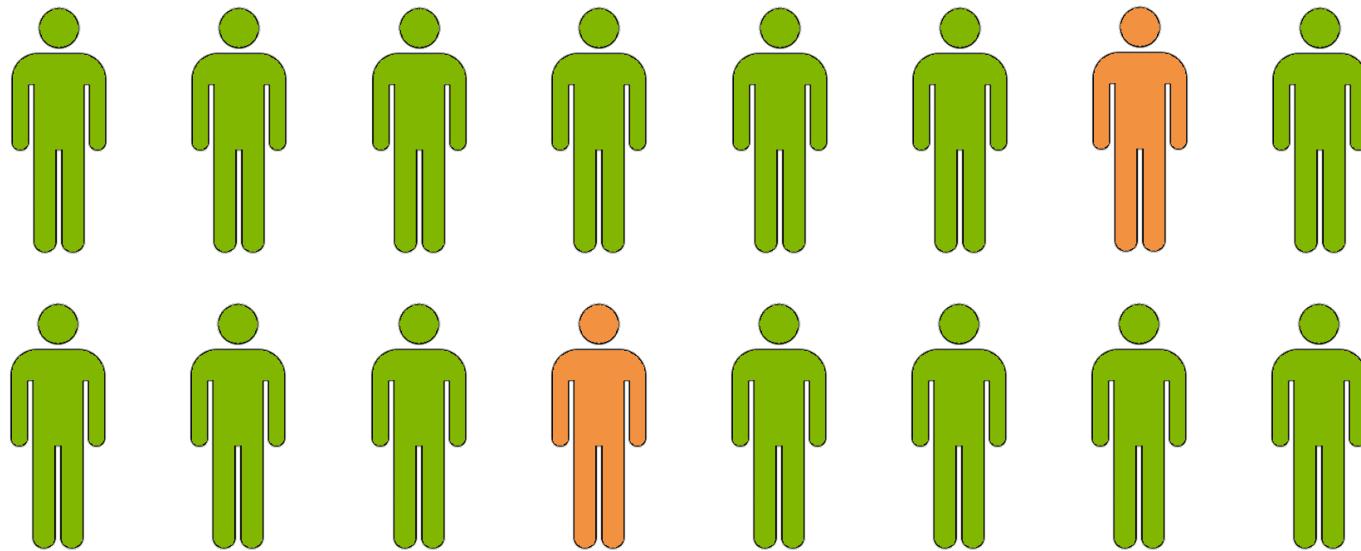


Group Testing

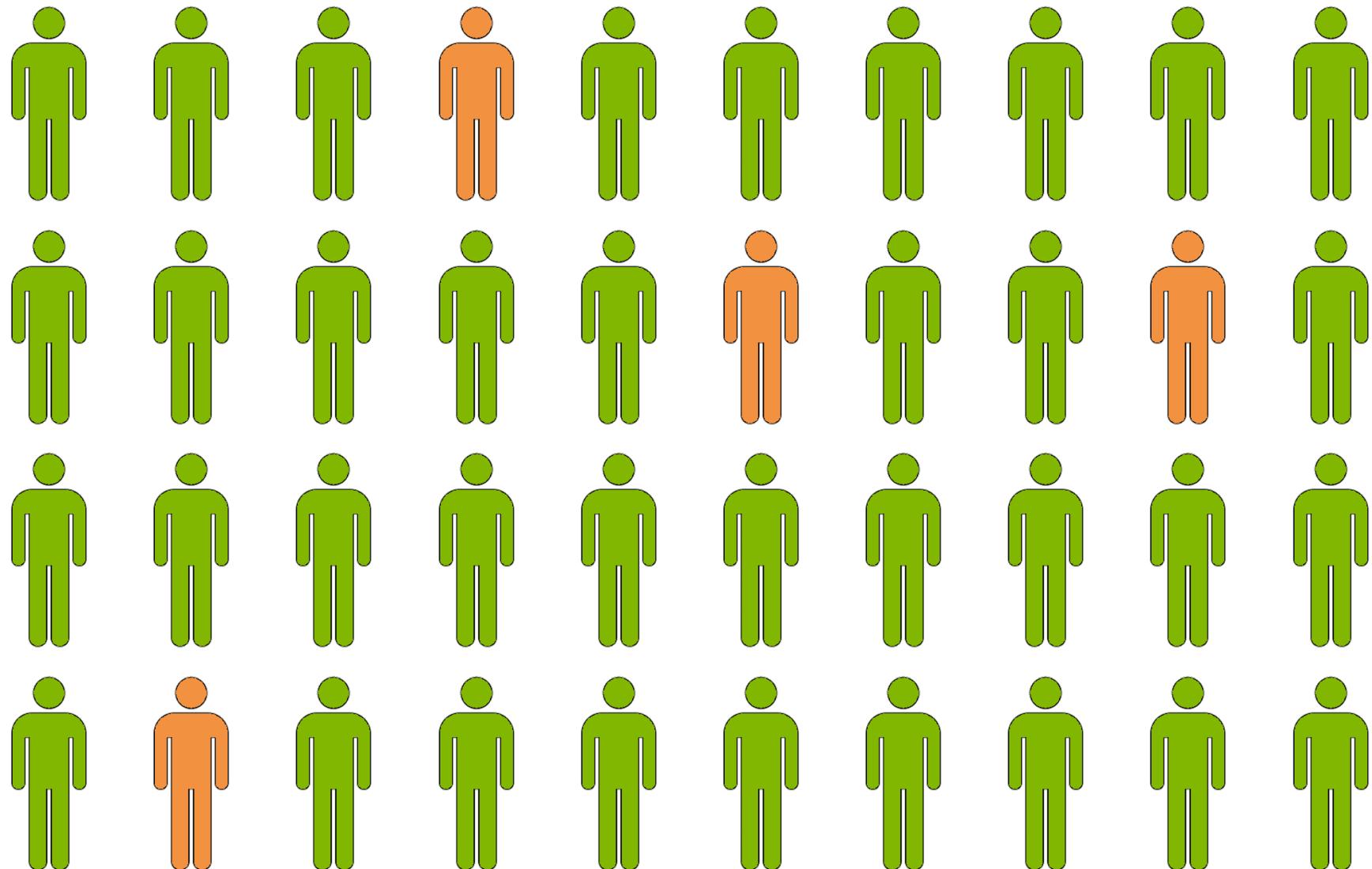
and the Coronavirus



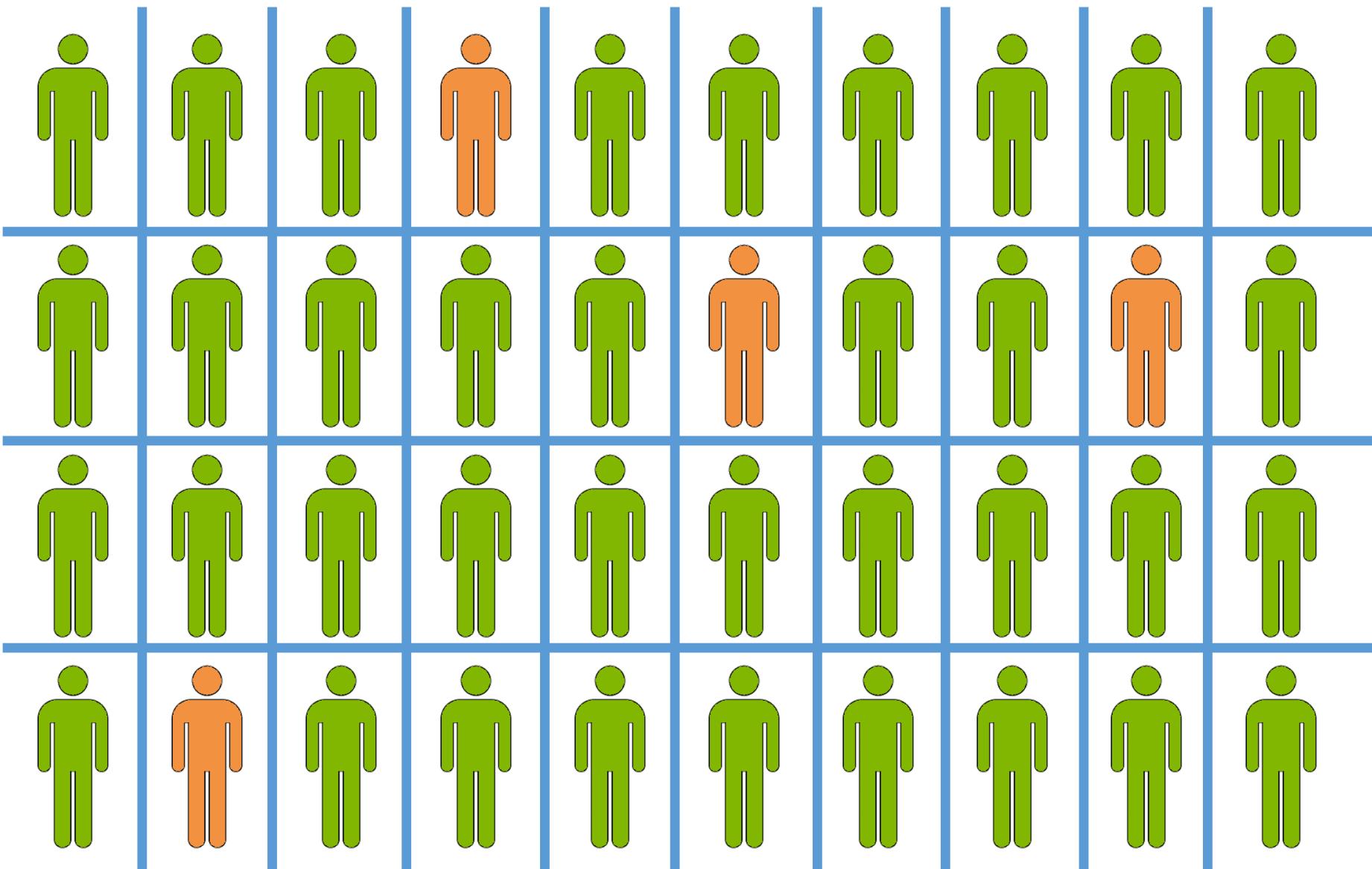
Matthew Aldridge
University of Leeds

Leeds Statistics seminar
April 2020

Group testing



Group testing

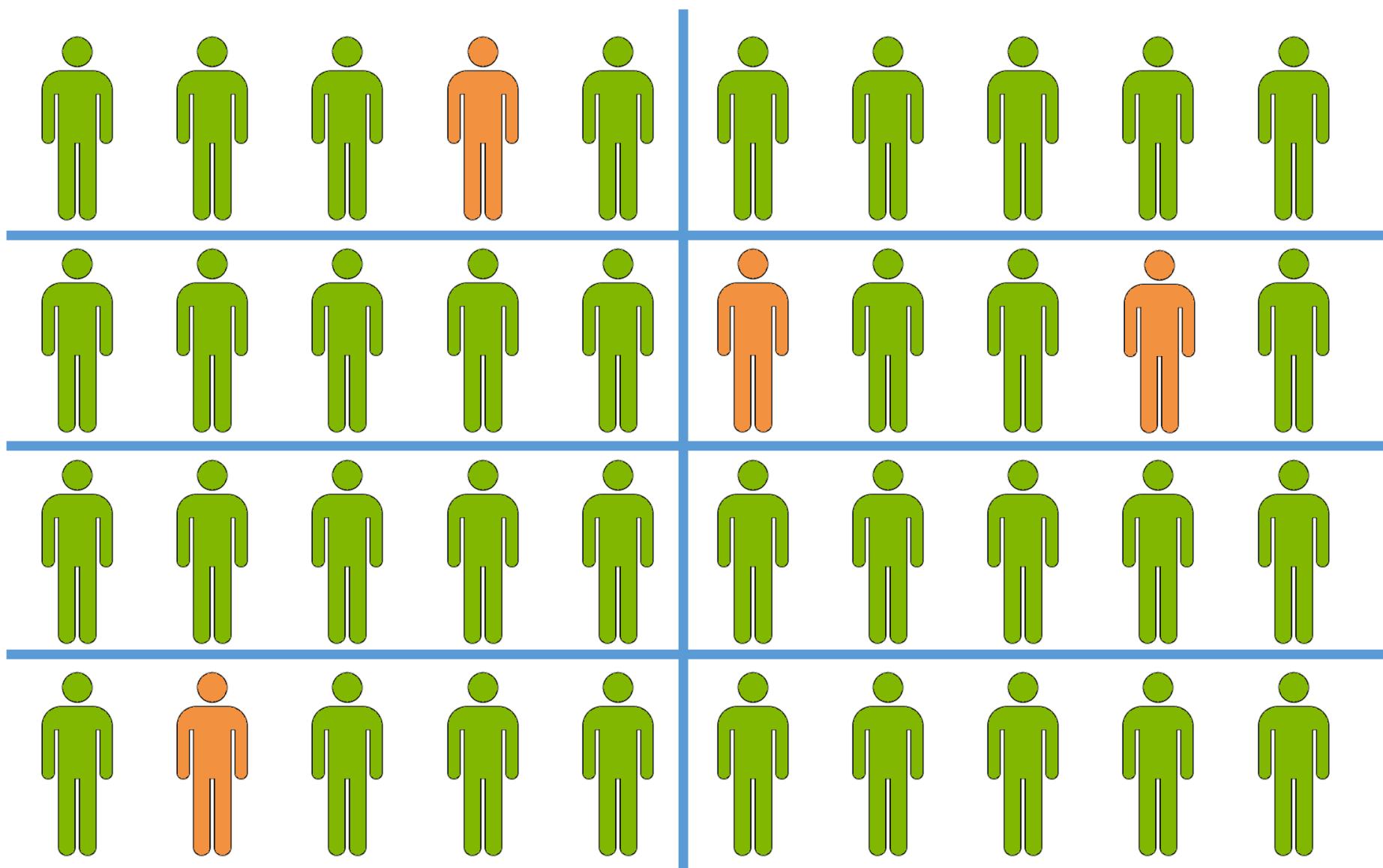


Group testing

Individual testing:

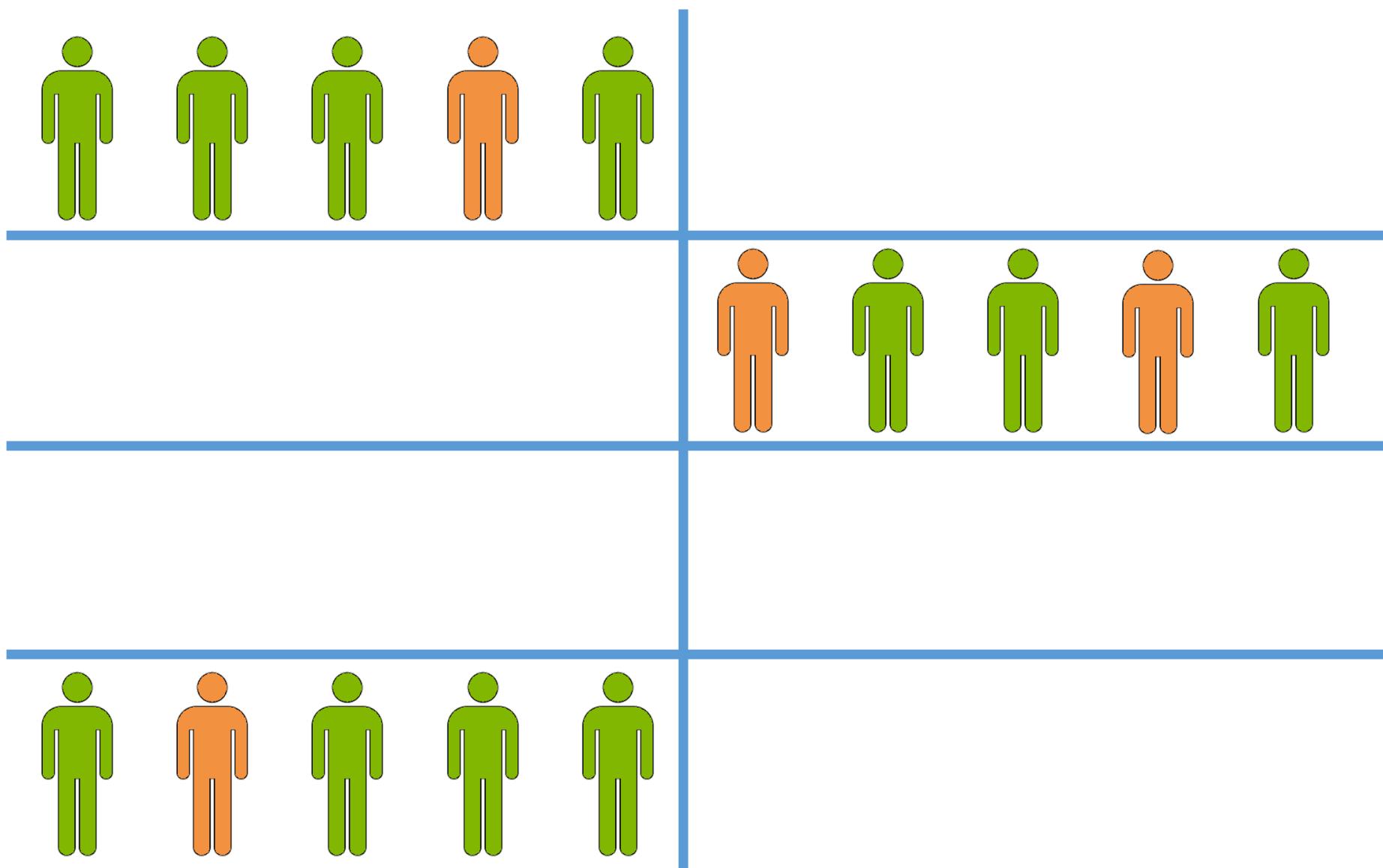
40 soldiers = **40 tests**

Group testing



Dorfman, 1943

Group testing



Dorfman, 1943

Group testing

Individual testing:

40 soldiers = **40 tests**

Dorfman testing:

Stage 1: pooled testing:

8 groups of 5 soldiers = 8 tests

Stage 2: individual testing:

3 groups of 5 soldiers = 15 tests

Total: $8 + 15 = \mathbf{23 \text{ tests}}$

Group testing

n items (soldiers)

p prevalence of “defective items”
(proportion of soldiers with syphilis)

T tests: “Does this group of items contain at least one defective item?” (blood tests)

Group testing

n items

p prevalence

T tests

Given n and p ,
how big does T have to be
to reliably work out
which items were defective?

Part 1

The theory of group testing
(well-established mathematics)

Part 2

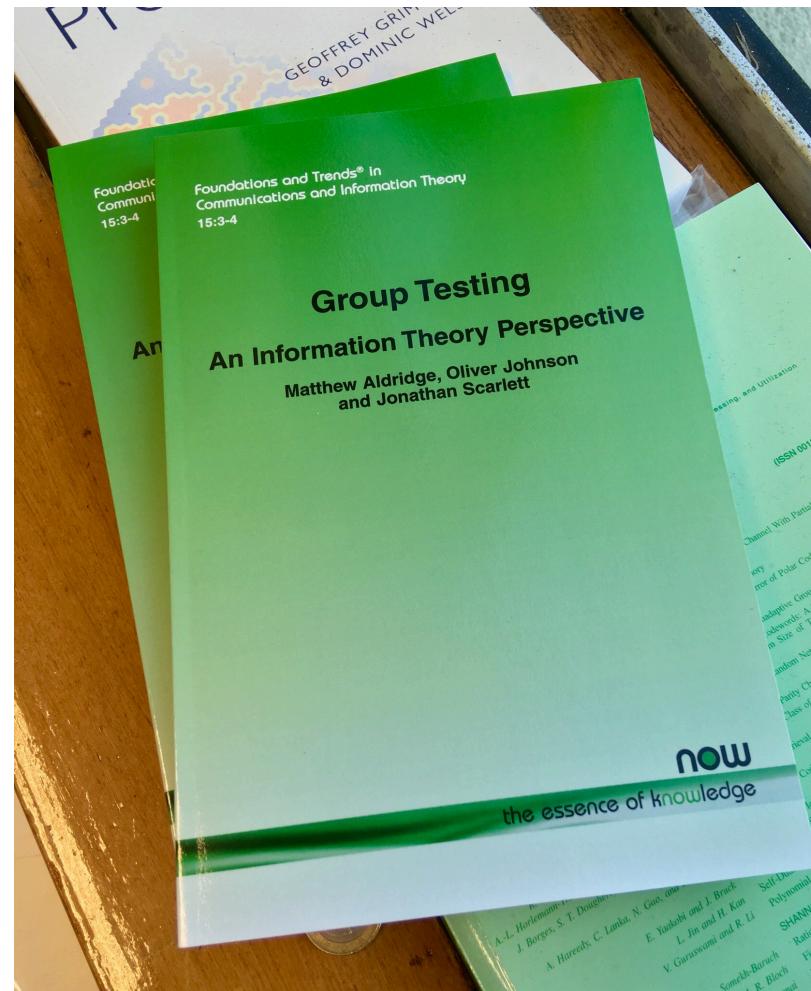
Group testing for the coronavirus
(open questions, speculative)

Part 1

The theory of group testing

M Aldridge, O Johnson and J Scarlett
Group Testing: An Information Theory Perspective
Foundations and Trends in Communications
and Information Theory, 2019

Preprint:
arXiv:1902.06002



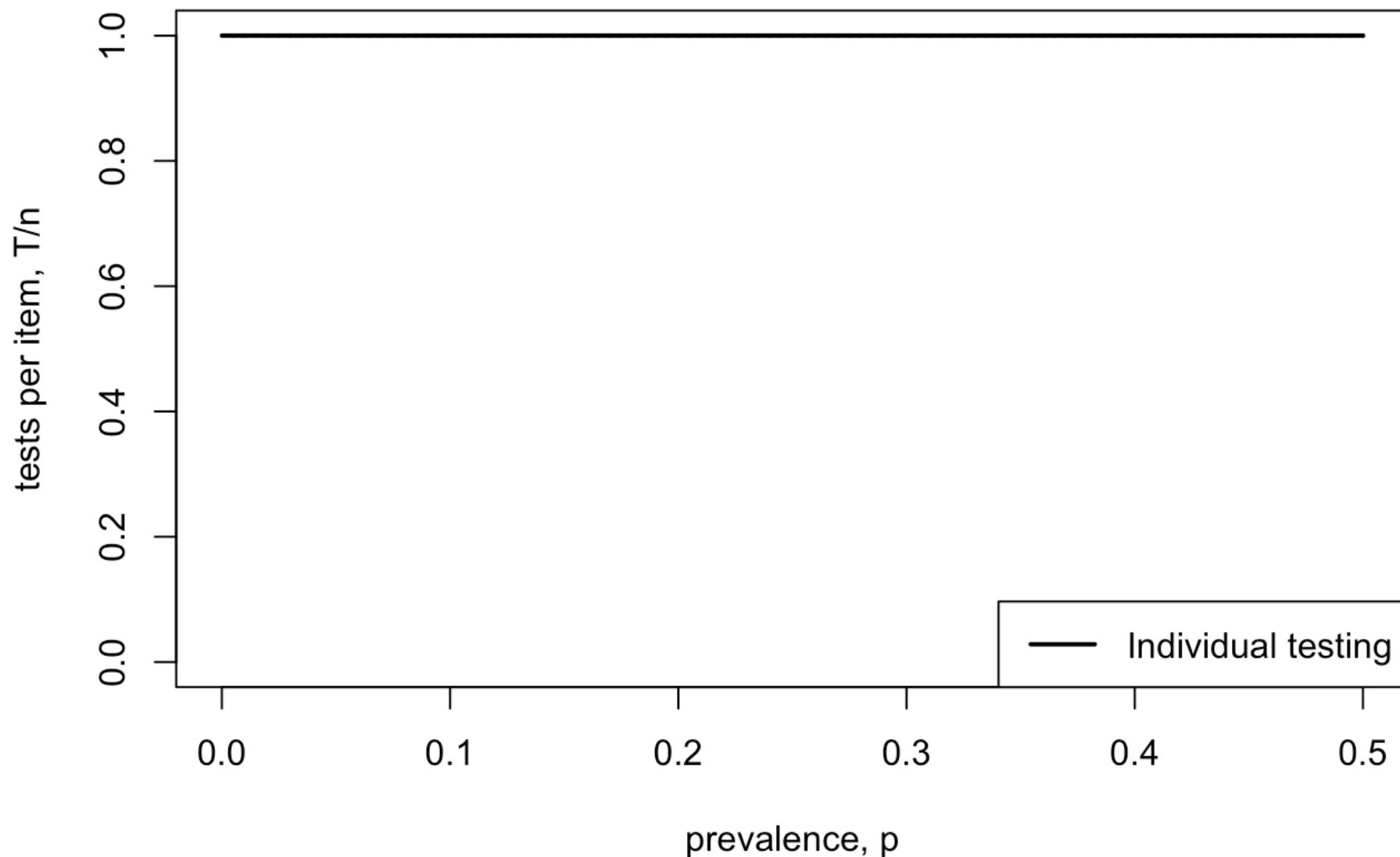
Individual testing

Test each item individually.

n items requires n tests

1 test per item

Group testing: tests per item (combinatorial)



Lower bound

Standard combinatorial and/or information theoretic bounds tell us we need at least

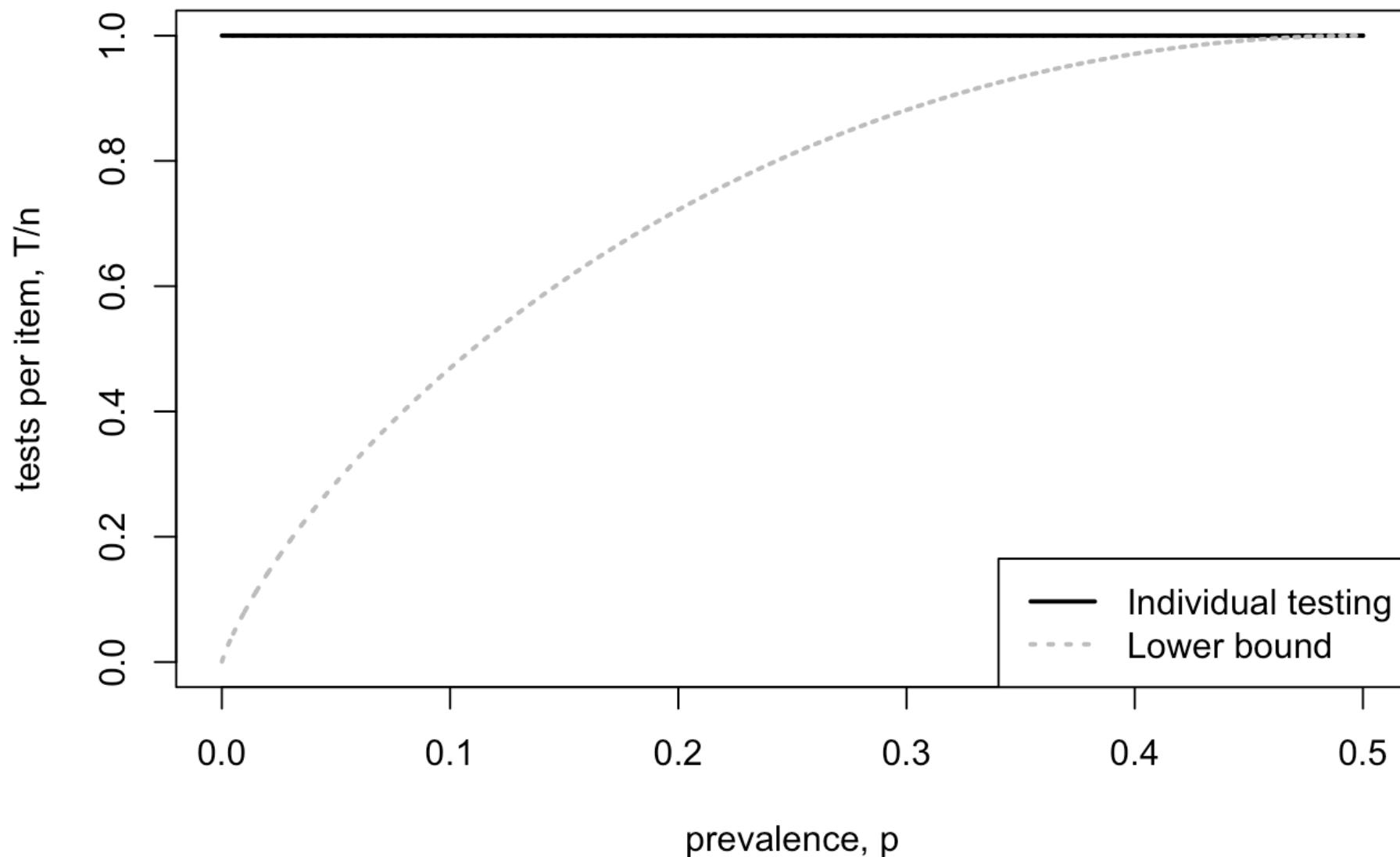
$$T \geq H(p)n$$

tests, where

$$H(p) = p \log_2 \frac{1}{p} + (1 - p) \log_2 \frac{1}{1 - p}$$

is the binary entropy

Group testing: tests per item (combinatorial)



“Combinatorial” group testing

Exactly pn defective items
unrealistic model

Look at worst-case
number of test required
guaranteed performance
even with small n

“Probabilistic” group testing

Each item is defective
IID with probability p
more realistic model

Look at average-case
number of tests required
average-case behavior
representative only for large n

“Combinatorial” group testing

Exactly pn defective items
unrealistic model

Look at worst-case
number of test required
guaranteed performance
even with small n

“Probabilistic” group testing

Each item is defective
IID with probability p
more realistic model

Look at average-case
number of tests required
average-case behavior
representative only for large n

Dorfman's algorithm

(Dorfman, 1943)

Split into n/s sets of size s

For each set:

Test the whole set.

If the test is **positive**:

test each item individually to find which items are defective and nondefective

If the test is **negative**:

all items are nondefective.

Dorfman's algorithm

(Dorfman, 1943)

Each set is tested once together.

There are n/s sets.

Each test containing a defective item is tested s times.

If all p_n defective items are in different sets,
this is pns tests.

Total number of tests:

$$\frac{n}{s} + pns = \left(\frac{1}{s} + ps\right)n$$

Dorfman's algorithm

Total number of tests:

$$\frac{n}{s} + pns = \left(\frac{1}{s} + ps\right)n$$

What size should we pick the sets to be?

What s minimizes this equation?

Dorfman's algorithm

Total number of tests:

$$\frac{n}{s} + pns = \left(\frac{1}{s} + ps\right)n$$

What size should we pick the sets to be?

What s minimizes this equation?

Quick calculus exercise: $\sqrt{1/p}$

Dorfman's algorithm

Total number of tests:

$$\frac{n}{s} + pns = \left(\frac{1}{s} + ps\right)n$$

What size should we pick the sets to be?

What s minimizes this equation?

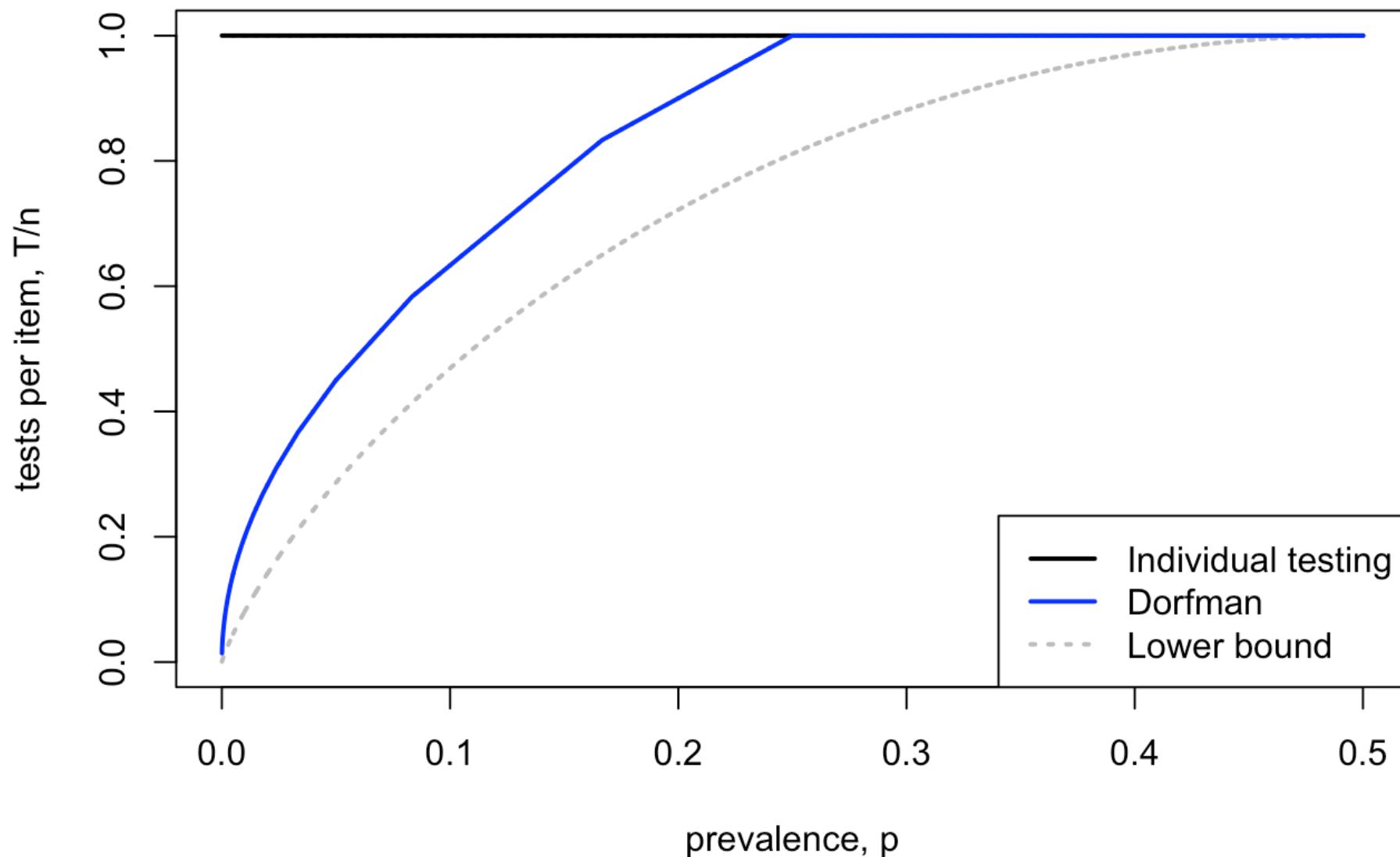
Quick calculus exercise: $\sqrt{1/p}$

But it needs to be an integer:

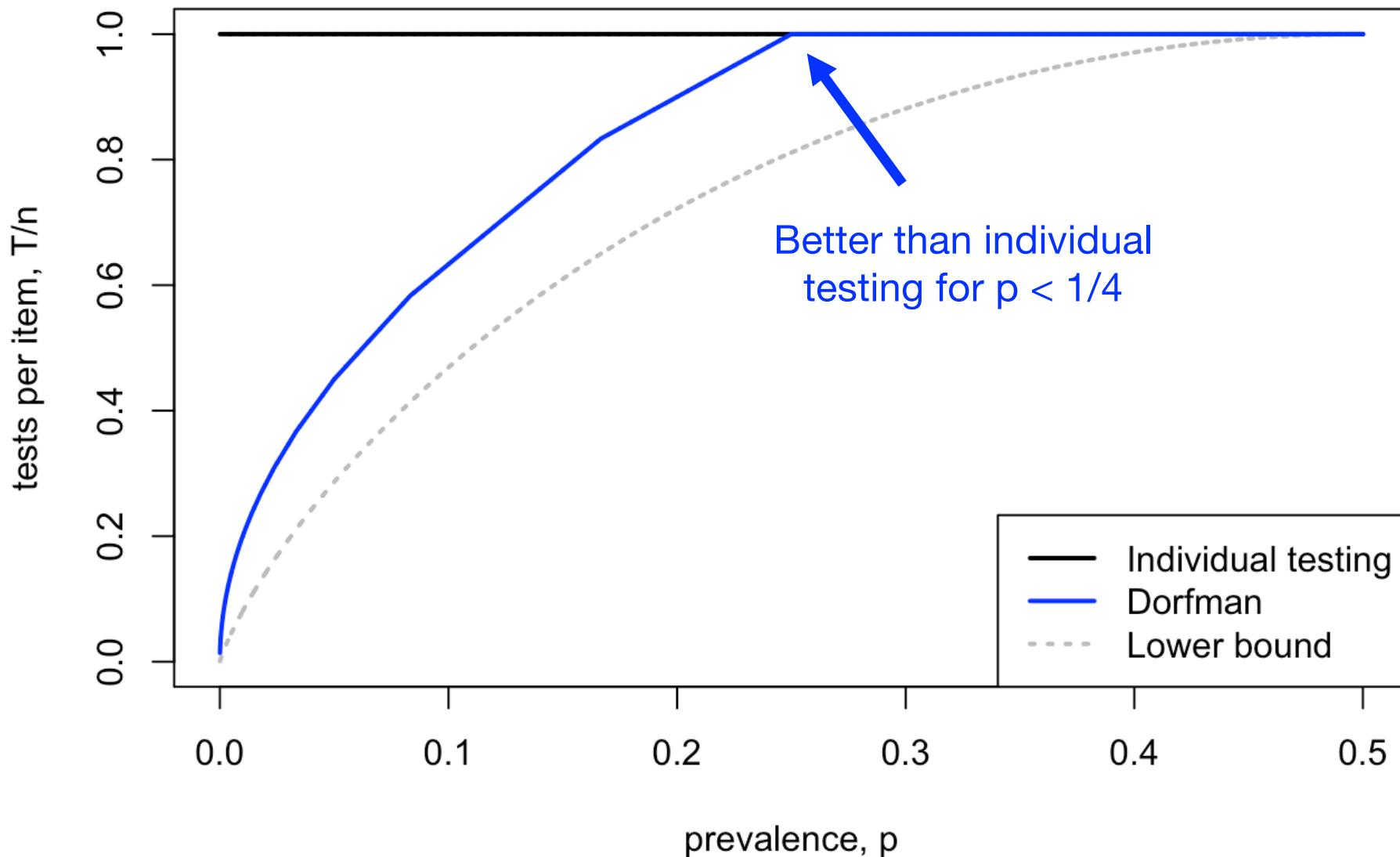
$$\left\lceil \sqrt{\frac{1}{p} + \frac{1}{4}} \right\rceil$$

← nearest integer

Group testing: tests per item (combinatorial)



Group testing: tests per item (combinatorial)



Rate

(Baldassini–Johnson–Aldridge, 2013)

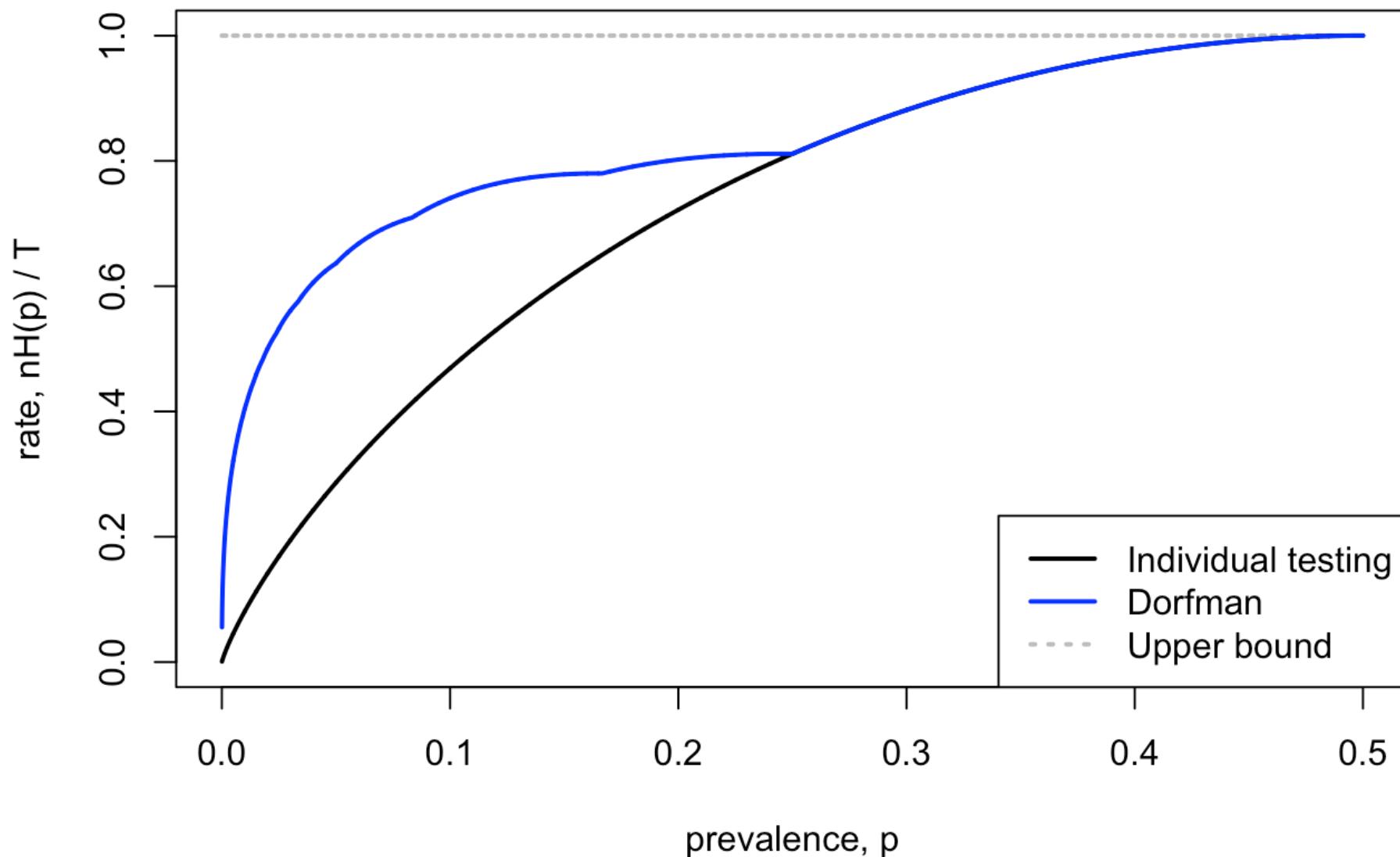
We often prefer to plot the **rate** $H(p)n/T$.

The rate tells us:

How many “bits of information” we learn from each test

How close we are to the entropy bound

Group testing rates (combinatorial)



Dorfman's algorithm

(Dorfman, 1943)

Split into n/s sets of size s

For each set:

Test the whole set.

If the test is **positive**:

test each item individually to find which items are defective and nondefective

If the test is **negative**:

all items are nondefective.

Dorfman's algorithm

(Dorfman, 1943)

Split into n/s sets of size s

For each set:

Test the whole set.

If the test is **positive**:

test each item individually to find which items are defective and nondefective

If the test is **negative**:

all items are nondefective.

*obviously
correct*

Dorfman's algorithm

(Dorfman, 1943)

Split into n/s sets of size s

seems
fine

For each set:

Test the whole set.

If the test is **positive**:

test each item individually to find which items are defective and nondefective

If the test is **negative**:

all items are nondefective.

obviously
correct

Dorfman's algorithm

(Dorfman, 1943)

Split into n/s sets of size s

seems
fine

For each set:

Test the whole set.

a bit crude

~~If the test is positive.~~

~~test~~ each item individually to find which items are defective and nondefective

If the test is ~~positive~~:

all items are nondefective.

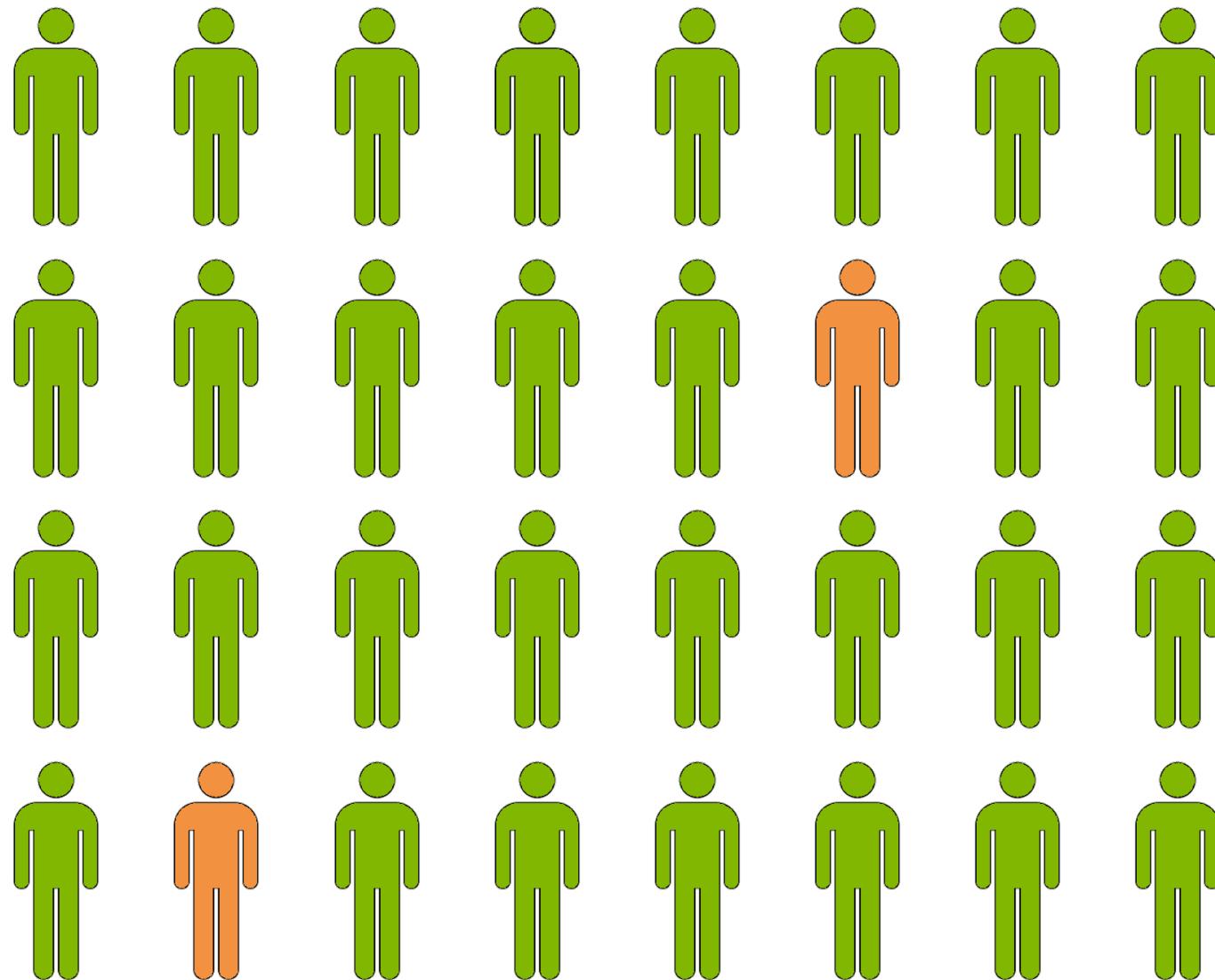
obviously
correct

Binary splitting

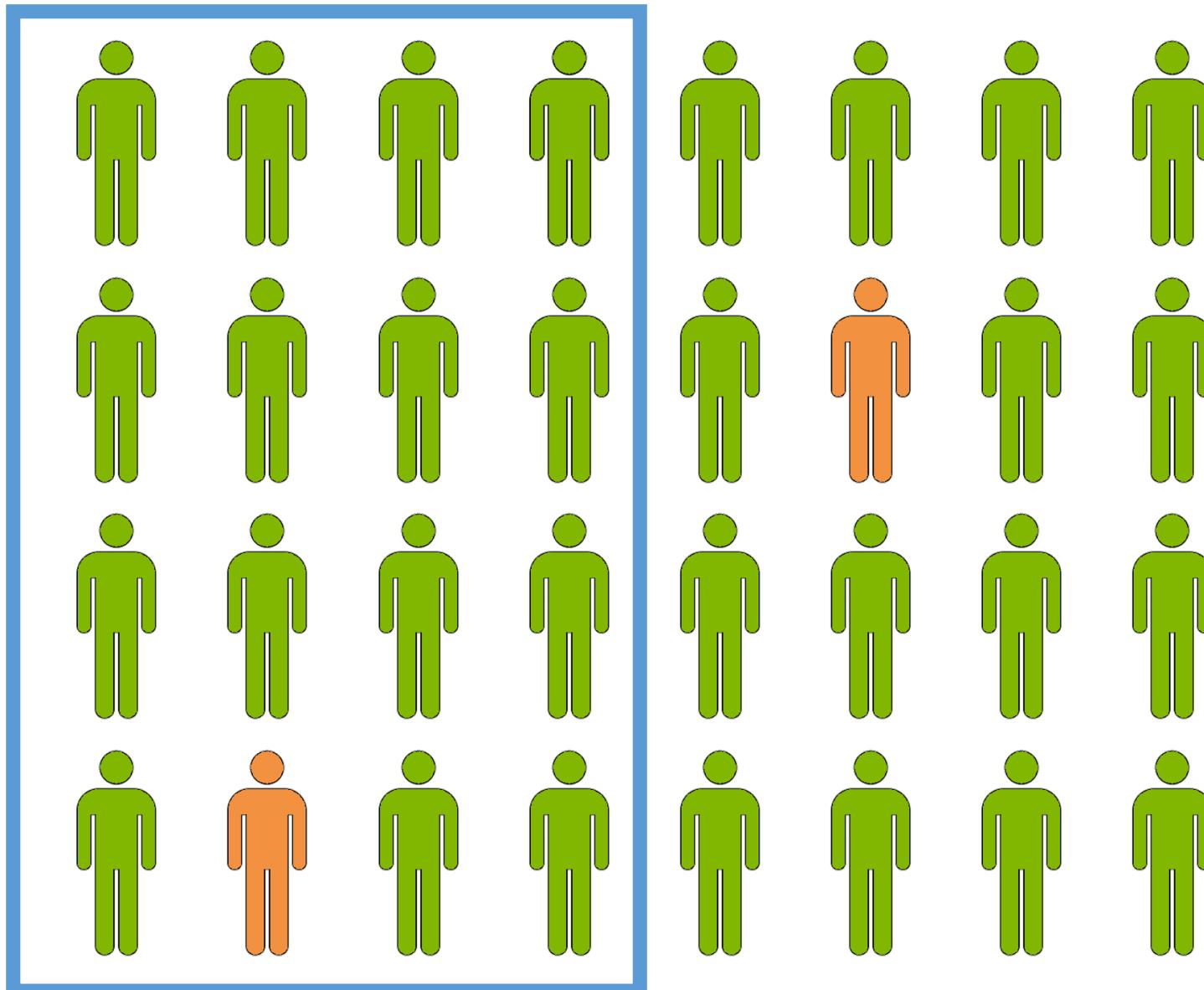
(Sobel & Groll, 1959)

Keep splitting the set in half,
keeping a half that has
a defective item in it

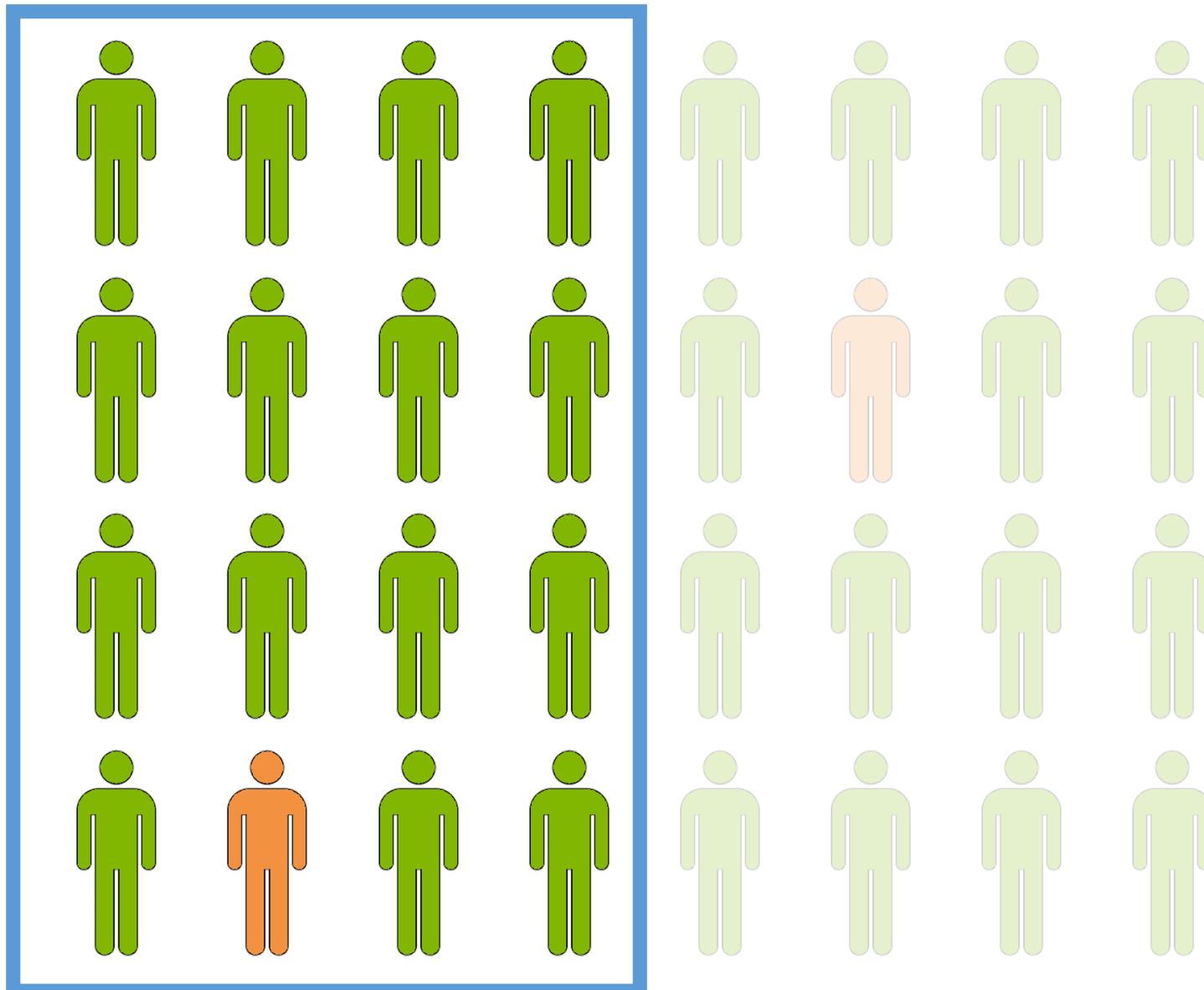
Binary splitting



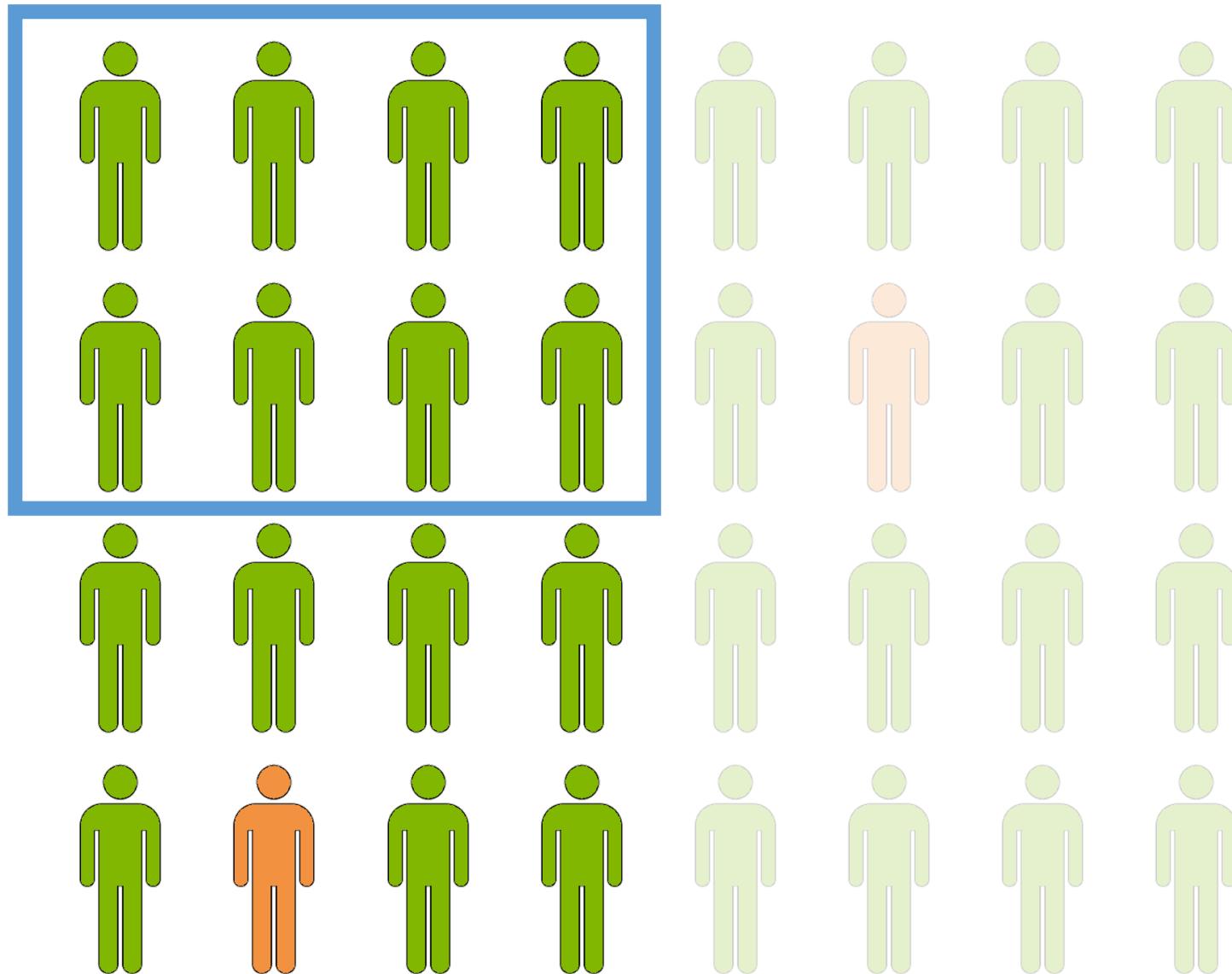
Binary splitting



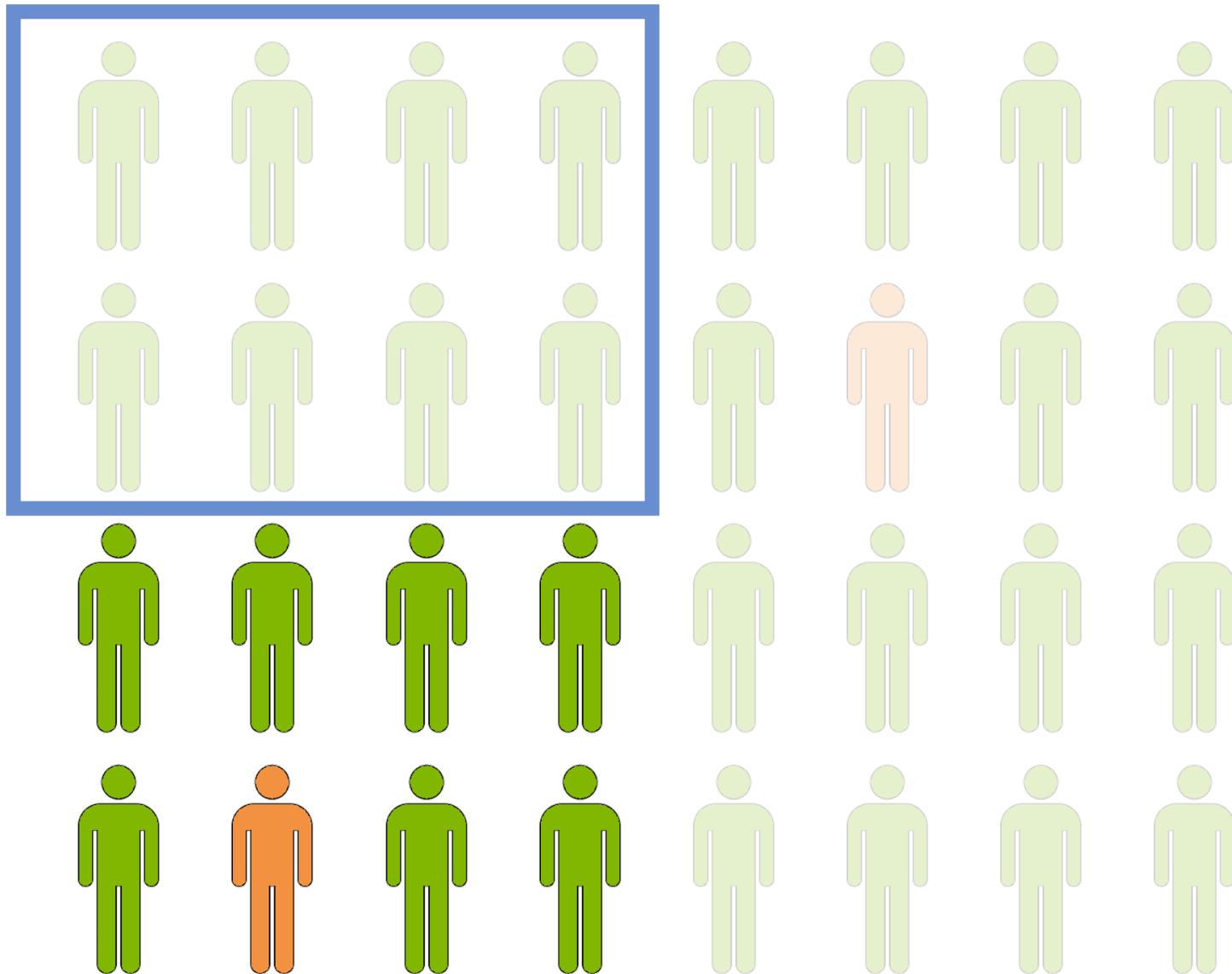
Binary splitting



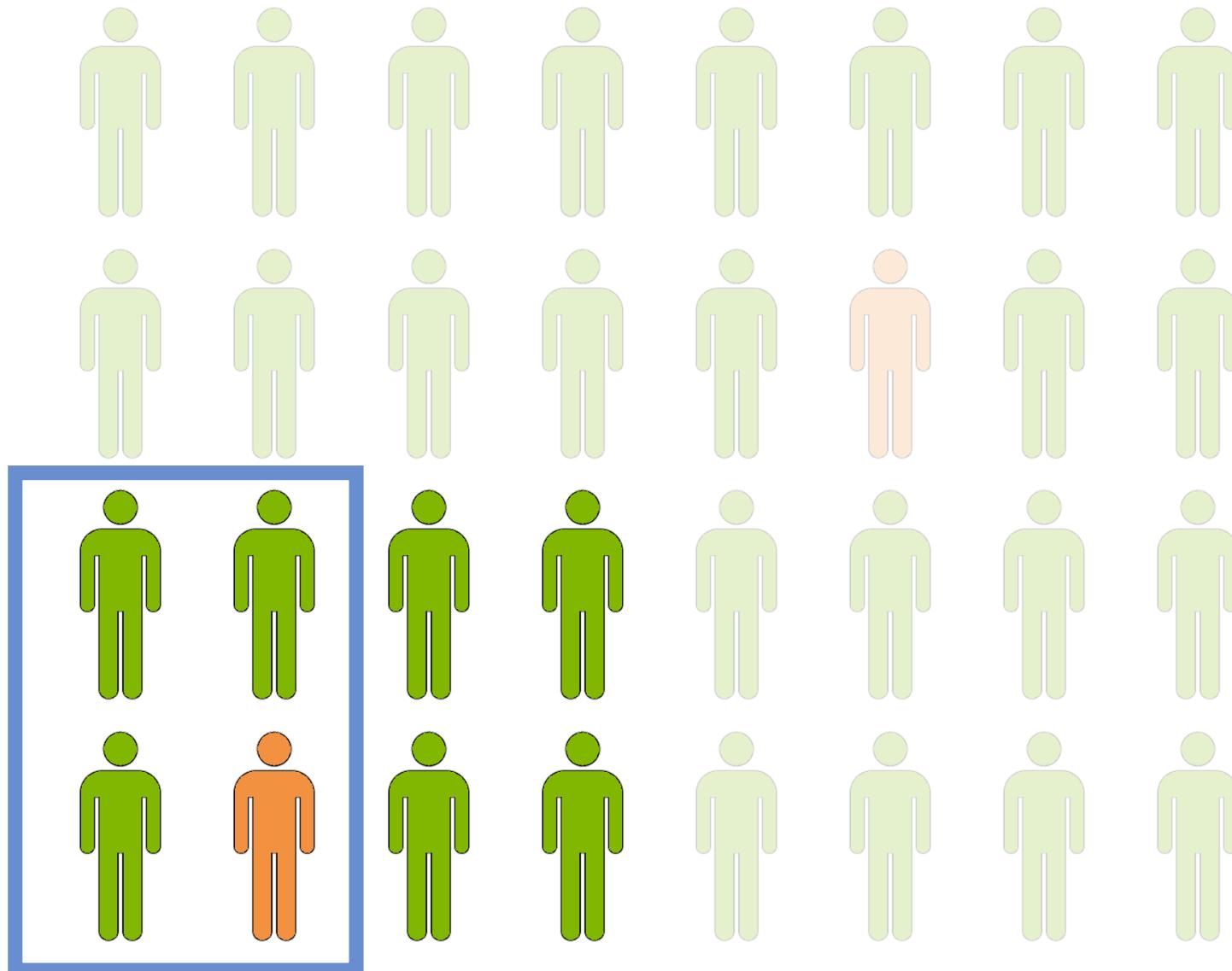
Binary splitting



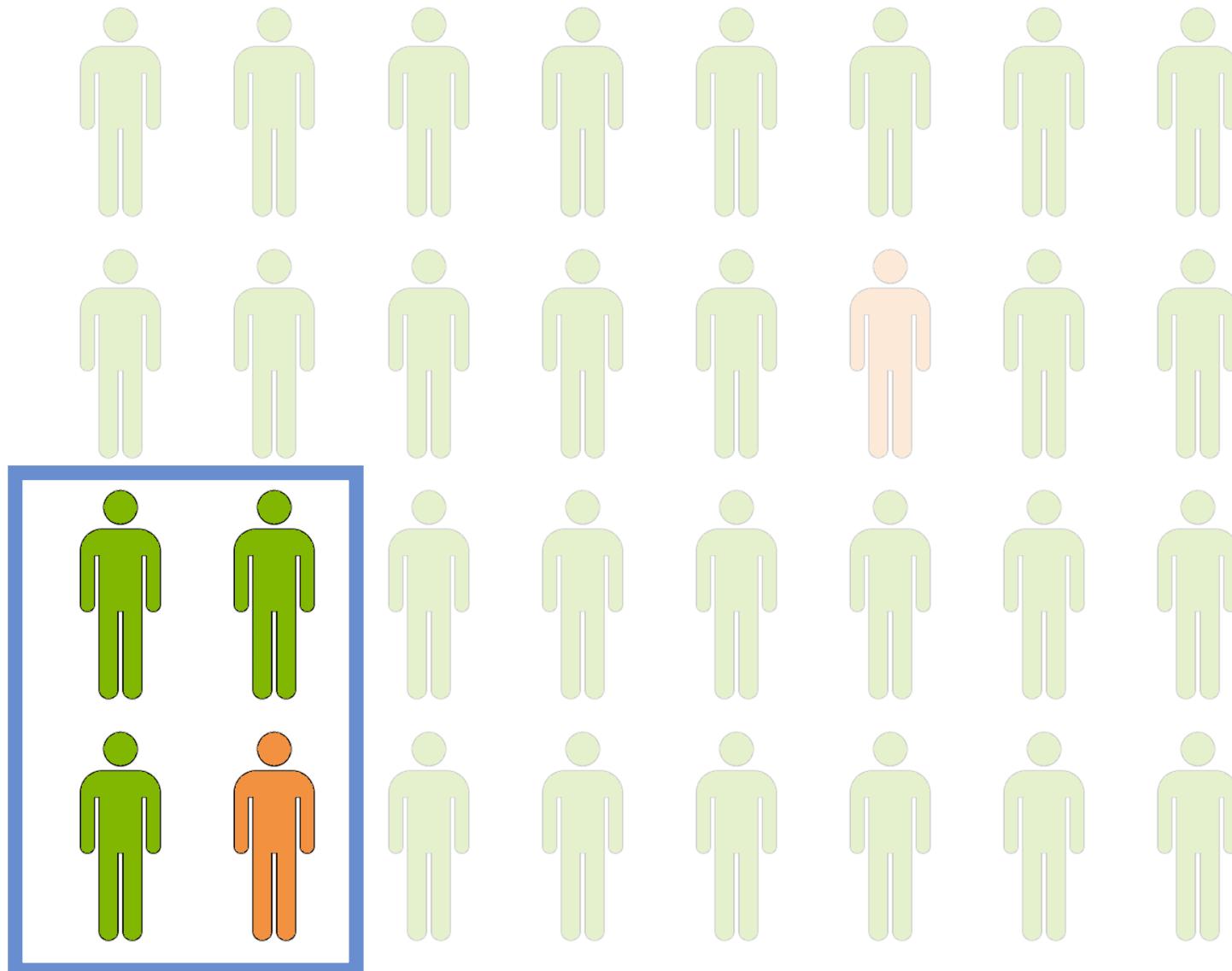
Binary splitting



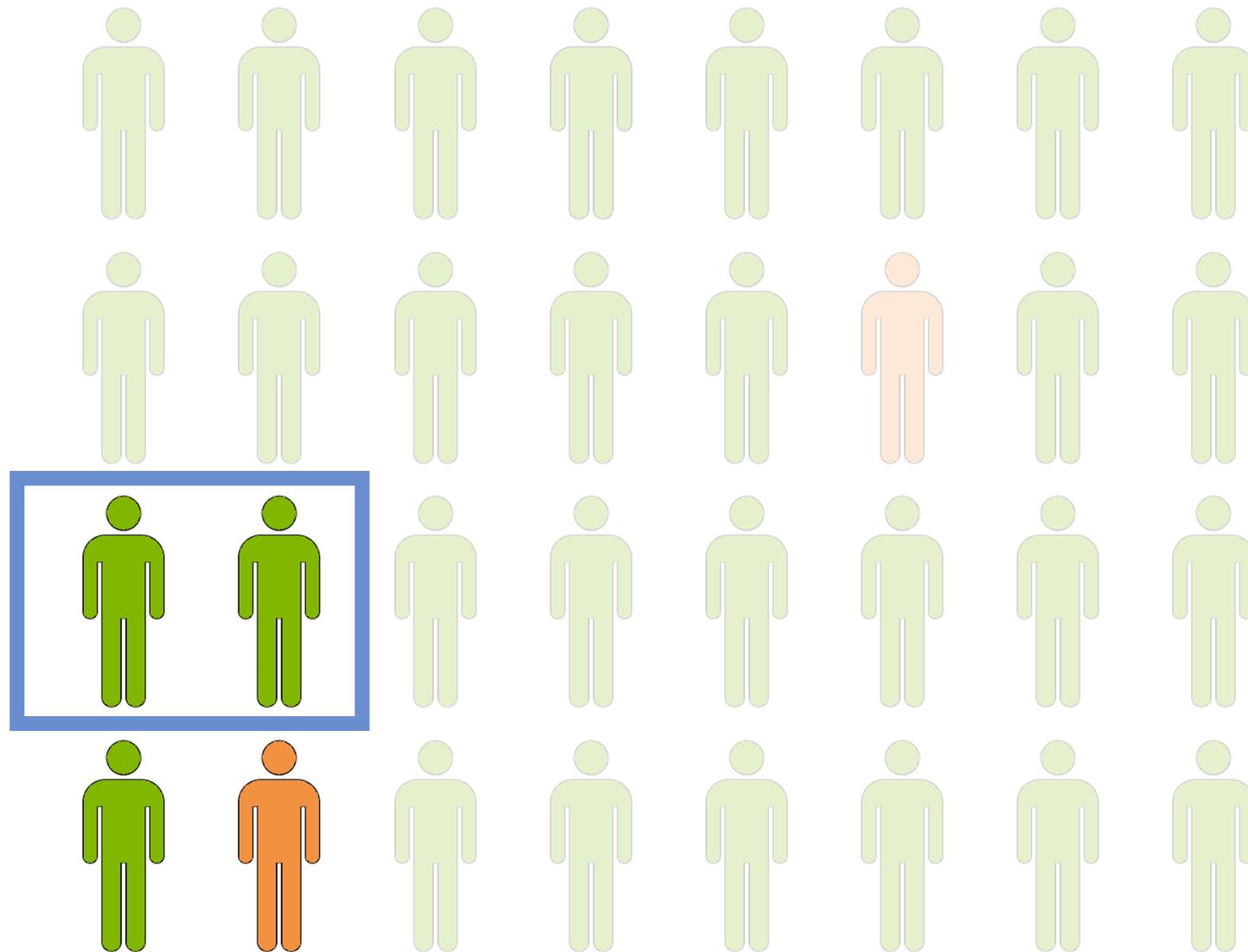
Binary splitting



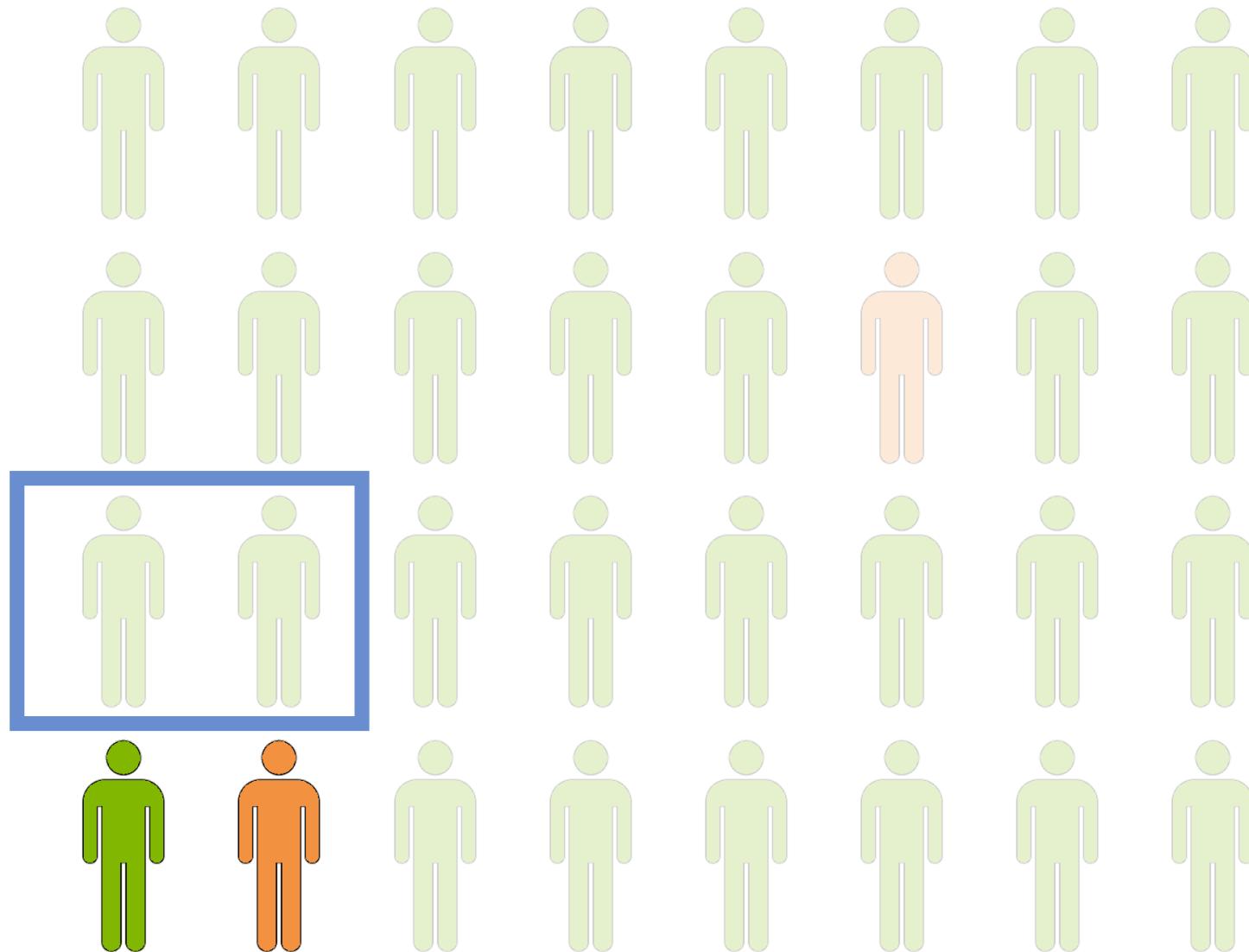
Binary splitting



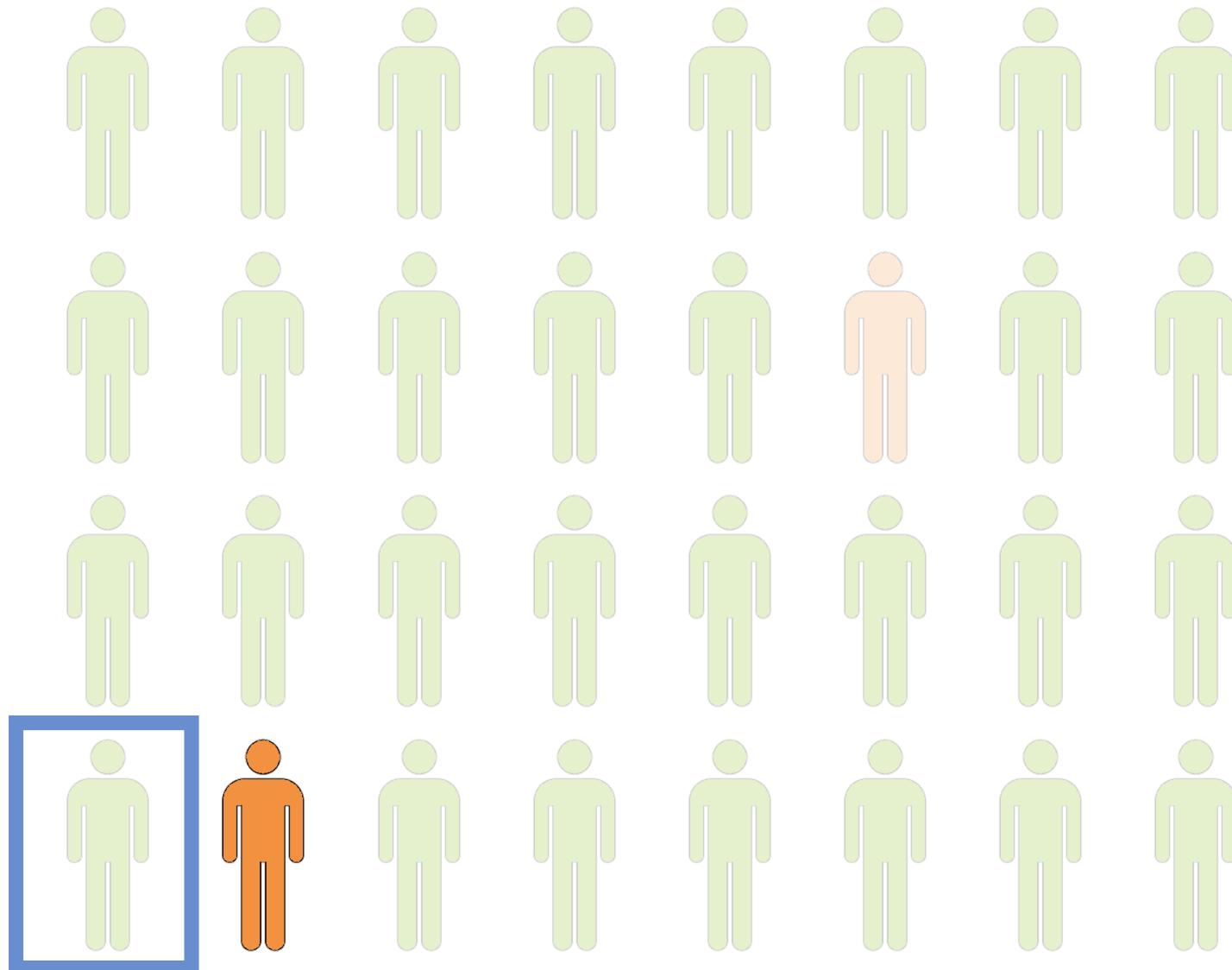
Binary splitting



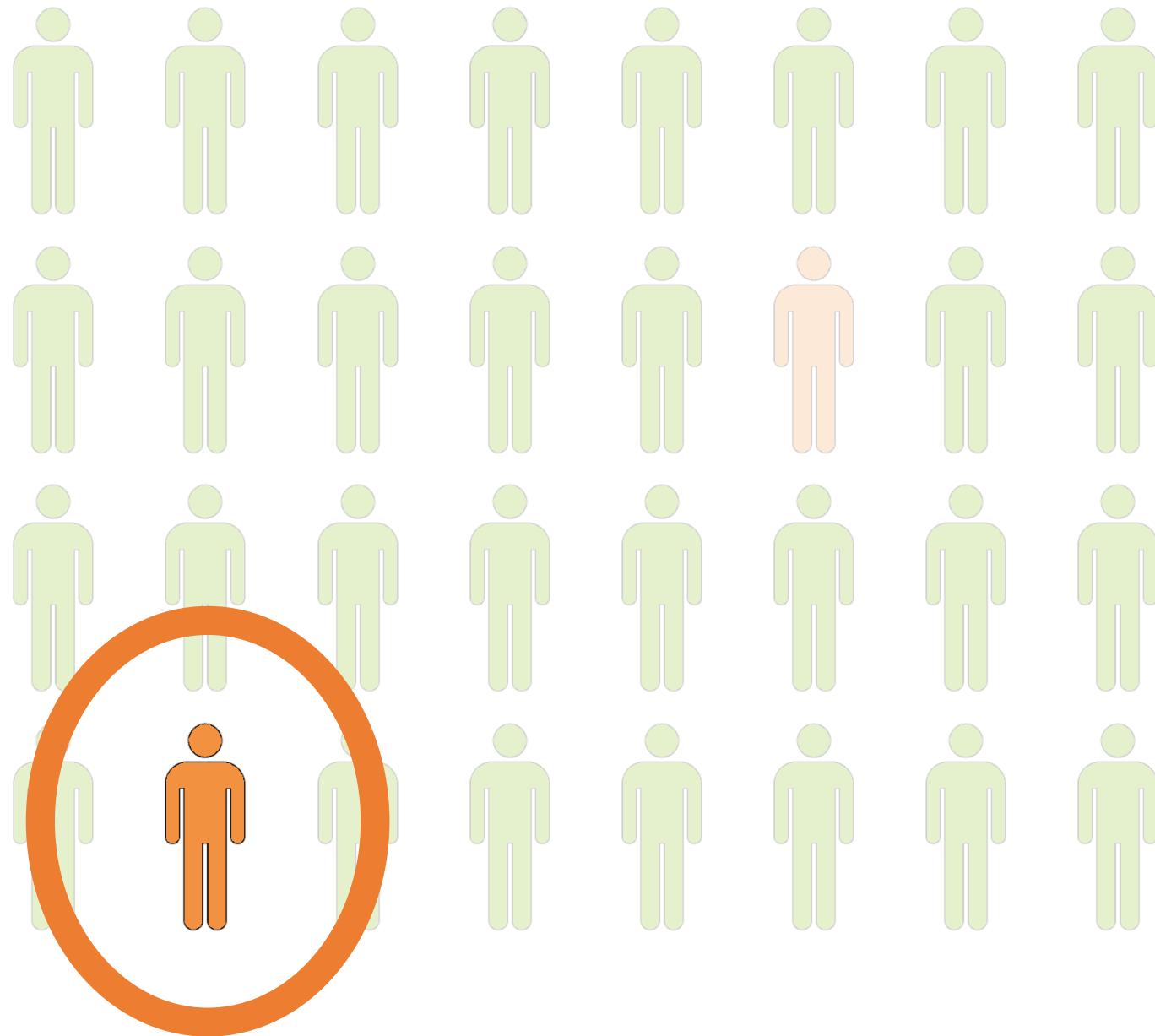
Binary splitting



Binary splitting



Binary splitting



Binary splitting

(Sobel & Groll, 1959)

Find a defective item from a set A containing at least one defective:

- 1)** If $|A| = 1$, that item is defective. Halt.
- 2)** Let B consist of half the elements of A . Test B .

If the test is positive:

 set $A := B$, and return to 1)

If the test is negative:

 set $A := B \setminus A$, and return to 1)

Binary splitting

(Sobel & Groll, 1959)

Find a defective item from a set A containing at least one defective:

Requires $\log_2 |A|$ tests

as that's the number of times
we have to cut in half
to get down to a single item

Generalized binary splitting

(Hwang, 1972; Zaman–Pippenger, 2016, Aldridge 2019)

Pick a set of size s

Test the whole set.

If the test is **positive**:

binary split to find a defective item

If the test is **negative**:

all items are nondefective.

Generalized binary splitting

(Hwang, 1972; Zaman–Pippenger, 2016, Aldridge 2019)

Each time through the algorithm we find
either

s nondefective items in 1 test

or

1 defective item

and up to $s - 1$ nondefective items
in $1 + \log_2 s$ tests

Generalized binary splitting

(Hwang, 1972; Zaman–Pippenger, 2016, Aldridge 2019)

Each time through the algorithm we find
either

s nondefective items in 1 test

or

1 defective item

and up to $s - 1$ nondefective items
in $1 + \log_2 s$ tests

Generalized binary splitting

(Aldridge 2019)

Each time through the algorithm we find
either

s nondefective items in 1 test

or

1 defective item

~~and up to $s - 1$ nondefective items~~

in $1 + \log_2 s$ tests

*"worst-case":
assume we're
always unlucky*

Generalized binary splitting

(Aldridge 2019)

So we require:

1 test for every s nondefectives

$1 + \log_2 s$ tests for every defective

Total number of tests:

$$\begin{aligned} T &= \frac{1}{s} (1 - p)n + (1 + \log_2 s)pn \\ &= \left(\frac{1}{s} + \left(-\frac{1}{s} + 1 + \log_2 s \right)p \right) n \end{aligned}$$

Generalized binary splitting

(Aldridge 2019)

Total number of tests:

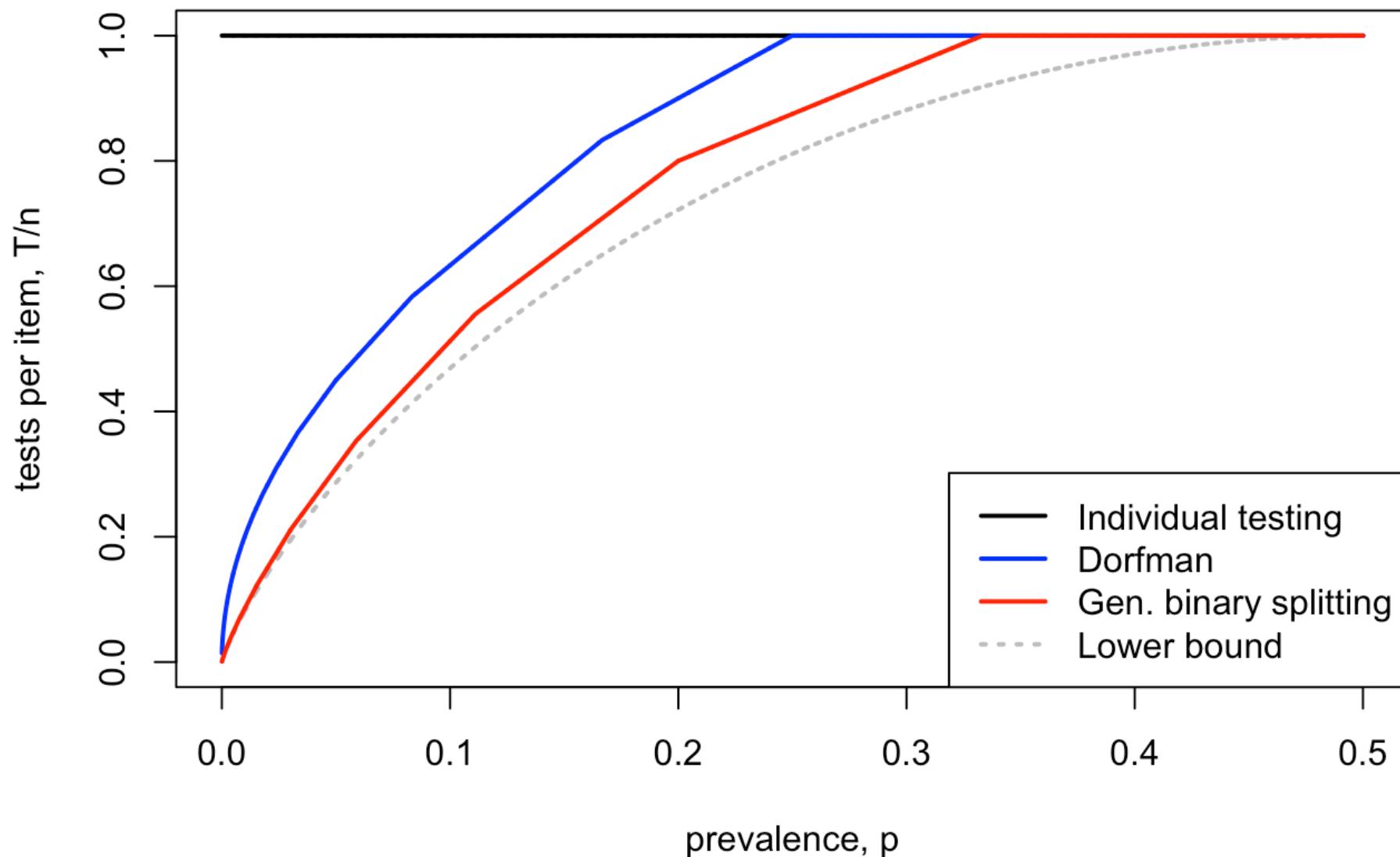
$$T = \left(\frac{1}{s} + \left(-\frac{1}{s} + 1 + \log_2 s \right) p \right) n$$

Optimal size of set is (exercise):

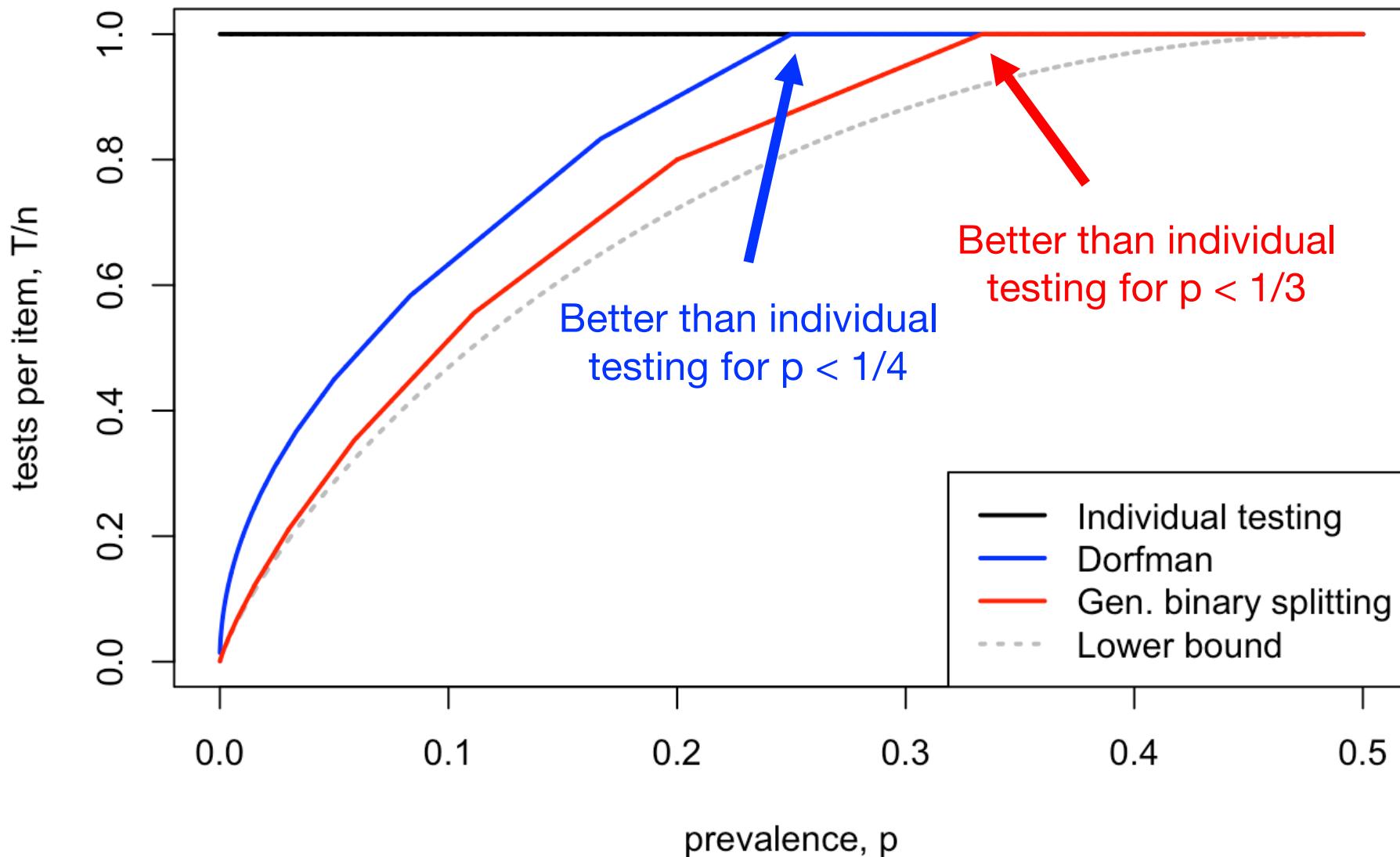
$$s = \left\lfloor \frac{1}{p} - 1 \right\rfloor_2$$

← Round down to power of 2 (useful for binary splitting)

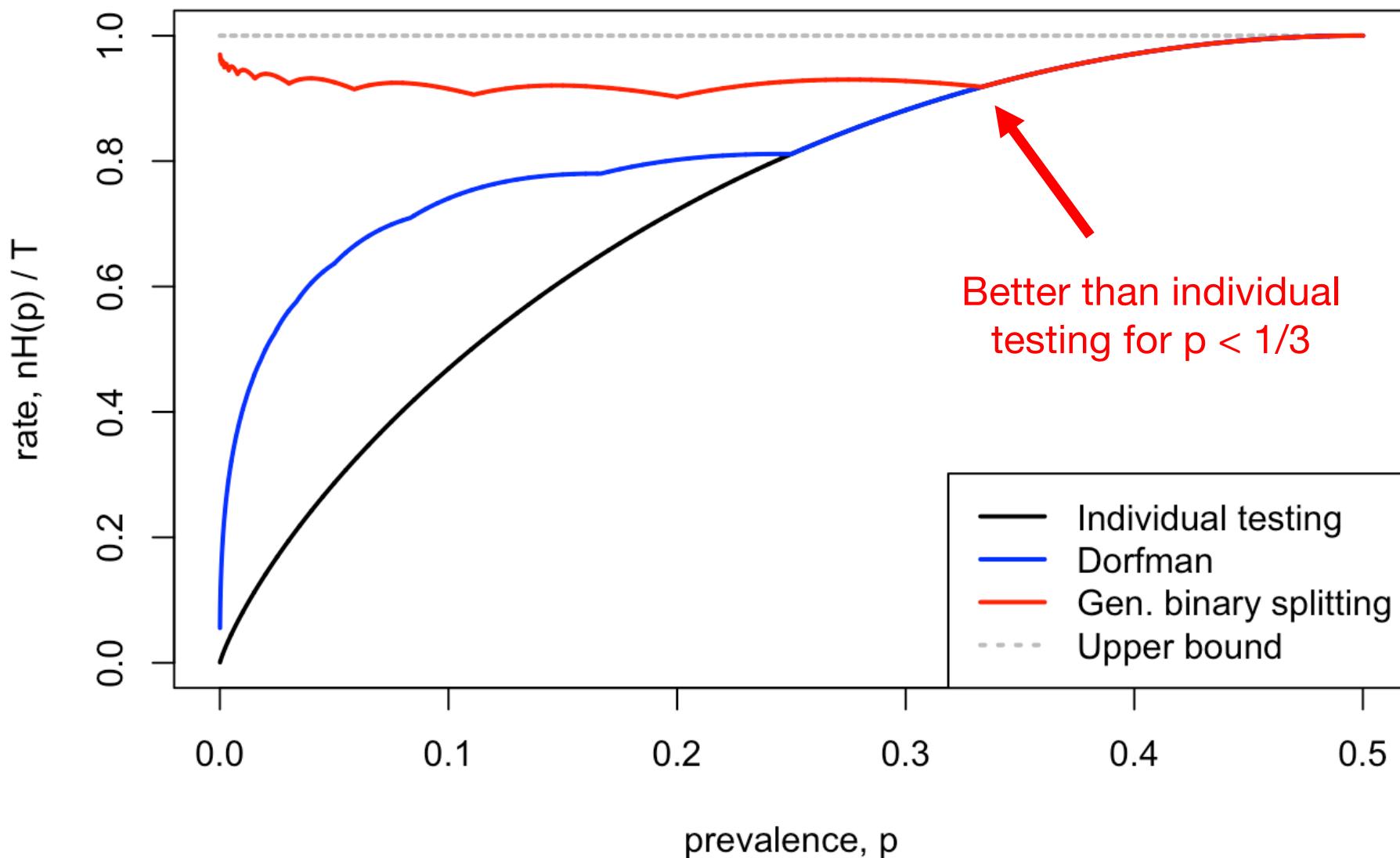
Group testing: tests per item



Group testing: tests per item



Group testing rates



Hu–Hwang–Wang conjecture

(Hu–Hwang–Wang, 1981)

Improving on individual testing
for $p < 1/3$ is “best possible”:

For $p \geq 1/3$, individual testing is optimal
for combinatorial group testing.

Hu–Hwang–Wang conjecture

(Hu–Hwang–Wang, 1981)

Improving on individual testing
for $p < 1/3$ is “best possible”:

For $p \geq 1/3$, individual testing is optimal
for combinatorial group testing.

Best result (Riccio–Colborn, 2000):

For $p \geq 0.369$, individual testing is optimal
for combinatorial group testing.

“Combinatorial” group testing

Exactly pn defective items
unrealistic model

Look at worst-case
number of test required
guaranteed performance
even with small n

“Probabilistic” group testing

Each item is defective
IID with probability p
more realistic model

Look at average-case
number of tests required
average-case behavior
representative only for large n

Dorfman: average case

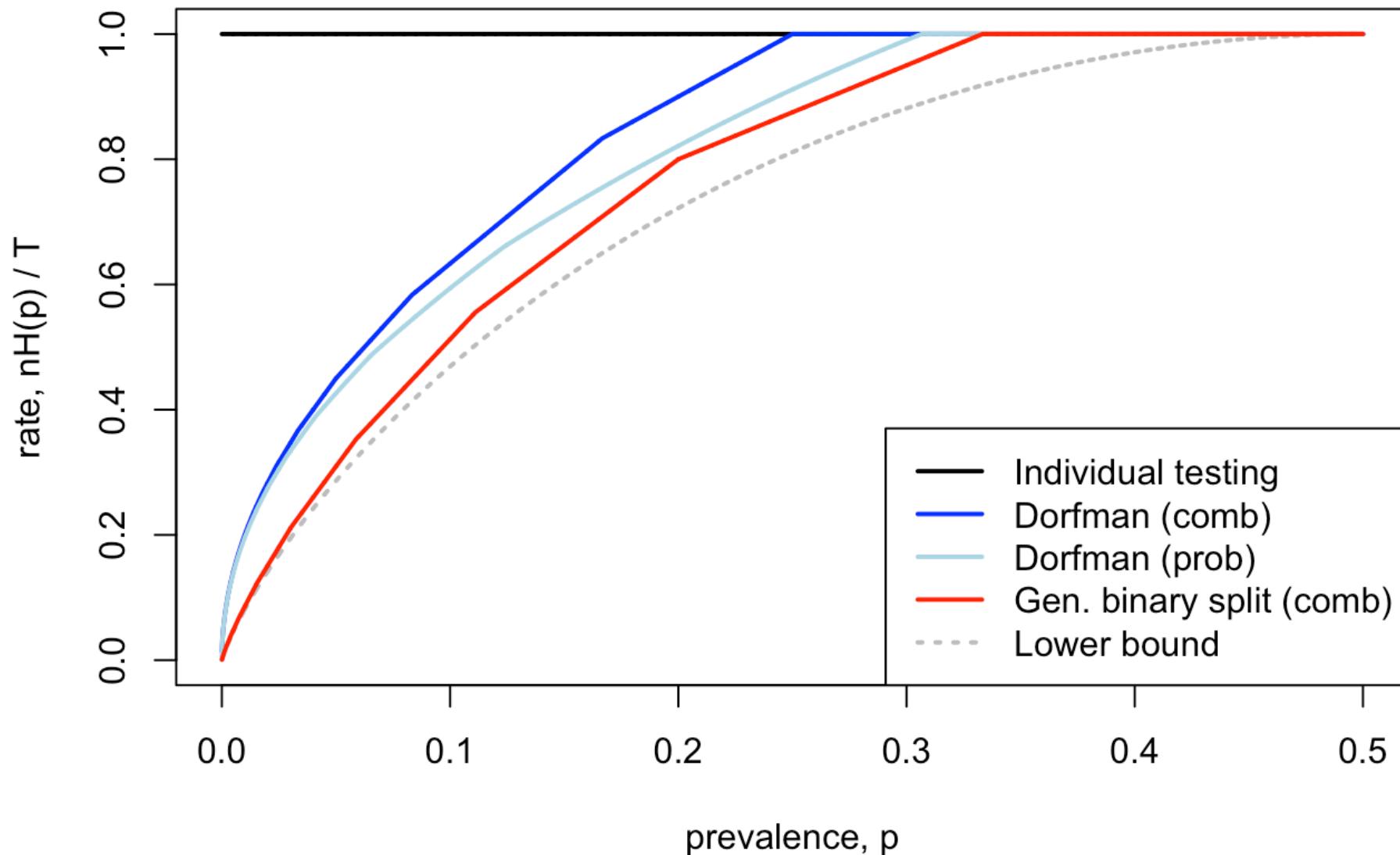
A set of size s contains:

no defectives with probability $(1 - p)^s$

at least one defective with probability $1 - (1 - p)^s$

$$\begin{aligned} T &= \frac{n}{s} (1 + (1 - (1 - p)^s)s) \\ &= \left(\frac{1}{s} + 1 - (1 - p)^s \right) n \end{aligned}$$

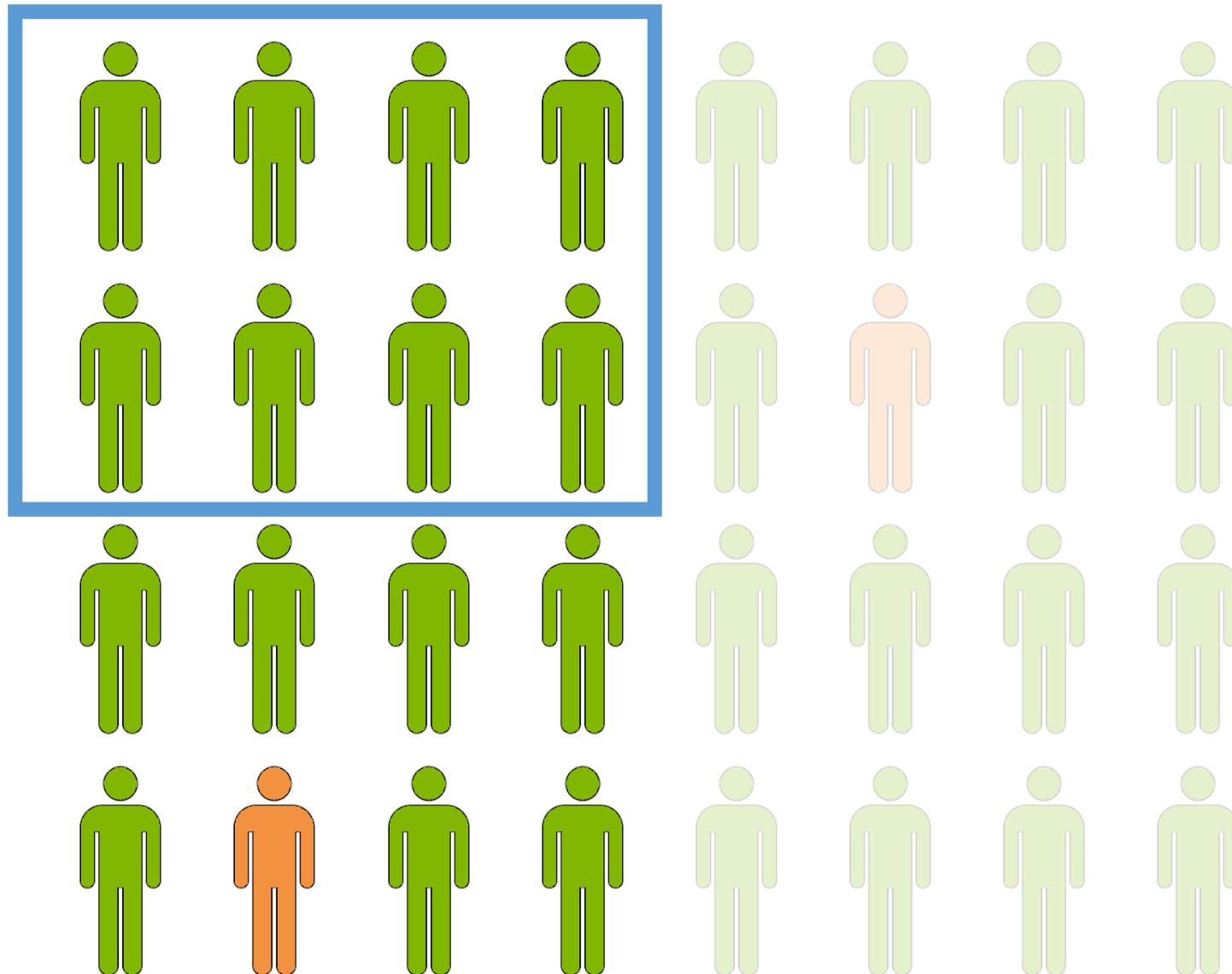
Group testing tests per item



Generalized binary splitting: average case

During binary splitting stages,
we are likely find some nondefectives,
which we can remove immediately.

Binary splitting

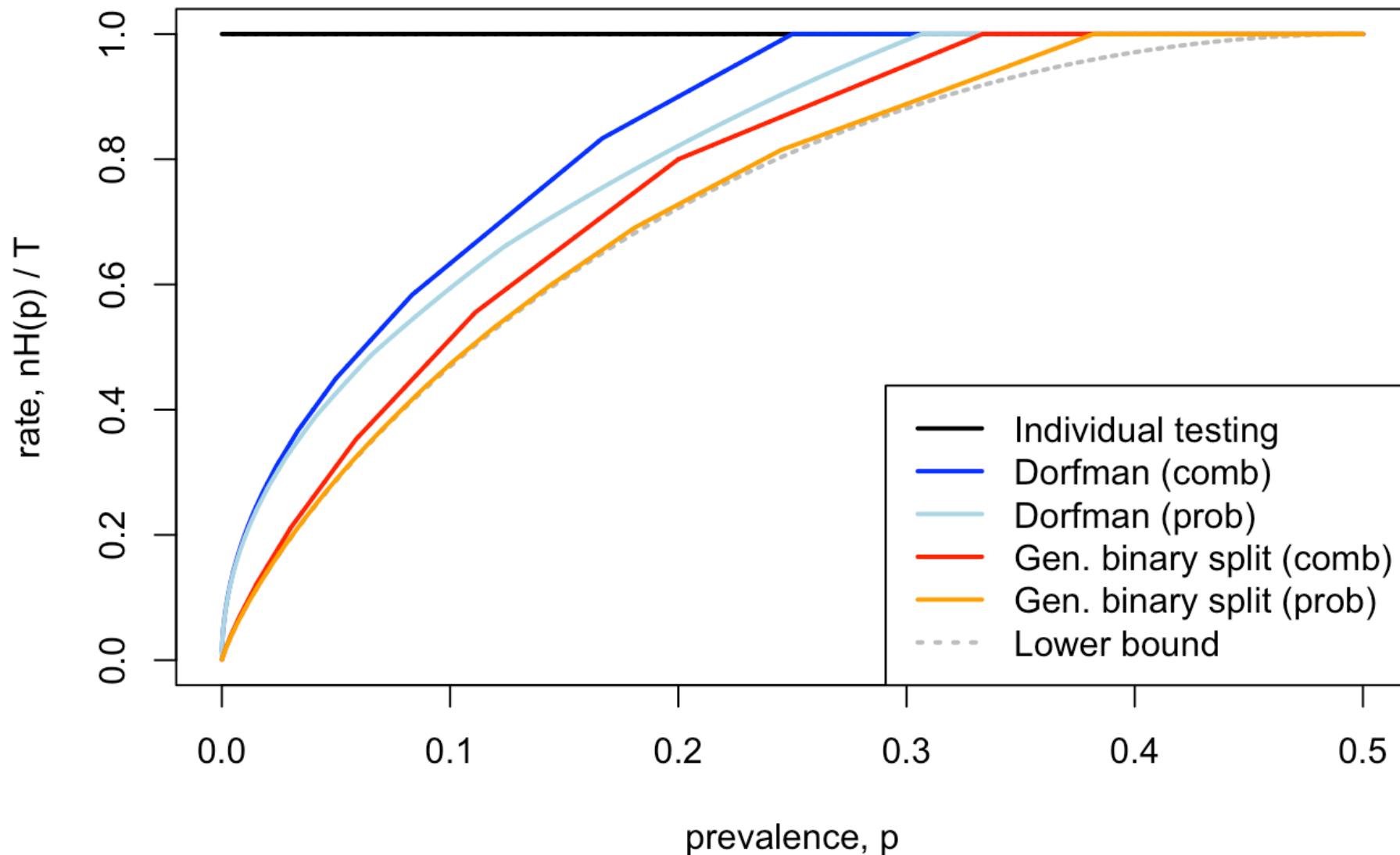


Generalized binary splitting: average case

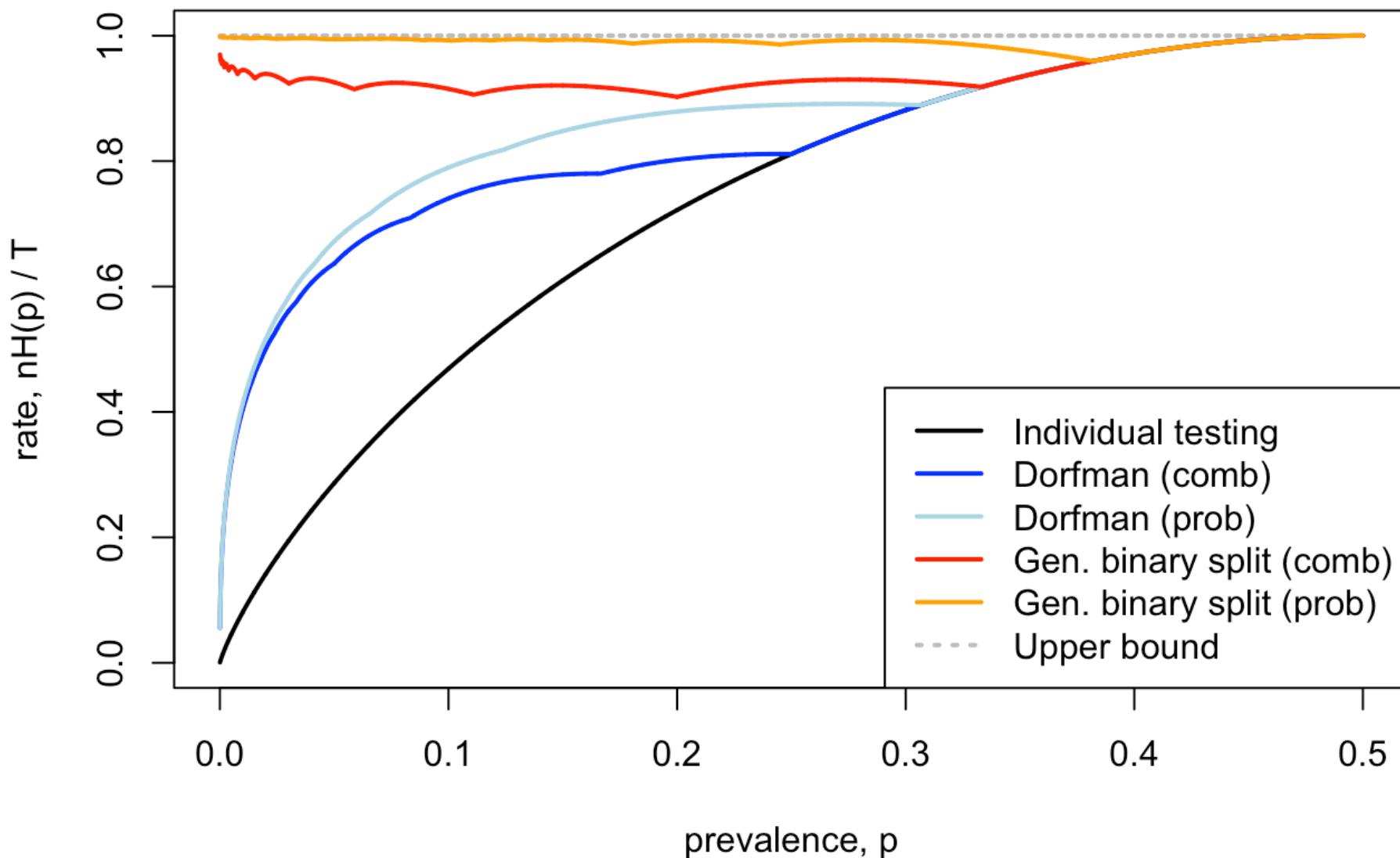
During binary splitting stages,
we are likely find some nondefectives,
which we can remove immediately.

We get a complicated expression,
involving binary splits of sets
whose size isn't a power of 2.
(see Aldridge, 2019)

Group testing tests per item



Group testing rates



Conclusions

When $p > 35\%$
individual testing is optimal

Conclusions

When $p > 35\%$

individual testing is optimal

When $p < 25\%$

group testing gives significant benefits,
even with the simple Dorfman algorithm

Conclusions

When $p > 35\%$

individual testing is optimal

When $p < 25\%$

group testing gives significant benefits,
even with the simple Dorfman algorithm

When $p < 5\%$

Dorfman is highly suboptimal, and
generalized binary splitting very close to optimal

Part 2

Group testing for the coronavirus

Why the UK is struggling to scale up coronavirus testing

Shift in policy has left British labs facing global competition for equipment



BBC

Your account



News

Sport

Weather

iPlayer

Sounds

NEWS

Home | UK | World | Business | Politics | Tech | Science | Health | Family & Education

US & Canada

Coronavirus: White House concedes US lacks enough test kits

6 March 2020

Share

In fact.

News > Health

Coronavirus: Is the UK testing enough people?

- 1 How have other countries responded to the coronavirus threat?
- 2 How many people are being tested in the UK?
- 3 Why are numbers lower than might be expected?
- 4 What does it mean for health workers?
- 5 What is being done to increase testing in the UK?

Harry Cockburn | 1 day ago



WIRED

Technology | Science | Culture | Gear | Business | Politic

Biology

Why isn't the UK testing more people for coronavirus?

There are two broad types of tests for coronavirus. Manufacturers of the kits are vastly ramping up their production cycles

It seems there are fewer coronavirus tests available than the number of people it would be useful to test:

Could group testing help make better use of these limited tests?

EDITORS' PICK | 2,004 views | Mar 29, 2020, 09:00am EDT

Group Testing Is Our Surefire Secret Weapon Against Coronavirus



Laurence Kotlikoff
Cont.
Taxes

The Washington Post

Democracy Dies in Darkness

≡

Coronavirus Live updates U.S. map World map

PostEverything • Perspective

A temporary coronavirus testing fix: Use each kit on 50 people at a time.

We don't have enough tests. Group testing offers a way to make best use of them.

It seems there are fewer coronavirus tests available than the number of people it would be useful to test:

Could group testing help make better use of these limited tests?

It seems there are fewer coronavirus tests available than the number of people it would be useful to test:

Could group testing help make better use of these limited tests?

Was group testing useful for testing soldiers for syphilis?

Was group testing useful for testing soldiers for syphilis?

“Unfortunately, this very promising idea of group blood samples for syphilis screen was not actually put to use. The main reason [...] was that the test was no longer accurate when as few as eight or nine samples were pooled.” (Du-Hwang, 1999)

Model assumption 1

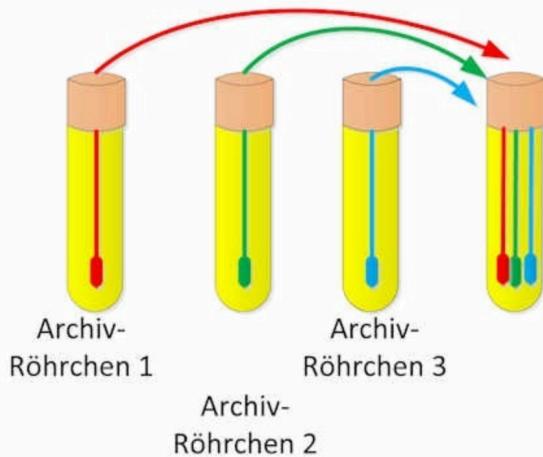
Pools containing a single positive sample
and any number of negative samples
still give a positive result:
there is no “dilution effect”.

Pool testing of SARS-CoV-02 samples increases worldwide test capacities many times over

30. März 2020

< :: >

Mini-Pool Methode



Evaluation of COVID-19 RT-qPCR test in multi-sample pools

Idan Yelin, Noga Aharony, Einat Shaer-Tamar, Amir Argoetti, Esther Messer, Dina Berenbaum, Einat Shafran, Areen Kuzli, Nagam Gandali, Tamar Hashimshony, Yael Mandel-Gutfreund, Michael Halberthal, Yuval Geffen, Moran Szwarcwort-Cohen, Roy Kishony

doi: <https://doi.org/10.1101/2020.03.26.20039438>

This article is a preprint and has not been peer-reviewed [what does this mean?]
It reports new medical research that has yet to be evaluated and so should not be used to guide clinical practice.

Abstract

Info/History

Metrics

Preview PDF

Abstract

The recent emergence of SARS-CoV-2 lead to a current pandemic of unprecedented levels. Though diagnostic tests are fundamental to the ability to detect and respond, many health systems are already experiencing shortages of reagents associated with this test. Here, testing a pooling approach for the standard RT-qPCR test, we find that a single positive sample can be detected even in pools of up to 32 samples, with an estimated false negative rate of 10%. Detection of positive samples diluted in even up to 64 samples may also be attainable, though may require additional amplification cycles. As it uses the standard protocols, reagents and equipment, this pooling method can be applied immediately in current clinical testing laboratories. We hope that such implementation of a pool test for COVID-19 would allow expanding current screening capacities thereby enabling the expansion of detection in the community, as well as in close integral groups, such as hospital departments, army units, or factory shifts.

Coronavirus tests

Antigen test

Tests if you currently have the virus

UK currently testing about 10,000 per day: either ill people in hospital or healthcare workers

Antibody test

Tests if you have had the virus in the past and are now immune

Could be available soon?

Coronavirus tests

Antigen test

Tests if you currently have the virus

UK currently testing about 10,000 per day: either ill people in hospital or healthcare workers

Antibody test

Tests if you have had the virus in the past and are now immune

Could be available soon?

Coronavirus tests

When $p > 35\%$

individual testing is optimal

When $p < 25\%$

group testing gives significant benefits

Coronavirus tests

When $p > 35\%$

individual testing is optimal

When $p < 25\%$

group testing gives significant benefits

So testing ill people in hospital
probably won't give any advantage.

Group testing is more likely to be useful for
future screening of the wider population.

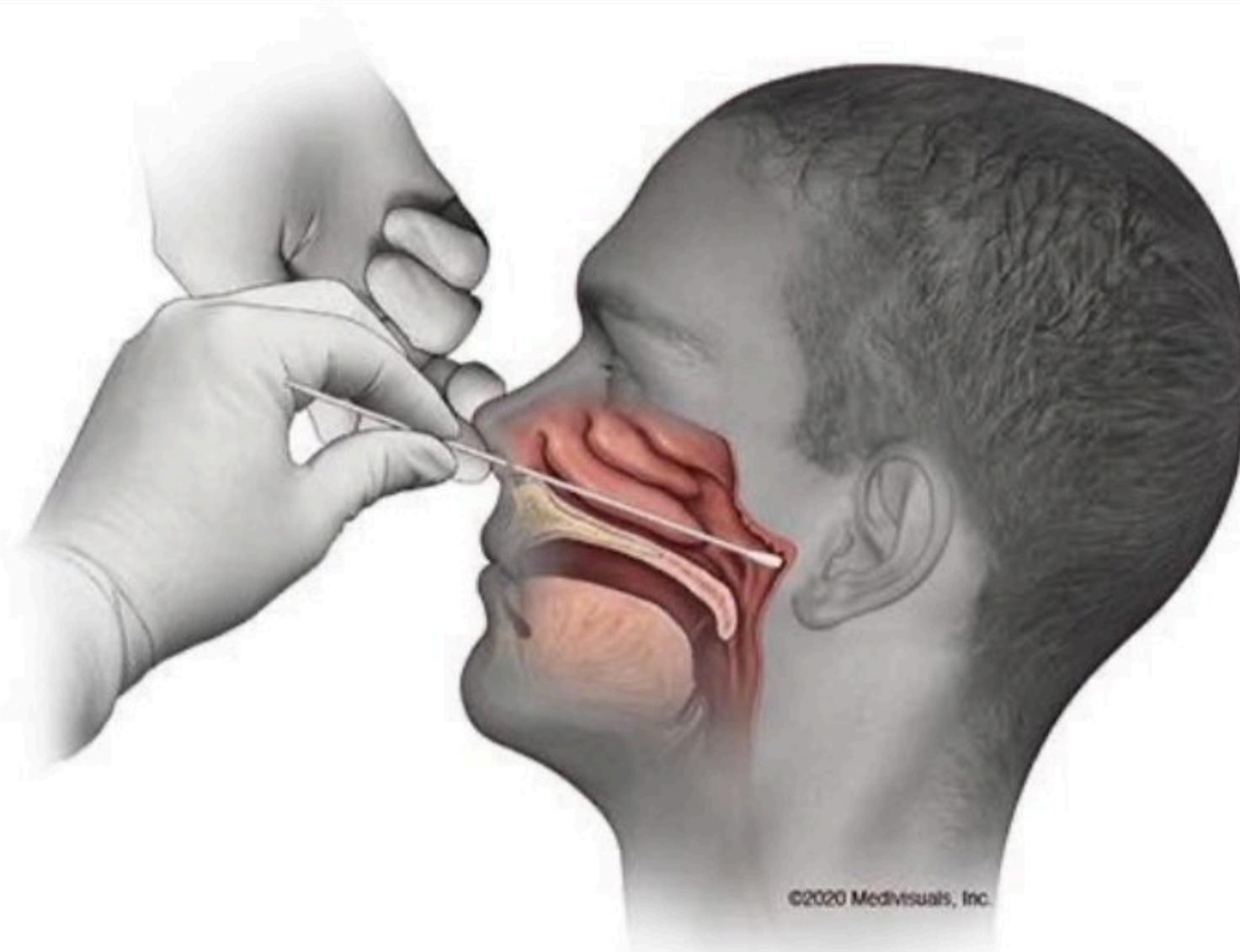
(What about healthcare workers?)

Coronavirus antigen test

PCR (polymerase chain reaction):

Take a swab from the nose or throat.

Coronavirus antigen test



©2020 MediVisuals, Inc.

Coronavirus antigen test

PCR (polymerase chain reaction):

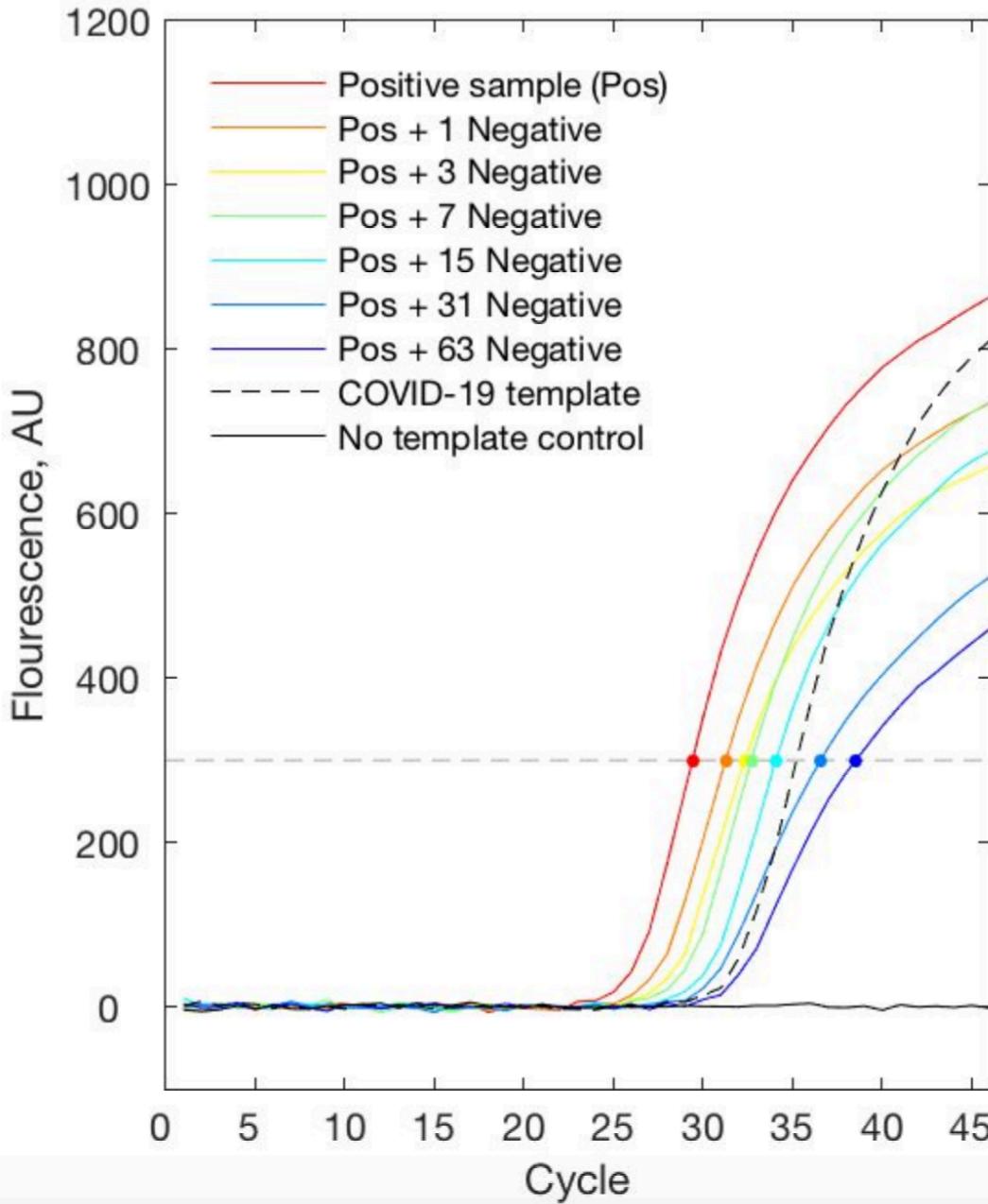
Take a swab from the nose or throat.

Amplify the RNA of the virus by repeatedly mixing the sample with chemicals (“reagents”) and spinning the sample in a centrifuge.

Add short DNA fragments (“primers”) that attach to the genetic material of the coronavirus.

Use a dye so that you can see whether this new DNA is built or not.

Coronavirus antigen test



Pooled samples require more cycles of amplification than individual tests.

But single positive samples are still detectable in pools of size 32 or even 64.

Coronavirus antigen test

Each experiment was replicated 10 times.

1 of the 10 experiments was poor:
only the pools of sizes 1, 2, 4 worked
pools of sizes 8, 16, 32, 64 failed

In 2 more of the 10 experiments:
the pool of size 64 didn't quite reach
the required level within 45 cycles:

can probably be fixed with a few more cycles
or reducing the required level

Coronavirus antigen test

Each experiment was replicated 10 times.

1 of the 10 experiments was poor:
only the pools of sizes 1, 2, 4 worked
pools of sizes 8, 16, 32, 64 failed

In 2 more of the 10 experiments:
the pool of size 64 didn't quite reach
the required level within 45 cycles:

Model assumption 1

Pools containing a single positive sample
and any number of negative samples
still give a positive result:
there is no “dilution effect”.

Probably OK for small pools.
For larger pools, we may need to consider
the possibility of dilution errors

Model assumption 2

The prevalence p is known exactly.

Model assumption 2

The prevalence p is known exactly.

Table 1: Posterior model estimates of percentage of total population infected as of 28th March 2020.

Country	% of total population infected (mean [95% credible interval])
Austria	1.1% [0.36%-3.1%]
Belgium	3.7% [1.3%-9.7%]
Denmark	1.1% [0.40%-3.1%]
France	3.0% [1.1%-7.4%]
Germany	0.72% [0.28%-1.8%]
Italy	9.8% [3.2%-26%]
Norway	0.41% [0.09%-1.2%]
Spain	15% [3.7%-41%]
Sweden	3.1% [0.85%-8.4%]
Switzerland	3.2% [1.3%-7.6%]
United Kingdom	2.7% [1.2%-5.4%]

(Flaxman–Mishra–Gandy et al, 2020)

Model assumption 2

The prevalence p is known exactly.

Solution A

Check the results are robust
to errors in estimating the prevalence

Assumed prevalence: 2.7%
Actual prevalence: 2.7%

Dorfman (comb.): pools of size 6
Tests-per-item: 0.33
3.0x better than individual testing

Gen. binary splitting (prob.): pools of size 25
Tests-per-item: 0.18
5.6x better than individual testing

Assumed prevalence: 2.7%
Actual prevalence: 5.4%

Dorfman (comb.): pools of size 6
Tests-per-item: 0.49
2.0x better than individual testing

Gen. binary splitting (prob.): pools of size 25
Tests-per-item: 0.32
3.1x better than individual testing

Assumed prevalence: 2.7%
Actual prevalence: 1.2%

Dorfman (comb.): pools of size 6
Tests-per-item: 0.24
4.2x better than individual testing

Gen. binary splitting (prob.): pools of size 25
Tests-per-item: 0.10
9.8x better than individual testing

Model assumption 2

The prevalence p is known exactly.

Solution B

Use group testing ideas
to estimate the prevalence

Solution B

Use group testing ideas
to estimate the prevalence

Statistical Modeling, Causal Inference, and Social Science

[HOME](#) | [BOOKS](#) | [BLOGROLL](#) | [SPONSORS](#)

« Let's do preregistered replication studies of the cognitive effects of air pollution—not because we think existing studies are bad, but because we think the topic is important and we want to understand it better.

I'm still struggling to understand hypothesis testing . . . leading to a more general discussion of the role of assumptions in statistics »

“For the cost of running 96 wells you can test 960 people and accurate assess the prevalence in the population to within about 1%. Do this at 100 locations around the country and you'd have a spatial map of the extent of this epidemic today. . . and have this data by Monday.”

Posted by [Andrew](#) on 27 March 2020, 5:03 pm

Daniel Lakeland writes:

COVID-19 is tested for using real-time reverse-transcriptase PCR (rt-rt-PCR). This is basically just a fancy way of saying they are detecting the presence of the RNA by converting it to DNA and amplifying it. It has already been shown by people in Israel that you can combine material from at least 64 swabs and still reliably detect the presence of the RNA.

Solution B

Use group testing ideas
to estimate the prevalence

If you sample (disjoint) sets of size s ,
the proportion of tests that are positive is

$$1 - (1 - p)^s.$$

Solve for p .

Solution B

Use group testing ideas
to estimate the prevalence

If you sample (disjoint) sets of size s ,
the proportion of tests that are positive is
 $1 - (1 - p)^s$.
Solve for p .

Only accurate when the proportion is close to 1/2:
Try a range of values of s .

Model assumption 3

The only goal is to minimise the number of tests.

Model assumption 3

The only goal is to minimise the number of tests.

What's the binding constraint?

Lab space?

Availability of reagents?

Availability of swabs?

Opportunity to take samples?

Model assumption 3

The only goal is to minimise the number of tests.

What's the binding constraint?

Lab space?

Availability of reagents?

Availability of swabs?

Opportunity to take samples?

Model assumption 4

We're quite happy to wait a long time for results

Model assumption 4

We're quite happy to wait a long time for results

Number of stages:

Individual testing: 1

Dorfman: 2

Generalised binary splitting: about $1 + \log_2 1/p$

Model assumption 4

We're quite happy to wait a long time for results

Number of stages:

Individual testing: 1

Dorfman: 2

Generalised binary splitting: about $1 + \log_2 1/p$

Gen. binary split with smaller than optimal sets?

Gen. binary split for a bit, then Dorfman?

Is group testing useful when testing for the coronavirus?

Is group testing useful when testing for the coronavirus?



Gov. Ricketts provides update on coronavirus testing

Last week, Ricketts said capacity was 200 a day. Now it's over 600 a day. The governor says testing capacity is ramping up because the state lab can now pool samples.

That means they can put five nasal swabs into one test tube and instead of testing them separately, do five at a time.

"If that one sample comes back negative, we know all five of those are negative, so we have just saved four tests. Now if that test comes back positive, then we will have to come back and retest all five of those again to figure out which one was positive," Ricketts said.

Group testing: tests per item

