NHS National Institute for Health Research

NIHR Statistical Meeting
13th February 2012

KING'S College LONDON

Sample Size by design in complex intervention trials

Toby Prevost
Unit for Medical Statistics
KCL

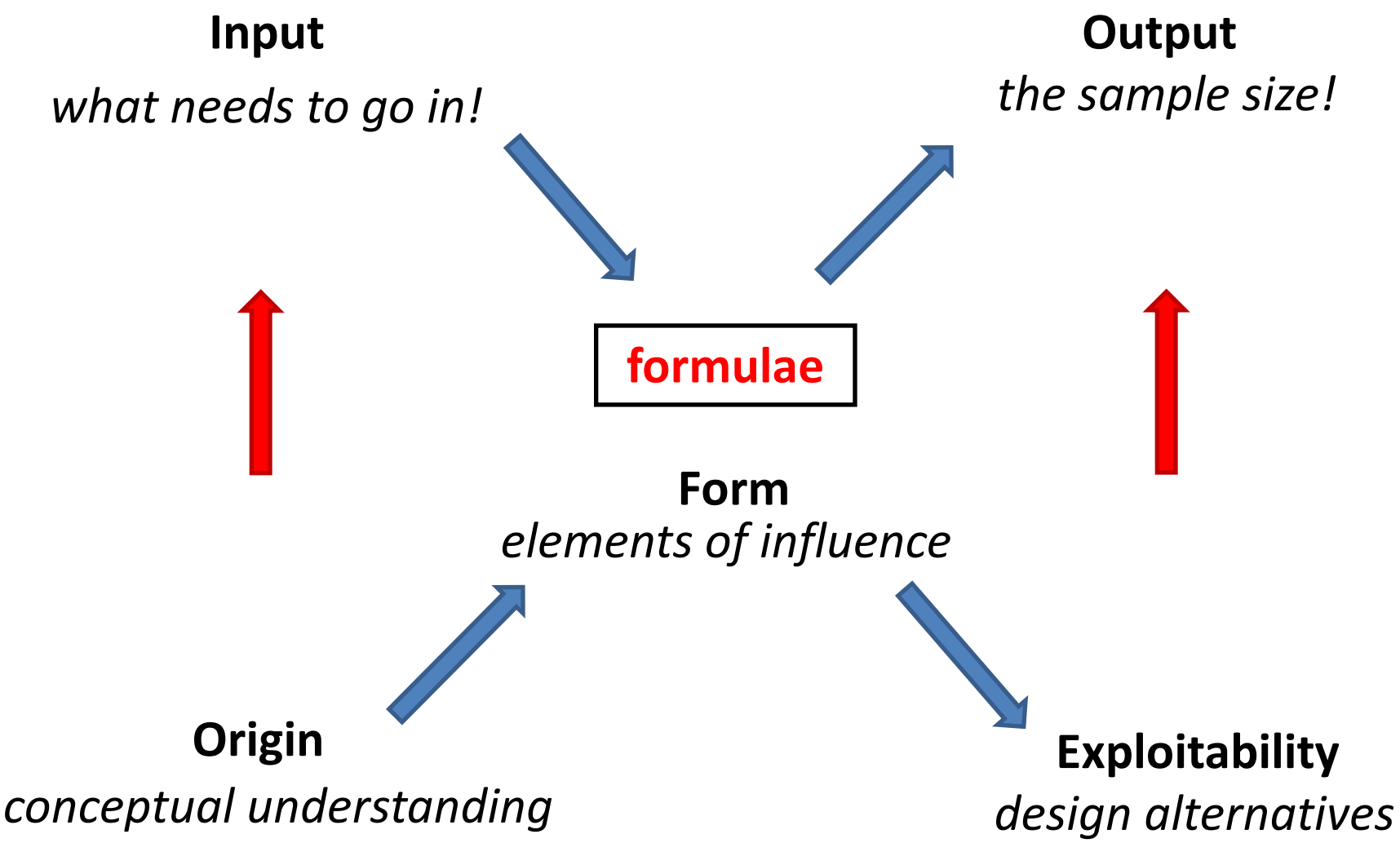
Fred Smith — University of Southampton

Chris Palmer — University of Cambridge

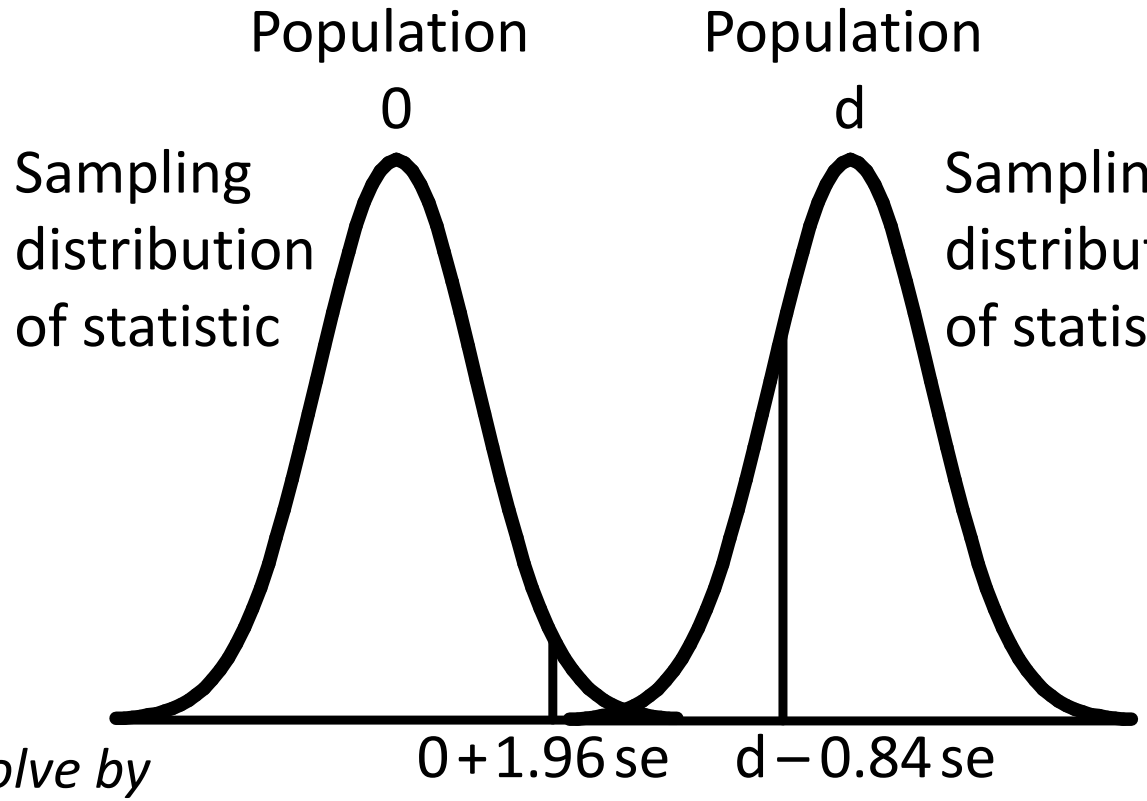
Ann-Louise Kinmonth — University of Cambridge

Simon Thompson — MRC Biostatistics Unit

What can sample size formulae tell us?

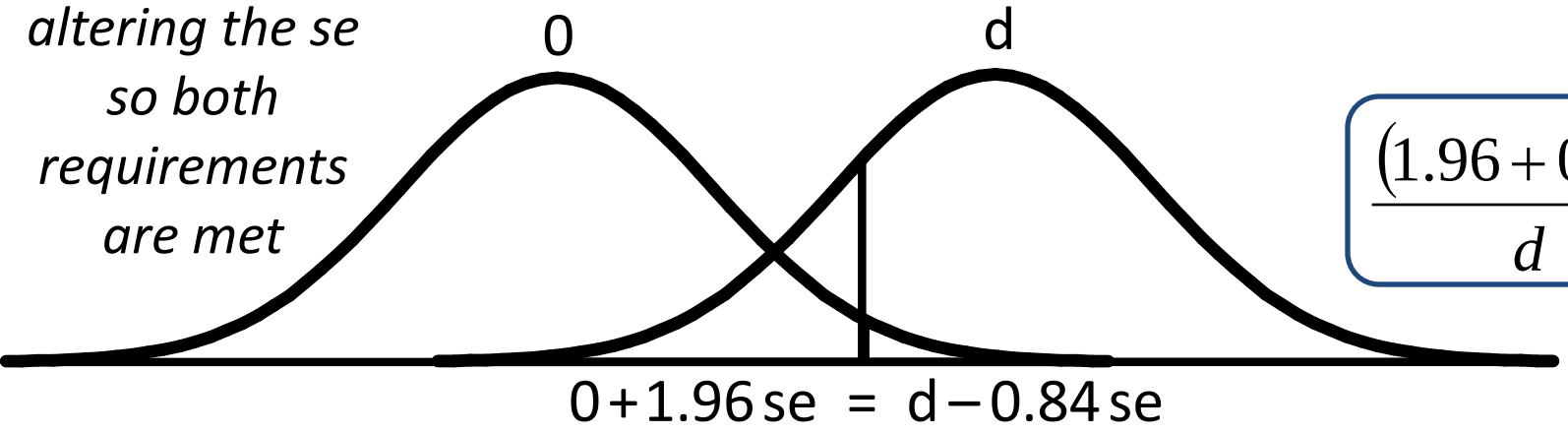


Formulae remind us – origins, concepts, understanding



Sampling Variability
Distributions
Normality
CLT
Estimates
Standard Error
Critical values
Powerful tests
Decisions

Resolve by altering the se so both requirements are met



$$\frac{(1.96+0.84)}{d}se = 1$$

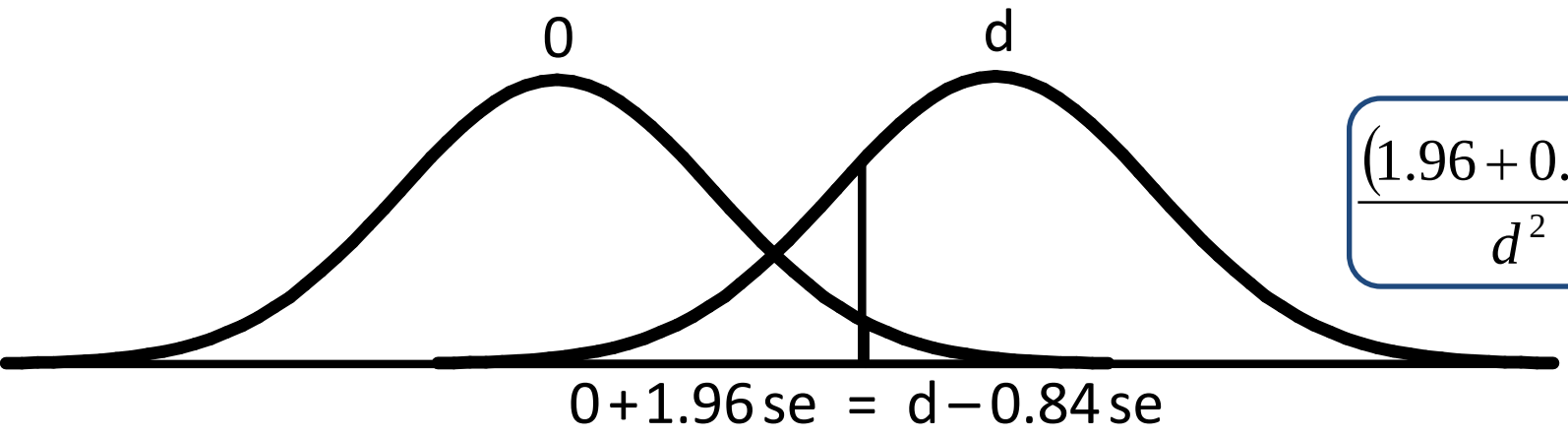
Comparing formulae - shows cost to se^2 of sub/groups

Mean/proportion; 1 group $$se^2 = sd^2\left[\frac{1}{n}\right] \qquad se = \frac{sd}{\sqrt{n}}$$

Difference; 2 groups (n per group) $$se^2 = sd^2\left[\frac{1}{n}+\frac{1}{n}\right] \qquad se = \sqrt{2}\,\frac{sd}{\sqrt{n}}$$

Difference in difference (Subgroup; interaction) $$se^2 = sd^2\left[\frac{1}{\frac{1}{2}n}+\frac{1}{\frac{1}{2}n}+\frac{1}{\frac{1}{2}n}+\frac{1}{\frac{1}{2}n}\right] \qquad se = \sqrt{8}\,\frac{sd}{\sqrt{n}}$$

Subgroup – assessed with test for interaction – detected with factorial design



$$\frac{(1.96+0.84)^2}{d^2}\,se^2 = 1$$

Factorial Design – two interventions/factors

- to test for ***interaction*** – needs the much larger size
- mainly used for ***efficiency*** – *buy 1 get 1 free*
  - estimate 1 factor combined over *levels* of the other

Montgomery AA, Astin AP, Peters TJ *Reporting of factorial trials of* ***complex interventions*** *in community settings: a systematic review.* Trials 2011;12:179

*~ 70% (38/53) based sample size on analysis "**at the margins**"*

- assumes no [population] interaction between interventions

A *sample* may provide very different intervention effect estimates at each level of the other factor "***inside the table***"

2 x 2 factorial example: Dunger et al. BMC Paediatrics 2009;9:79

Adolescent Type 1 Diabetes Cardio-Renal Intervention Trial

Allocation	No Statin	Statin	Ace factorial
No Ace	**125** **Neither**	**125** **Statin only**	250
Ace	**125** **Ace only**	**125** **Both**	250
Statin factorial	250	250	**500**

Outcome = log Albumin Creatinine Ratio (ACR); interpreted in geometric means

93% power to detect "***in the margin***" 25% reduction in ACR.

Sensitivity: 82% power to detect "***inside the table***" 30% reduction

Uses alternatives for ***power*** and "***d***"

Provides for consequence of interaction or chance

Choosing "*d*" - the detectable effect size

Consort on sample size:

- a balance between ***medical and statistical*** considerations
- large enough to detect a difference ***deemed clinically important*** with high power, if such a difference exists
- ***SD*** is an element of the calculation

Don't nominate/calculate a "d" that you alone consider important

Do ask questions to establish "worthwhile" and "plausible" ranges

Do ask for relevant good quality evidence (for "d" and sd)

Avoid side-stepping with Cohen's effect size

Cohen's effect size

$$= \frac{d}{sd}$$

Choices:

0.2 → "Small"
0.5 → "Medium"
0.8 → "Large"

- hides importance of 'd'
- Hides SD raised from measurement error

Explore longer recruitment, extra centre(s), design options b4 ***d***

Formulae have common form - with a design constant

Mean/proportion; 1 group $se^2 = 1 \bullet \frac{sd^2}{n}$

Difference; 2 groups $se^2 = 2 \bullet \frac{sd^2}{n}$

Difference in difference (Subgroup; interaction) $se^2 = 8 \bullet \frac{sd^2}{n}$

Difference; 2 unequal groups (in a:1 ratio) $se^2 = 2\frac{(a+1)^2}{4a} \bullet \frac{sd^2}{n}$

3:1 (**a=3**)
increases by 1/3

14:1 (**a=14**)
equates to
test for interaction

Elements of se^2

Outcome variability

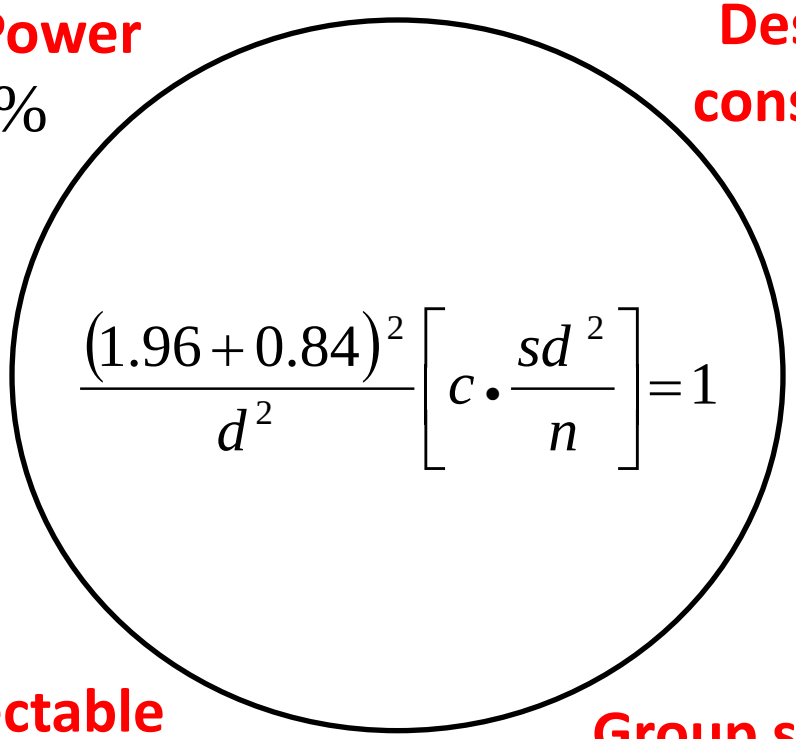
$$se^2 = c \bullet \frac{sd^2}{n}$$

Design constant

Group size

$$\frac{(1.96+0.84)^2}{d^2} se^2 = 1$$

Balancing the elements of influence



Power

$90\% \leftrightarrow 80\%$

Design constant

$a = 3 \leftrightarrow a = 1$

Significance level

fixed

$$\frac{(1.96+0.84)^2}{d^2}\left[c \bullet \frac{sd^2}{n}\right]=1$$

Outcome variability

Adjust for baseline

$$sd^2 \rightarrow sd^2\left(1-r^2\right)$$

r = 0.5

Detectable effect size

Group size

- Power / Adjusted SD offer <u>some</u> (25%) flexibility
- Though "***n***" is the <u>main</u> 'purchaser' of "***d***"
- *NB: Design constant can have appreciable effect*

Compare designs: ratio of sampling variances = design effect

Mean/proportion; 1 group $se^2 = 1 \bullet \frac{sd^2}{n}$ **(2)**

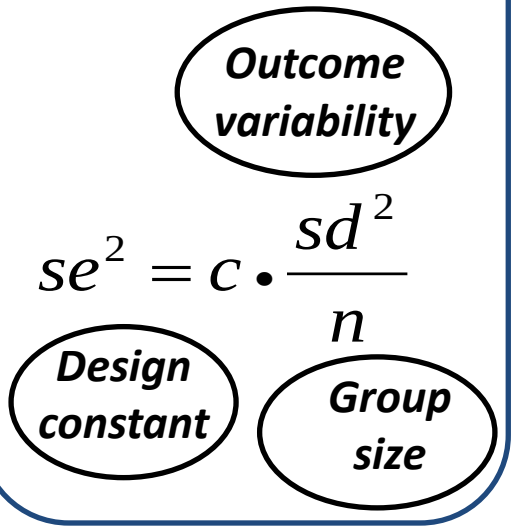
$$\frac{(1.96+0.84)^2}{d^2} se^2 = 1$$

Difference; 2 groups $se^2 = 2 \bullet \frac{sd^2}{n}$

Difference in difference (Subgroup; interaction) $se^2 = 8 \bullet \frac{sd^2}{n}$

Difference; 2 unequal groups (in a:1 ratio) $se^2 = 2\frac{(a+1)^2}{4a} \bullet \frac{sd^2}{n}$

Elements of se^2



Clusters of size *m*; 1 group
Intracluster correlation ρ $se^2 = [1+(m-1)\rho] \bullet \frac{sd^2}{n}$ **(1)**

Clustering introduces an unknown design constant

Design Effect = **(1) / (2)**
(**Kish**, 1951)

$$deff = 1+(m-1)\rho$$

Design effects (*deff*) due to clustering in surveys and trials

cluster mean

Single cluster of m subjects

$$se^2 = \frac{1}{m^2} V\left(y_1 + y_2 + \ldots y_m\right)$$

$$= \frac{1}{m^2}\left[m\, sd^2 + \binom{m}{2} 2\rho\, sd^2 \right]$$
$$\text{Cov}$$

$$= \frac{sd^2}{m}\underbrace{\left[1+(m-1)\rho\right]}_{deff}$$

survey mean

k independent clusters (n=km)

$$se^2 = \frac{sd^2}{n}\underbrace{\left[1+(m-1)\rho\right]}_{deff}$$

difference in trial means

2 groups each of k clusters

$$se^2 = 2\frac{sd^2}{n}\underbrace{\left[1+(m-1)\rho\right]}_{deff}$$

Inflate ***n*** from simple design by ***deff*** by inflating ***k*** (clusters) not ***m*** (size).

Inflating ***m*** increases the ***deff*** itself, further increasing ***m***, ***deff***, ***m***, ***deff*** …

complex intervention trial

with only 1 group clustered

$$se^2 = \frac{sd^2}{n} + \frac{sd^2}{n}\left[1+(m-1)\rho\right]$$

Campbell MJ *Stat Meth Med Res* 2000;9:81-94 **-** Cluster RCTs in general practice research

Eldridge S, Ashby D, Kerry S *Int J of Epi* 2006;35:1292-1300 – clusters of unequal size

What is the influence of this Design effect (*deff*)?

$$deff = 1 + (m-1)\rho$$ (product form)

ρ is fortunately typically...

- ***large*** when m is naturally ***small***
  "close-knit" - e.g. **household**

  see *m = 1.2 deff ≤ m bounded!*

- ***small*** when m is naturally ***large***
  e.g. **regions**

- **intermediate** (0.01 to 0.05)
  in **School Classes**, **General Practices**

  m=21, ρ=0.01 → *deff*=1.2
  ρ=0.05 → *deff=2* *double!*

It may be more efficient...

- to use **more clusters**
  with **reduced *m***
  via shorter recruitment time
  or subsample within cluster
  (depends on costs etc etc)

- to **avoid clustered design**, if possible,
  accepting probably relatively small
  bias from ***contamination (arguably!)***

Torgerson DJ *BMJ* 2001;322:355 - Cluster randomisation answers contamination?
Okoumunne OC, Gulliford MC et al *HTA* 1999;3:5 - evaluating area-wide interventions

Imprecision in variance estimates: '*impact*' on SS & analysis

sd^2: 95% CI by study size

n	95% CI Rel. to estimate
20	0.55 to 2.00
30	0.60 to 1.75
50	0.70 to 1.50
100	**0.75 to 1.33**
250	0.85 to 1.20

ρ: Upper 5% Confidence limit

For sample sizes of ***km***=200

k *clusters*	*of size* ***m***	Upper 5% CI with $\hat{\rho} = 0.01$
40	5	0.11
20	10	0.10
8	25	0.11
4	50	0.19

Estimated ← Applies backwards

Anticipated → Applies forwards

Use ***previous*** $\hat{\rho}$ based on many clusters/studies

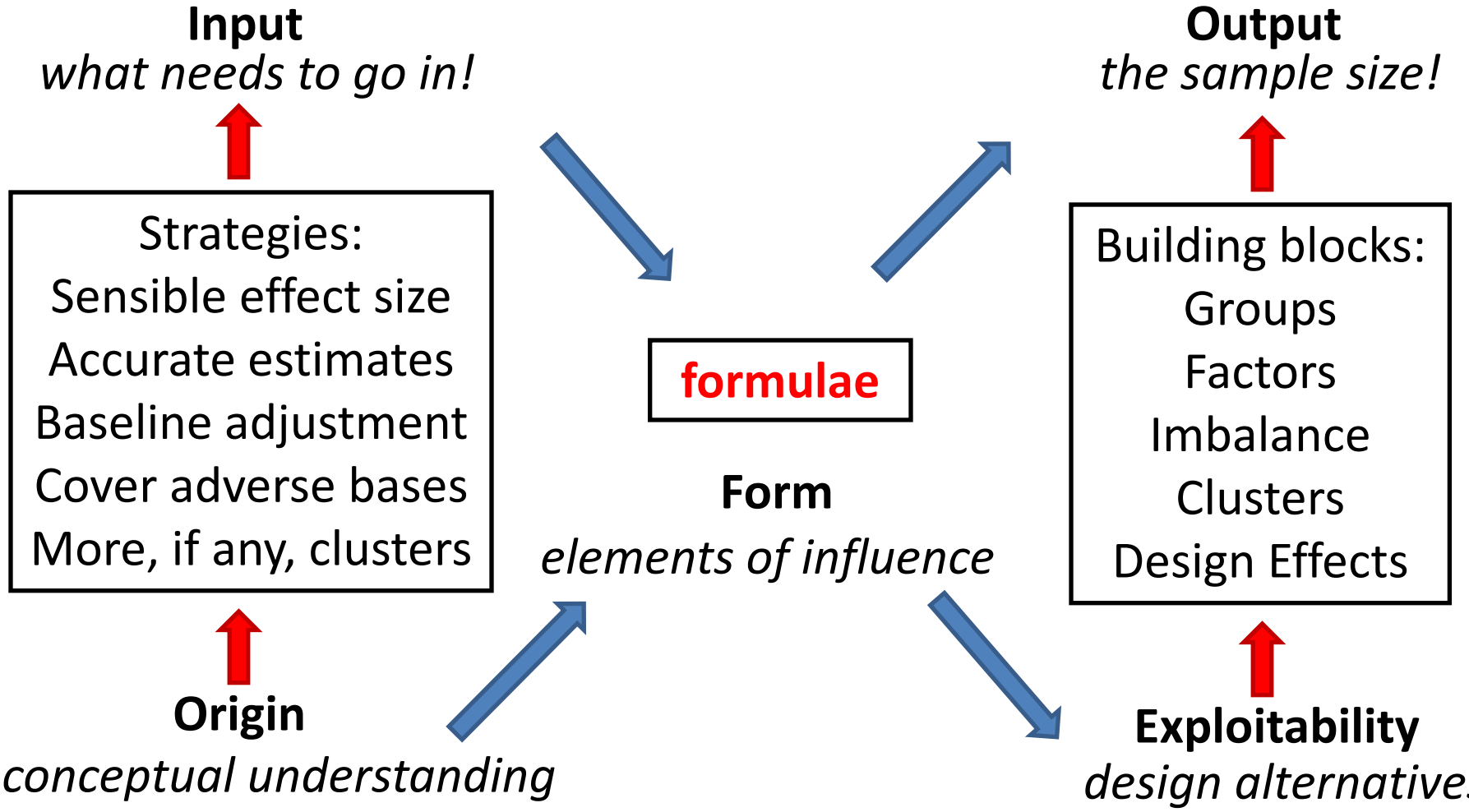
Reduce ***deff*** with reduced size clusters

Reduce '***impact***' by recruiting more clusters

Vickers AJ *Journal of Clin Epi* 2003;56:717-20 – The SD & Underpowering in RCTs

Turner R, Prevost AT, Thompson SG Allowing for imprecision of the intracluster correlation coefficient in the design of cluster RCTs *Stat Med* 2004;23:1195-1214

Sample Size by design - formulae inform better inputs



Donner A, Klar N (2000) Design & analysis of Cluster RCTs in health research, Arnold.
Julious SA (2009) Sample Sizes for Clinical Trials, CRC Press.
Bland JM The tyranny of power: ...a better way to calculate SS? *BMJ* 2009;339:b3985

@RSS on 26th March: Factorial / Two-stage / Stepped Wedge designs: why choose?