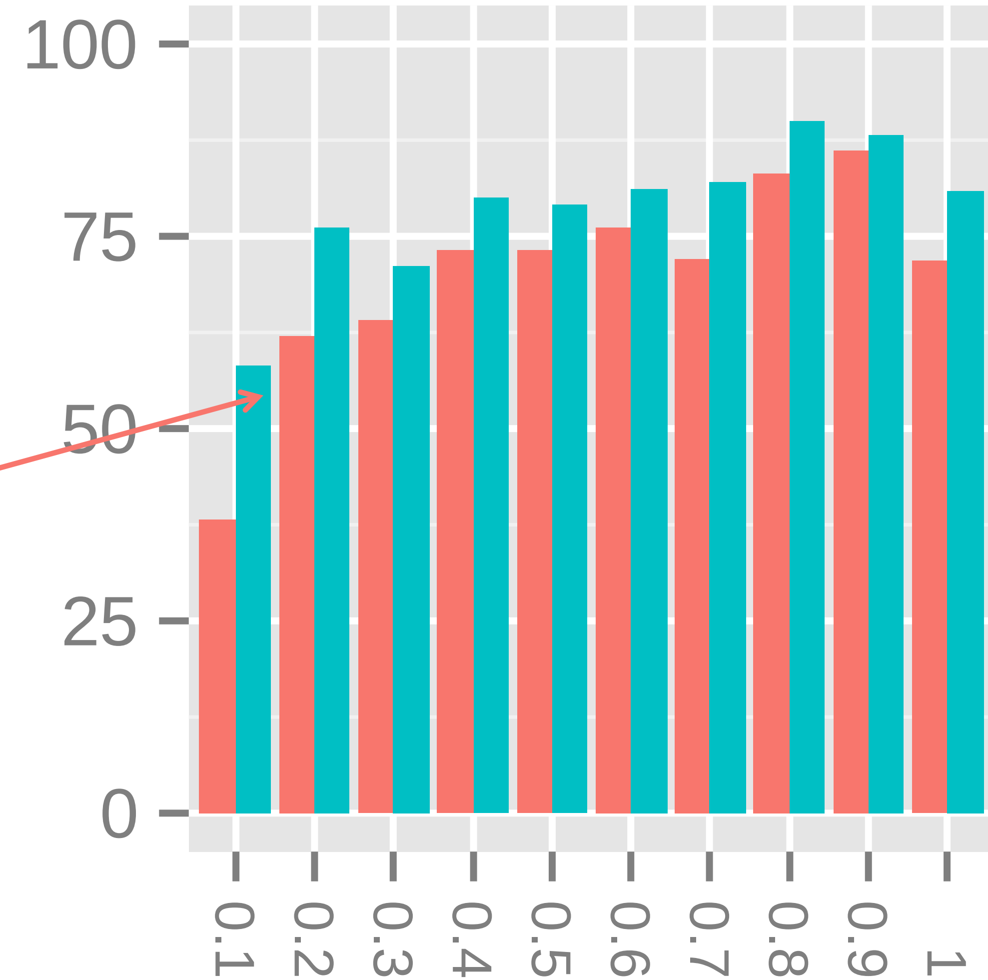


moderate **SS** and **N**

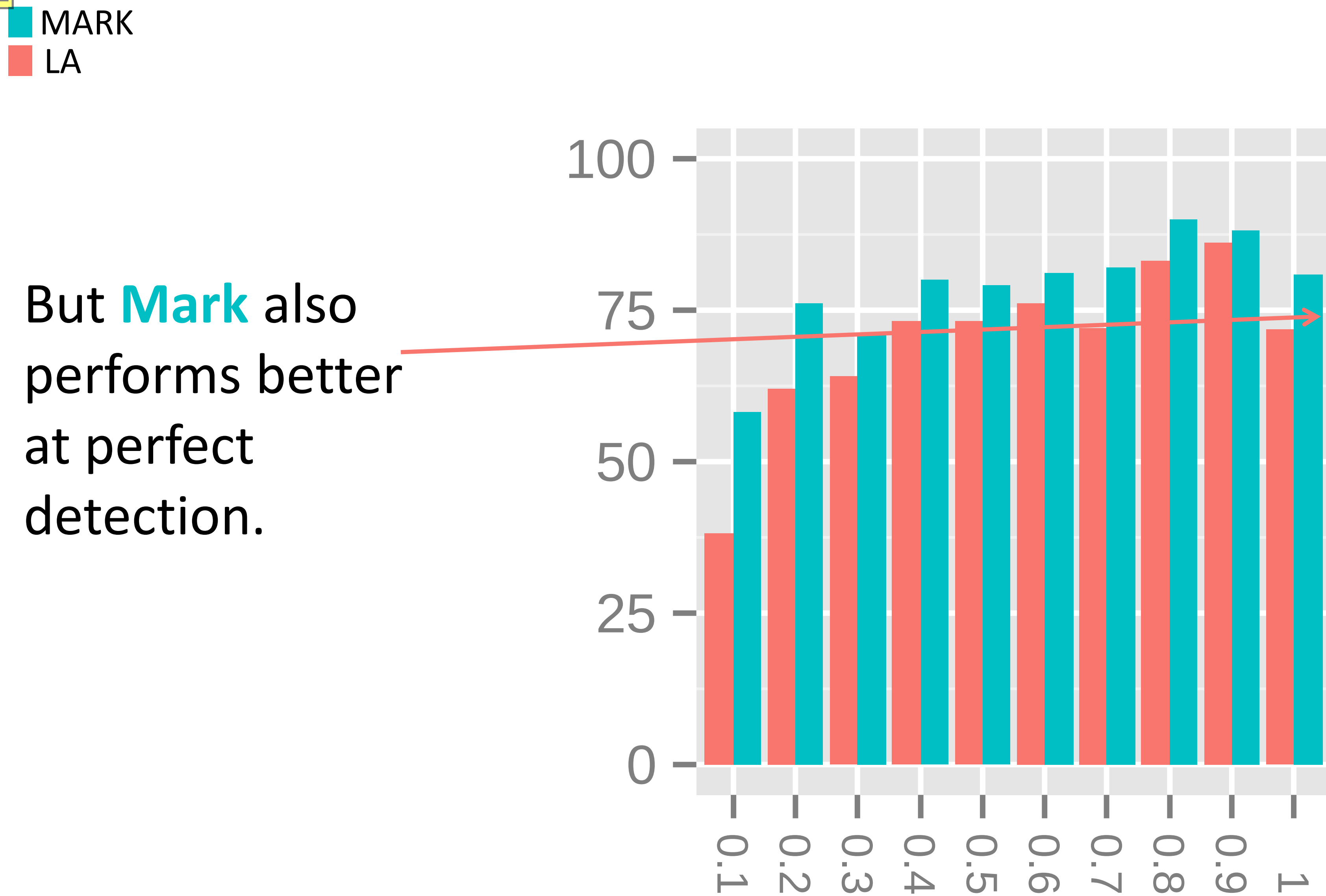




Decreasing test performance with **decreasing** p

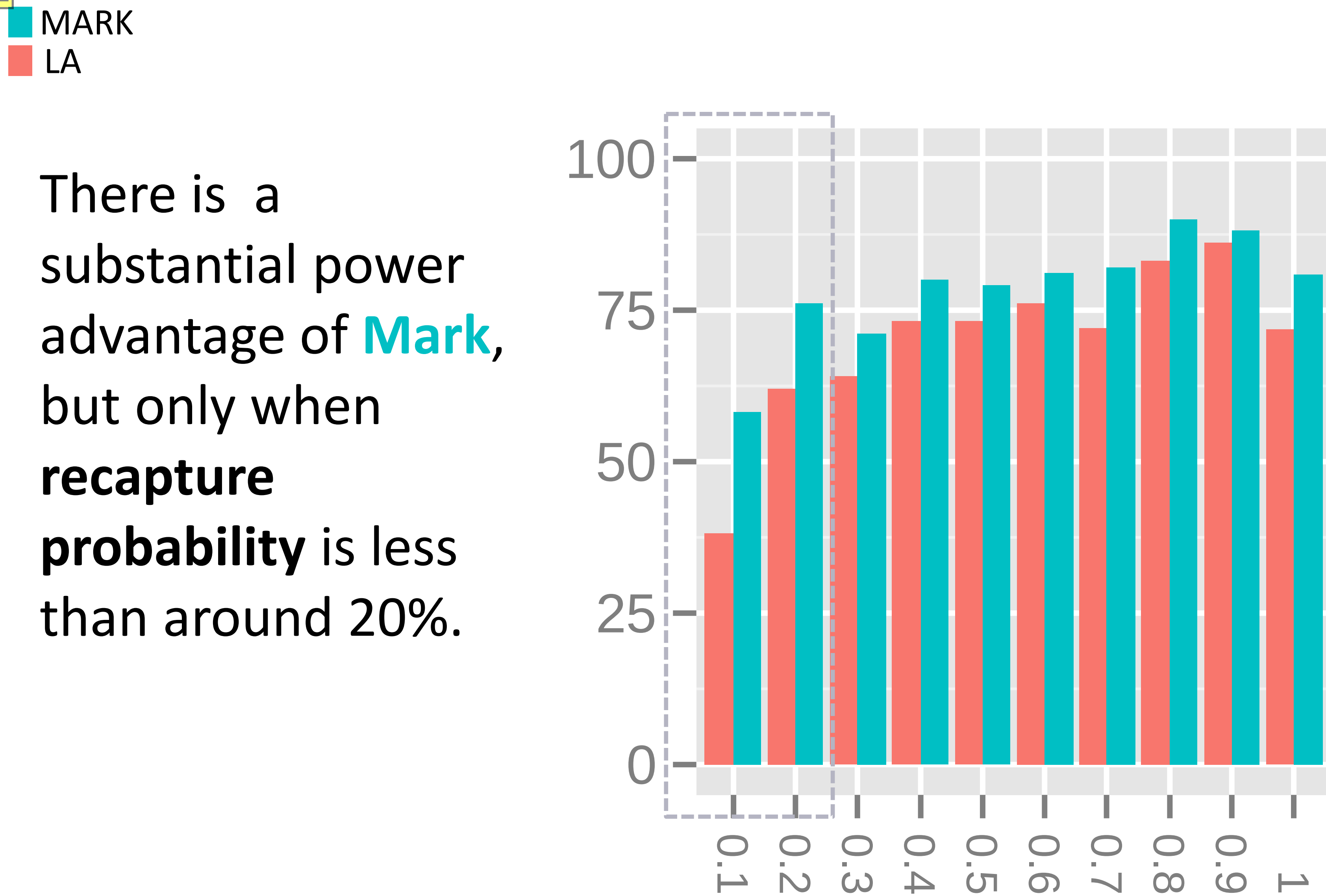


Mark performs better than **LA** at low p



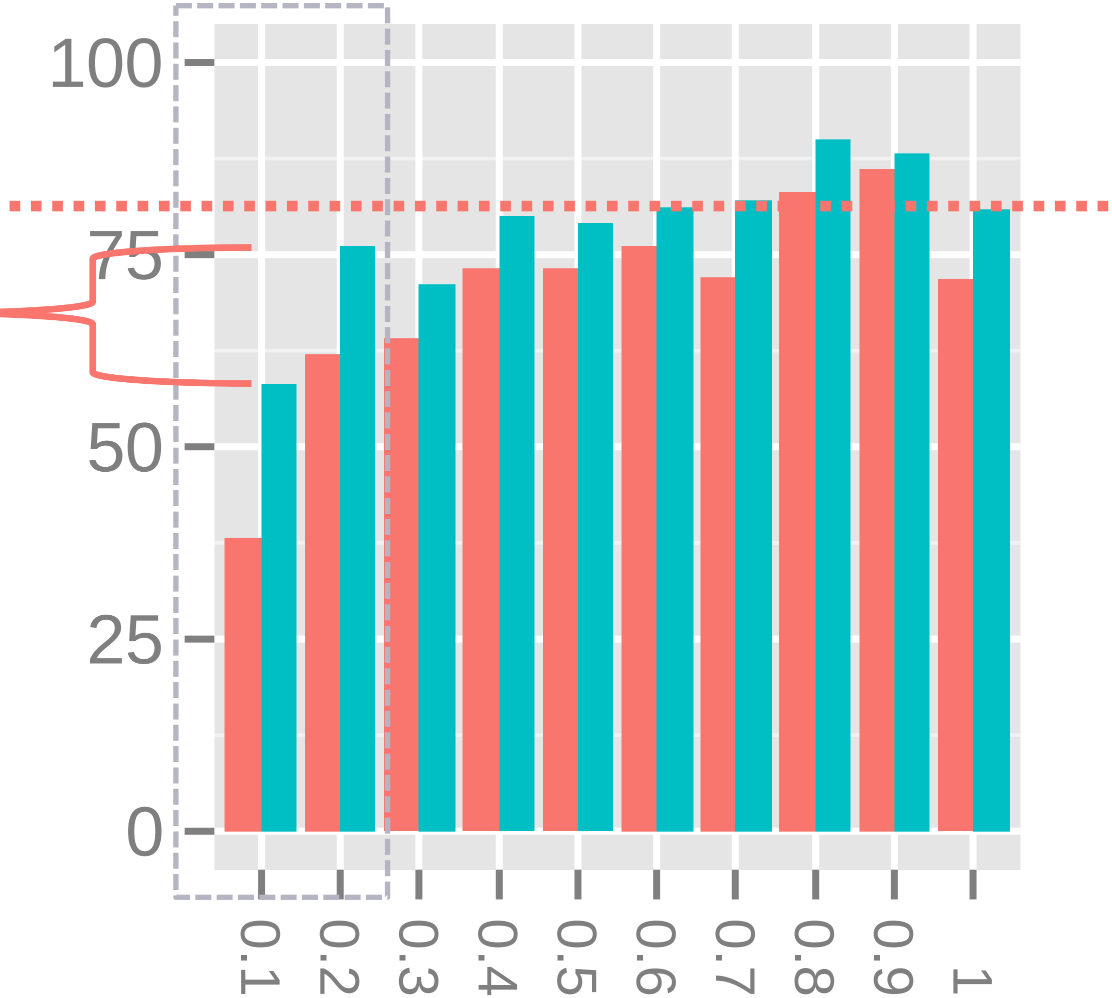
But **Mark** also performs better at perfect detection.

There is a substantial power advantage of **Mark**, but only when **recapture probability** is less than around 20%.

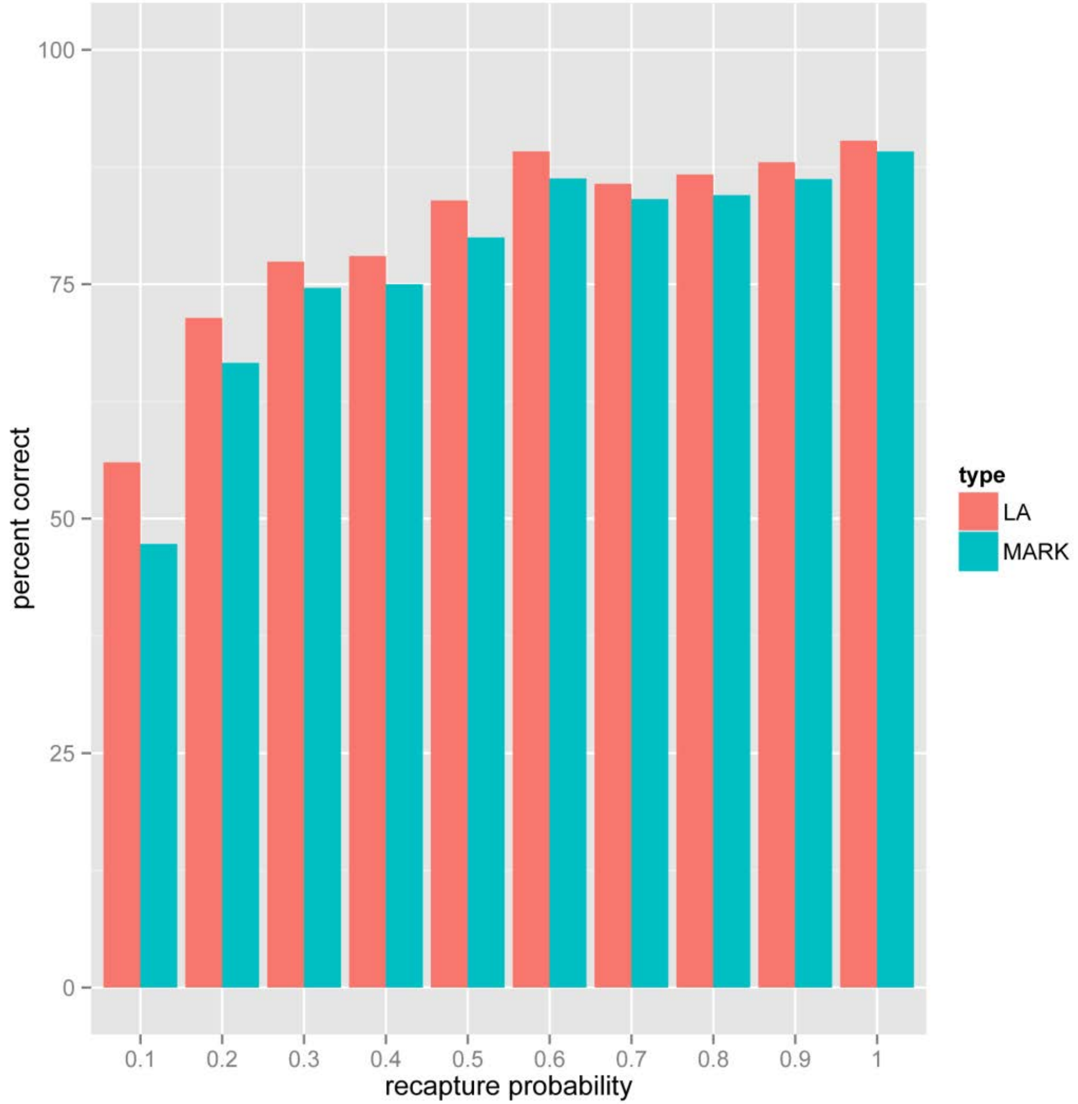


MARK
LA

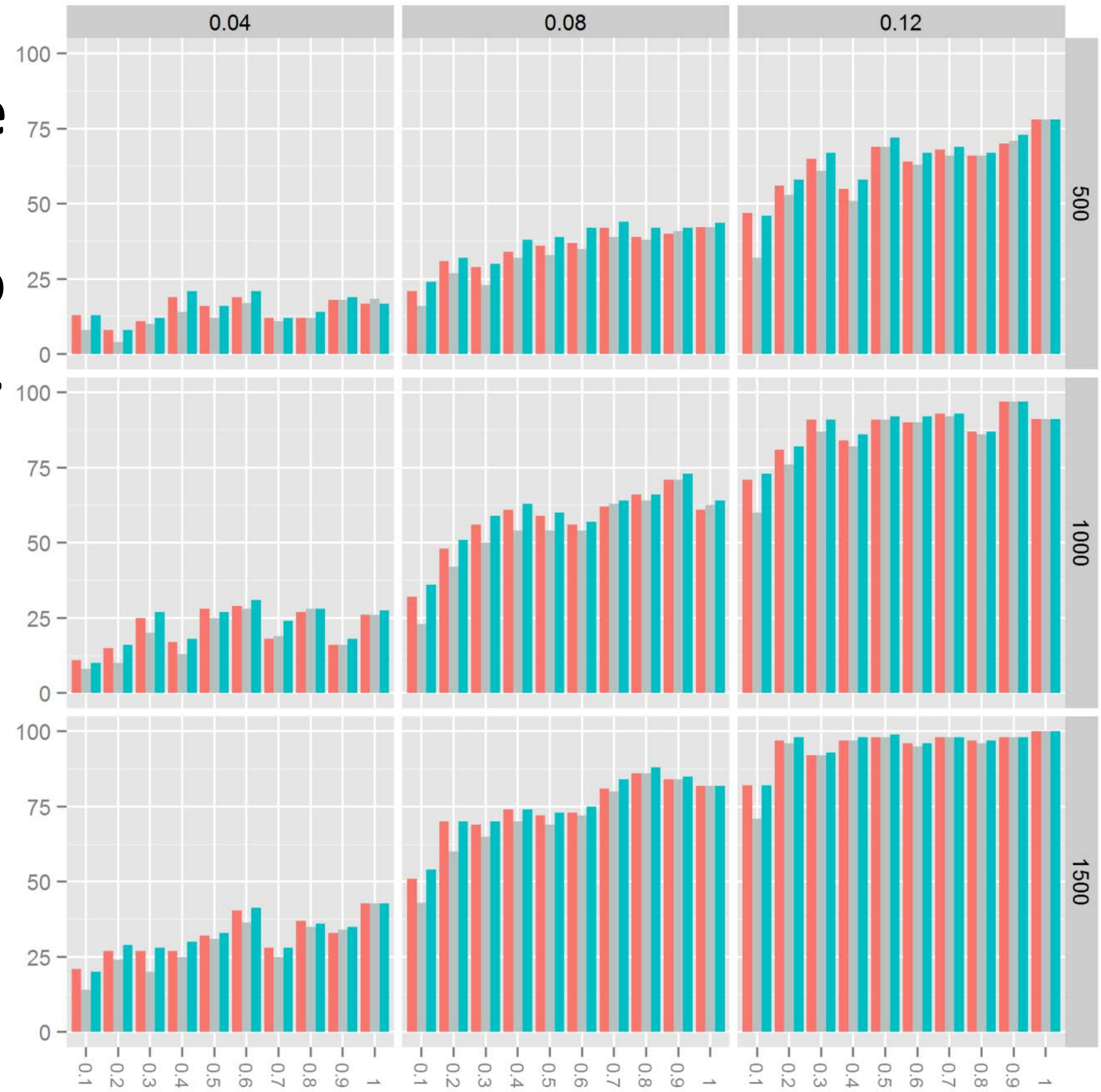
Decreasing **recapture probability** reduces the power of both statistical tests, including **Mark**.

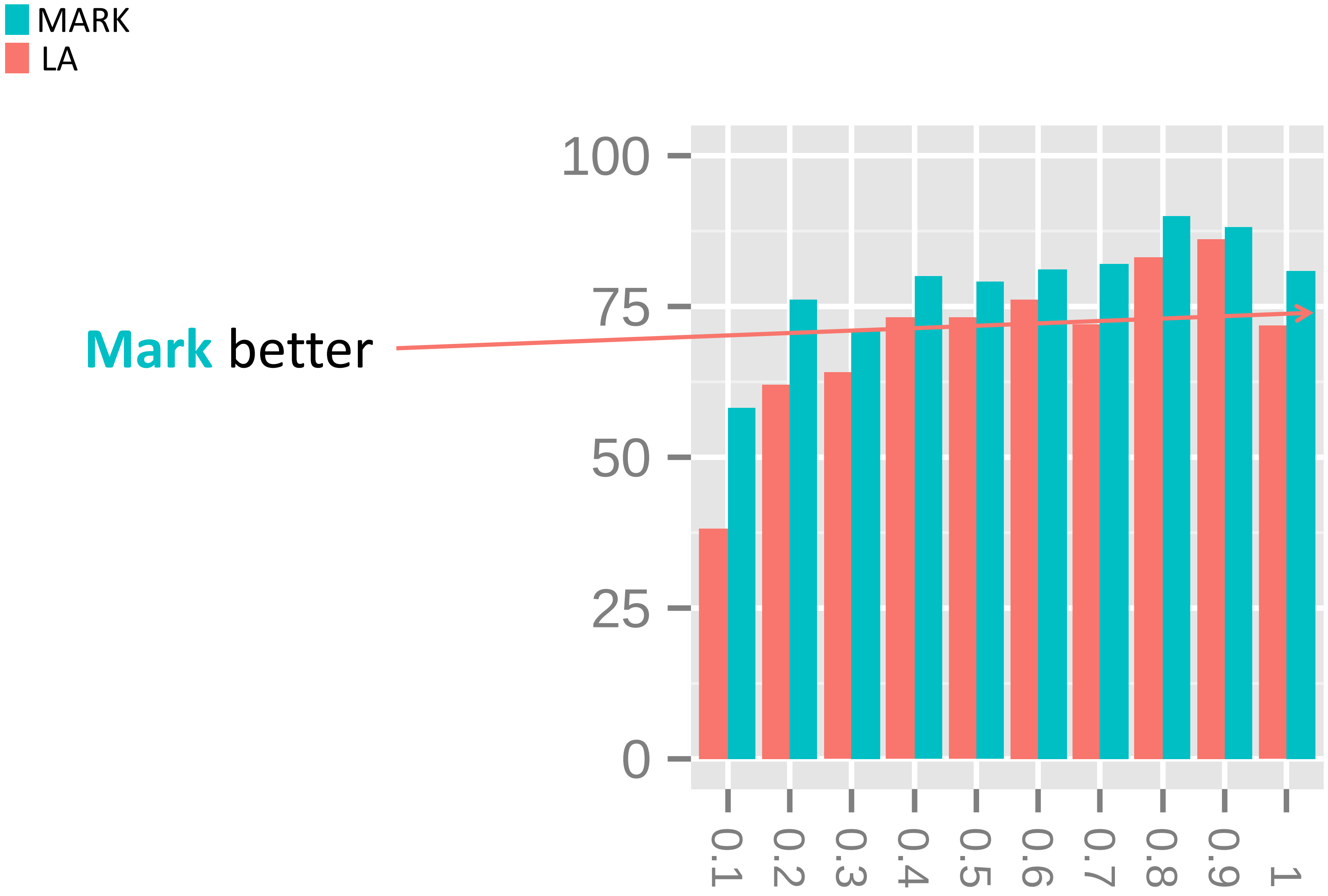


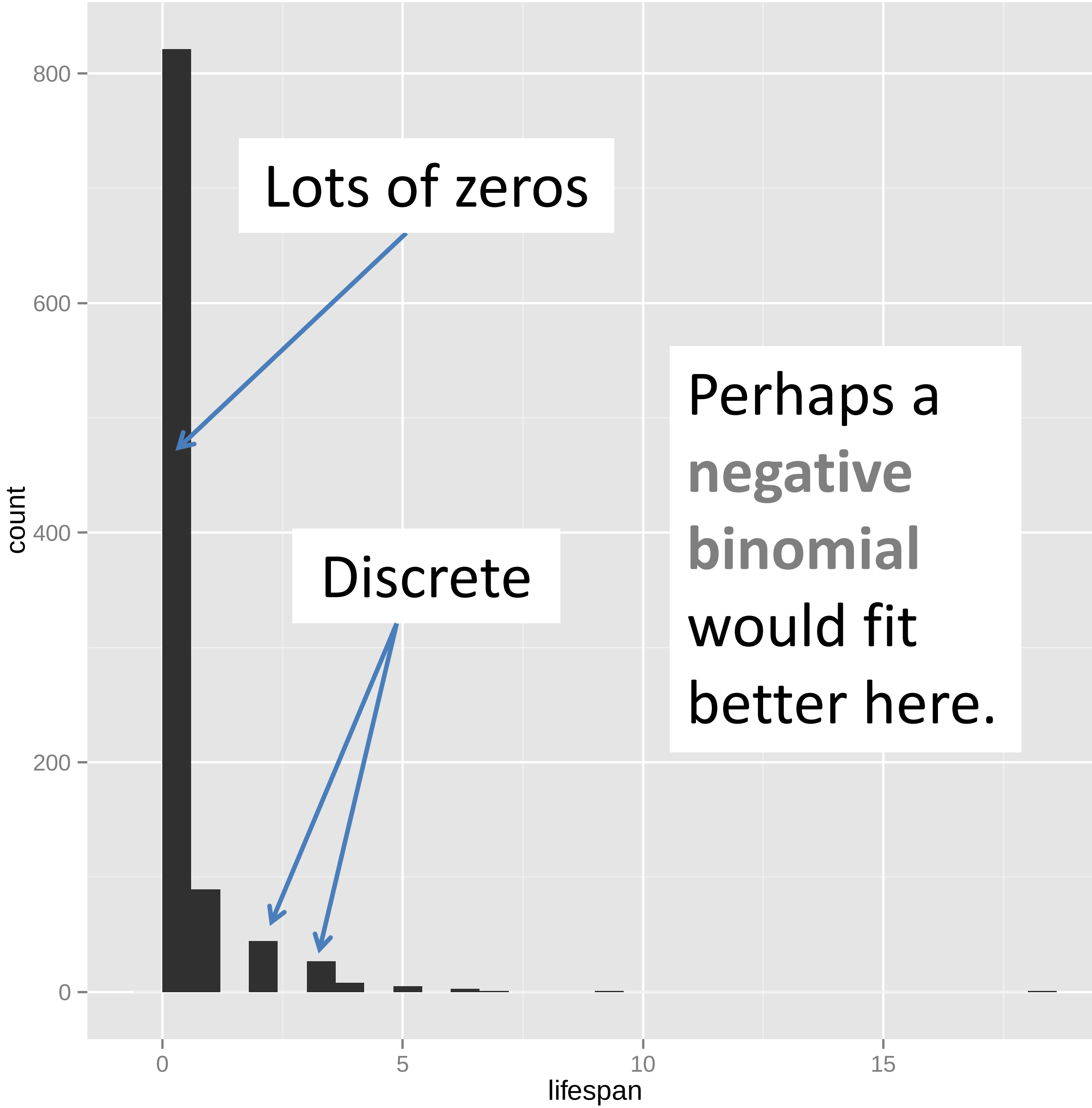
Disruptive Selection

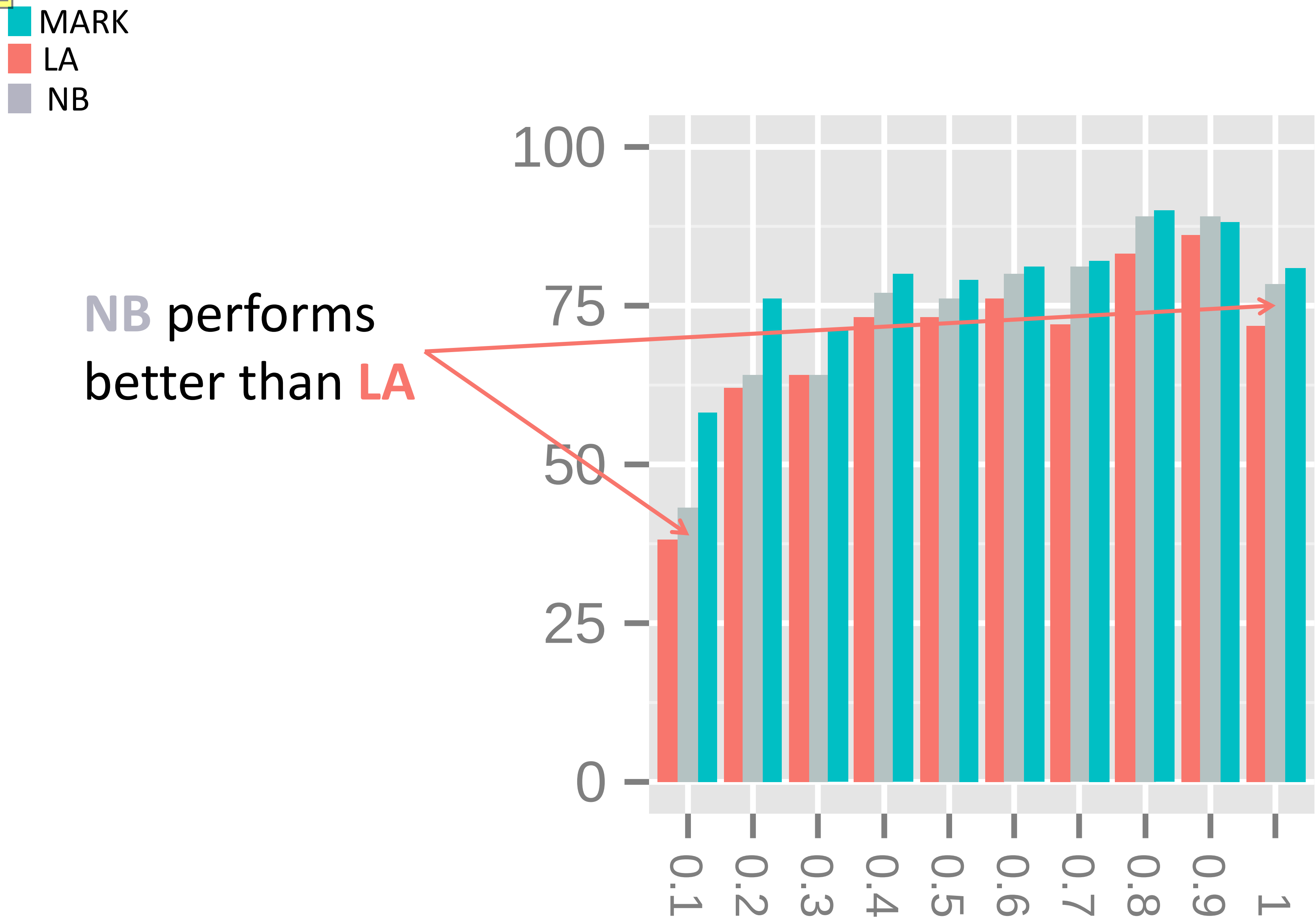


Decreasing **recapture probability** also reduces the power to infer **linear selection**.

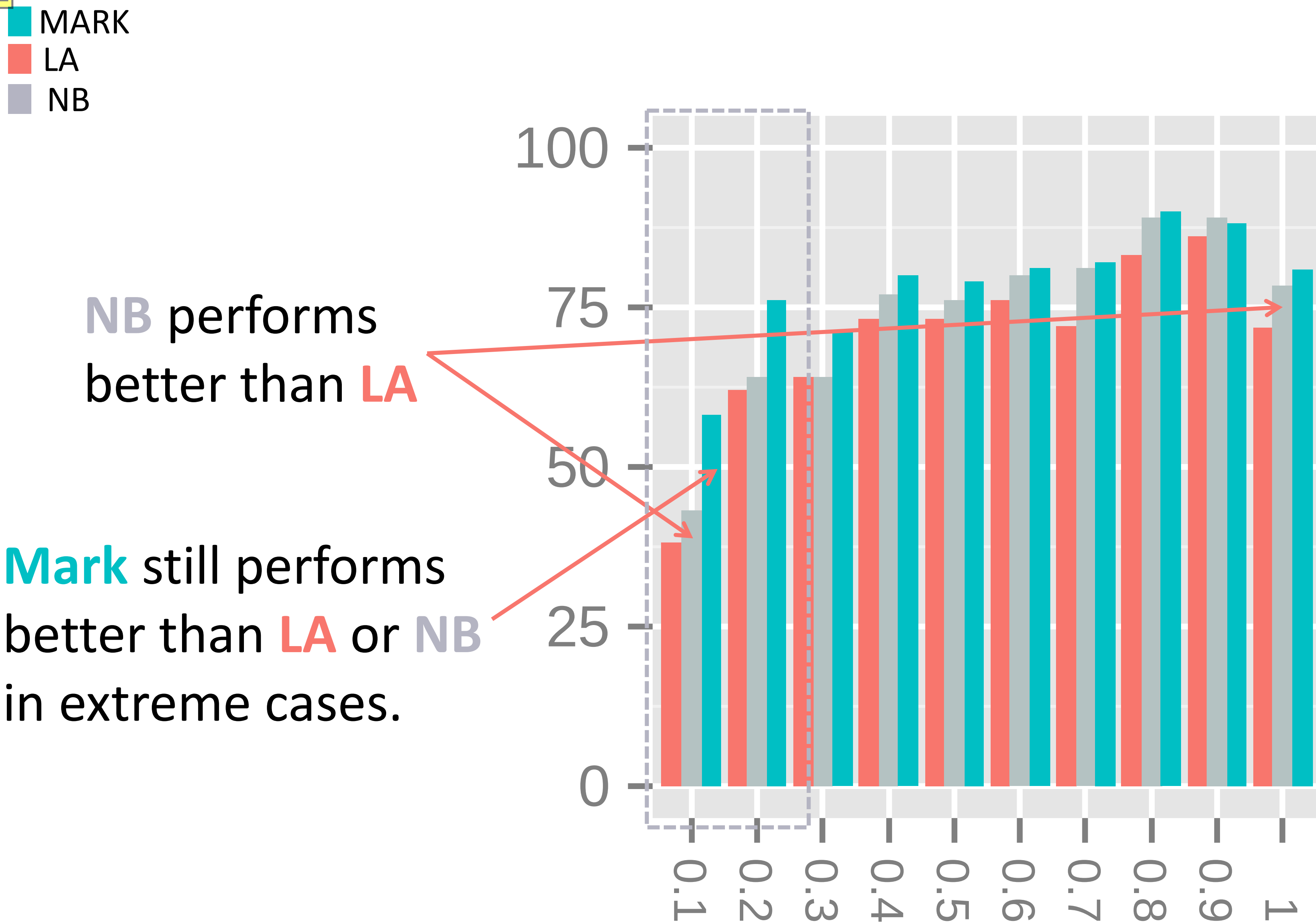



MARK
LA
Mark better
100
75
50
25
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


Lots of zeros
Discrete
Perhaps a **negative binomial** would fit better here.
800
600
400
200
0
count
0
5
10
15
lifespan



NB performs better than **LA**



NB performs better than LA

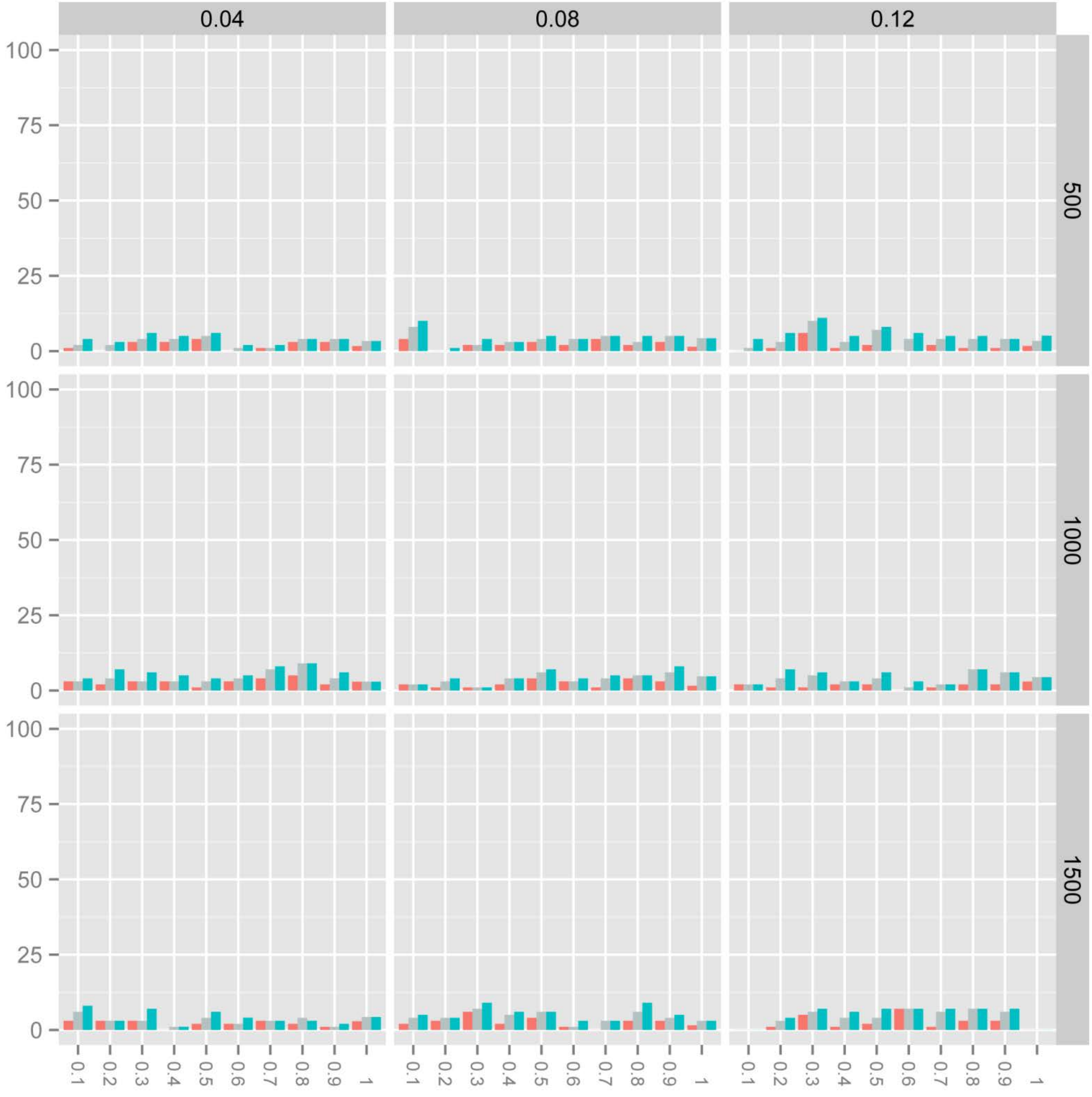
Mark still performs better than LA or NB in extreme cases.

Conclusions

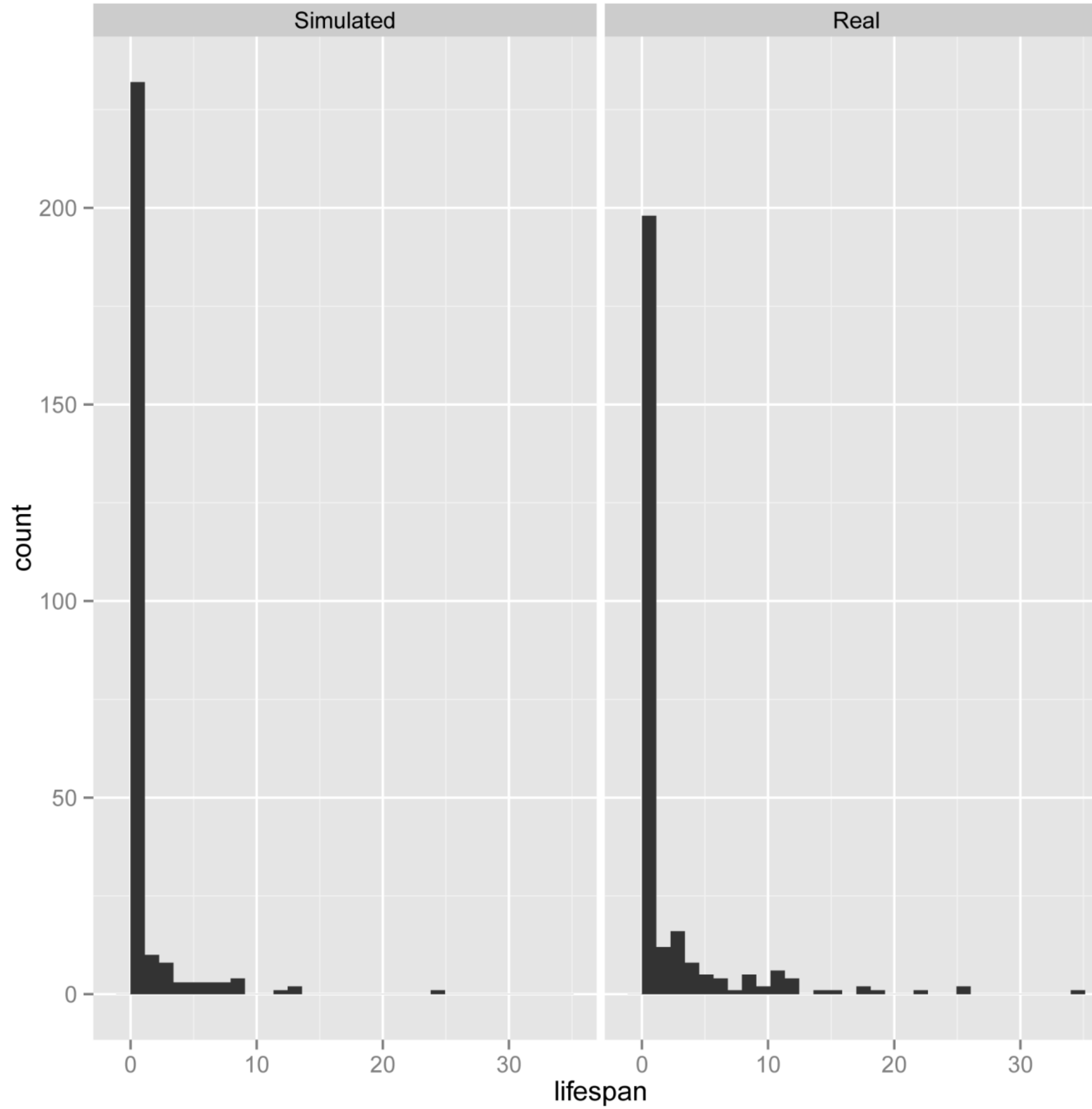
- **Low** recapture probability = **low** power
- **Mark** gives more power to detect **SS**.
- But **Mark** seems to not increase our power in other selection scenarios, such as **linear** and **disruptive** selection.
- It seems no current **statstical magic** can save us from **low power**.

Conclusions

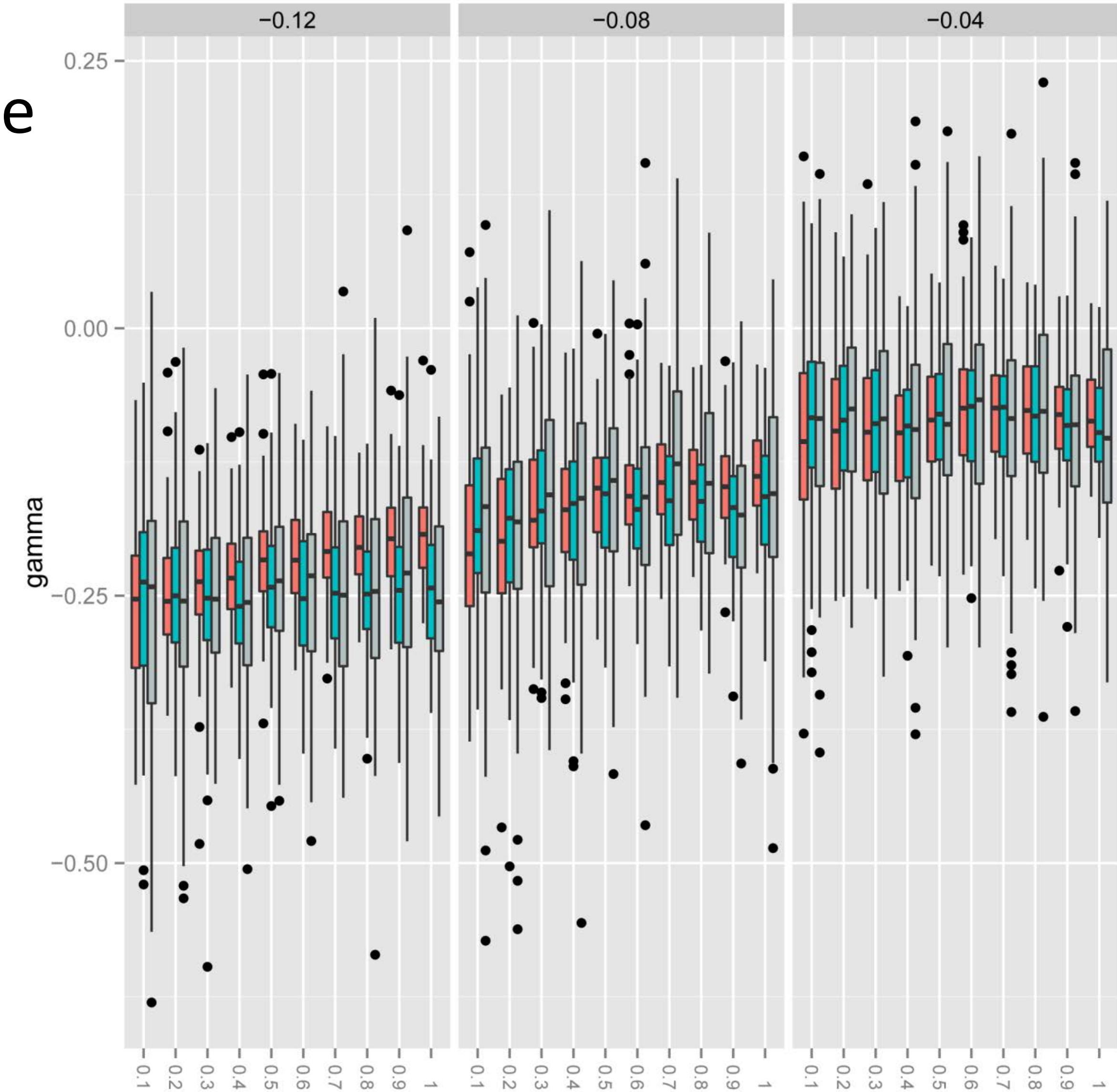
- **Low** recapture probability = **low** power
- Mark gives more power to detect **SS**.
- But Mark seems to not increase our power in other selection scenarios, such as **linear** and **disruptive** selection.
- It seems no current **statstical magic** can save us from **low power**.
- Thank You!


0.04
0.08
0.12
500
1000
1500
100
75
50
25
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

This could be explained by bad assumptions.

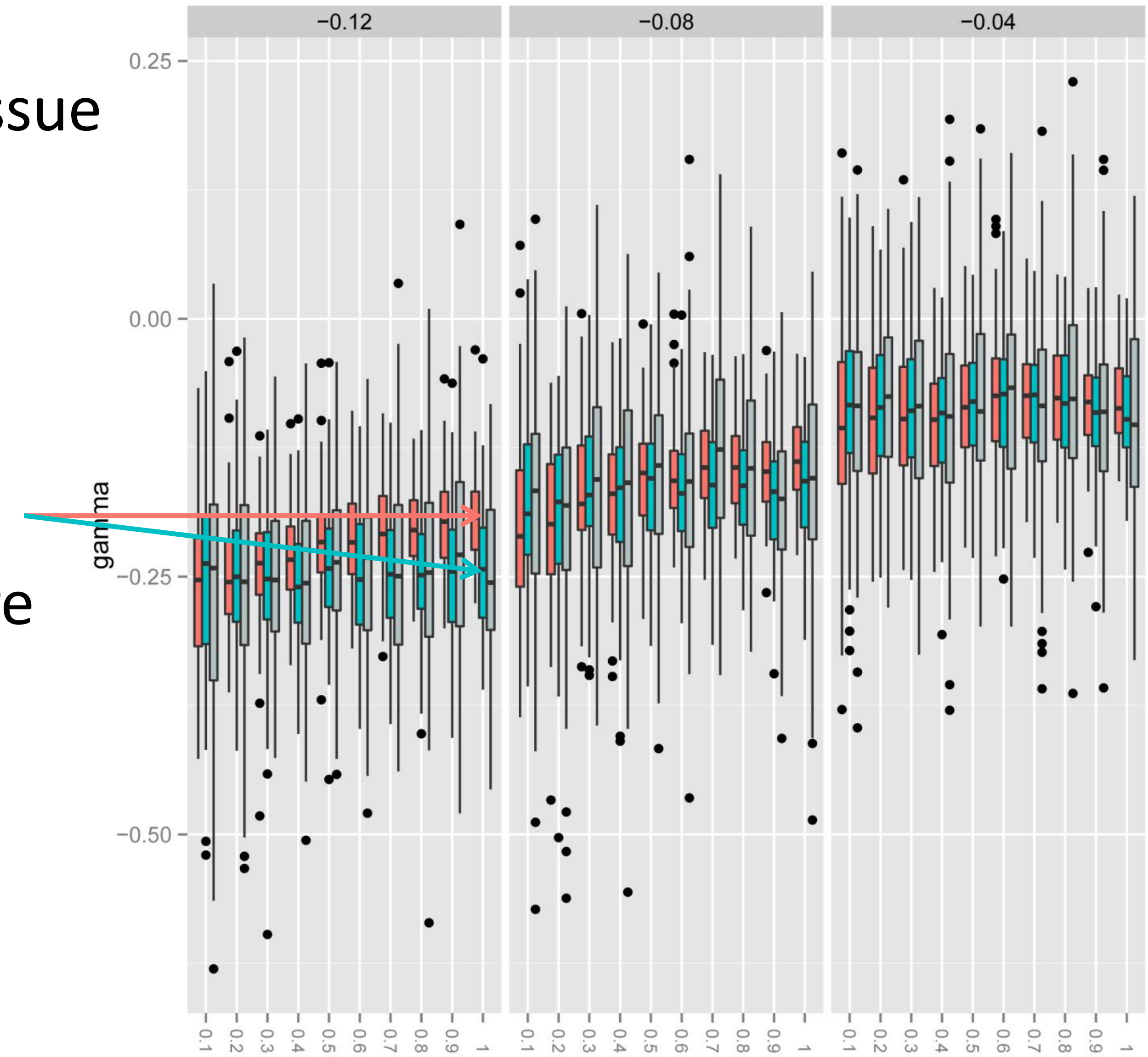


There is also the issue of **estimates**.



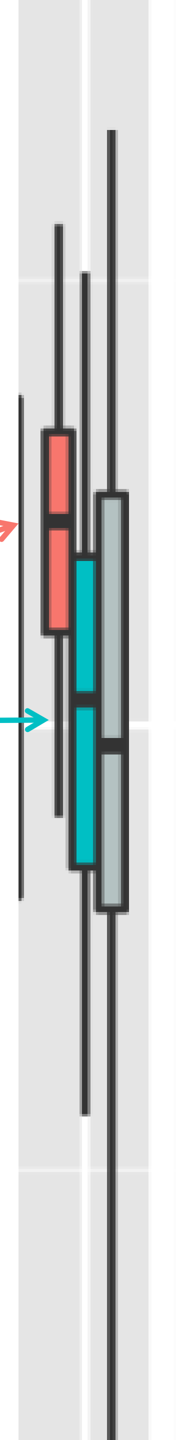
There is also the issue of **estimates**.

Even at perfect detection, LA and Mark estimates are don't agree.



There is also the issue of **estimates**.

Even at perfect detection, **LA** and **Mark** estimates don't agree.



Dis-advantages of MARK

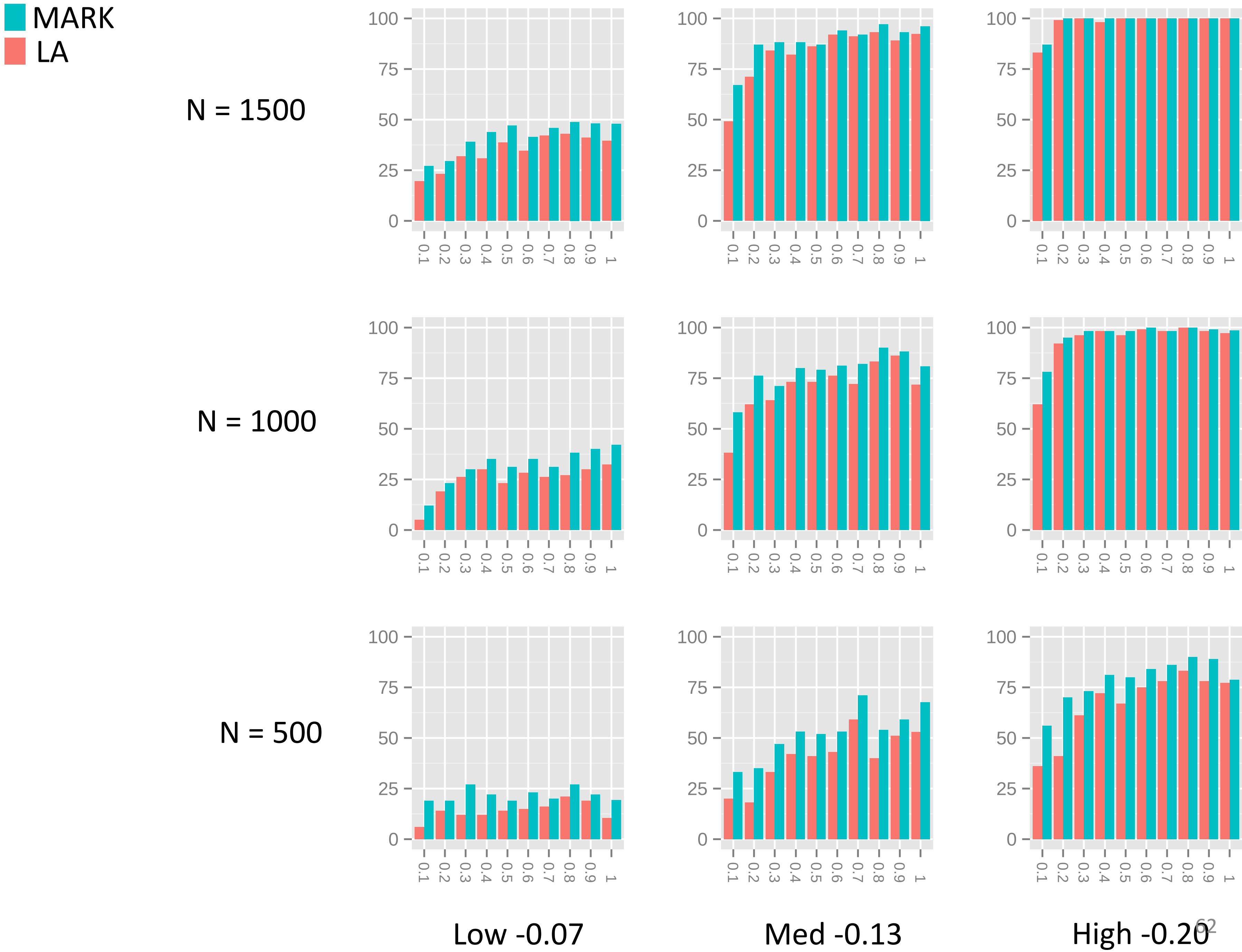
- Need to create a capture history
- **Is complex**, exists a 900 page book explaining the method.
- Performs simarly to other methods at moderate recapture probability.

Advantages of MARK

- Has higher statistical power in extreme cases.
- Can control for **trait dependedence** on p.

Advantages of LA

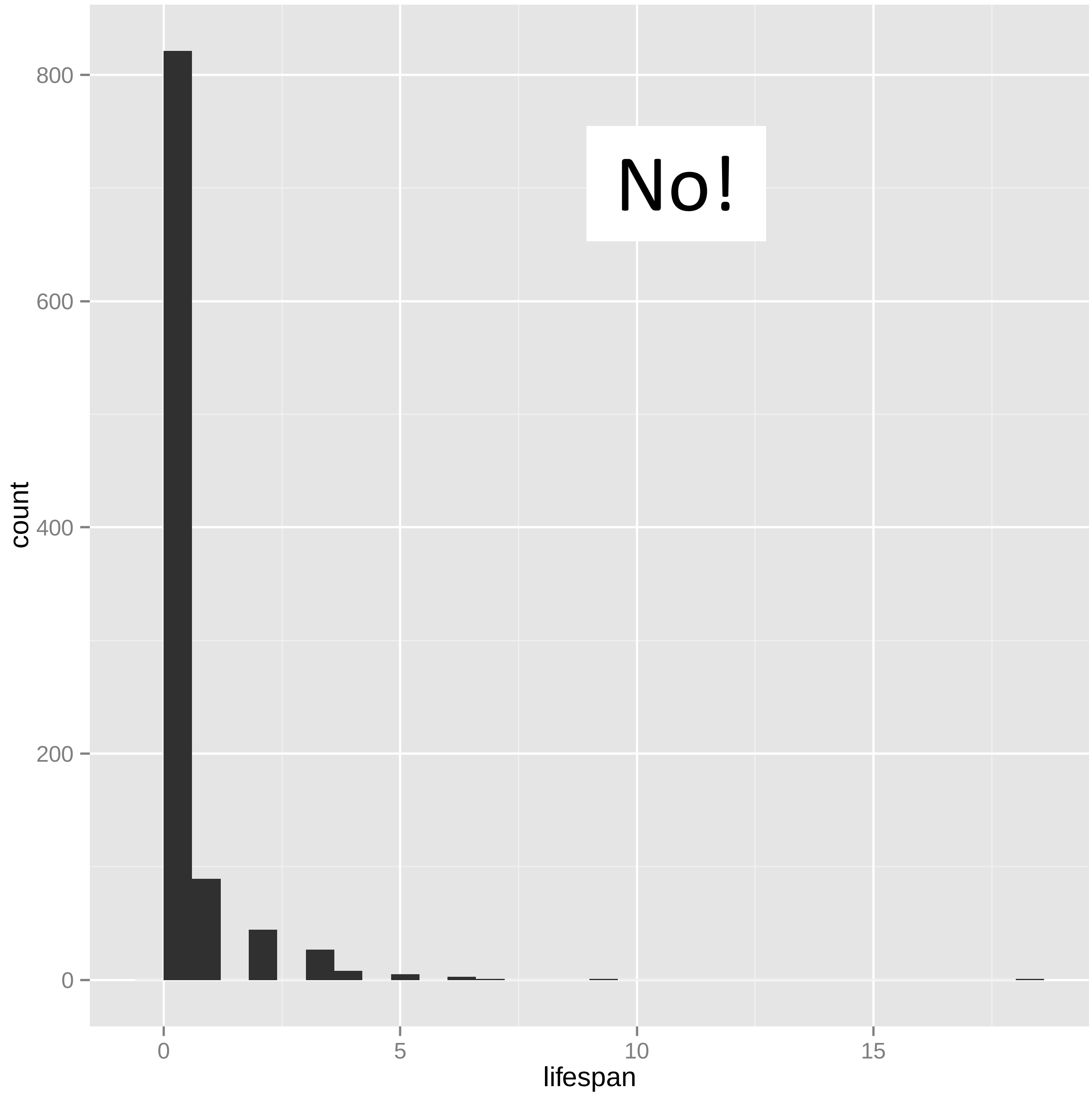
- **simple**
- long history of use
- clear theoretical interpretation
- gradients can be compared directly with past studies

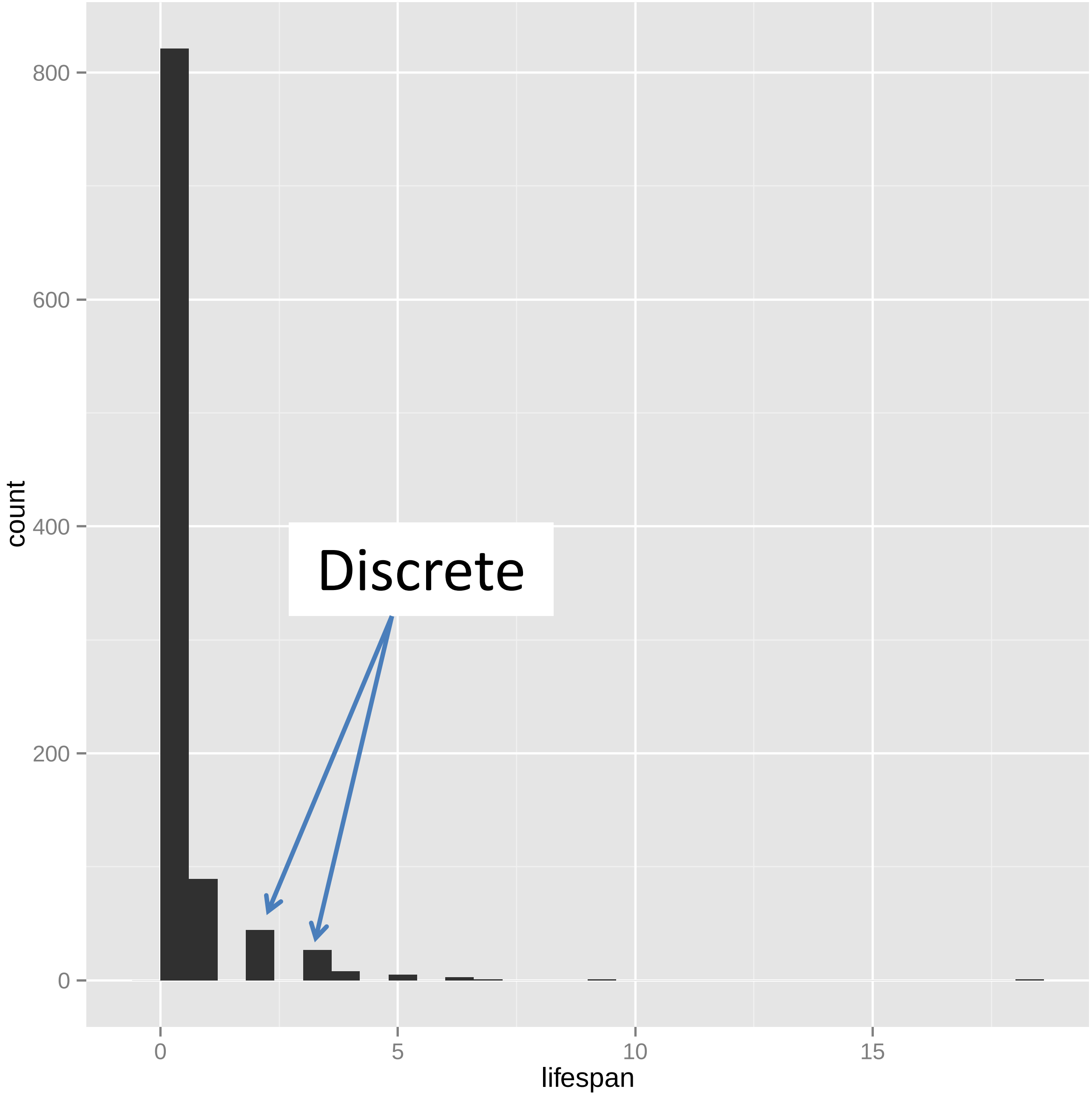

MARK
LA
N = 1500
N = 1000
N = 500
100
75
50
25
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Low -0.07
Med -0.13
High -0.20

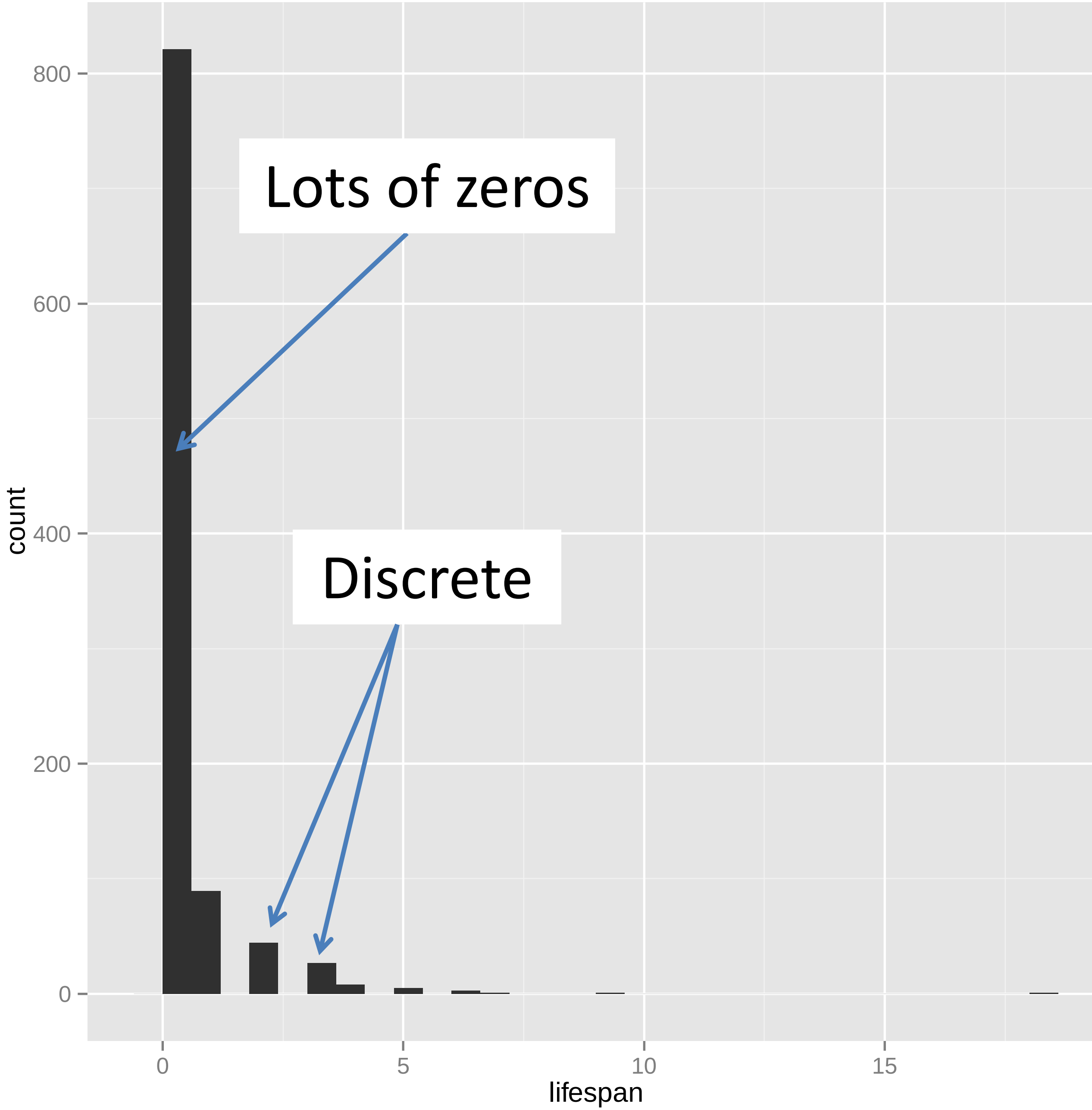
Why is this?

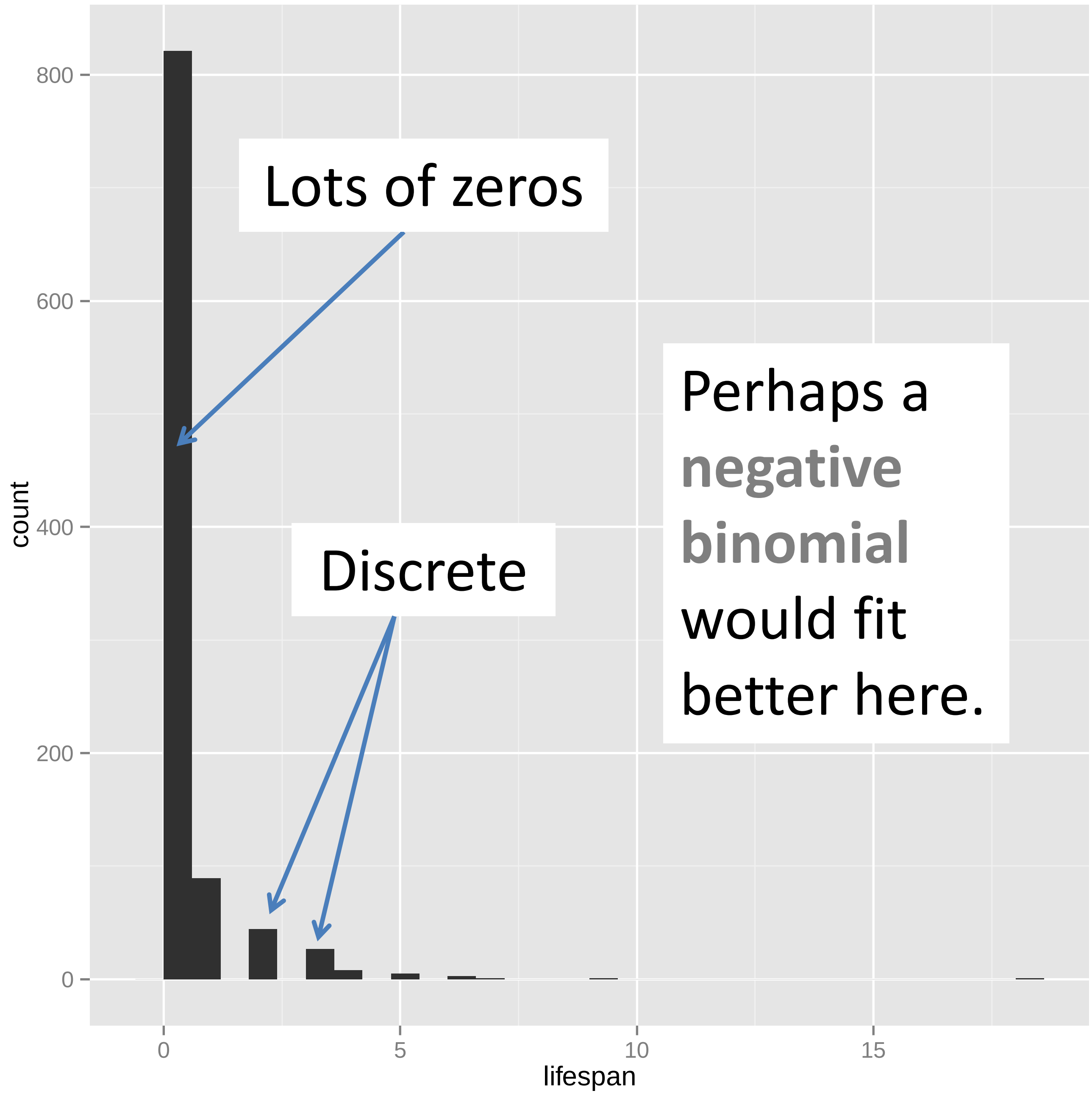
LA assumes a normally distributed lifespan

Is lifespan normally distributed in our simulated data?


No!
count
800
600
400
200
0
0
5
10
15
lifespan

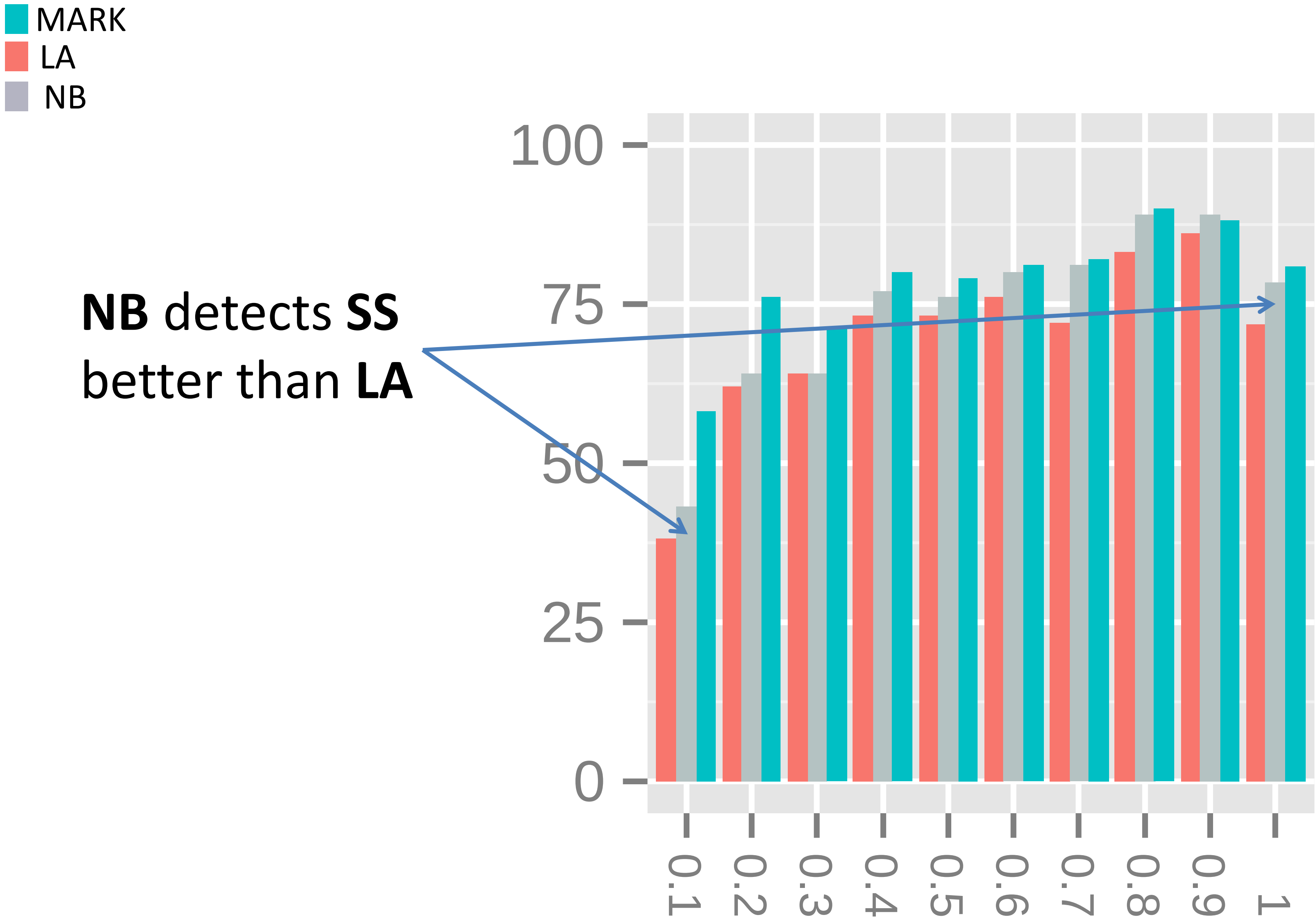

800
600
400
200
0
count
Discrete
0
5
10
15
lifespan

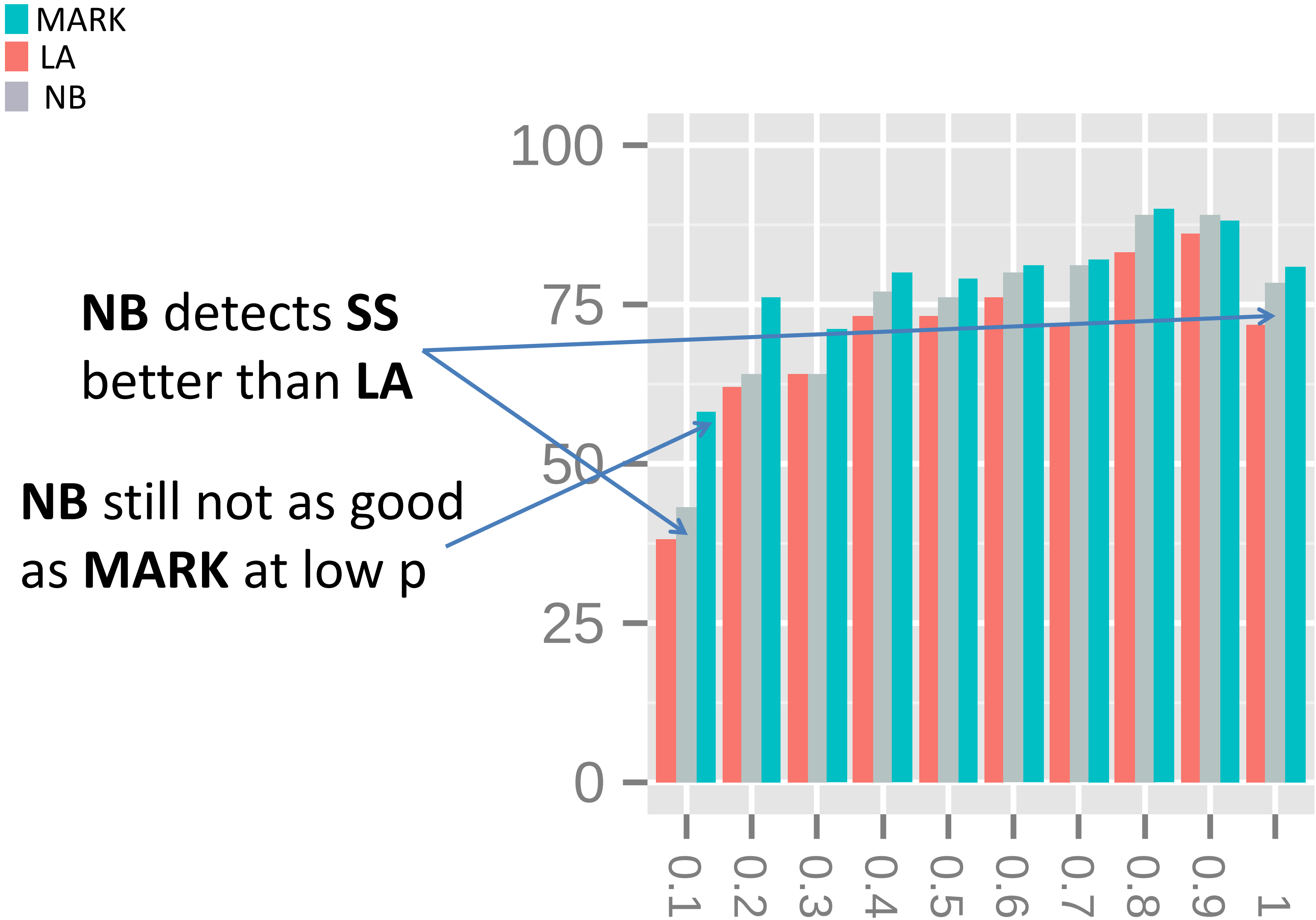

Lots of zeros
Discrete
count
lifespan
800
600
400
200
0
0
5
10
15


800
600
400
200
0
count
0
5
10
15
lifespan
Lots of zeros
Discrete
Perhaps a **negative binomial** would fit better here.

negative binomial distribution **(NB)**

NB detects **SS** better than **LA**

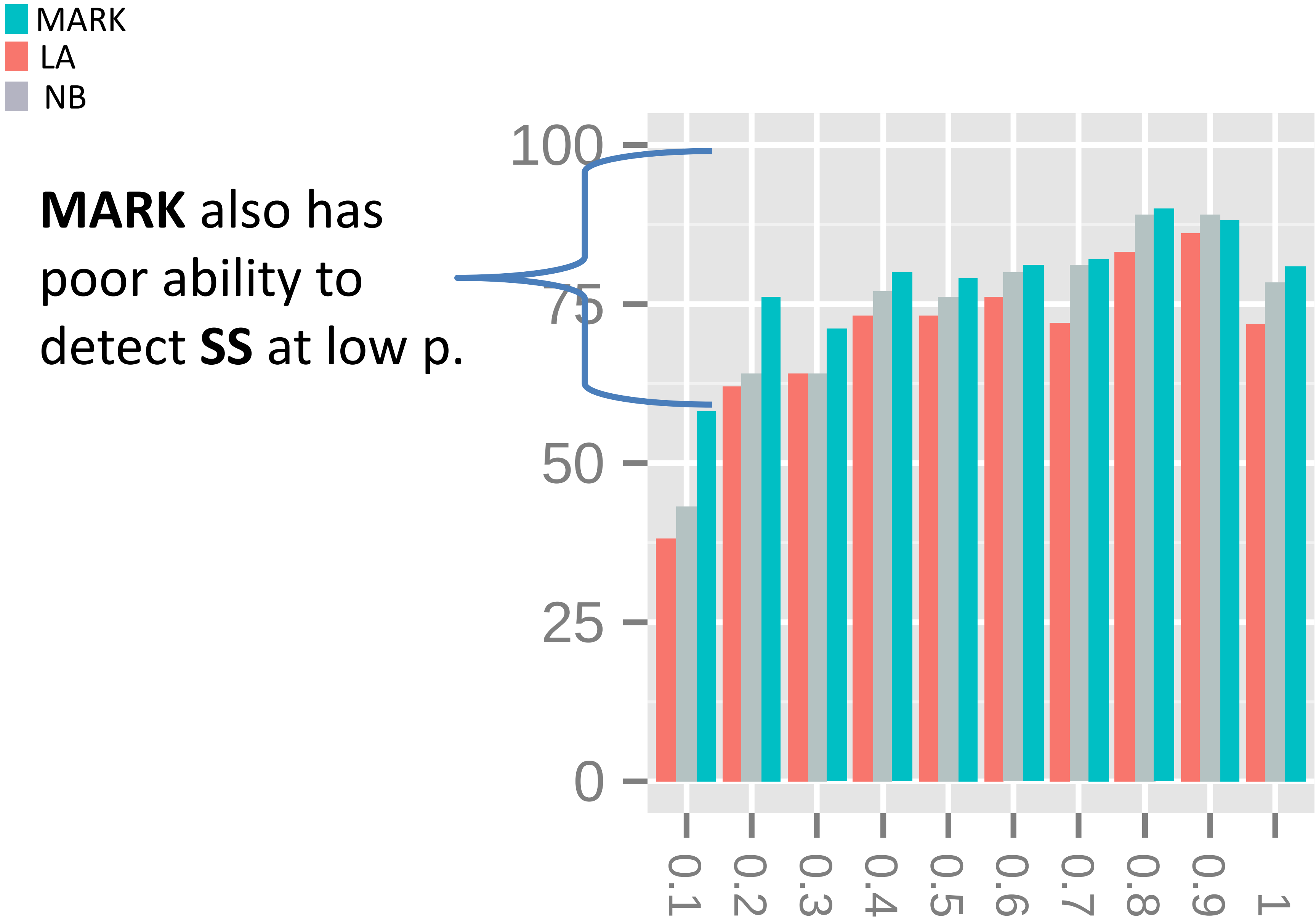


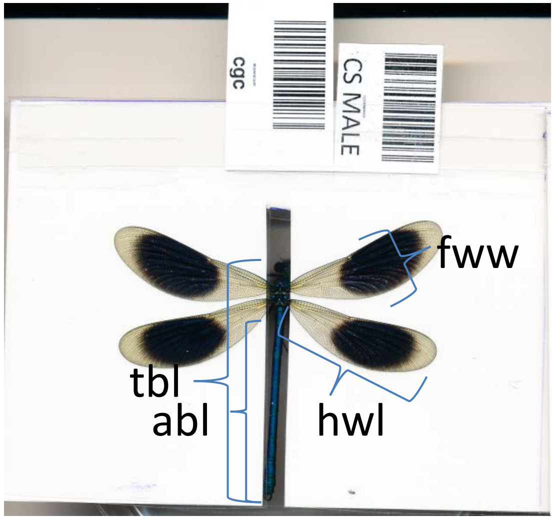


NB detects **SS** better than **LA**

NB still not as good as **MARK** at low p

MARK also has poor ability to detect **SS** at low p.

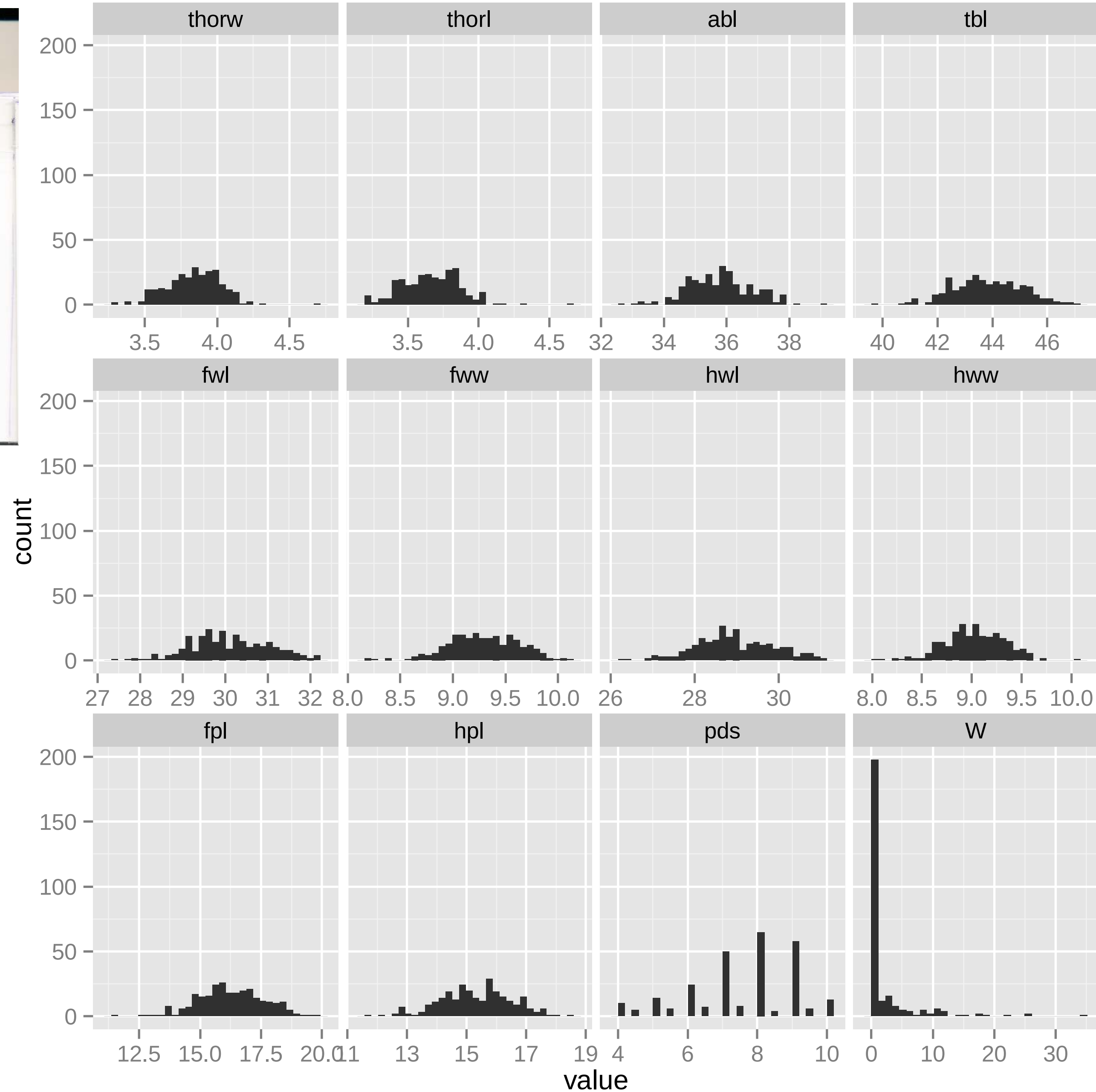


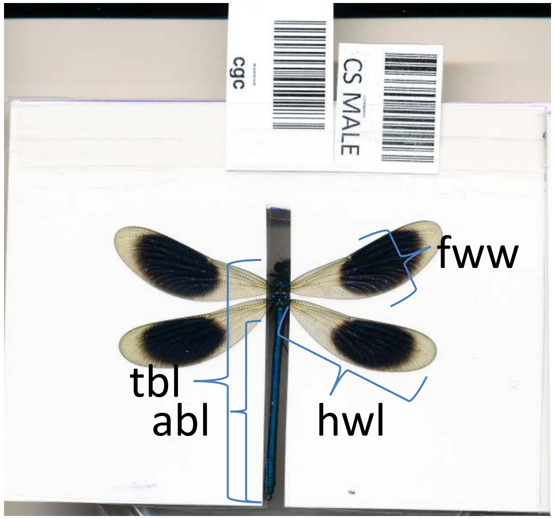


Male

Calopteryx splendens

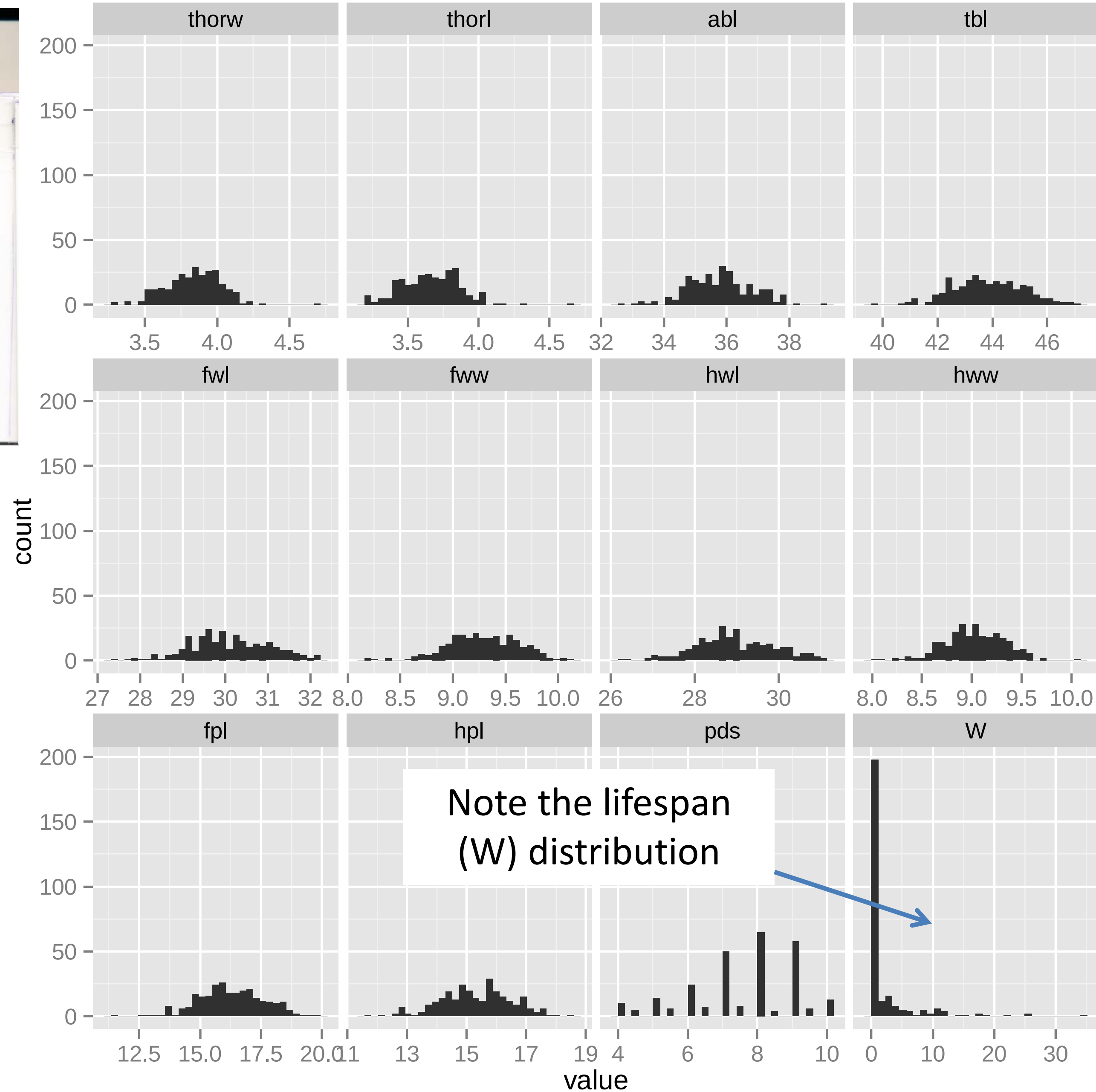
- 11 traits
- N = 324
- p ~ 0.1
- **low recapture probability**

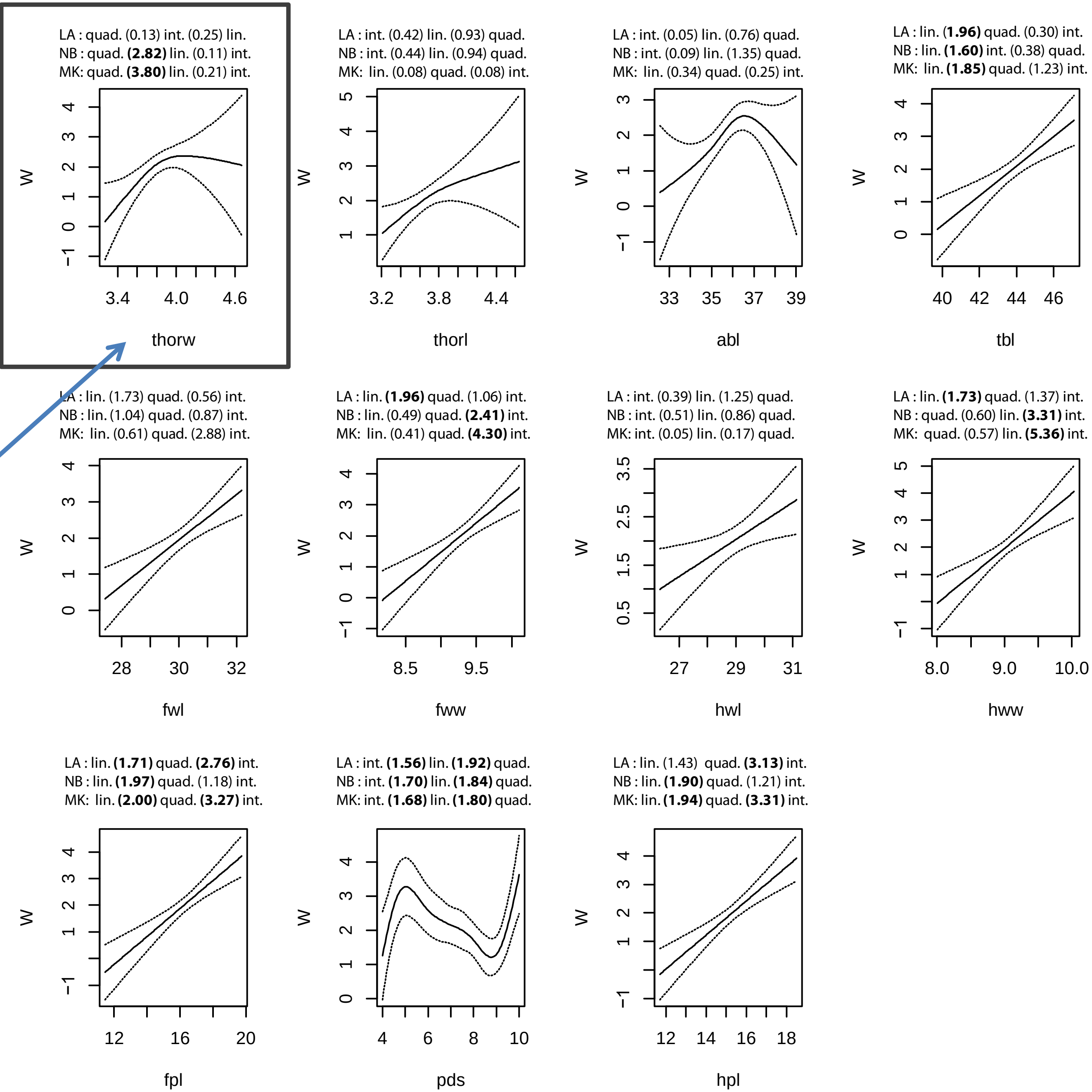




Male
Calopteryx splendens

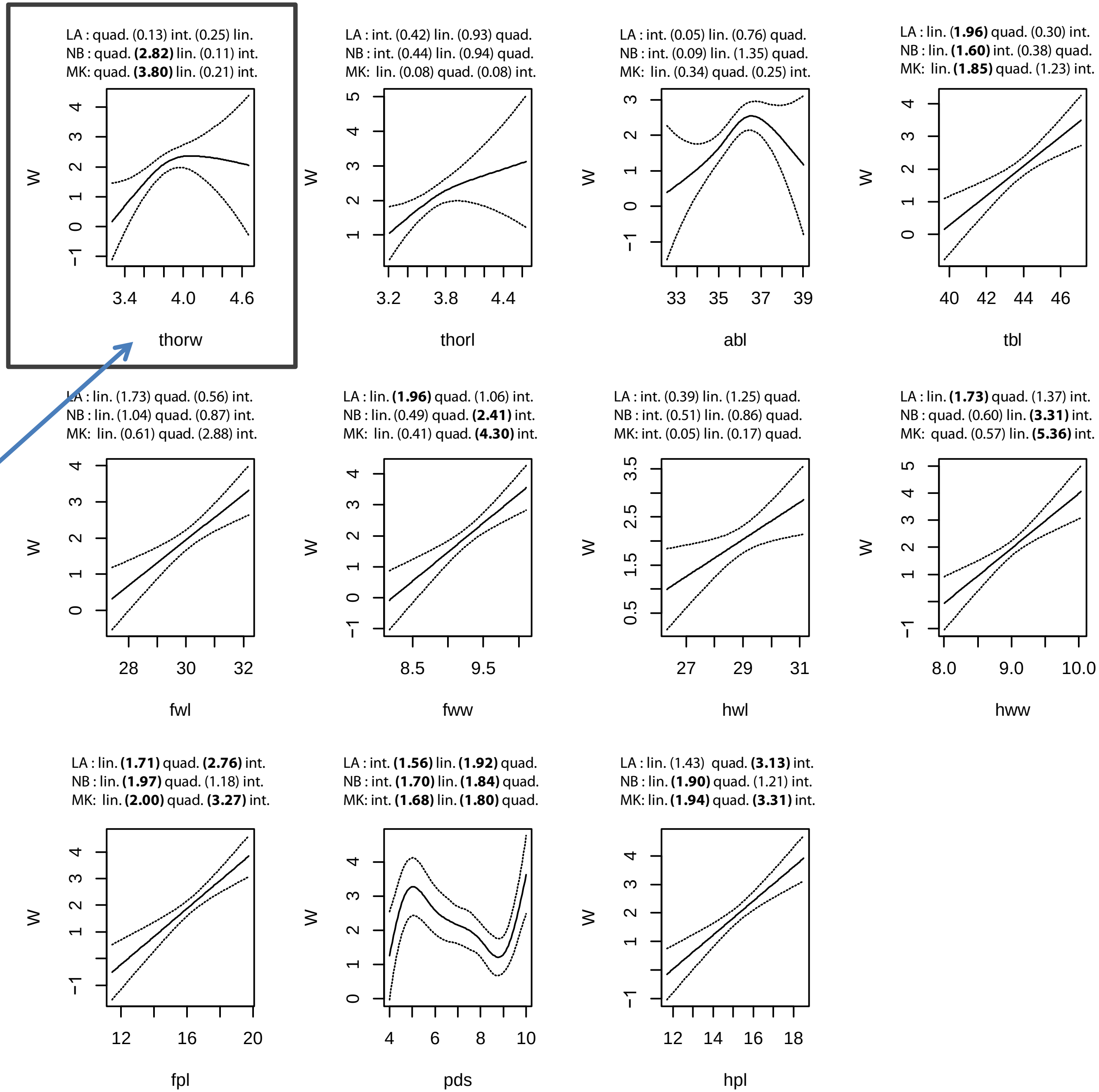
- 11 traits
- N = 324
- p ~ 0.1
- **low recapture probability**



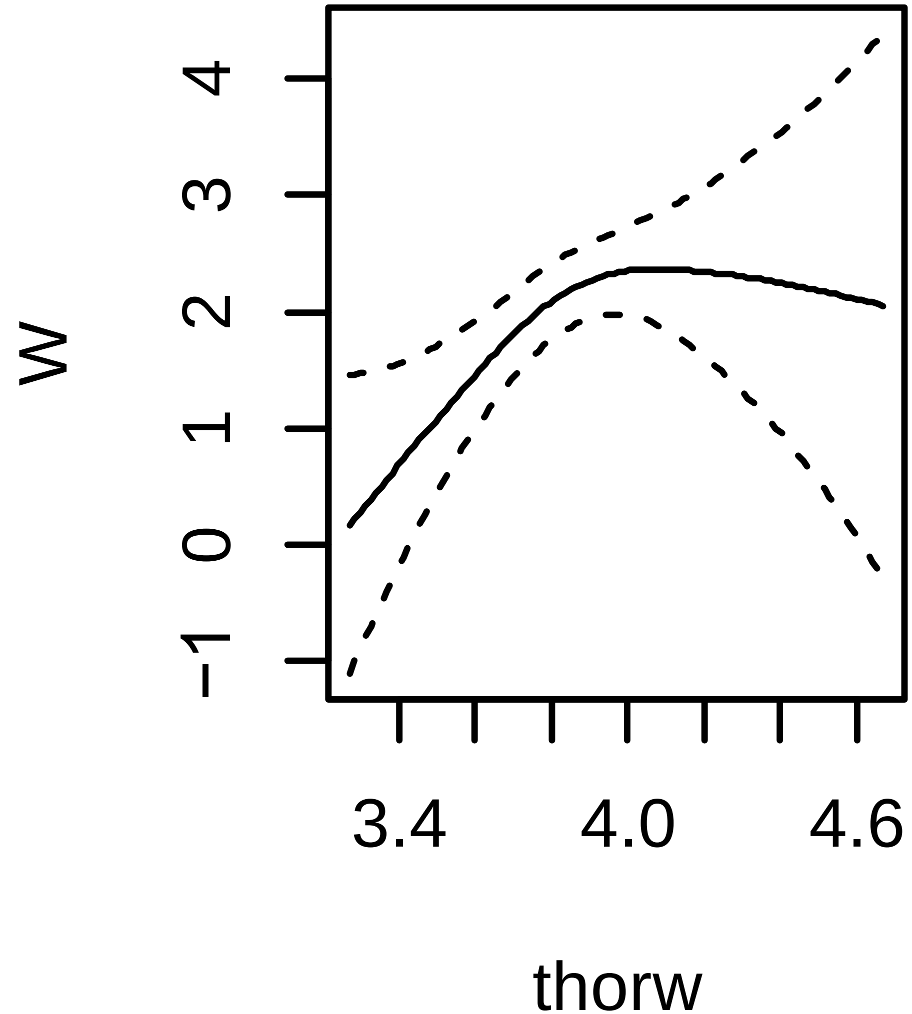


Out of all 11 traits tested, only one showed **SS**

Let’s zoom into Thorax width



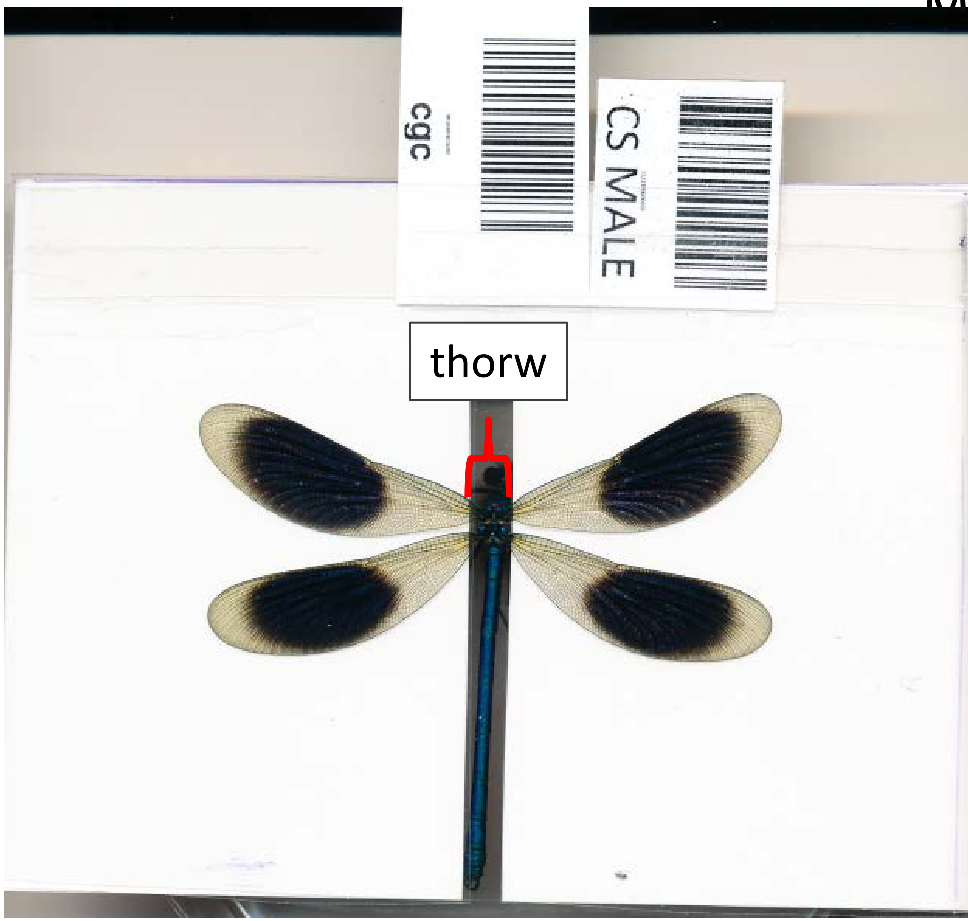
LA : quad. (0.13) int. (0.25) lin.
NB : quad. **(2.82)** lin. (0.11) int.
MK: quad. **(3.80)** lin. (0.21) int.

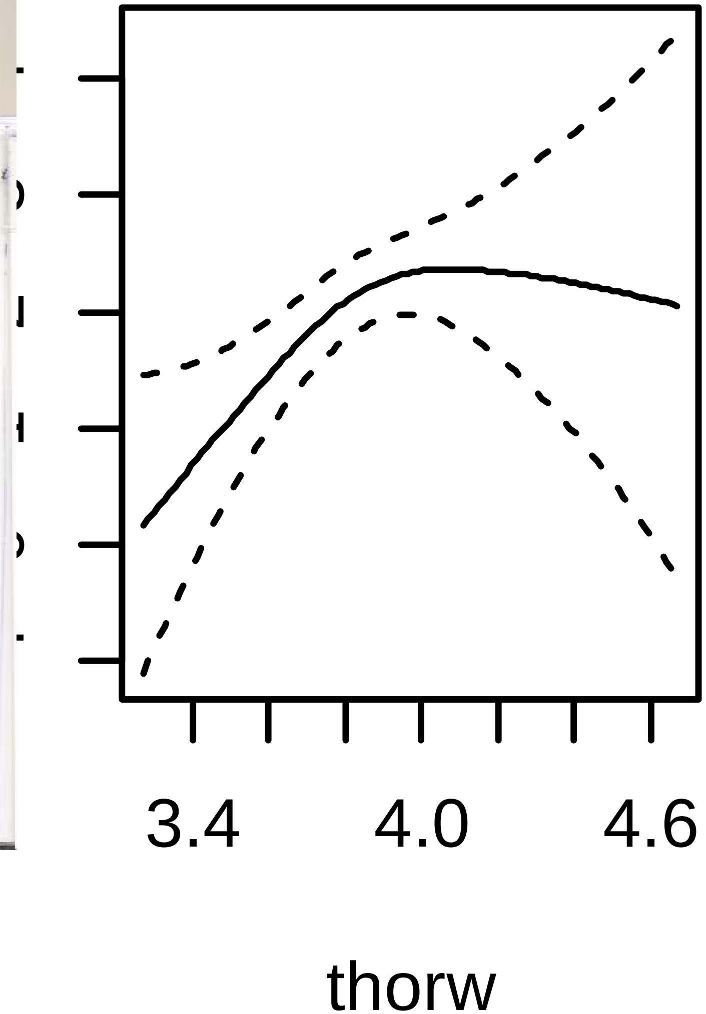


LA : quad. (0.13) int. (0.25) lin.

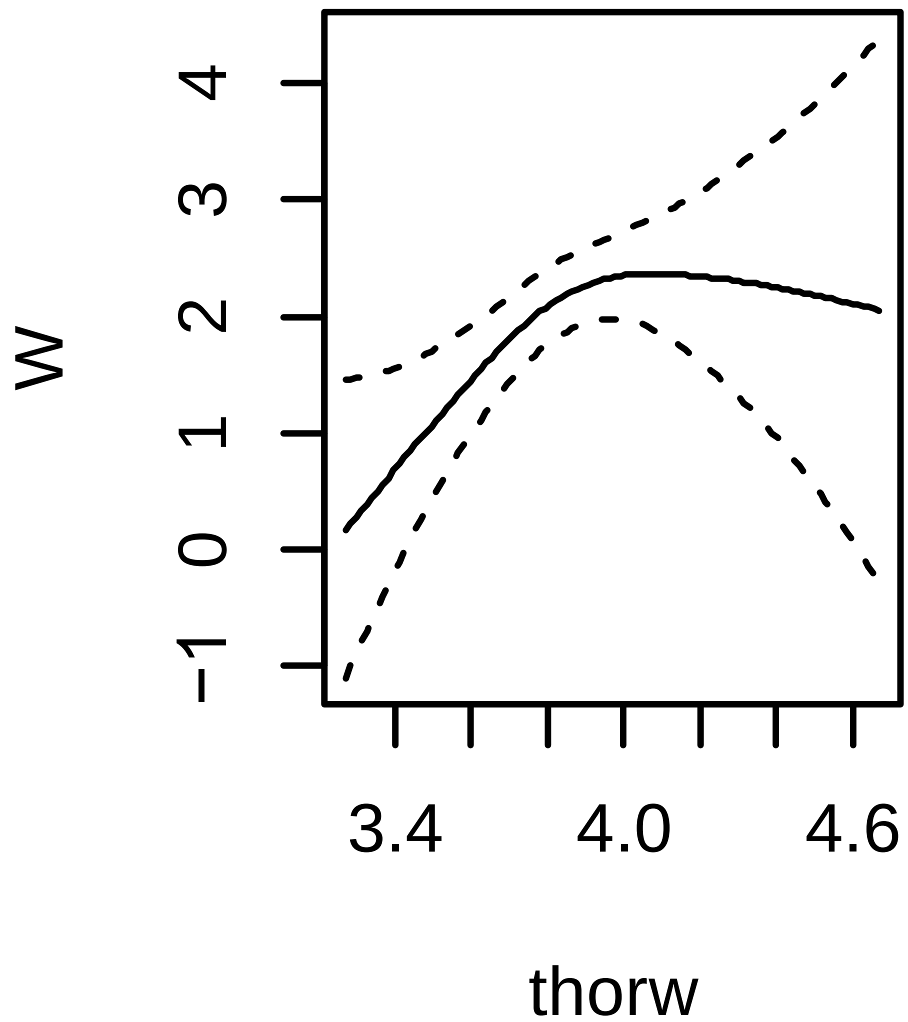
NB : quad. **(2.82)** lin. (0.11) int.

MK: quad. **(3.80)** lin. (0.21) int.



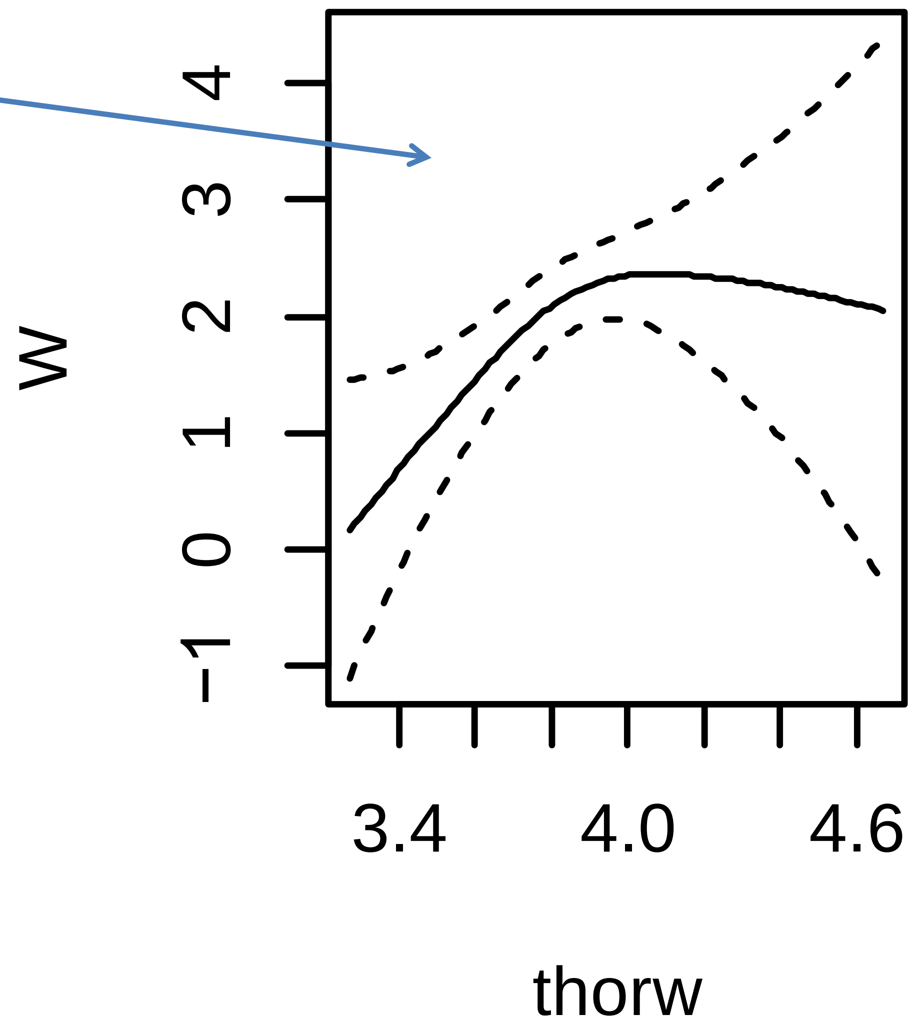


LA : quad. (0.13) int. (0.25) lin.
NB : quad. **(2.82)** lin. (0.11) int.
MK: quad. **(3.80)** lin. (0.21) int.



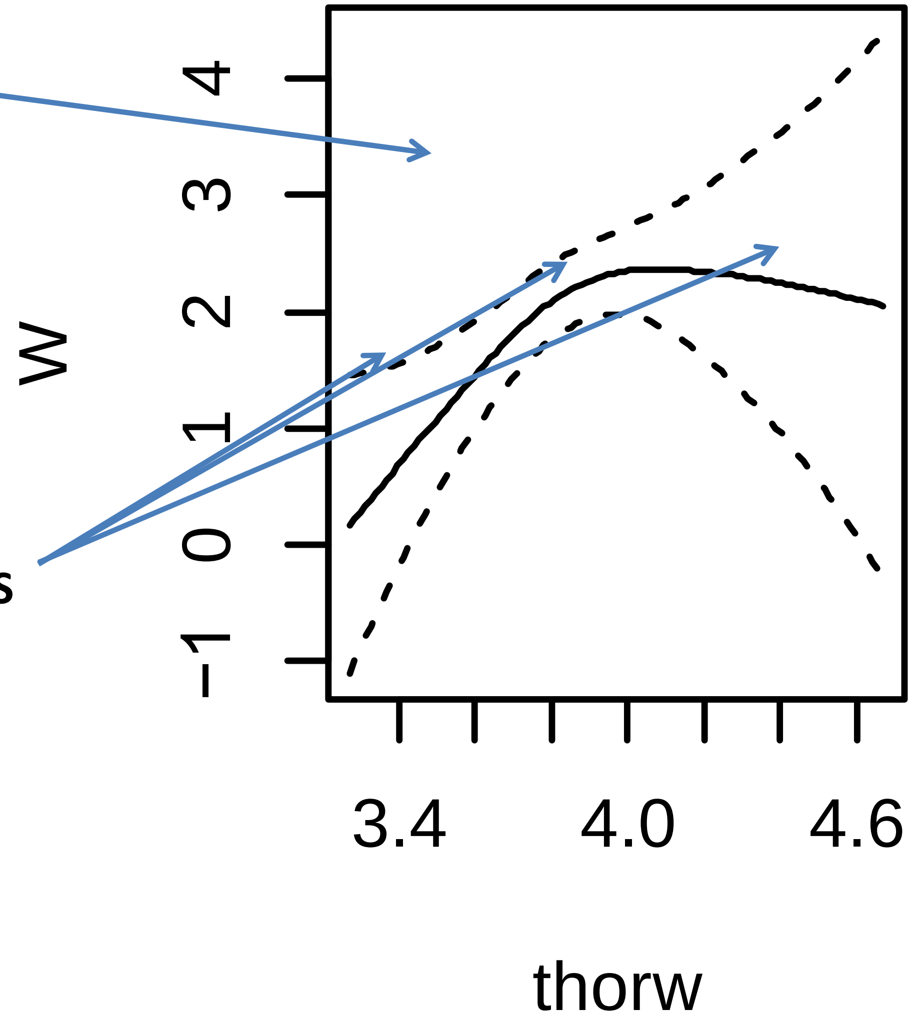
LA : quad. (0.13) int. (0.25) lin.
NB : quad. **(2.82)** lin. (0.11) int.
MK: quad. **(3.80)** lin. (0.21) int.

The main graph shows a spline of the minimum **lifespan vs thorax width**



LA : quad. (0.13) int. (0.25) lin.
NB : quad. **(2.82)** lin. (0.11) int.
MK: quad. **(3.80)** lin. (0.21) int.

The main graph shows a spline of the minimum **lifespan** vs **thorax width**

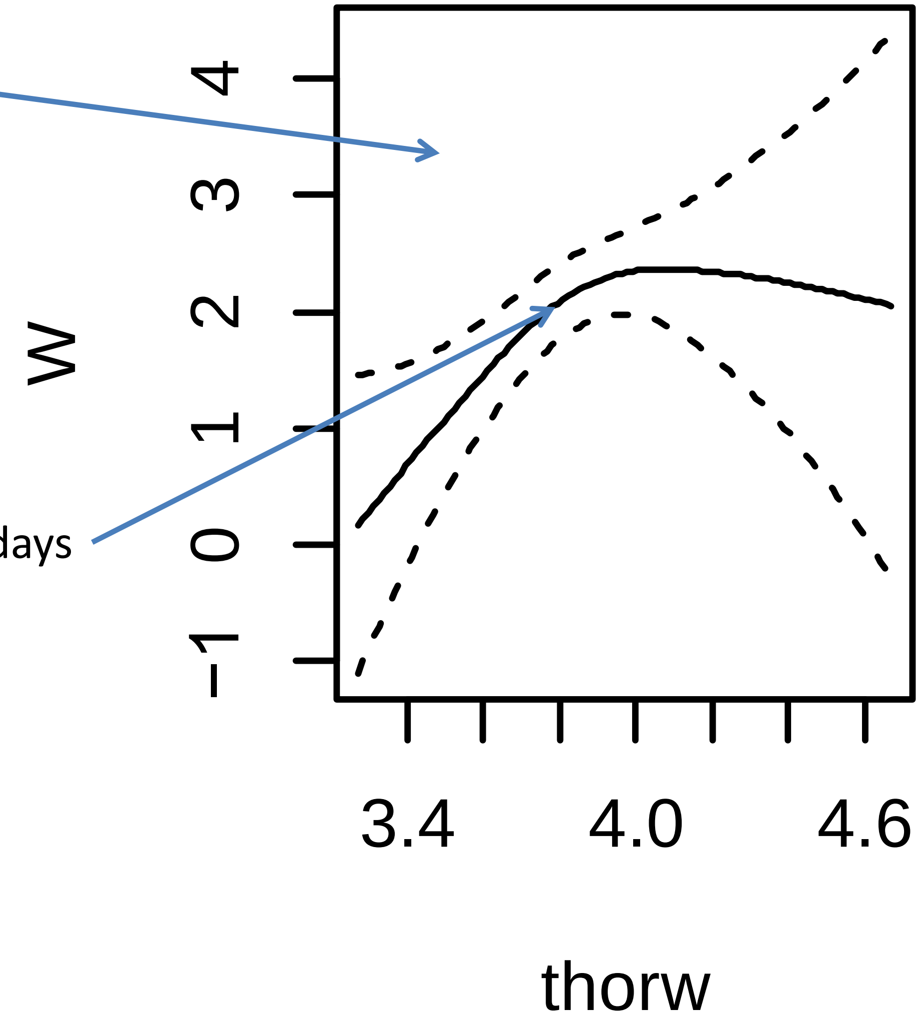


Note the **SS**

LA : quad. (0.13) int. (0.25) lin.
NB : quad. **(2.82)** lin. (0.11) int.
MK: quad. **(3.80)** lin. (0.21) int.

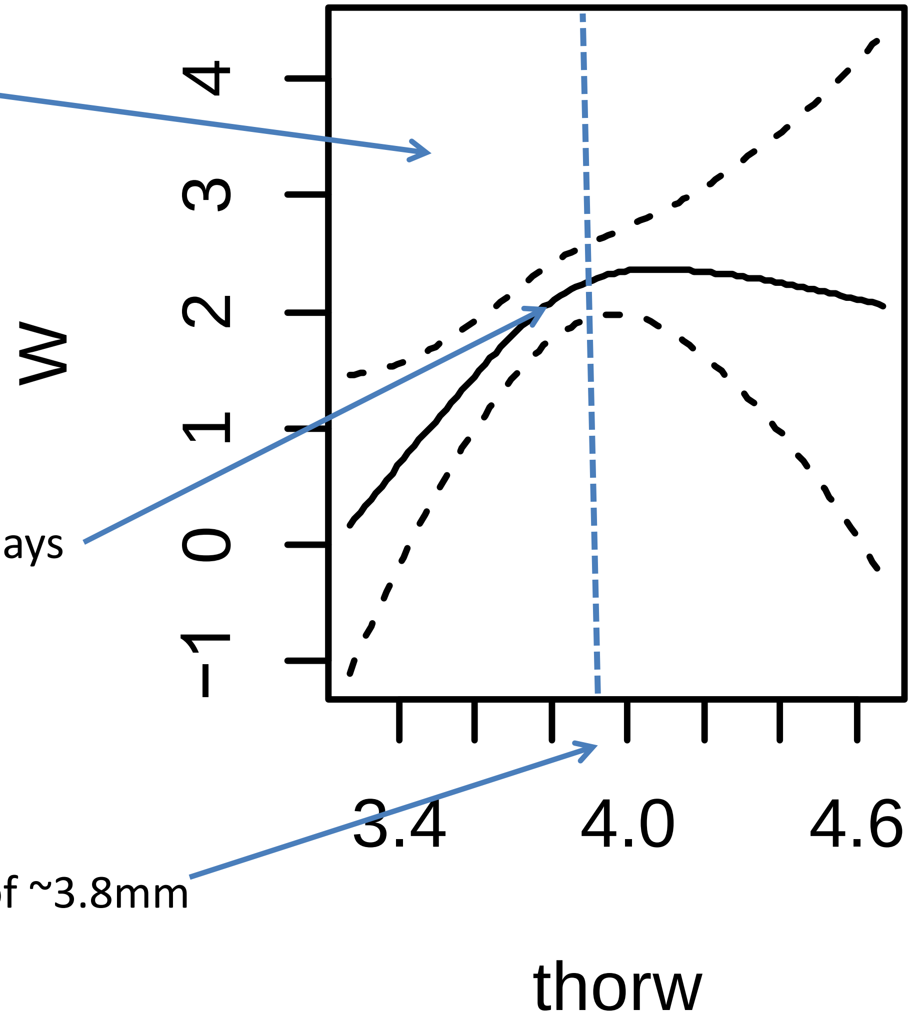
The main graph shows a spline of the minimum lifespan vs Thorax Width

We have a mean lifespan of 2 days



LA : quad. (0.13) int. (0.25) lin.
NB : quad. **(2.82)** lin. (0.11) int.
MK: quad. **(3.80)** lin. (0.21) int.

The main graph shows a spline of the minimum lifespan vs Thorax Width

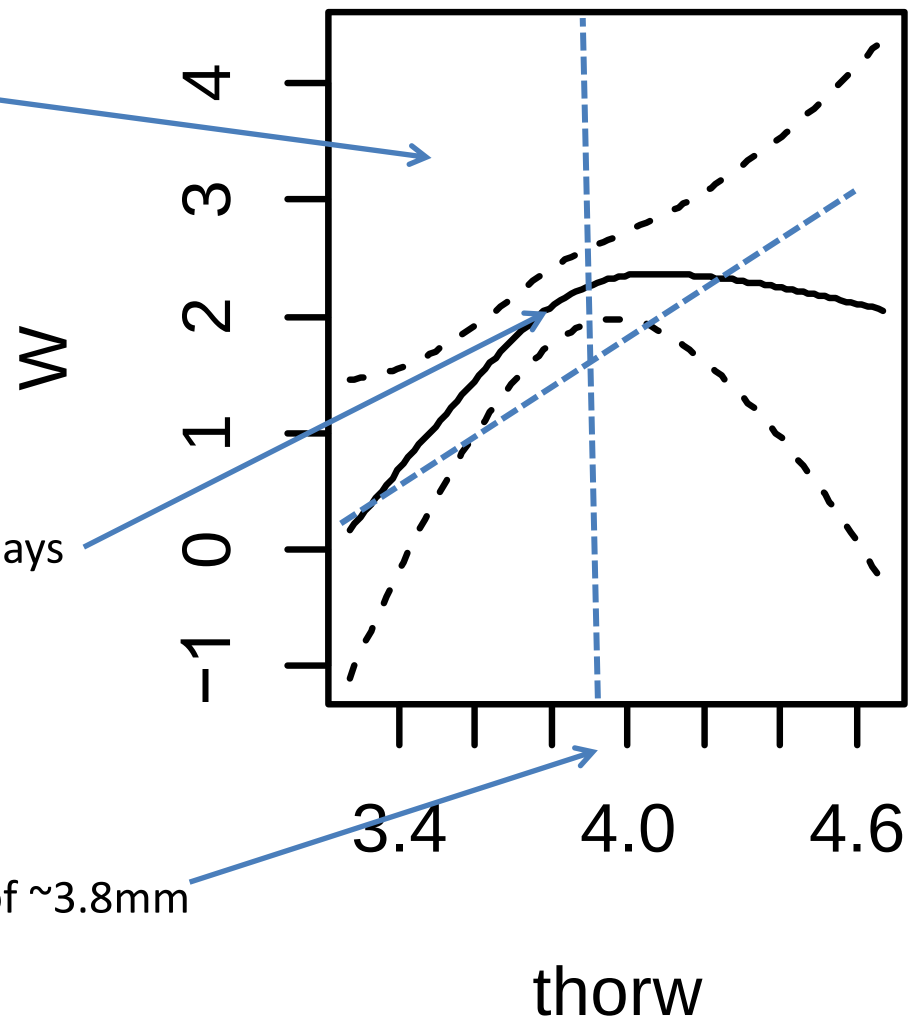


We have a mean lifespan of 2 days

We have a mean thorw of ~3.8mm

LA : quad. (0.13) int. (0.25) lin.
NB : quad. **(2.82)** lin. (0.11) int.
MK: quad. **(3.80)** lin. (0.21) int.

The main graph shows a spline of the minimum lifespan vs Thorax Width

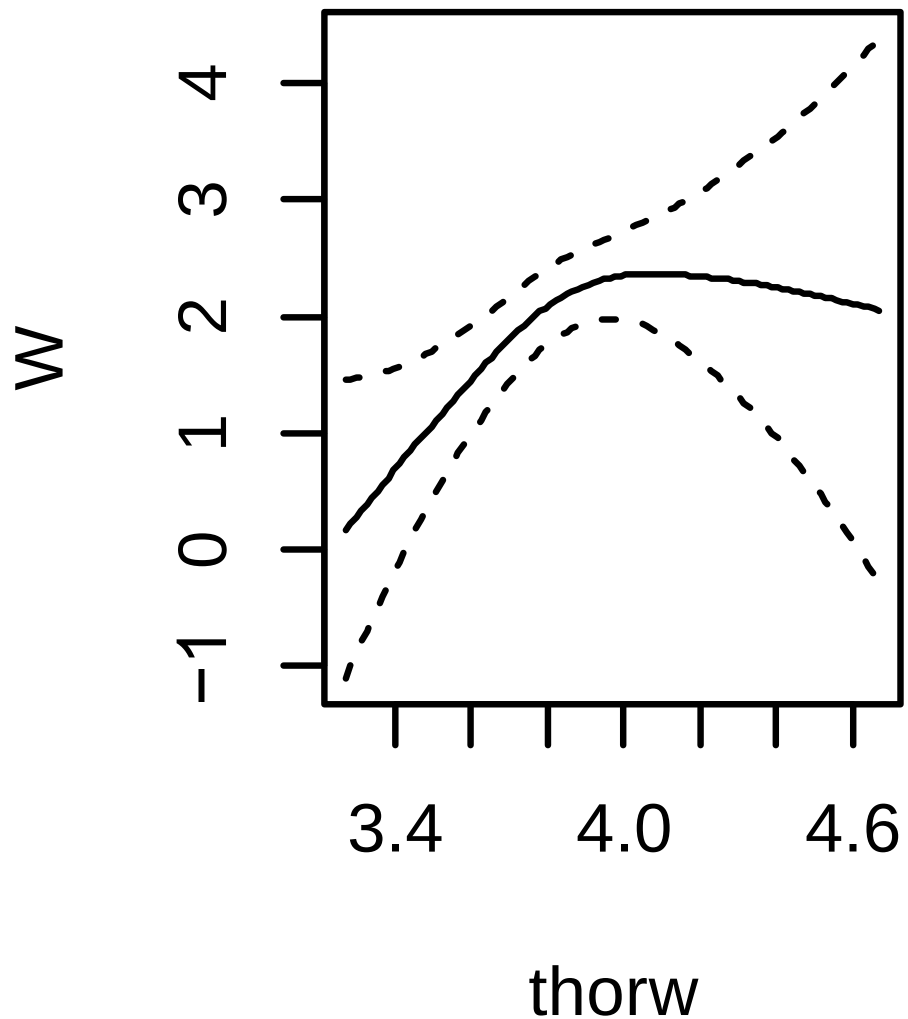


We have a mean lifespan of 2 days

We have a mean thorw of ~3.8mm

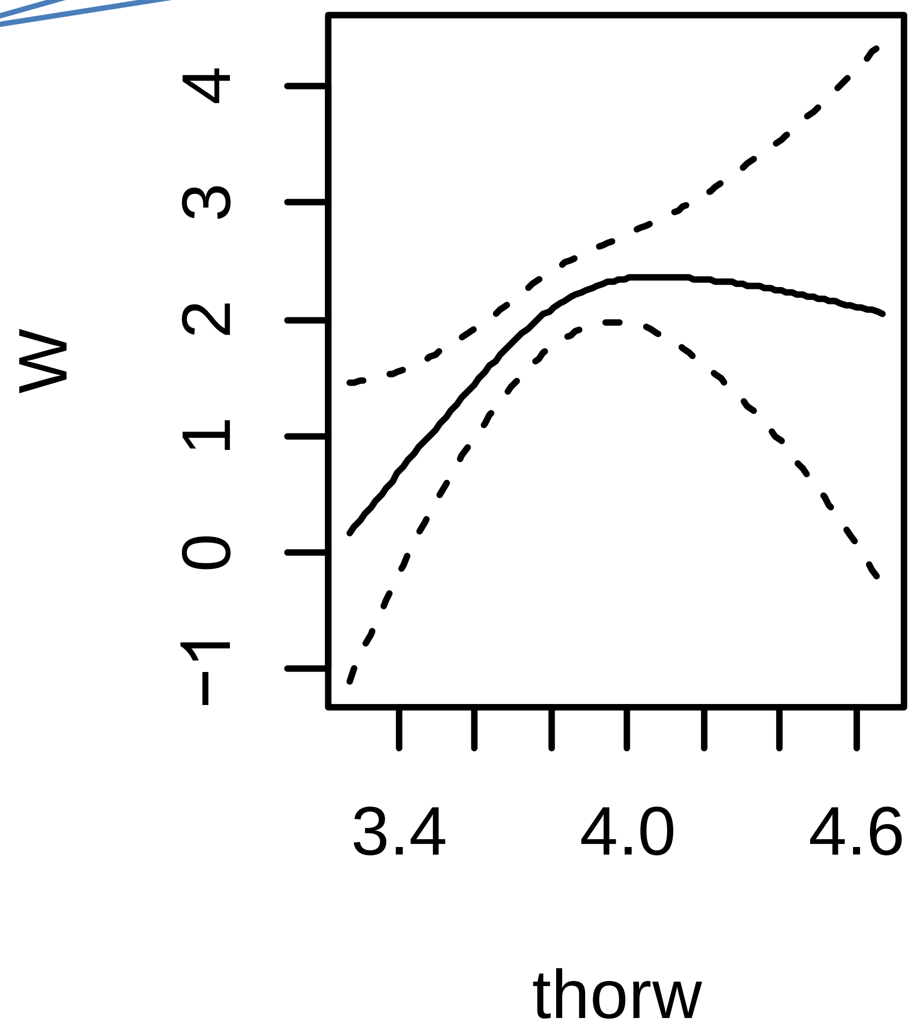
LA : quad. (0.13) int. (0.25) lin.
NB : quad. **(2.82)** lin. (0.11) int.
MK: quad. **(3.80)** lin. (0.21) int.

The text here shows us the performance of the tests



LA : quad. (0.13) int. (0.25) lin.
NB : quad. **(2.82)** lin. (0.11) int.
MK: quad. **(3.80)** lin. (0.21) int.

The numbers in parantheses show the **delta AIC** to the next best model (higher is more significant)

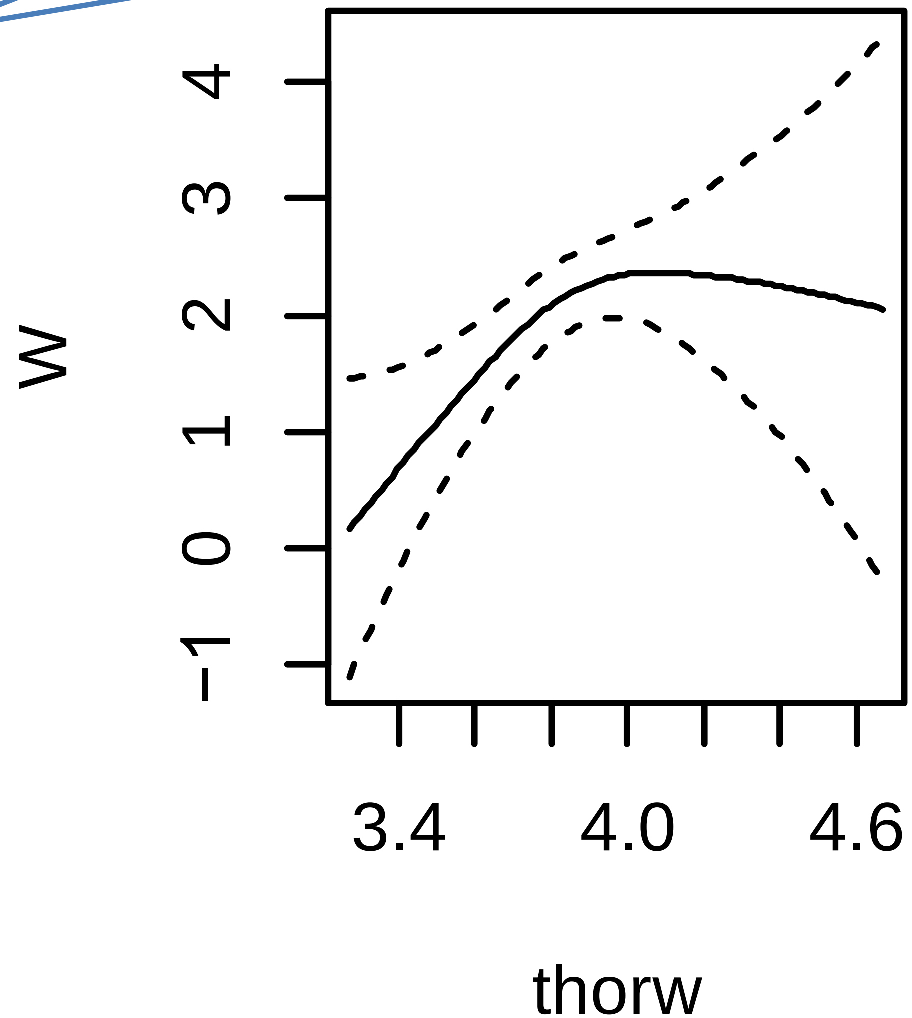


LA : quad. (0.13) int. (0.25) lin.

NB : quad. **(2.82)** lin. (0.11) int.

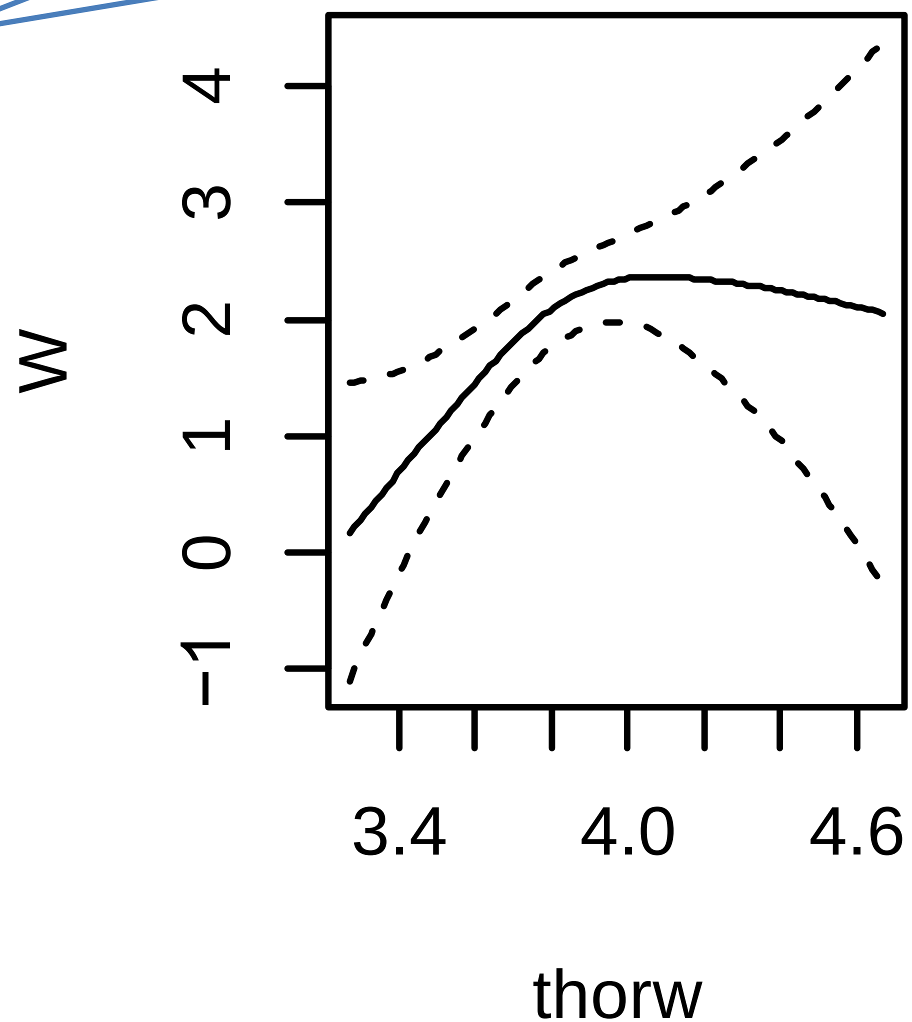
MK: quad. **(3.80)** lin. (0.21) int.

All tests detected **SS**, but not all detected significant **SS**



LA : quad. (0.13) int. (0.25) lin.
NB : quad. **(2.82)** lin. (0.11) int.
MK: quad. **(3.80)** lin. (0.21) int.

All tests detected **SS**, but not all detected significant **SS**



Mark had the highest delta AIC, and therefore the highest power to detect **SS**

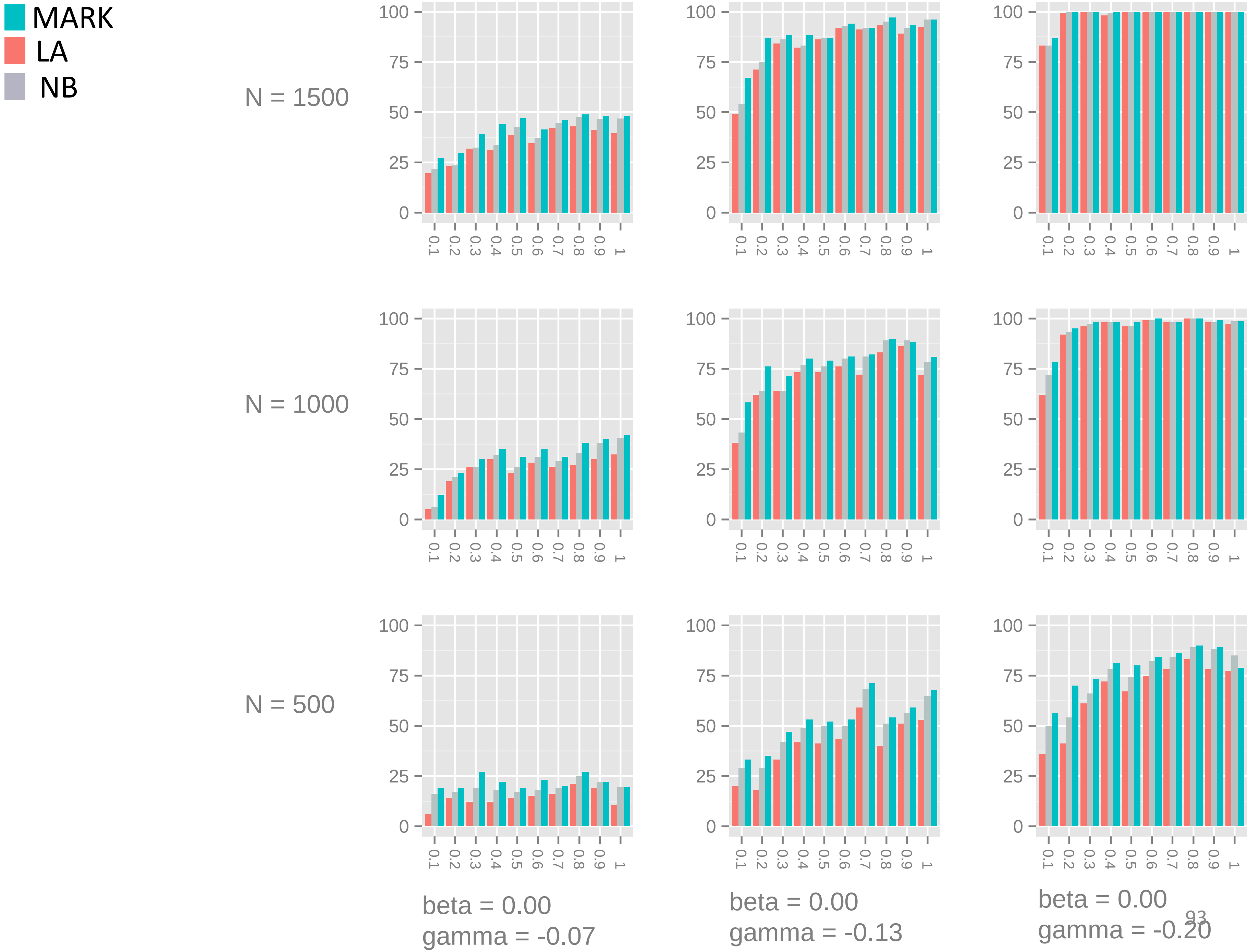
NB had the second highest power

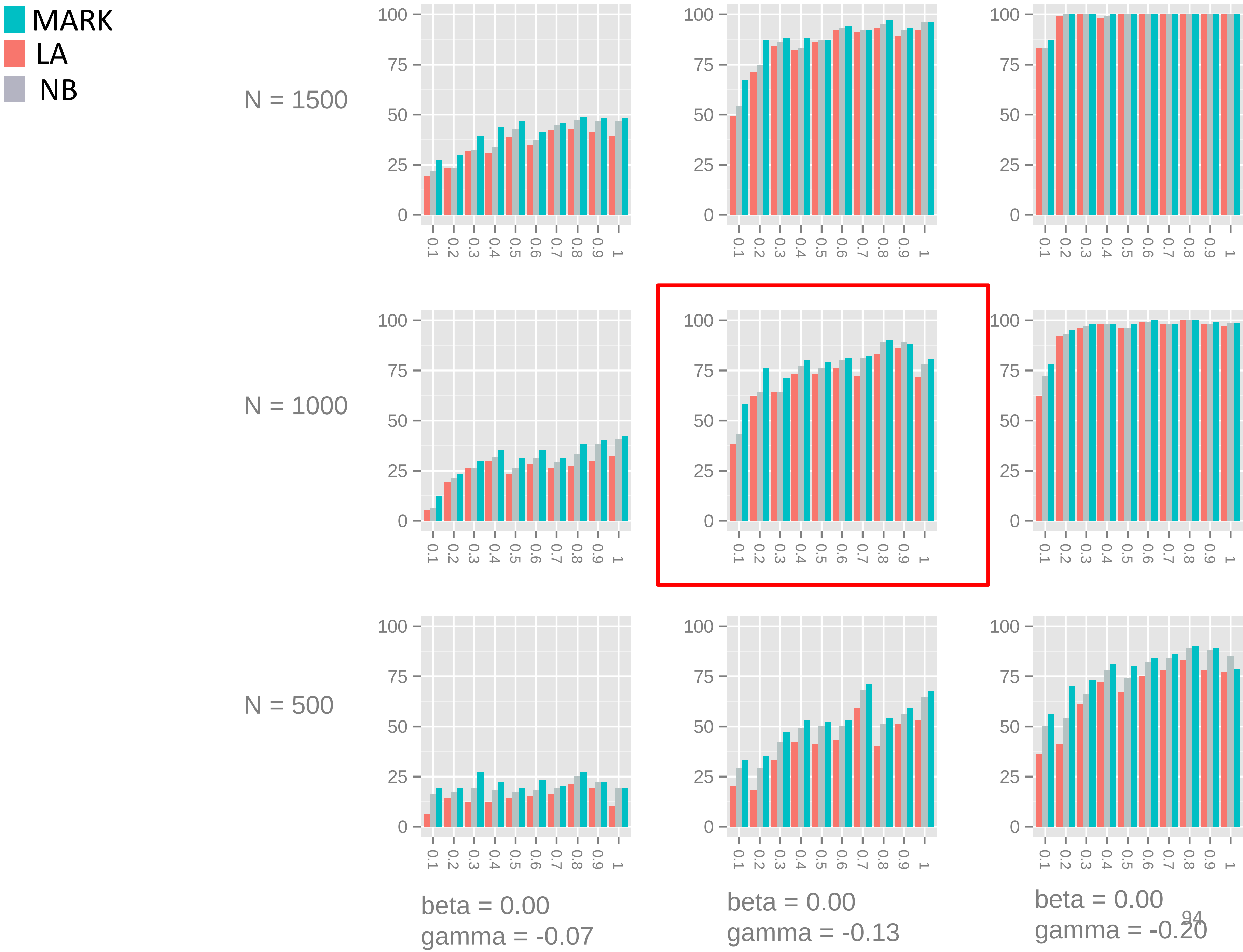
LA had the lowest power

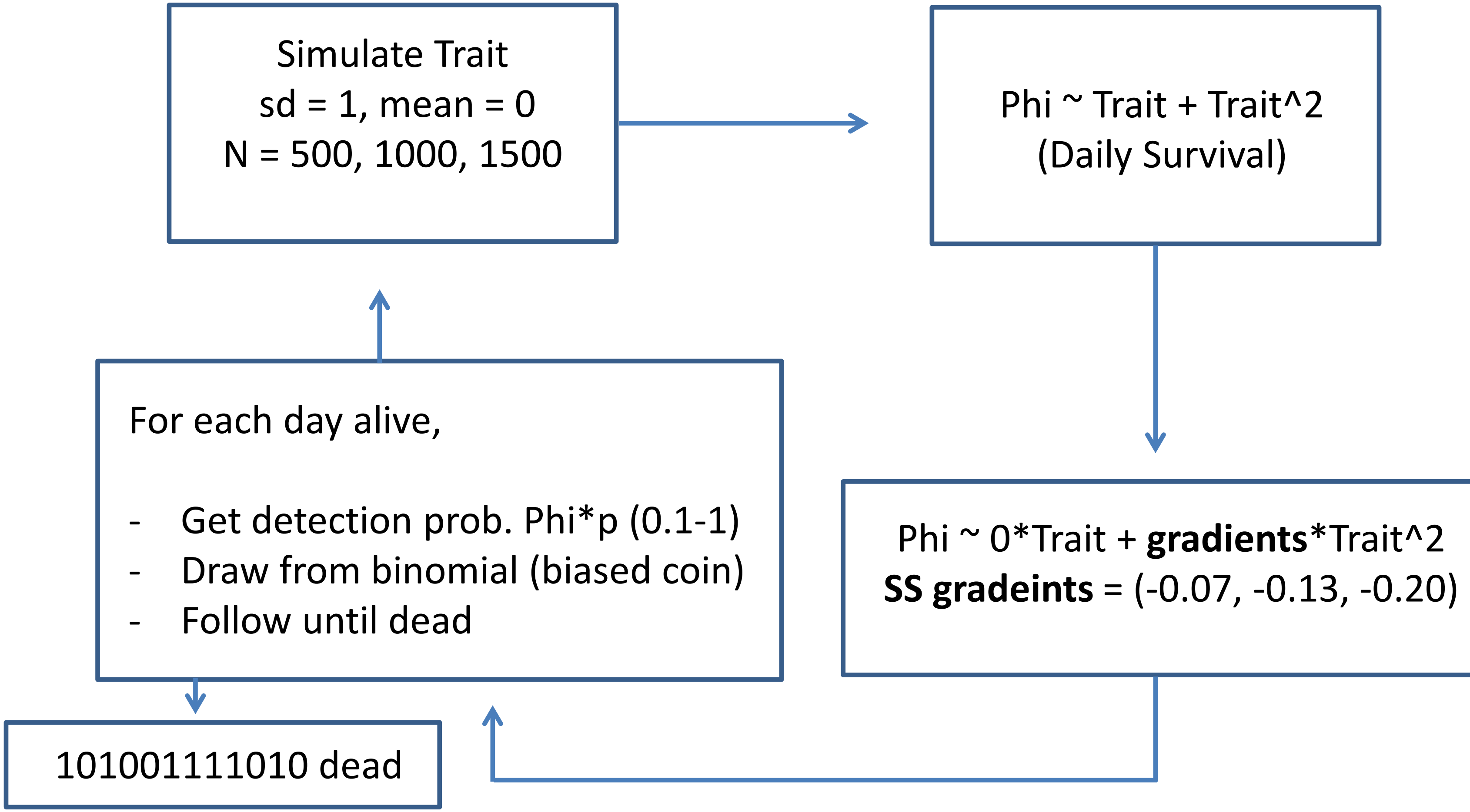
Why not just use **MARK**?

Dis-advantages of MARK

- Need to create a capture history
- Need to perform a model selection procedure to test significance
- Need to control for over-dispersion
- **Is complex**, exists a 900 page book explaining the method.
- Very few selection studies using MARK, early adopter penalty.
- Performs simarly to other methods at moderate p.


MARK
LA
NB
N = 1500
N = 1000
N = 500
0
25
50
75
100
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
beta = 0.00
gamma = -0.07
beta = 0.00
gamma = -0.13
beta = 0.00
gamma = -0.20


MARK
LA
NB
N = 1500
N = 1000
N = 500
100
75
50
25
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
beta = 0.00
gamma = -0.07
beta = 0.00
gamma = -0.13
beta = 0.00
gamma = -0.20


Simulate Trait
sd = 1, mean = 0
N = 500, 1000, 1500
Phi ~ Trait + Trait^2
(Daily Survival)
Phi ~ 0*Trait + **gradients***Trait^2
SS gradeints = (-0.07, -0.13, -0.20)
For each day alive,
- Get detection prob. Phi*p (0.1-1)
- Draw from binomial (biased coin)
- Follow until dead
101001111010 dead

What I did…

- Constructed a simple mark recapture **simulator** in R

- Constructed a simple mark recapture simulator in R
- Generated **1000** simulated datasets for each combination of **SS** gradient (-0.07, -0.13, -0.20) and sample size (**N** = 500, 1000, and 1500 individuals).

- Constructed a simple mark recapture simulator in R
- Generated 1000 simulated datasets for each combination of SS gradient (-0.07, -0.13, -0.20) and sample size (500, 1000, and 1500 individuals).
- While also varying the **probability of being seen (p)** (from 0.1 to 1, by 0.1 increments)

- Constructed a simple mark recapture simulator in R
- Generated 1000 simulated datasets for each combination of SS gradient (-0.07, -0.13, -0.20) and sample size (500, 1000, and 1500 individuals).
- While also varying the probability of being seen (p) (from 0.1 to 1, by 0.1 increments)
- Tested **LA** vs **MARK** on each of the simulated datasets.

- Constructed a simple mark recapture simulator in R
- Generated 1000 simulated datasets for each combination of SS gradient (-0.07, -0.13, -0.20) and sample size (500, 1000, and 1500 individuals).
- While also varying the probability of being seen (p) (from 0.1 to 1, by 0.1 increments)
- Tested LA vs MARK on each of the simulated datasets.
- Examined these tests' ability to detect **SS** by comparing the **AIC of a linear model vs a quadratic model**.

- Constructed a simple mark recapture simulator in R
- Generated 1000 simulated datasets for each combination of SS gradient (-0.07, -0.13, -0.20) and sample size (500, 1000, and 1500 individuals).
- While also varying the probability of being seen (p) (from 0.1 to 1, by 0.1 increments)
- Tested LA vs MARK on each of the simulated datasets.
- Examined these tests' ability to detect SS by comparing the AIC of a linear model vs a quadratic model.
- Finally, I looked at a **real dataset** of damselfly survival in the same way.