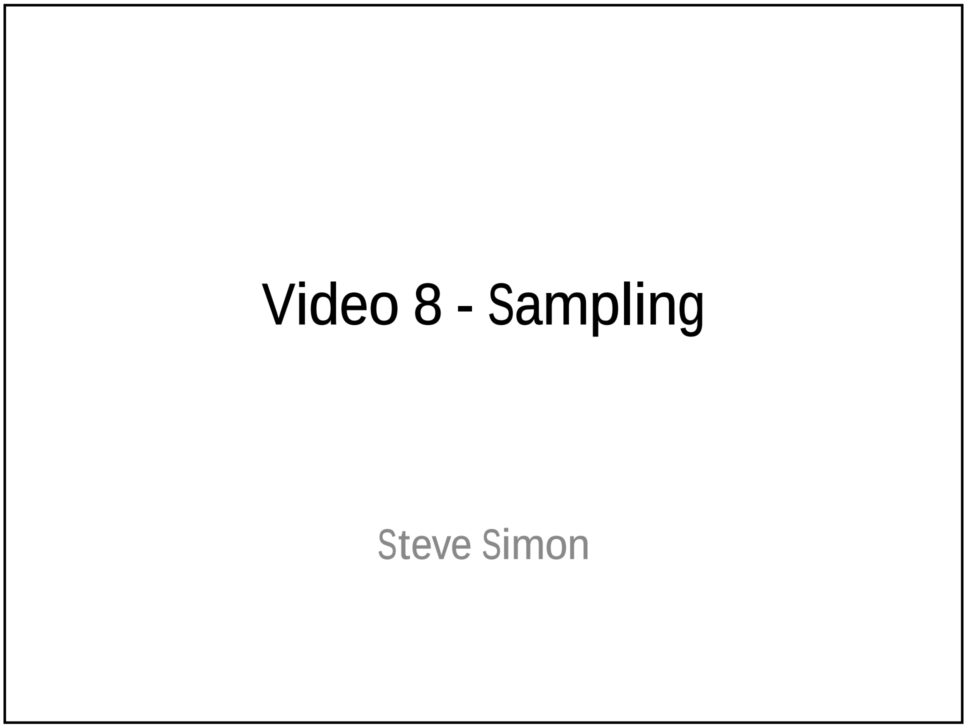
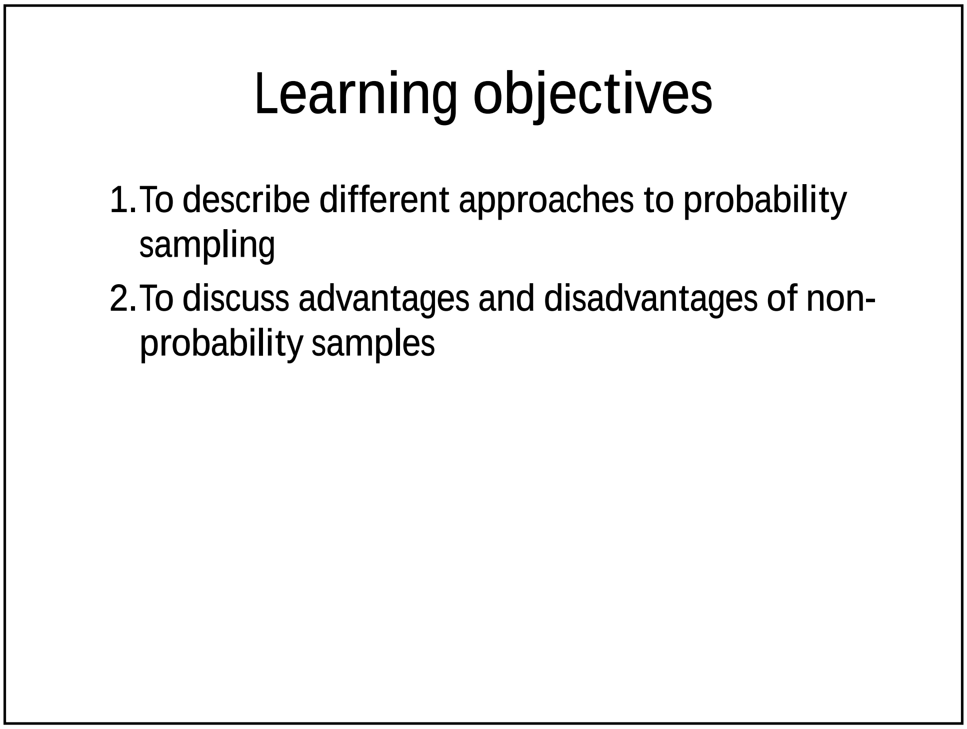


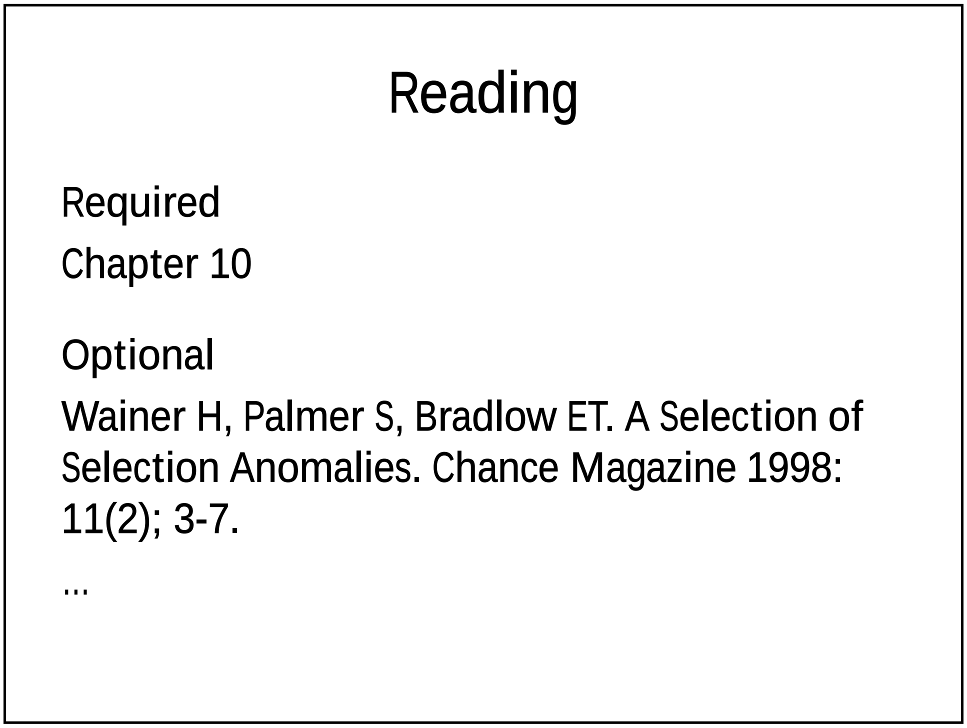
Video 8 - Sampling

Steve Simon



Learning objectives

1. To describe different approaches to probability sampling
2. To discuss advantages and disadvantages of non-probability samples



Reading

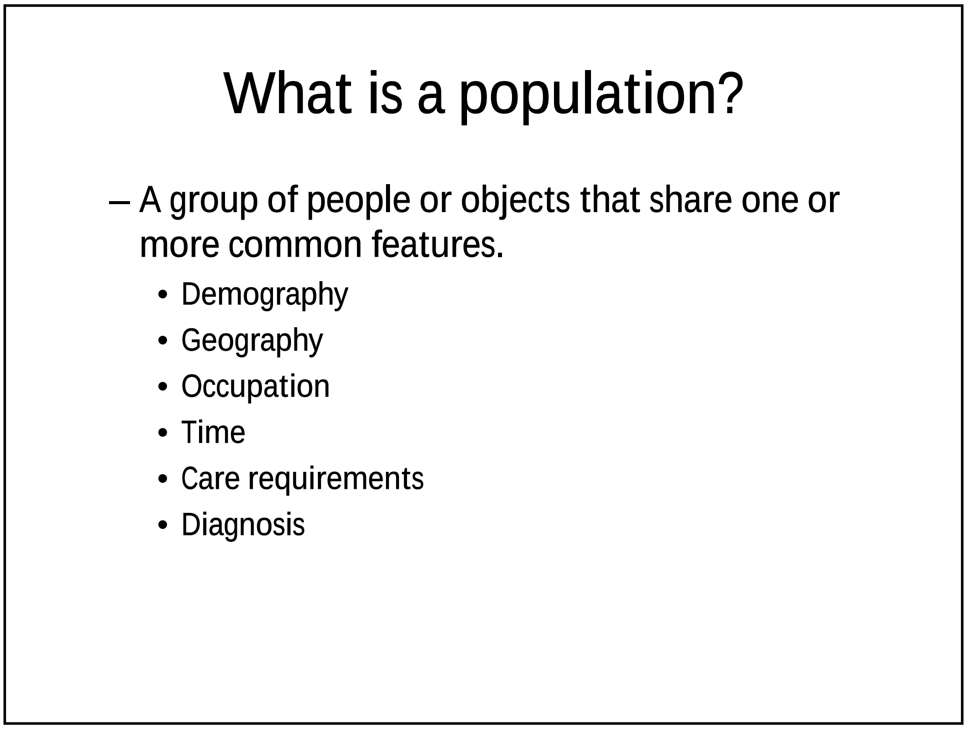
Required

Chapter 10

Optional

Wainer H, Palmer S, Bradlow ET. A Selection of Selection Anomalies. Chance Magazine 1998: 11(2); 3-7.

…



What is a population?

– A group of people or objects that share one or more common features.
  - Demography
  - Geography
  - Occupation
  - Time
  - Care requirements
  - Diagnosis

What is a sampling frame

- – Physical list
  - Ideally everyone or almost everyone in population
  - Used to draw your sample
- – Expensive, not always available.
- – Example: Master Address File.

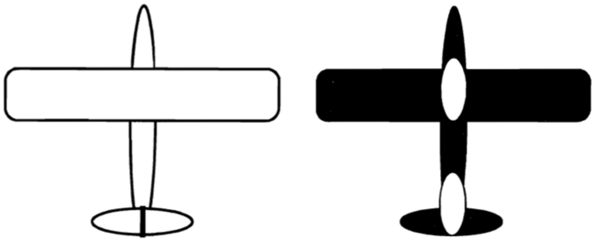
What is a sample?

- – A sample is a subset of a population
- – Representativeness more important than size
- – Reasons for sampling
  - Expense
  - Time
  - Quality control

Two major types of samples

- Random sample
  - Everyone has known non-zero probability
- Non-random sample
  - Different selection probabilities
  - Some may have zero selection probability

Extreme example: World War II bombers

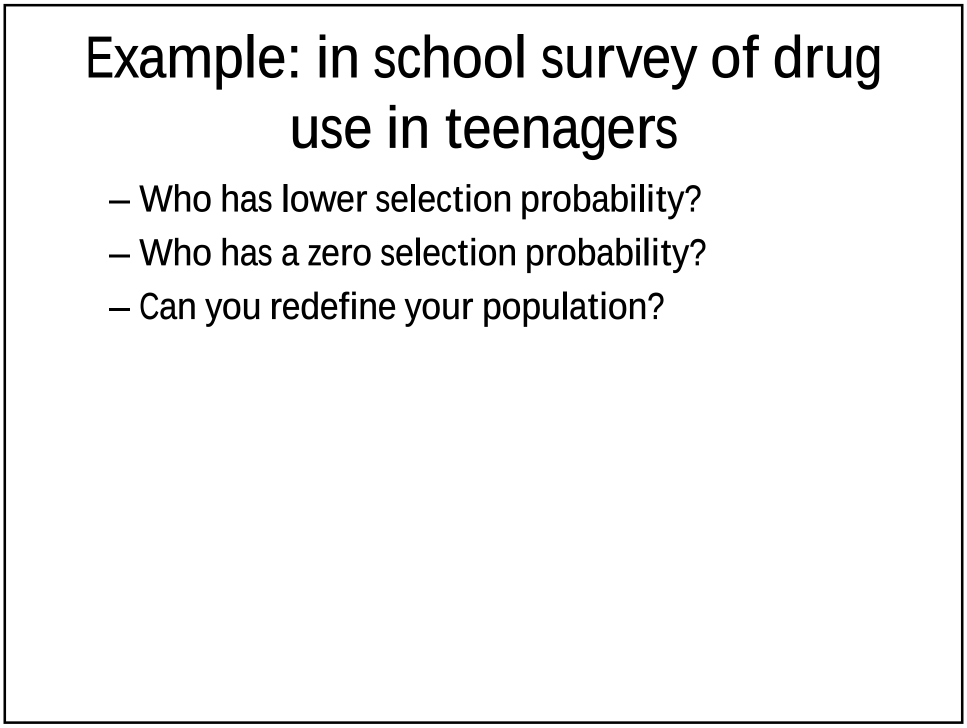


An outline of a plane.

A depiction of a plane
with shading indicating
where returning planes
had been shot.

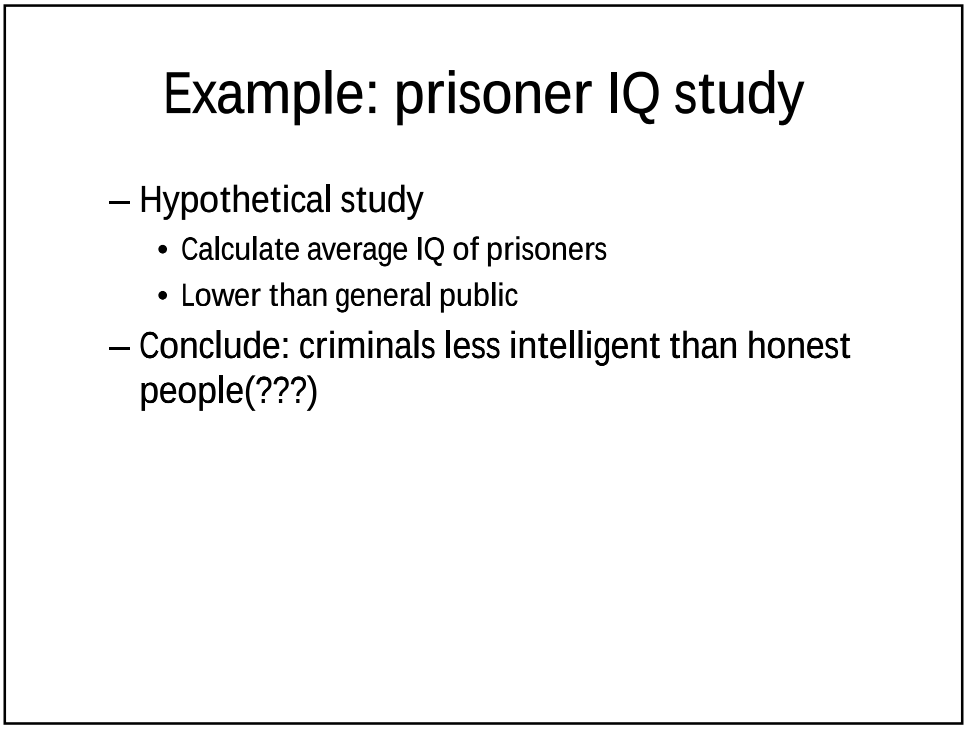
Figure 6. A schematic representation of Abraham Wald's ingenious scheme to investigate where to armor aircraft.

Image of bomber with indication of damage



Example: in school survey of drug use in teenagers

- Who has lower selection probability?
- Who has a zero selection probability?
- Can you redefine your population?



Example: prisoner IQ study

- Hypothetical study
  - Calculate average IQ of prisoners
  - Lower than general public
- Conclude: criminals less intelligent than honest people(???)

Take a break here

– What have we learned so far?
  - Definitions: population, sampling frame, sample
  - Distinction between random sample and a non-random sample

– What is coming next?
  - Different types of probability samples
  - How to generate a random sample

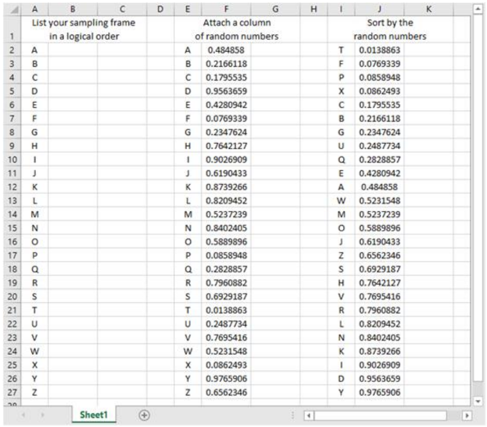
Sampling

– Sampling designs – Probability sampling
  - Simple random sampling
  - Systematic sampling
  - Stratified sampling
  - Cluster sampling

How to draw a simple random sample

1. List the sampling frame in a logical order
2. Attach a column of random numbers
3. Sort by the column of random numbers
4. Select your sample, starting at the top

Simple random sample using Microsoft Excel



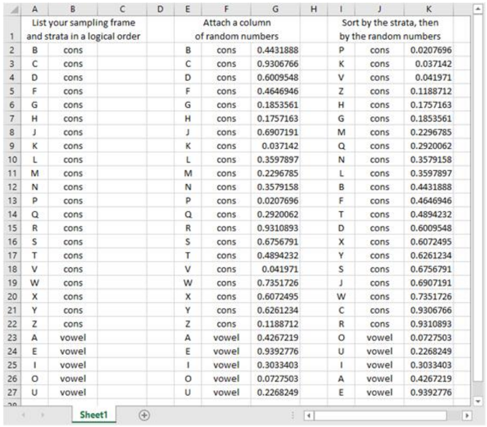
	A	B	C	D	E	F	G	H	I	J	K
1	List your sampling frame in a logical order				Attach a column of random numbers				Sort by the random numbers		
2	A				A	0.484858			T	0.0138863	
3	B				B	0.2166118			F	0.0769339	
4	C				C	0.1795535			P	0.0858948	
5	D				D	0.9563659			X	0.0862493	
6	E				E	0.4280942			C	0.1795535	
7	F				F	0.0769339			B	0.2166118	
8	G				G	0.2347624			G	0.2347624	
9	H				H	0.7642127			U	0.2487734	
10	I				I	0.9026909			Q	0.2828857	
11	J				J	0.6190433			E	0.4280942	
12	K				K	0.8739266			A	0.484858	
13	L				L	0.8209452			W	0.5231548	
14	M				M	0.5237239			M	0.5237239	
15	N				N	0.8402405			O	0.5889896	
16	O				O	0.5889896			J	0.6190433	
17	P				P	0.0858948			Z	0.6562346	
18	Q				Q	0.2828857			S	0.6929187	
19	R				R	0.7960882			H	0.7642127	
20	S				S	0.6929187			V	0.7695416	
21	T				T	0.0138863			R	0.7960882	
22	U				U	0.2487734			L	0.8209452	
23	V				V	0.7695416			N	0.8402405	
24	W				W	0.5231548			K	0.8739266	
25	X				X	0.0862493			I	0.9026909	
26	Y				Y	0.9765906			D	0.9563659	
27	Z				Z	0.6562346			Y	0.9765906	

A spreadsheet illustrating simple random sampling

How to draw a stratified random sample

1. List the sampling frame and strata in a logical order
2. Attach a column of random numbers
3. Sort by the strata and the column of random numbers
4. Select your sample, starting at the top

Stratified random sample using Microsoft Excel



	A	B	C	D	E	F	G	H	I	J	K
1	List your sampling frame and strata in a logical order				Attach a column of random numbers				Sort by the strata, then by the random numbers		
2	B	cons			B	cons	0.4431888		P	cons	0.0207696
3	C	cons			C	cons	0.9306766		K	cons	0.037142
4	D	cons			D	cons	0.6009548		V	cons	0.041971
5	F	cons			F	cons	0.4646946		Z	cons	0.1188712
6	G	cons			G	cons	0.1853561		H	cons	0.1757163
7	H	cons			H	cons	0.1757163		G	cons	0.1853561
8	J	cons			J	cons	0.6907191		M	cons	0.2296785
9	K	cons			K	cons	0.037142		Q	cons	0.2920062
10	L	cons			L	cons	0.3597897		N	cons	0.3579158
11	M	cons			M	cons	0.2296785		L	cons	0.3597897
12	N	cons			N	cons	0.3579158		B	cons	0.4431888
13	P	cons			P	cons	0.0207696		F	cons	0.4646946
14	Q	cons			Q	cons	0.2920062		T	cons	0.4894232
15	R	cons			R	cons	0.9310893		D	cons	0.6009548
16	S	cons			S	cons	0.6756791		X	cons	0.6072495
17	T	cons			T	cons	0.4894232		Y	cons	0.6261234
18	V	cons			V	cons	0.041971		S	cons	0.6756791
19	W	cons			W	cons	0.7351726		J	cons	0.6907191
20	X	cons			X	cons	0.6072495		W	cons	0.7351726
21	Y	cons			Y	cons	0.6261234		C	cons	0.9306766
22	Z	cons			Z	cons	0.1188712		R	cons	0.9310893
23	A	vowel			A	vowel	0.4267219		O	vowel	0.0727503
24	E	vowel			E	vowel	0.9392776		U	vowel	0.2268249
25	I	vowel			I	vowel	0.3033403		I	vowel	0.3033403
26	O	vowel			O	vowel	0.0727503		A	vowel	0.4267219
27	U	vowel			U	vowel	0.2268249		E	vowel	0.9392776

A spreadsheet illustrating stratified random sampling

Take another break here

– What have we learned so far?
  - Types of probability samples
  - How to draw a random sample

– What is coming up next?
  - Different types of non-probability samples
  - How to allocate treatments randomly

Sampling

– Sampling designs – Nonprobability sampling
  - Convenience sampling
  - Quota sampling
  - Purposive sampling
  - Purposeful sampling
  - Snowball sampling

Example of a purposive sample

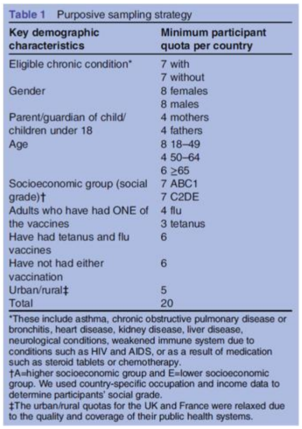


Table 1 Purposive sampling strategy

Key demographic characteristics	Minimum participant quota per country
Eligible chronic condition*	7 with 7 without
Gender	8 females 8 males
Parent/guardian of child/ children under 18	4 mothers 4 fathers
Age	8 18–49 4 50–64 6 ≥65
Socioeconomic group (social grade)†	7 ABC1 7 C2DE
Adults who have had ONE of the vaccines	4 flu 3 tetanus
Have had tetanus and flu vaccines	6
Have not had either vaccination	6
Urban/rural‡	5
Total	20

*These include asthma, chronic obstructive pulmonary disease or bronchitis, heart disease, kidney disease, liver disease, neurological conditions, weakened immune system due to conditions such as HIV and AIDS, or as a result of medication such as steroid tablets or chemotherapy.

†A=higher socioeconomic group and E=lower socioeconomic group. We used country-specific occupation and income data to determine participants' social grade.

‡The urban/rural quotas for the UK and France were relaxed due to the quality and coverage of their public health systems.

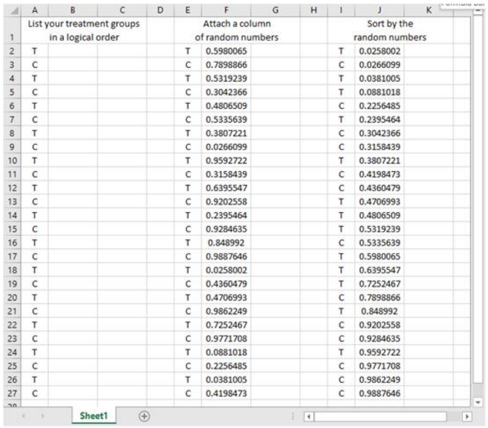
Table describing purposive sampling strategy

Randomizing treatments within a convenience sample

Many studies use a convenience sample, which may hamper external validity, but they randomly assign treatment or control conditions within the convenience sample, which helps with internal validity. The process works much like the process of drawing a simple random sample.

1. List your treatment groups in a logical order
2. Attach a column of random numbers
3. Sort by the column of random numbers
4. Allocate treatment groups, starting at the top of the list

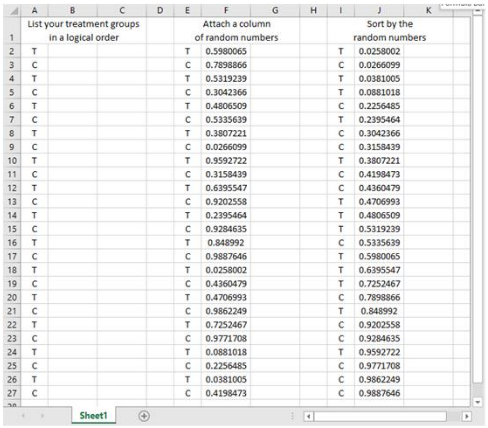
Randomizing treatment allocation using Microsoft Excel



	A	B	C	D	E	F	G	H	I	J	K
1	List your treatment groups in a logical order				Attach a column of random numbers				Sort by the random numbers		
2	T				T	0.5980065			T	0.0258002	
3	C				C	0.7898866			C	0.0266099	
4	T				T	0.5319239			T	0.0381005	
5	C				C	0.3042366			T	0.0881018	
6	T				T	0.4806509			C	0.2256485	
7	C				C	0.5335639			T	0.2395464	
8	T				T	0.3807221			C	0.3042366	
9	C				C	0.0266099			C	0.3158439	
10	T				T	0.9592722			T	0.3807221	
11	C				C	0.3158439			C	0.4198473	
12	T				T	0.6395547			C	0.4360479	
13	C				C	0.9202558			T	0.4706993	
14	T				T	0.2395464			T	0.4806509	
15	C				C	0.9284635			T	0.5319239	
16	T				T	0.848992			C	0.5335639	
17	C				C	0.9887646			T	0.5980065	
18	T				T	0.0258002			T	0.6395547	
19	C				C	0.4360479			T	0.7252467	
20	T				T	0.4706993			C	0.7898866	
21	C				C	0.9862249			T	0.848992	
22	T				T	0.7252467			C	0.9202558	
23	C				C	0.9771708			C	0.9284635	
24	T				T	0.0881018			T	0.9592722	
25	C				C	0.2256485			C	0.9771708	
26	T				T	0.0381005			C	0.9862249	
27	C				C	0.4198473			C	0.9887646	

A spreadsheet illustrating random treatment allocation

Randomizing a crossover trial using Microsoft Excel



	A	B	C	D	E	F	G	H	I	J	K
1	List your treatment groups in a logical order				Attach a column of random numbers				Sort by the random numbers		
2	T				T	0.5980065			T	0.0258002	
3	C				C	0.7898866			C	0.0266099	
4	T				T	0.5319239			T	0.0381005	
5	C				C	0.3042366			T	0.0881018	
6	T				T	0.4806509			C	0.2256485	
7	C				C	0.5335639			T	0.2395464	
8	T				T	0.3807221			C	0.3042366	
9	C				C	0.0266099			C	0.3158439	
10	T				T	0.9592722			T	0.3807221	
11	C				C	0.3158439			C	0.4198473	
12	T				T	0.6395547			C	0.4360479	
13	C				C	0.9202558			T	0.4706993	
14	T				T	0.2395464			T	0.4806509	
15	C				C	0.9284635			T	0.5319239	
16	T				T	0.848992			C	0.5335639	
17	C				C	0.9887646			T	0.5980065	
18	T				T	0.0258002			T	0.6395547	
19	C				C	0.4360479			T	0.7252467	
20	T				T	0.4706993			C	0.7898866	
21	C				C	0.9862249			T	0.848992	
22	T				T	0.7252467			C	0.9202558	
23	C				C	0.9771708			C	0.9284635	
24	T				T	0.0881018			T	0.9592722	
25	C				C	0.2256485			C	0.9771708	
26	T				T	0.0381005			C	0.9862249	
27	C				C	0.4198473			C	0.9887646	

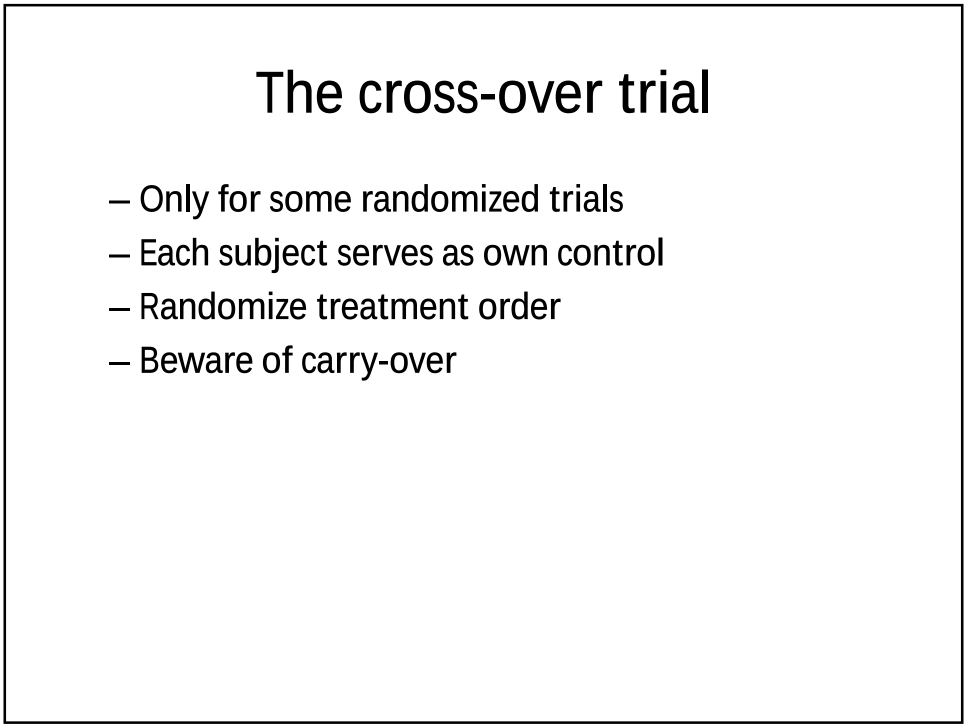
A spreadsheet illustrating random allocation of treatment order

Matching and pairing

- Improved precision
- Logistical issues
- Works for both randomized and observational studies

The logistics of matching

- Not obvious
- Simplest solution: greedy matching
- Unpaired patients are lost to your analysis
  - Extra precision from pairing
  - Loss of precision from loss of the unpaired.



The cross-over trial

- – Only for some randomized trials
- – Each subject serves as own control
- – Randomize treatment order
- – Beware of carry-over