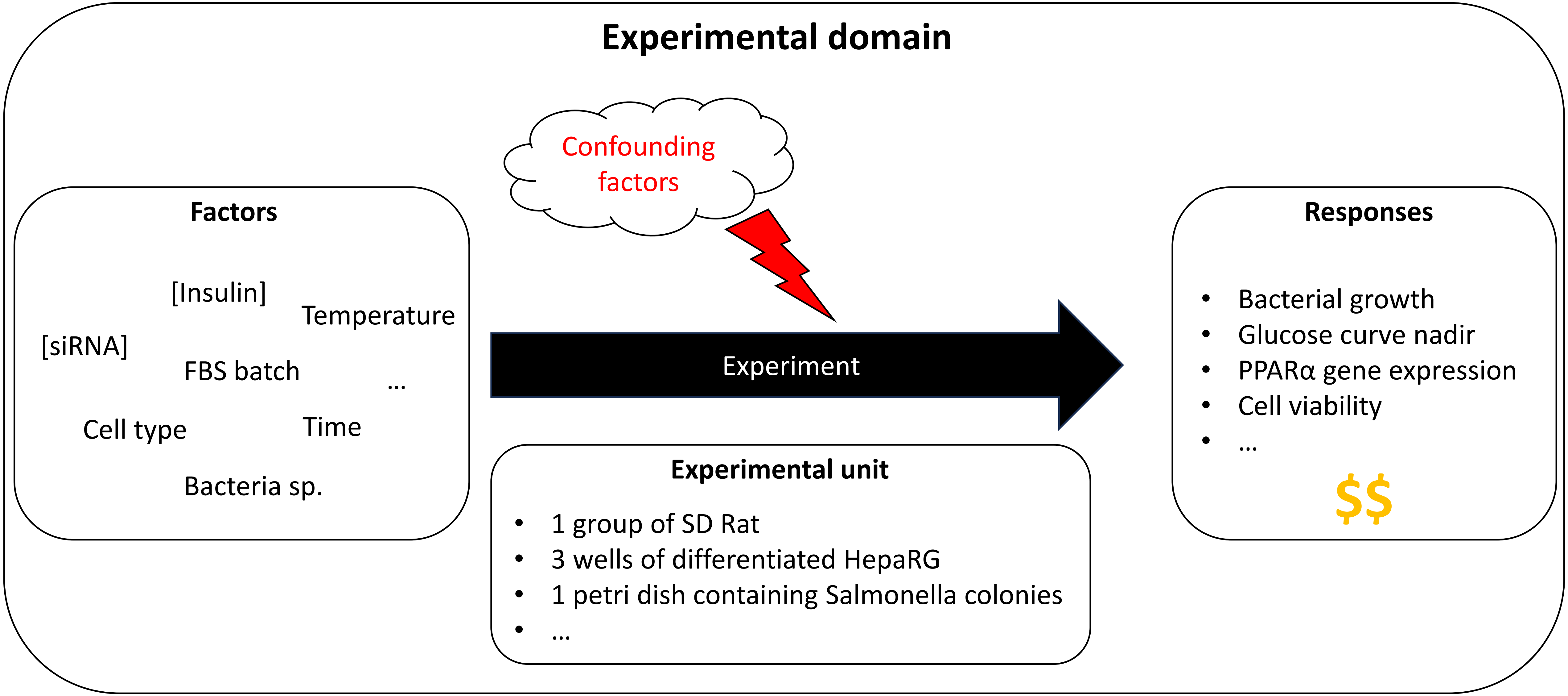
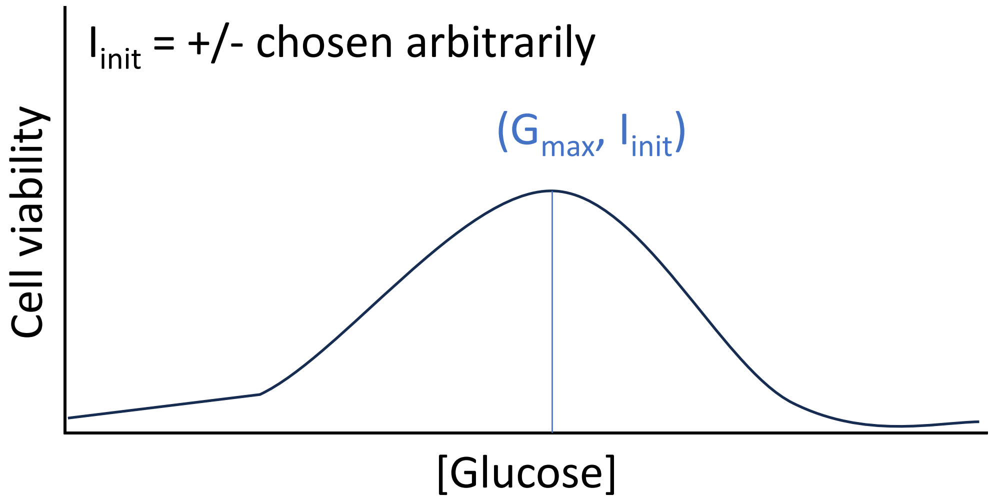
Introduction to Design of Experiments (DoE): work smarter, work cheaper

2023/11/09, NPhDA Technical meetings – Kevin Maunand

What is an experimental domain



Classic approach : One factor at a time

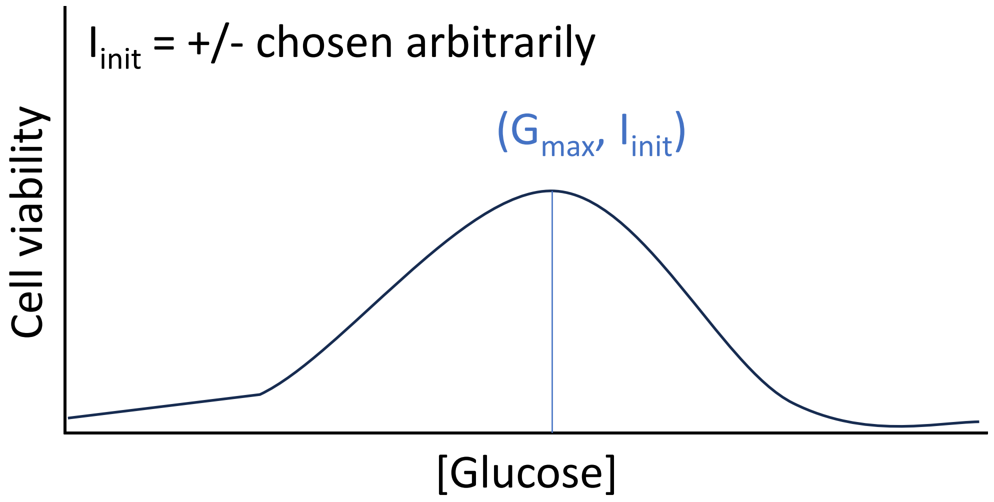


A. Dose-response to [Glucose], [Insulin] fixed at I_{init}

8 measures

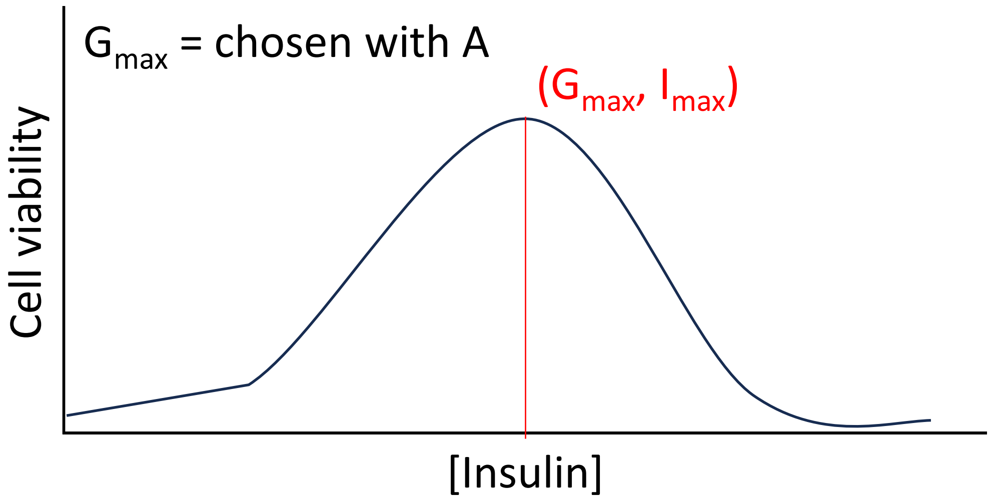
A : Viability is optimal at <u>G_{max} in initial [Insulin] condition (I_{init})</u>

Classic approach : One factor at a time



A. Dose-response to [Glucose], [Insulin] fixed at I_{init}

8 measures

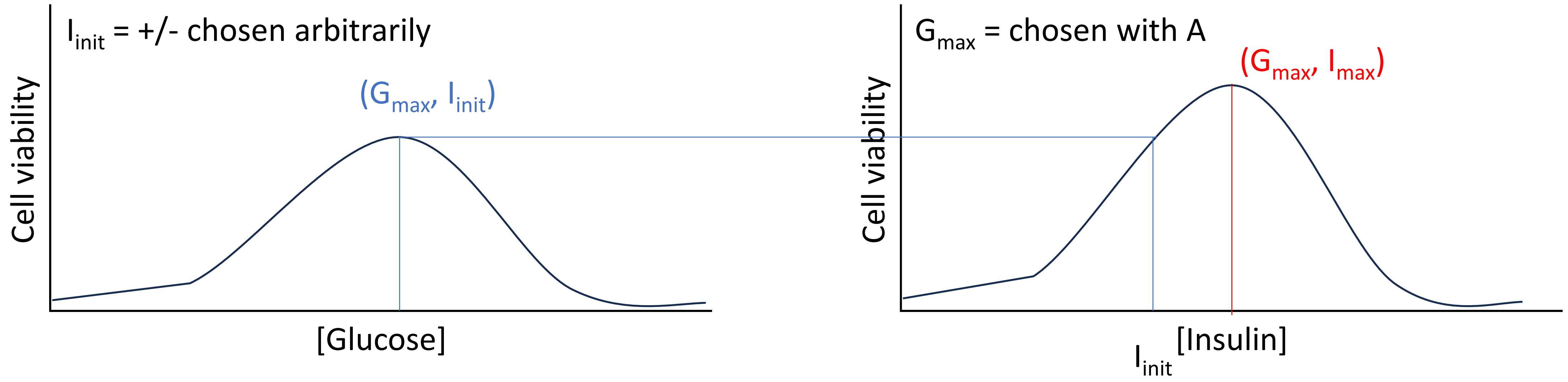


B. Dose-response to [Insulin], [Glucose] fixed at G_{max}

8 measures

A : Viability is optimal at <u>G_{max} in initial [Insulin] condition (I_{init})</u>

Classic approach : One factor at a time



A. Dose-response to [Glucose], [Insulin] fixed at I_{init}

8 measures

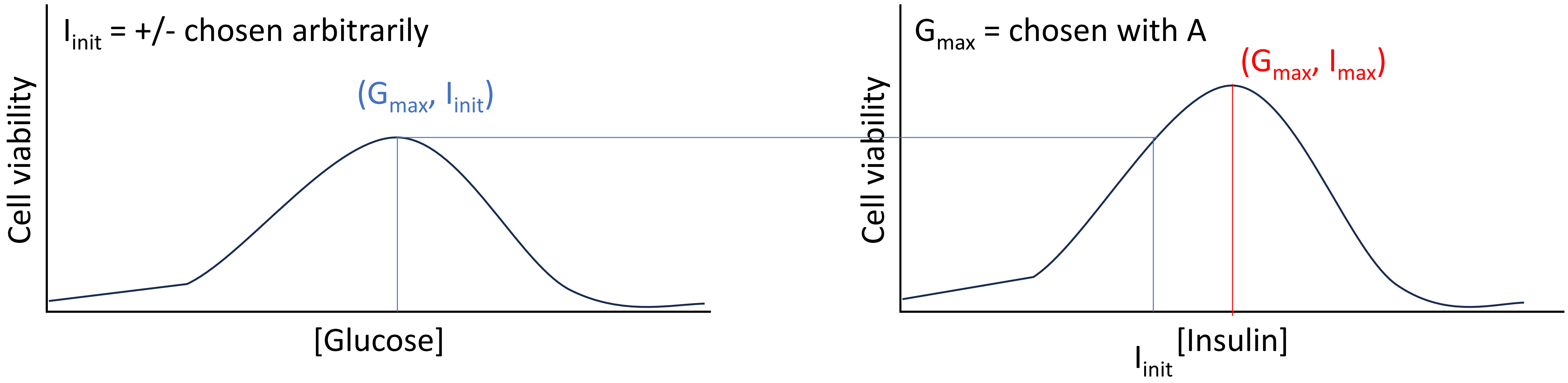
B. Dose-response to [Insulin], [Glucose] fixed at G_{max}

8 measures

A : Viability is optimal at <u>G_{max} in initial [Insulin] condition (I_{init})</u>

B : When we change [Insulin] at G_{max}, viability is not optimal at I_{init}.

Classic approach : One factor at a time



A. Dose-response to [Glucose], [Insulin] fixed at I_{init}

8 measures

B. Dose-response to [Insulin], [Glucose] fixed at G_{max}

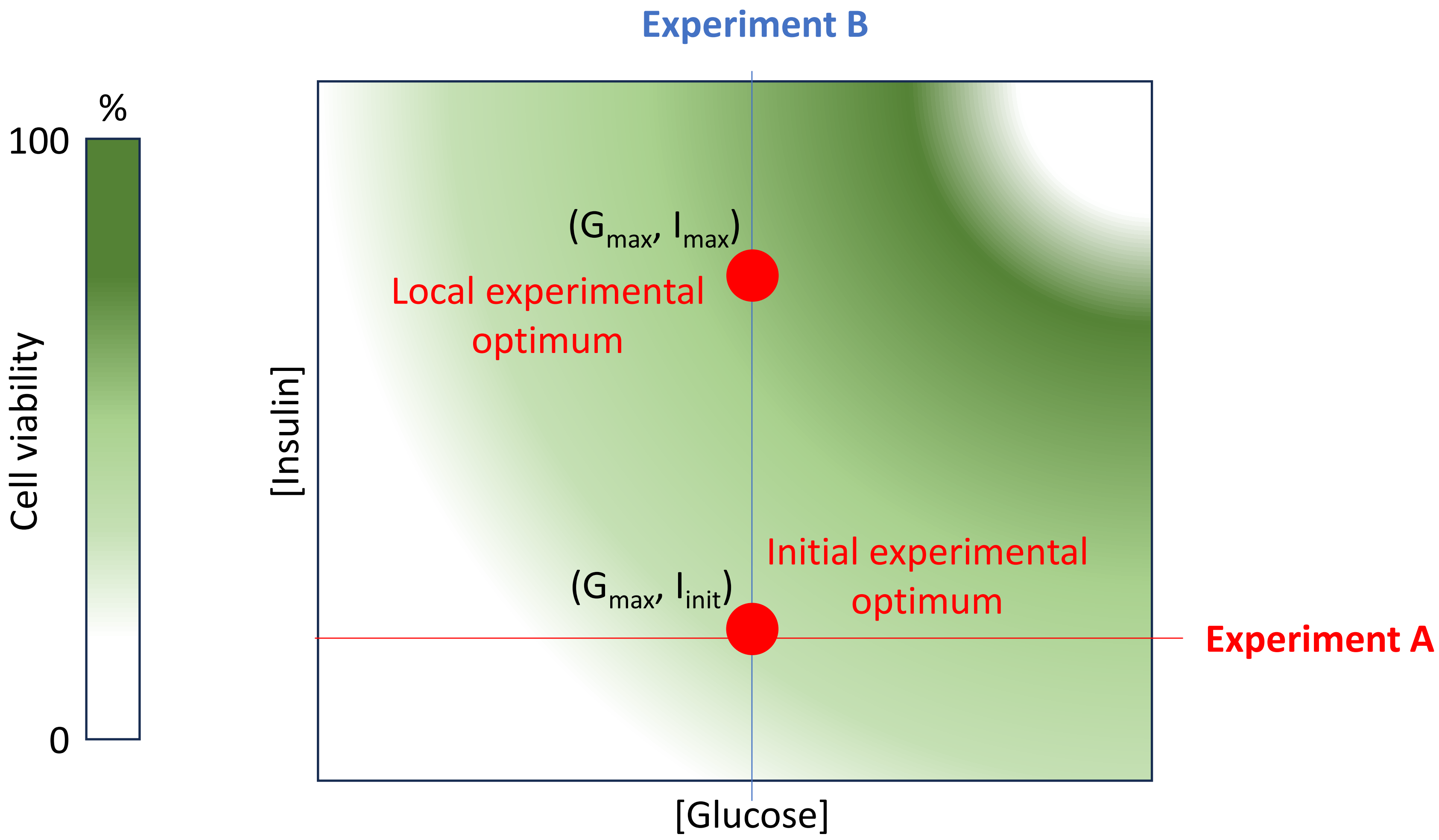
8 measures

A : Viability is optimal at G_{max} in initial [Insulin] condition (I_{init})

B : When we change [Insulin] at G_{max}, viability is not optimal at I_{init}.

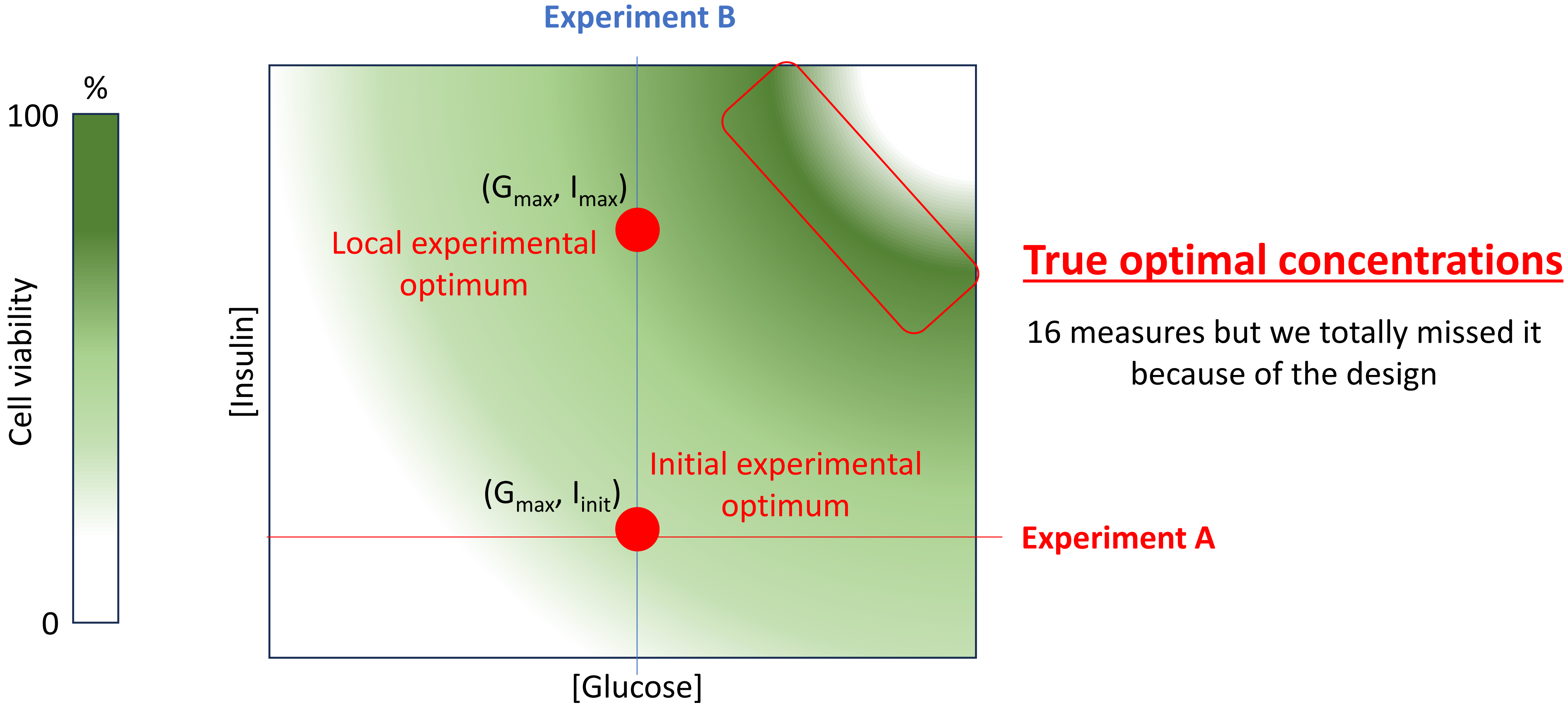
→ Is the new optimum (G_{max}, I_{max}) really the optimal condition?

Caveats of "one factor at a time" approach



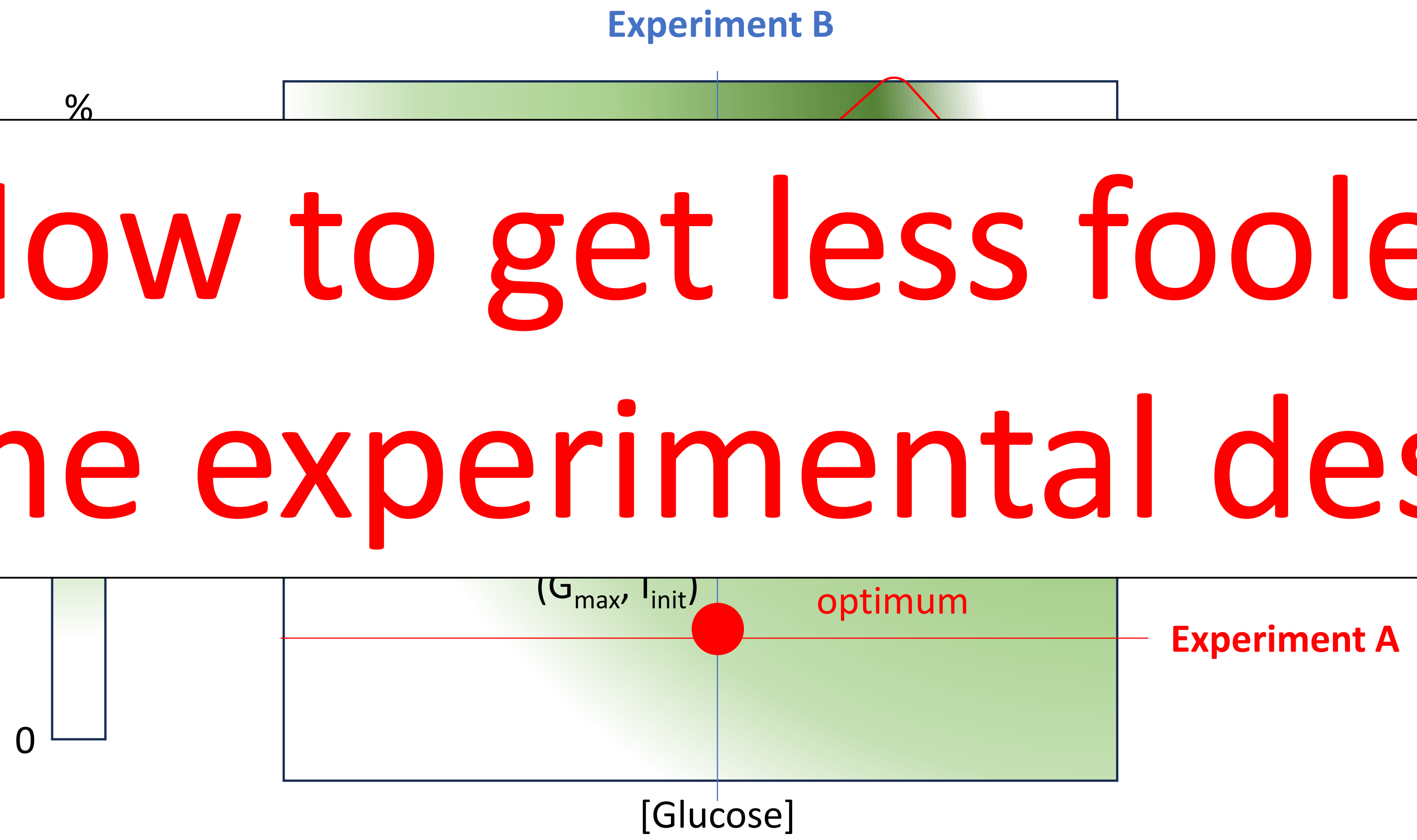
Cartoon surface response of cell viability

Caveats of “one factor at a time” approach



Cartoon surface response of cell viability

Caveats of "one factor at a time" approach

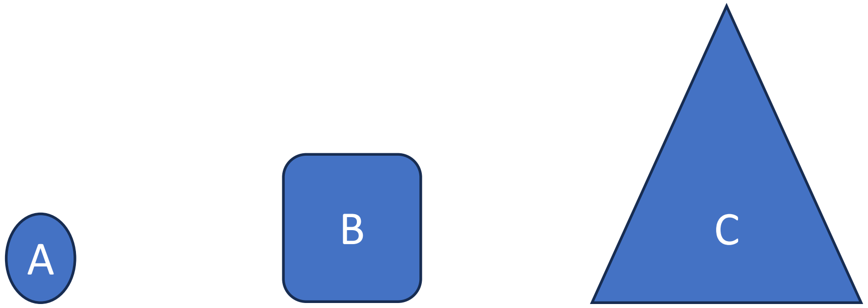


How to get less fooled by the experimental design?

Cartoon surface response of cell viability

Design of experiments interest: Yates pan balance

Yates pan balance: a classic example

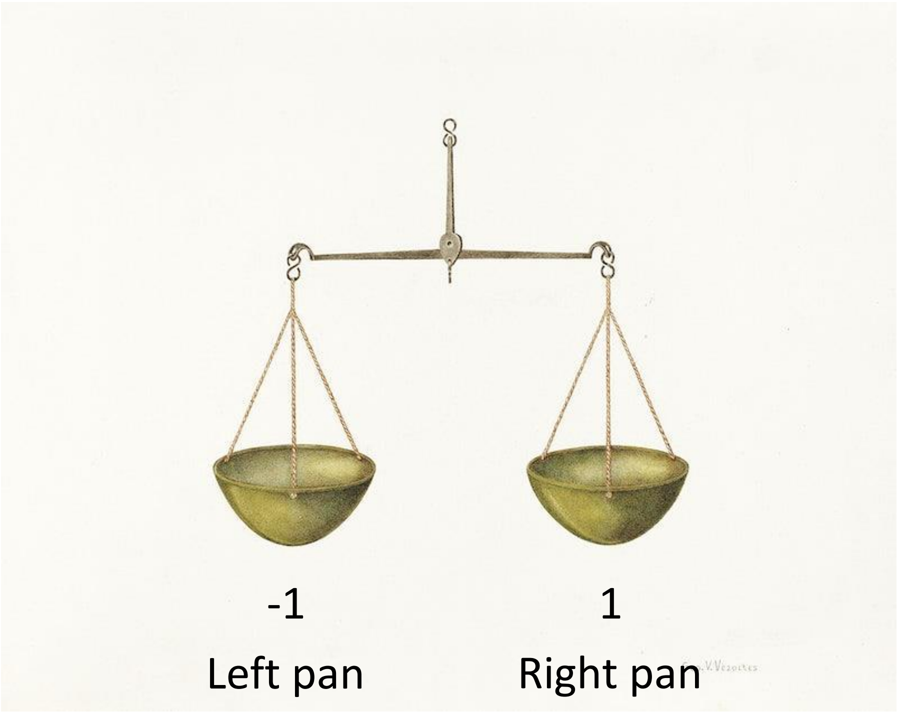


What is the weight of A, B, and C?
Which is the heaviest?

Precision must be very low (error = +/- 1 g)

Empty balance does not show 0 g

Budget : 800 € (100 €/measure)



0 = not on balance

Yates pan balance : 1st strategy “kitchen”

1st strategy : only one measure per object, one object at a time (n=1)

N°	Left pan	Right pan	Y (masse)
1	-	-	1 g
2	-	A	3 g
3	-	B	4 g
4	-	C	10 g

Yates pan balance : 1st strategy “kitchen”

N°	0 = cte	A	B	C
1	1	0	0	0
2	1	1	0	0
3	1	0	1	0
4	1	0	0	1

Y (masse)
1 g
3 g
4 g
10 g

$$\begin{cases} M_0 = Y1 \\ M_0 + M_A = Y2 \\ M_0 + M_B = Y3 \\ M_0 + M_C = Y4 \end{cases}$$

M_A= 3-1 = 2 g
M_B= 4-1 = 3 g
M_C= 10-1 = 9 g

Yates pan balance : 1st strategy "kitchen"

N°	0 = cte	A	B	C
1	1	0	0	0
2	1	1	0	0
3	1	0	1	0
4	1	0	0	1

Y (masse)
1 g
3 g
4 g
10 g

$$\begin{cases} M_0 = Y1 \\ M_0 + M_A = Y2 \\ M_0 + M_B = Y3 \\ M_0 + M_C = Y4 \end{cases}$$

➡

$M_A = 3-1 = 2$ g
$M_B = 4-1 = 3$ g
$M_C = 10-1 = 9$ g

$V(X-Y) = V(X) + V(Y) – 2 * cov(X,Y)$

$V(aX-b) = a^2 * V(X)$

➡

$V(M_0) = V(Y1) = \sigma^2$

$V(M_A) = V(Y1) + V(Y2) = 2\sigma^2$

Very inaccurate measure, price = 400 €

Can we be more accurate?

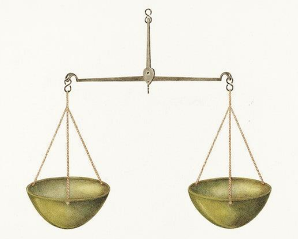
Measure more times!

Yates pan balance: 2nd strategy “do more”

2nd strategy : two measures per object, one object at a time (n=2)

N°	Left pan	Right pan
1	-	-
2	-	A
3	-	B
4	-	C
1’	-	-
2’	-	A
3’	-	B
4’	-	C

Yates pan balance: 2nd strategy “do more”



N°	0 = cte	A	B	C
1	1	0	0	0
2	1	1	0	0
3	1	0	1	0
4	1	0	0	1
1’	1	0	0	0
2’	1	1	0	0
3’	1	0	1	0
4’	1	0	0	1

Moy (Y) (masse)
1 g
3 g
4 g
10 g

$$\begin{cases} M_0 = \frac{Y1 + Y1'}{2} \\ M_0 + M_A = \frac{Y2 + Y2'}{2} \\ M_0 + M_B = \frac{Y3 + Y3'}{2} \\ M_0 + M_C = \frac{Y4 + Y4'}{2} \end{cases}$$

M_A= 3-1 = 2 g
M_B= 4-1 = 3 g
M_C= 10-1 = 9 g

Yates pan balance: 2nd strategy "do more"

N°	0 = cte	A	B	C
1	1	0	0	0
2	1	1	0	0
3	1	0	1	0
4	1	0	0	1
1'	1	0	0	0
2'	1	1	0	0
3'	1	0	1	0
4'	1	0	0	1

Moy (Y) (masse)
1 g
3 g
4 g
10 g

$$\begin{cases} M_0 = \frac{Y1 + Y1'}{2} \\ M_0 + M_A = \frac{Y2 + Y2'}{2} \\ M_0 + M_B = \frac{Y3 + Y3'}{2} \\ M_0 + M_C = \frac{Y4 + Y4'}{2} \end{cases}$$

M_A= 3-1 = 2 g
M_B= 4-1 = 3 g
M_C= 10-1 = 9 g

V(X-Y) = V(X) + V(Y) – 2 * cov(X,Y)

V(aX-b) = a² * V(X)

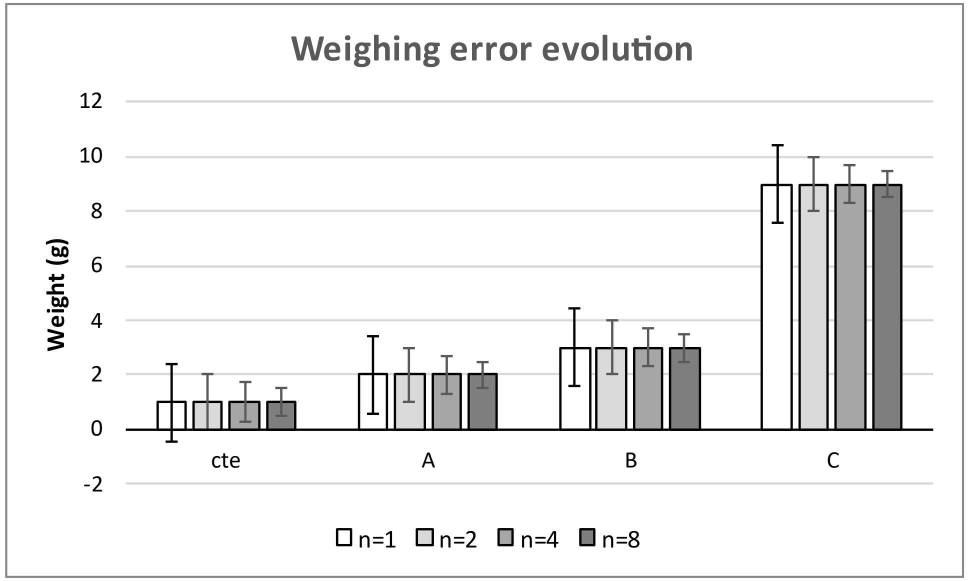
$$V(M_0) = \frac{V(Y1)+V(Y1')}{2^2} = \sigma^2/2$$

$$V(M_A) = \frac{V(Y2)+V(Y2')}{2^2} + V(M_0) = \sigma^2$$

Less inaccurate measure, price = 800 €

Yates pan balance: 2nd strategy "do more"

Measures numbers	4 (n=1)	8 (n=2)	16 (n=4)	32 (n=8)
$V(M_A)$	$2\sigma^2$	σ^2	$\sigma^2/2$	$\sigma^2/4$
Price	400 €	800 €	1600 €	3200 €



Can we do the same for cheaper?

Measure smarter!

Yates pan balance: 3rd strategy “Smart”

3rd strategy : Measure each object 2 times by pairs, (“n”=2)

N°	Left pan	Right pan
1	-	-
2	-	A + B
3	-	A + C
4	-	B + C

Y (masse)
Y1
Y2
Y3
Y4

Yates pan balance: 3rd strategy "Smart"

N°	0 = cte	A	B	C
1	1	0	0	0
2	1	1	1	0
3	1	1	0	1
4	1	0	1	1

Y (masse)
1 g
6 g
12 g
13 g

$M_0 = Y1$

$M_0 + M_A + M_B = Y2$

$M_0 + M_A + M_C = Y3$

$M_0 + M_B + M_C = Y4$

$M_A = (Y2+Y3-Y4-Y1)/2 = 2$ g

$M_B = (Y2+Y4-Y3-Y1)/2 = 3$ g

$M_C = (Y4+Y3-Y2-Y1)/2 = 9$ g

Yates pan balance: 3rd strategy “Smart”

N°	0 = cte	A	B	C
1	1	0	0	0
2	1	1	1	0
3	1	1	0	1
4	1	0	1	1

Y (masse)
1 g
6 g
12 g
13 g

$M_0 = Y1$

$M_0 + M_A + M_B = Y2$

$M_0 + M_A + M_C = Y3$

$M_0 + M_B + M_C = Y4$

$M_A = (Y2+Y3-Y4-Y1)/2 = 2$ g

$M_B = (Y2+Y4-Y3-Y1)/2 = 3$ g

$M_C = (Y4+Y3-Y2-Y1)/2 = 9$ g

V(X-Y) = V(X) + V(Y) – 2 * cov(X,Y)

V(aX-b) = a² * V(X)

$V(M_0) = V(Y1) = \sigma^2$

$$V(M_A) = \frac{V(Y2)+V(Y3)+V(Y4)+V(Y1)}{2^2} = \sigma^2$$

As inaccurate as before, price = 400 €

Can we be more accurate AND cheaper?

Measure smarter-er!

Yates pan balance : 4th strategy “smarter”

4th strategy : Measure each object 4 times in 3-tuples (“n” = 4)

N°	Left pan	Right pan
1	A + B + C	-
2	A	B + C
3	B	A + C
4	C	A + B

Y (masse)
Y1
Y2
Y3
Y4

Yates pan balance : 4th strategy “smarter”

N°	0 = cte	A	B	C
1	1	-1	-1	-1
2	1	-1	1	1
3	1	1	-1	1
4	1	1	1	-1

Y (masse)
-13 g
11 g
9 g
-3 g

$$\begin{cases} M_0 - M_A - M_B - M_C = Y1 \\ M_0 - M_A + M_B + M_C = Y2 \\ M_0 + M_A - M_B + M_C = Y3 \\ M_0 + M_A + M_B - M_C = Y4 \end{cases}$$

$M_0 = (Y1+Y2+Y3+Y4)/4 = 1\ g$
$M_A = (-Y1-Y2+Y3+Y4)/4 = 2\ g$
$M_B = (-Y1+Y2-Y3+Y4)/4 = 3\ g$
$M_C = (-Y1+Y2+Y3-Y4)/4 = 9\ g$

Yates pan balance : 4th strategy “smarter”

N°	0 = cte	A	B	C
1	1	-1	-1	-1
2	1	-1	1	1
3	1	1	-1	1
4	1	1	1	-1

Y (masse)
-13 g
11 g
9 g
-3 g

$$\begin{cases} M_0 - M_A - M_B - M_C = Y1 \\ M_0 - M_A + M_B + M_C = Y2 \\ M_0 + M_A - M_B + M_C = Y3 \\ M_0 + M_A + M_B - M_C = Y4 \end{cases}$$

$M_0 = (Y1+Y2+Y3+Y4)/4 = 1$ g
$M_A = (-Y1-Y2+Y3+Y4)/4 = 2$ g
$M_B = (-Y1+Y2-Y3+Y4)/4 = 3$ g
$M_C = (-Y1+Y2+Y3-Y4)/4 = 9$ g

$V(X-Y) = V(X) + V(Y) - 2 * cov(X,Y)$

$V(aX-b) = a^2 * V(X)$

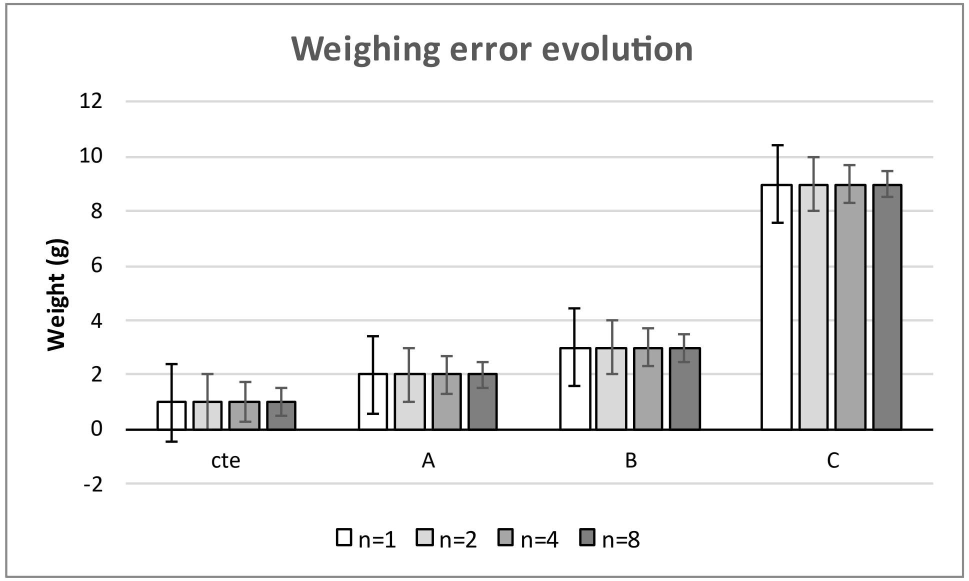
$$V(M_0) = \frac{V(Y1)+V(Y2)+V(Y3)+V(Y4)}{4^2} = \sigma^2/4$$

$$V(M_A) = \frac{V(Y2)+V(Y3)+V(Y4)+V(Y1)}{4^2} = \sigma^2/4$$

Least inaccurate measure, price = 400 €

Yates pan balance: 2nd strategy "do more"

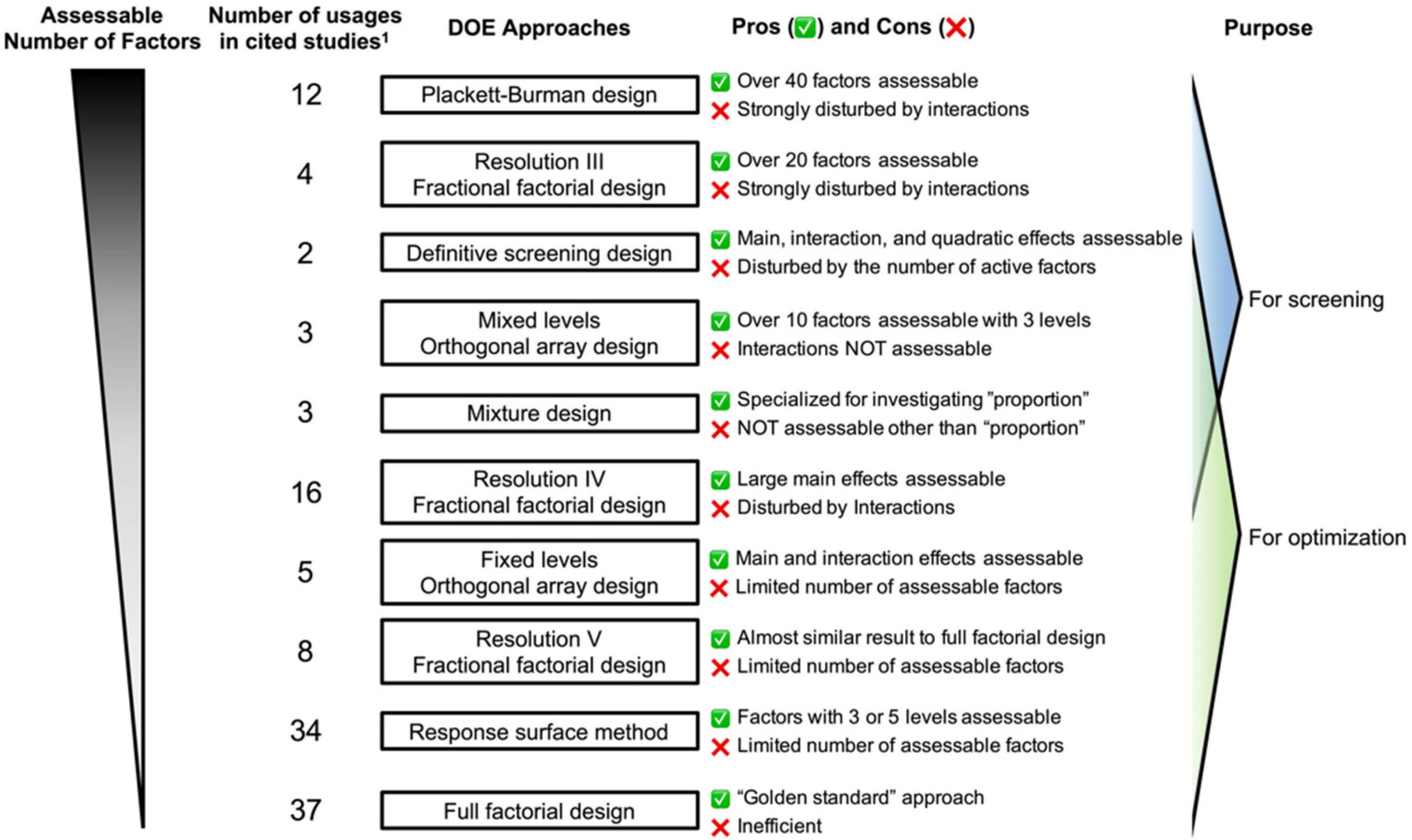
Measures numbers	4 (n=1)	8 (n=2)	16 (n=4)	32 (n=8)	4th strategy: 4 ("n"=4)
$V(M_A)$	$2\sigma^2$	σ^2	$\sigma^2/2$	$\sigma^2/4$	$\sigma^2/4$
Price	400 €	800 €	1600 €	3200 €	400 €



What if I think my factors interact with each other or I have a lot of factors?

→ Choose the good plan

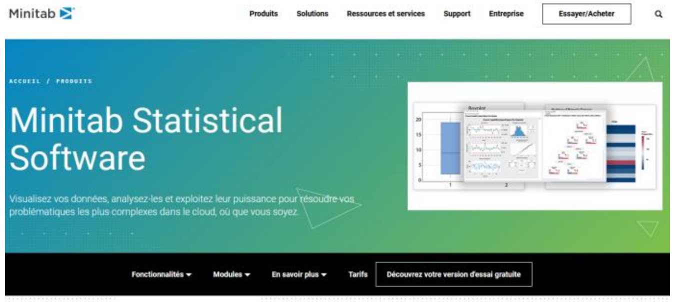
How to choose your DoE?



Clever experimental designs: shortcuts for better iPSC differentiation, Ryota Yasui et al., Cells, 2021

https://doi.org/10.3390/cells10123540

Resources



RStudio®

PubMed


DEFINITIVE SCREENING IN THE PRESENCE OF SECOND-ORDER EFFECTS

m = 4

1	0+--
2	0-++
3	-0-+
4	+0+-
5	--0-
6	++0+
7	-++0
8	+--0
9	0000

m = 5

1	0++--
2	0--++
3	+0--+
4	-0++-
5	+-0+-
6	-+0-+
7	+-+0+
8	-+-0-
9	++++0
10	----0
11	00000

m = 6

1	0+----
2	0-++++
3	+0-++-
4	-0+--+
5	--0+--
6	++0-++
7	-++0+-
8	+--0-+
9	+-+-0-
10	-+-+0+
11	++++-0
12	----+0
13	000000

m = 7

1	0+-+-+-
2	0-+-+-+
3	-0+-++-
4	+0-+--+
5	+-0++++
6	-+0----
7	+--0+--
8	-++0-++
9	--++0--
10	++--0++
11	-+-++0+
12	+-+--0-
13	+++++-0
14	-----+0
15	0000000

m = 8

1	0-++-+++
2	0+--+---
3	-0-++++-
4	+0+----+
5	--0++--+
6	++0--++-
7	+-+0++--
8	-+-0--++
9	--+-0-+-
10	++-+0+-+
11	+---+0++
12	-+++-0--
13	-++-++0+
14	+--+--0-
15	+++++-+0
16	-----+-0
17	00000000

m = 9

1	0++++++++
2	0--------
3	+0+-+--+-
4	-0-+-++-+
5	-+0-+-+--
6	+-0+-+-++
7	--+0+---+
8	++-0-+++-
9	+-+-0++--
10	-+-+0--++
11	----+0+++
12	++++-0---
13	++--++0-+
14	--++--0+-
15	---+++-0-
16	+++---+0+
17	-++--+-+0
18	+--++-+-0
19	000000000

m = 10

1	0++-++++-+
2	0--+----+-
3	+0-++-++--
4	-0+--+--++
5	-+0---+---
6	+-0+++-+++
7	-++0+--++-
8	+--0-++--+
9	----0++++-
10	++++0----+
11	-+-++0+-++
12	+-+--0-+--
13	++----0+++
14	--++++0---
15	++++-++0+-
16	----+--0-+
17	++--++--0-
18	--++--++0+
19	+-+-+-+-+0
20	-+-+-+-+-0
21	0000000000

m = 11

1	0-+-----+-+
2	0+-+++++-+-
3	-0--+----++
4	+0++-++++--
5	--0++++---+
6	++0----+++-
7	---0-++-++-
8	+++0+--+--+
9	+--+0+-++++
10	-++-0-+----
11	--++-0-+-+-
12	++--+0+-+-+
13	---++-0++--
14	+++--+0--++
15	-+++--+0+++
16	+---++-0---
17	-+---+-+0-+
18	+-+++-+-0+-
19	+-----++-0+
20	-+++++--+0-
21	++-+------0
22	--+-++++++0
23	00000000000

m = 12

1	0--+-+-+++-+
2	0++-+-+---+-
3	-0+++++++---
4	+0-------+++
5	++0-++-++-++
6	--0+--+--+--
7	+--0+-+-+--+
8	-++0-+-+-++-
9	++++0-++++++
10	----0+------
11	+-+-+0++-+-+
12	-+-+-0--+-+-
13	++++-+0----+
14	----+-0++++-
15	--+++--0--++
16	++---++0++--
17	+-++++--0++-
18	-+----++0--+
19	++-++--+-0--
20	--+--++-+0++
21	-+-++++--+0+
22	+-+----++-0-
23	+--+-+++--+0
24	-++-+---++-0
25	000000000000

FIGURE 1. Designs for $m = 4$ Through $m = 12$ Factors.

B. Jones and C. J. Nachtsheim, 2011, Journal of Quality Technology 43(1):1