Integrating genetic sequencing information into vaccine effectiveness analyses

Experience from I-MOVE influenza

Influenza (sub)clade-specific vaccine effectiveness

- Influenza VE by (sub)clade provides key evidence on how well the vaccines perform against currently circulating viruses
- Helps inform the WHO influenza vaccine strain selection committee
- We can do the same for SARS-CoV-2: variant-specific VE
  - Monitor vaccine performance
  - Provide information for composition of new vaccines

Influenza (sub)clade-specific vaccine effectiveness, selected I-MOVE publications

Table 3
Pooled adjusted seasonal vaccine effectiveness against influenza A(H3N2) clades and subclades, I-MOVE/I-MOVE+ primary care multicenter study, Europe, influenza seasons 2016–17 and 2017–18.

Season	Clade/subclade	N	Cases; vacc /Controls; vacc	Adjusted VE (%)	CI (%)
2016–17	3C.2a (all variants excluding 3C.2a1)	4492	252;28/4240;496	31	−12 to 57
	3C.2a3 N121K+S144K+(N122D+262 N)	4445	205;20/4240;496	37	−9 to 63
	3C.2a1 (all variants)	5007	767;99/4240;496	25	−1 to 44
	3C.2a1 (no mutations in A-E antigenic sites other than N171K)	4454	214;19/4240;496	56	20 to 76
	3C.2a1b N171K+N121K+K92R+H311Q	4397	157;30/4240;496	7	−56 to 45

TABLE 2
Pooled adjusted seasonal vaccine effectiveness against influenza A(H3N2), overall, by age groups, by clade and genetic variants, I-MOVE primary care multicentre study, Europe, influenza season 2018/19 (n = 5,802)

Age group	Outcome	n	Cases		Controls		Adjusted VE (%)	95% CI (%)
			All	Vaccinated	All	Vaccinated		
All ages	A(H3N2) clade 3C.2a1b + T131K[a]	2,582	131	15	2,451	329	57	16–78
15–64 years		1,468	81	7	1,387	141	51	−21 to 80
All ages	A(H3N2) clade 3C.2a1b + T135K[a]	2,764	203	40	2,561	342	7	−52 to 43
15–64 years		1,515	130	19	1,385	145	−7	−102 to 43

Influenza (sub)clade-specific vaccine effectiveness, selected I-MOVE publications

Table 3
Pooled adjusted seasonal vaccine effectiveness against influenza A(H3N2) clades and subclades, I-MOVE/I-MOVE+ primary care multicenter study, Europe, influenza seasons 2016–17 and 2017–18.

Season	Clade/subclade	N	Cases; vacc /Controls; vacc	Adjusted VE (%)	CI (%)
2016–17	3C.2a (all variants excluding 3C.2a1)	4492	252;28/4240;496	31	−12 to 57
	3C.2a3 N121K+S144K+(N122D+262 N)	4445	205;20/4240;496	37	−9 to 63
	3C.2a1 (all variants)	5007	767;99/4240;496	25	−1 to 44
	3C.2a1 (no mutations in A-E antigenic sites other than N171K)	4454	214;19/4240;496	56	20 to 76
	3C.2a1b N171K+N121K+K92R+H311Q	4397	157;30/4240;496	7	−56 to 45

Clade-specific VE, including specific mutations

TABLE 2
Pooled adjusted seasonal vaccine effectiveness against influenza A(H3N2), overall, by age groups, by clade and genetic variants, I-MOVE primary care multicentre study, Europe, influenza season 2018/19 (n = 5,802)

Age group	Outcome	n	Cases		Controls		Adjusted VE (%)	95% CI (%)
			All	Vaccinated	All	Vaccinated		
All ages	A(H3N2) clade 3C.2a1b + T131K[a]	2,582	131	15	2,451	329	57	16–78
15–64 years		1,468	81	7	1,387	141	51	−21 to 80
All ages	A(H3N2) clade 3C.2a1b + T135K[a]	2,764	203	40	2,561	342	7	−52 to 43
15–64 years		1,515	130	19	1,385	145	−7	−102 to 43

Influenza (sub)clade-specific vaccine effectiveness, selected I-MOVE publications

Table 3
Pooled adjusted seasonal vaccine effectiveness against influenza A(H3N2) clades and subclades, I-MOVE/I-MOVE+ primary care multicenter study, Europe, influenza seasons 2016–17 and 2017–18.

Season	Clade/subclade	N	Cases; vacc /Controls; vacc	Adjusted VE (%)	CI (%)
2016–17	3C.2a (all variants excluding 3C.2a1)	4492	252;28/4240;496	31	−12 to 57
	3C.2a3 N121K+S144K+(N122D+262 N)	4445	205;20/4240;496	37	−9 to 63
	3C.2a1 (all variants)	5007	767;99/4240;496	25	−1 to 44
	3C.2a1 (no mutations in A-E antigenic sites other than N171K)	4454	214;19/4240;496	56	20 to 76
	3C.2a1b N171K+N121K+K92R+H311Q	4397	157;30/4240;496	7	−56 to 45

Clade-specific VE, including specific mutations

TABLE 2
Pooled adjusted seasonal vaccine effectiveness against influenza A(H3N2), overall, by age groups, by clade and genetic variants, I-MOVE primary care multicentre study, Europe, influenza season 2018/19 (n = 5,802)

Age group	Outcome	n	Cases		Controls		Adjusted VE (%)	95% CI (%)
			All	Vaccinated	All	Vaccinated		
All ages	A(H3N2) clade 3C.2a1b+T131K[a]	2,582	131	15	2,451	329	57	16–78
15–64 years		1,468	81	7	1,387	141	51	−21 to 80
All ages	A(H3N2) clade 3C.2a1b+T135K[a]	2,764	203	40	2,561	342	7	−52 to 43
15–64 years		1,515	130	19	1,385	145	−7	−102 to 43

Indication a specific mutation is affecting VE

Process to measure variant-specific VE

- Selection of viruses to sequence
- Link genetic and epidemiological/clinical data
- Carry out analysis taking sampling frame into account

Selection of viruses to sequence

- To measure variant-specific VE:
  Representative viruses should be selected for sequencing
  - Not focussing on vaccinated, certain age groups, certain risk groups
- Depending on capacity
  - Sequence all viruses
  - Proportion of viruses

→ Depends on CT value (viral load) as well.
Some specimens technically difficult to sequence

Sequencing a proportion of viruses: How does this work?

- Early in the influenza season: 100% of viruses (technically feasible) sequenced
- Peak influenza season: 50% of viruses
- Keep note of sampling time periods:

Time period	First date*	Last date*	Sampling fraction used	Comments
1	05/11/2018	27/01/2019	1	week 45/2018 - 03/2019
2	28/01/2019	10/02/2019	0.50	week 04/2019 – week 06/2019
3	11/02/2019	07/04/2019	1	week 07/2019 - 14/2019

* Based on swab date, or date of receipt at lab, for example.

Sequencing a proportion of viruses: How does this work?

- Easier to do this retrospectively
  - Take all viruses from the sampling time period (weeks 4–6 2019): list by sample ID
  - Assign a random number to each virus, e.g. in Excel: RAND()

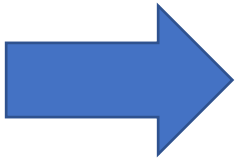
Sample ID	Random number
A-1863	0.996197911
A-1903	0.06806494
A-3071	0.466662581
B-3073	0.432270336
B-3074	0.735710169
B-3399	0.672602049
B-3439	0.711426169
B-4189	0.587049379
B-4202	0.808299908

- If using Excel, paste the random number as values (or they keep changing)
- Sort by the random number
- Select the desired number of viruses, e.g. if you have 12 viruses and plan to sequence 50%, select 6 viruses

Sequencing a proportion of viruses: How does this work?

Sample ID	Random number
A-1863	0.430552
A-1903	0.414266
A-3071	0.635639
B-3073	0.874289
B-3074	0.213095
B-3399	0.729659
B-3439	0.967658
B-4189	0.472435
B-4202	0.036803
C-4894	0.694327
C-5010	0.293686
C-5927	0.047221

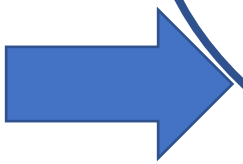
Sorting



Sample ID	Random number
B-3439	0.967658
B-3073	0.874289
B-3399	0.729659
C-4894	0.694327
A-3071	0.635639
B-4189	0.472435
A-1863	0.430552
A-1903	0.414266
C-5010	0.293686
B-3074	0.213095
C-5927	0.047221
B-4202	0.036803

Sequencing a proportion of viruses: How does this work?

Sample ID	Random number
A-1863	0.430552
A-1903	0.414266
A-3071	0.635639
B-3073	0.874289
B-3074	0.213095
B-3399	0.729659
B-3439	0.967658
B-4189	0.472435
B-4202	0.036803
C-4894	0.694327
C-5010	0.293686
C-5927	0.047221



Sample ID	Random number
B-3439	0.967658
B-3073	0.874289
B-3399	0.729659
C-4894	0.694327
A-3071	0.635639
B-4189	0.472435
A-1863	0.430552
A-1903	0.414266
C-5010	0.293686
B-3074	0.213095
C-5927	0.047221
B-4202	0.036803

50%

Sequencing a proportion of viruses: How does this work?

Sample ID	Random number
A-1863	0.430552
A-1903	0.414266
A-3071	0.635639
B-3073	0.874289
B-3074	0.213095
B-3399	0.729659
B-3439	0.967658
B-4189	0.472435
B-4202	0.036803
C-4894	0.694327
C-5010	0.293686
C-5927	0.047221

Sample ID	Random number
B-3439	0.967658
B-3073	0.874289
B-3399	0.729659
C-4894	0.694327
A-3071	0.635639
B-4189	0.472435
A-1863	0.430552
A-1903	0.414266
C-5010	0.293686
B-3074	0.213095
C-5927	0.047221
B-4202	0.036803

High CT value, cannot be sequenced

Sequencing a proportion of viruses: How does this work?

Sample ID	Random number
A-1863	0.430552
A-1903	0.414266
A-3071	0.635639
B-3073	0.874289
B-3074	0.213095
B-3399	0.729659
B-3439	0.967658
B-4189	0.472435
B-4202	0.036803
C-4894	0.694327
C-5010	0.293686
C-5927	0.047221

Sample ID	Random number
B-3439	0.967658
B-3073	0.874289
B-3399	0.729659
C-4894	0.694327
A-3071	0.635639
B-4189	0.472435
A-1863	0.430552
A-1903	0.414266
C-5010	0.293686
B-3074	0.213095
C-5927	0.047221
B-4202	0.036803

High CT value, cannot be sequenced

Take next virus in the sequence

Sequencing a proportion of viruses: How does this work?

Sample ID	Random number
A-1863	0.430552
A-1903	0.414266
A-3071	0.635639
B-3073	0
B-3074	0
B-3399	0.729659
B-3439	0.967658
B-4189	0.472435
B-4202	0.036803
C-4894	0.694327
C-5010	0.293686
C-5927	0.047221

Reasons for not sequencing should be documented

Sample ID	Random number
B-3439	0.967658
B-3073	0.874289
B-3399	0.729659
	0.694327
	0.635639
B-4189	0.472435
A-1863	0.430552
A-1903	0.414266
C-5010	0.293686
B-3074	0.213095
C-5927	0.047221
B-4202	0.036803

High CT value, cannot be sequenced

Take next virus in the sequence

Sequencing a proportion of viruses: How does this work?

- If you would like to sequence more, select from the list:

Sample ID	Random number
B-3439	0.967658
B-3073	0.874289
B-3399	0.729659
C-4894	0.694327
A-3071	0.635639
B-4189	0.472435
A-1863	0.430552
A-1903	0.414266
C-5010	0.293686
B-3074	0.213095
C-5927	0.047221
B-4202	0.036803

Sequencing a proportion of viruses: How does this work?

- If you would like to sequence more, select from the list:
- Do not add viruses from a different time period to the list, this won't be random. Create a new list

Week 07-12 2019	
Sample ID	**Random number**
E-3174	0.886907
E-3075	0.793435
F-4299	0.787123
E-3499	0.762857
C-8010	0.479988
D-1933	0.380644
D-3471	0.337251
F-4202	0.148071
E-3639	0.100943
D-1963	0.095506
C-6730	0.016687

Sample ID	Random number
B-3439	0.967658
B-3073	0.874289
B-3399	0.729659
C-4894	0.694327
A-3071	0.635639
B-4189	0.472435
A-1863	0.430552
A-1903	0.414266
C-5010	0.293686
B-3074	0.213095
C-5927	0.047221
B-4202	0.036803

Sequencing a proportion of viruses: How does this work?

- If you would like to sequence more, select from the list:
- Do not add viruses from a different time period to the list, this won't be random. Create a new list

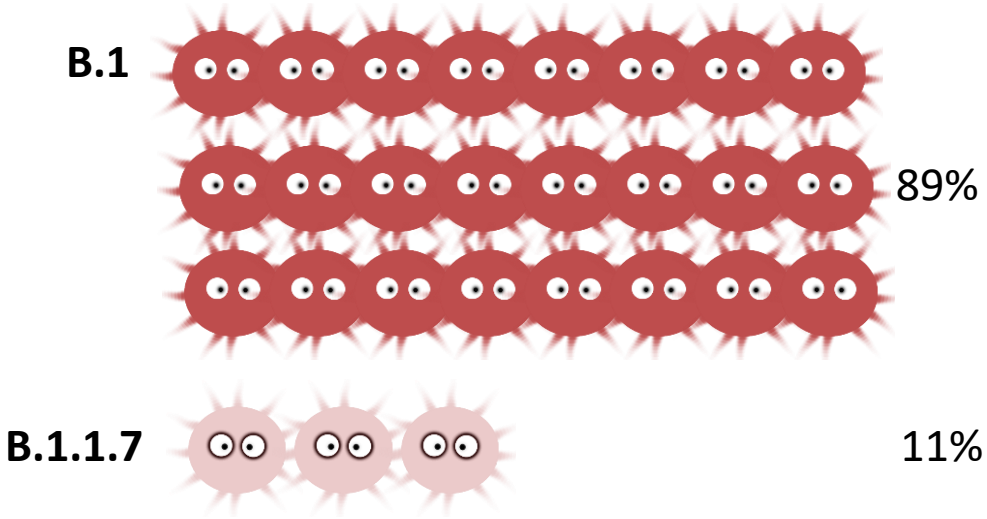
Week 07-12 20	
Sample ID	Random n
E-3174	
E-3075	
F-4299	
E-3499	0.762857
C-8010	0.479988
D-1933	0.380644
D-3471	0.337251
F-4202	0.148071
E-3639	0.100943
D-1963	0.095506
C-6730	0.016687

For influenza, you need separate lists for each influenza (sub)type, e.g. A(H1N1)pdm09, A(H3N2), B (for each time period)

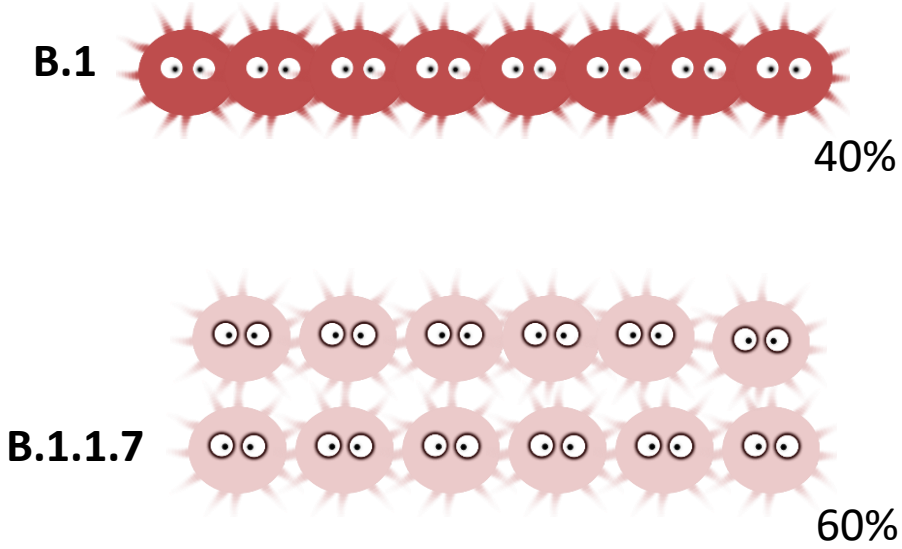
Sample ID	Random number
B-3439	0.967658
B-3073	0.874289
B-3399	0.729659
C-4894	0.694327
A-3071	0.635639
B-4189	0.472435
A-1863	0.430552
A-1903	0.414266
C-5010	0.293686
B-3074	0.213095
C-5927	0.047221
B-4202	0.036803

Why is the sampling time period important?

- Sampling time period 1:

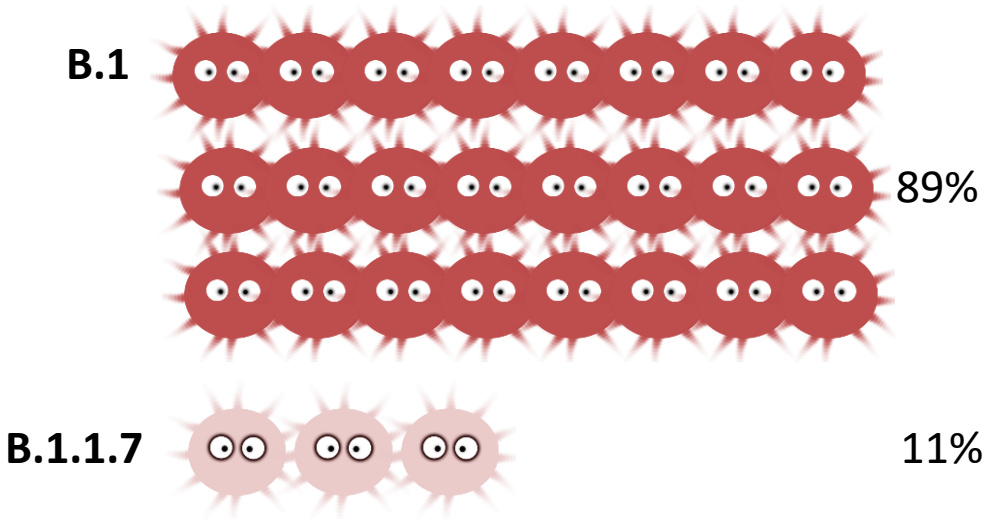


- Sampling time period 2:

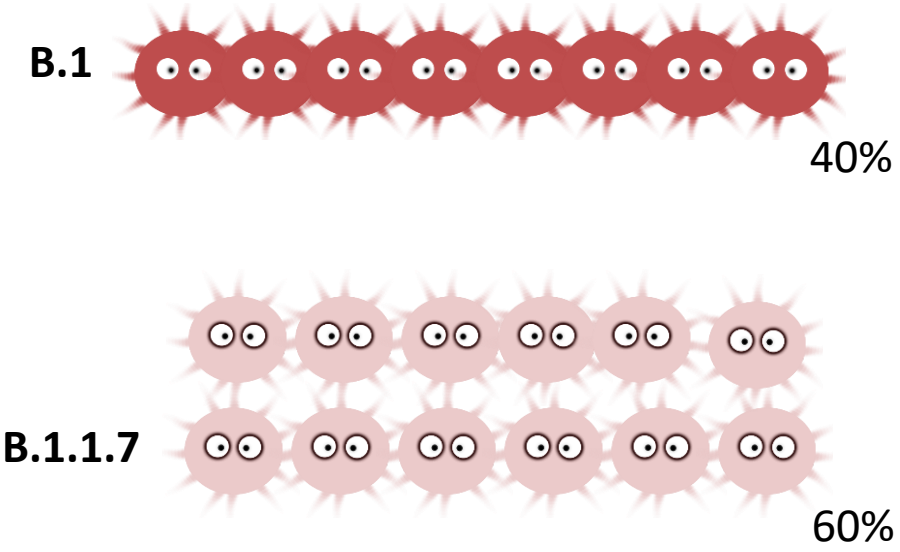


Why is the sampling time period important?

- Sampling time period 1:



- Sampling time period 2:



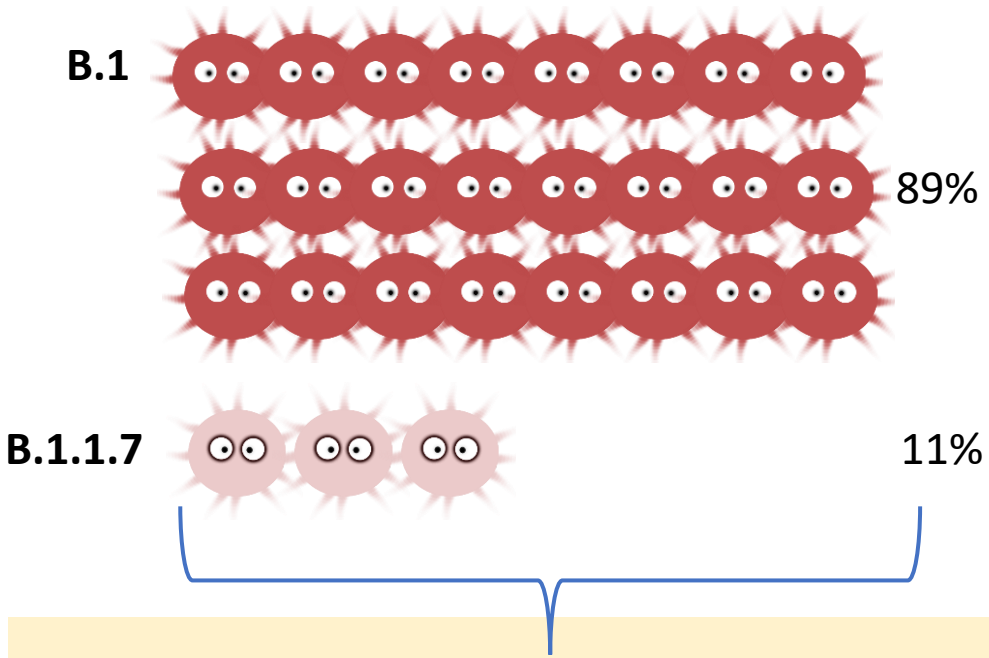
Total

B.1:
68% (32/47)

B.1.1.7:
32% (15/47)

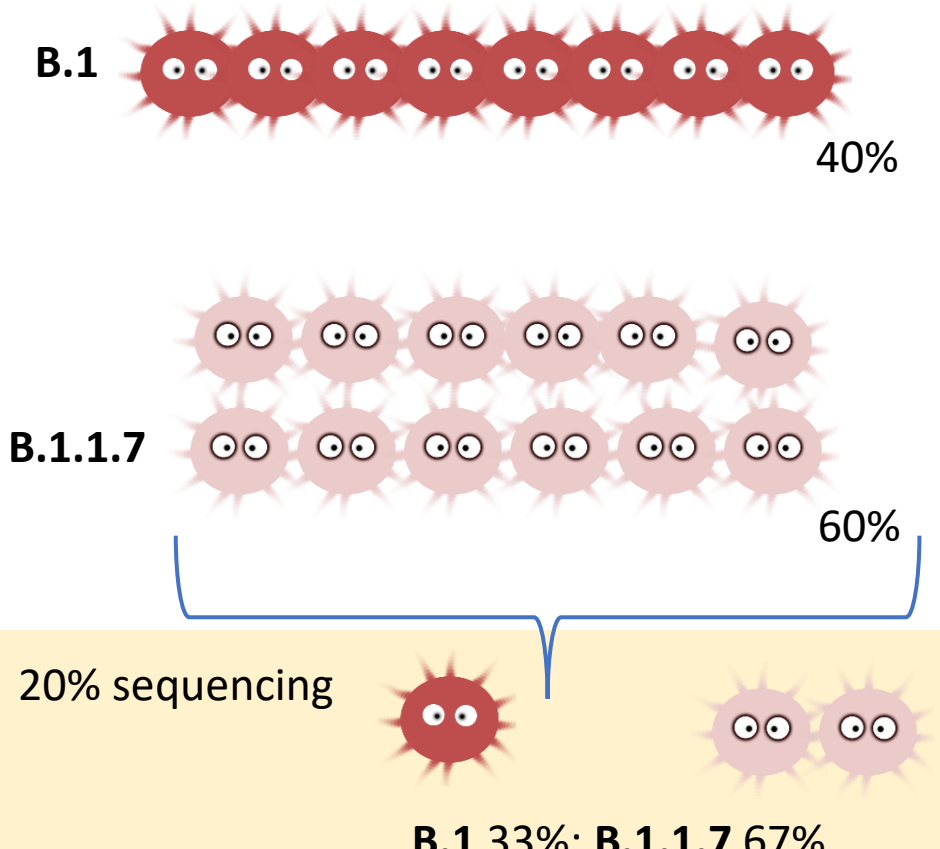
Why is the sampling time period important?

- Sampling time period 1:



B.1 89%; **B.1.1.7** 11%

- Sampling time period 2:



B.1 33%; **B.1.1.7** 67%

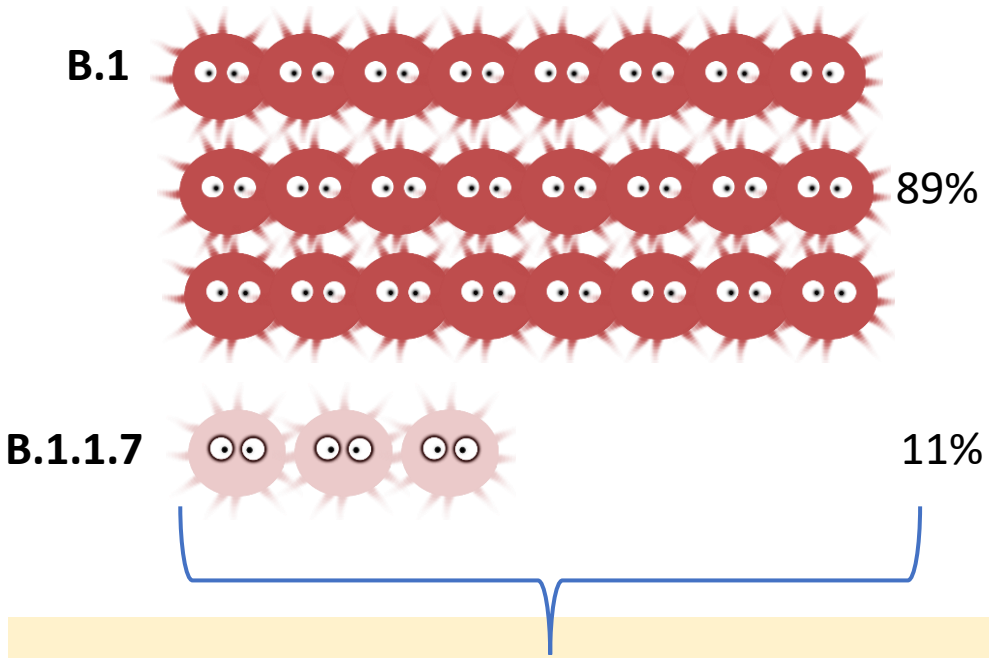
Total

B.1:
68% (32/47)

B.1.1.7:
32% (15/47)

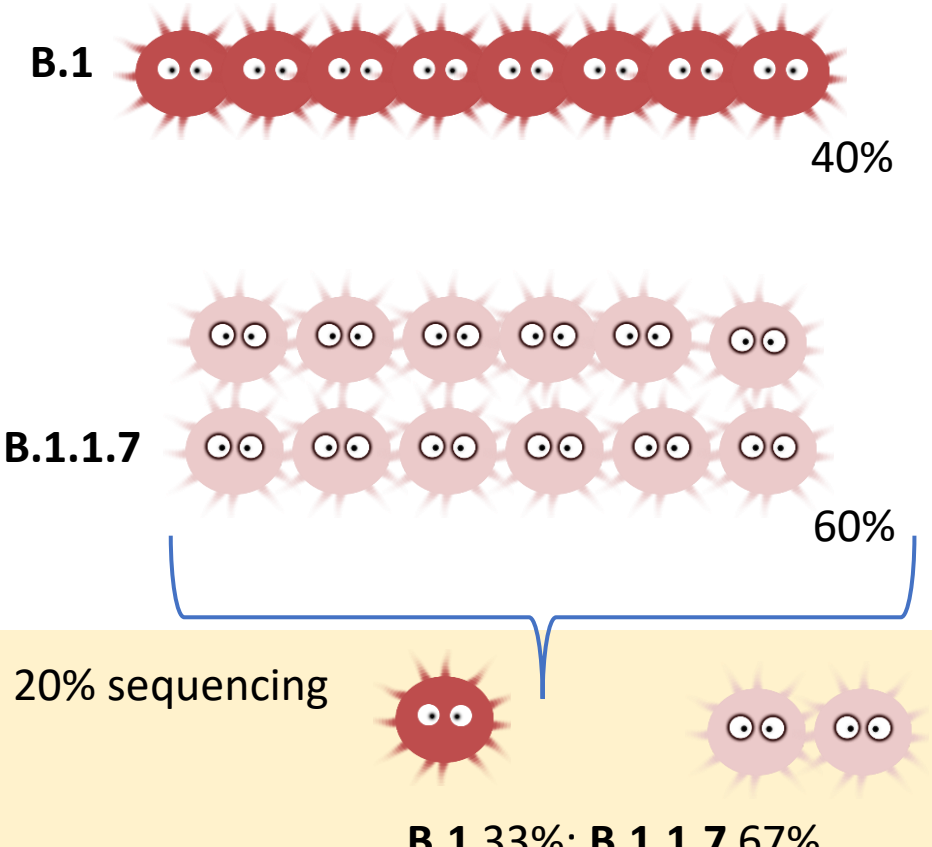
Why is the sampling time period important?

- Sampling time period 1:



B.1 89%; **B.1.1.7** 11%

- Sampling time period 2:



B.1 33%; **B.1.1.7** 67%

Total

B.1:
68% (32/47)

B.1.1.7:
32% (15/47)

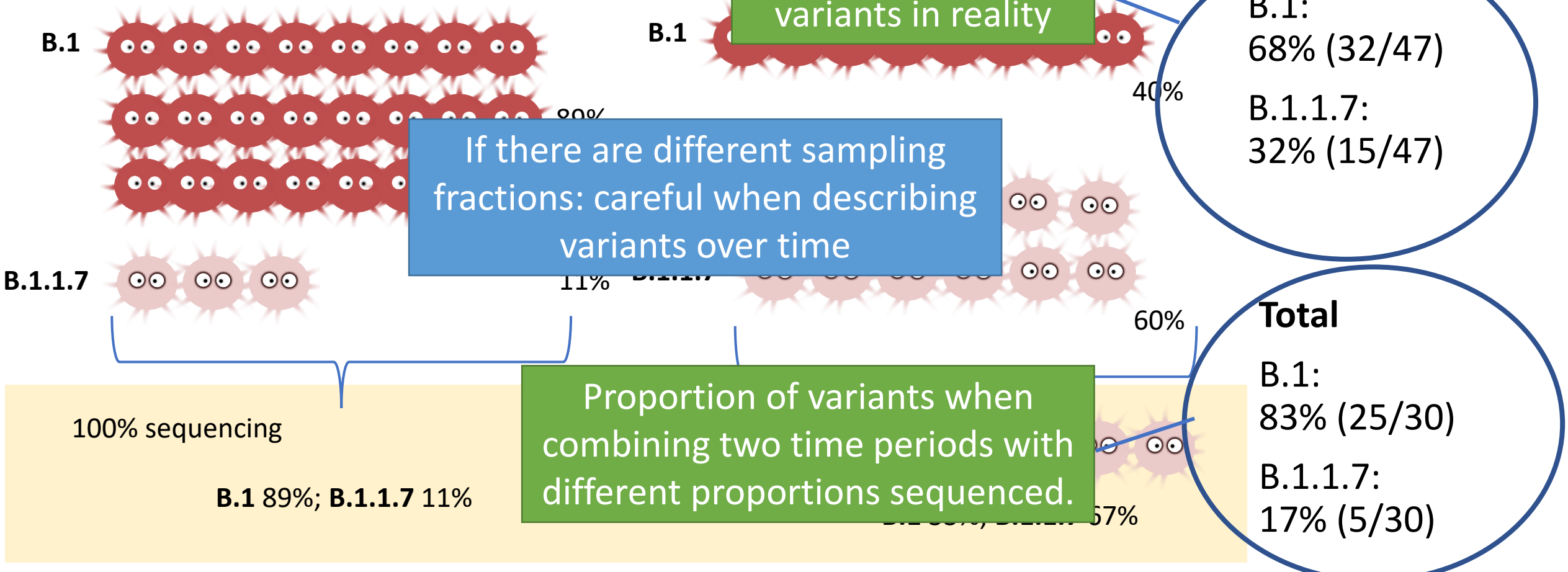
Total

B.1:
83% (25/30)

B.1.1.7:
17% (5/30)

Why is the sampling time period important?

- Sampling time period 1:
- Sa[illegible]2:



Proportions of variants in reality

Total

B.1:
68% (32/47)

B.1.1.7:
32% (15/47)

If there are different sampling fractions: careful when describing variants over time

Proportion of variants when combining two time periods with different proportions sequenced.

Total

B.1:
83% (25/30)

B.1.1.7:
17% (5/30)

Sampling time period and sampling fraction

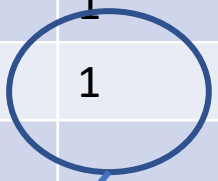
- Important to present results by sampling time period in a descriptive analysis, or use a weighting approach
- For VE studies, this matters less, unless the VE against a variant is different over time (e.g. waning immunity)
  - We recommend to weight analyses by the inverse of the sampling fraction
    - Viruses in sampling time periods with lower proportion sequenced have higher weight
      - Sampling period 1: 27 viruses, 27 sequenced → weight = 1
      - Sampling period 2: 20 viruses, 3 sequenced → weight = 20/3
    - Consult a statistician!
  - Remember to measure VE only during the period the variant was circulating (exclude controls/participants from earlier/later periods)

Linking lab and epi data (fictitious data)

Epi ID	Lab sample ID	PCR result	Vaccination status	Age	GISAID accession ID	Lineage	Time period
E-3174	2021-L-035	Pos	Unvacc	43	EPI_ISL_969715	B.1.1.7	1
E-3075	2021-L-036	Neg	Vacc	65			
F-4299	2021-L-037	Pos	Vacc	15	EPI_ISL_969815	B.1.1.7	1
E-3499	2021-B-105	Pos	Unvacc	24			1
E-3174	2021-A-021	Neg	Vacc	29			
E-3075	2021-A-022	Pos	Unvacc	51	EPI_ISL_969900	B.1.1.7	2

Linking lab and epi data (fictitious data)

Epi ID	Lab sample ID	PCR result	Vaccination status	Age	GISAID accession ID	Lineage	Time period
E-3174	2021-L-035	Pos	Unvacc	43	EPI_ISL_969715	B.1.1.7	1
E-3075	2021-L-036	Neg	Vacc	65			
F-4299	2021-L-037	Pos	Vacc	15	EPI_ISL_969815	B.1.1.7	1
E-3499	2021-B-105	Pos	Unvacc	24			1
E-3174	2021-A-021	Neg	Vacc	29			
E-3075	2021-A-022	Pos	Unvacc	51	EPI_ISL_969900	B.1.1.7	2



Every positive sample has a time period associated with it (even if not sequenced), to calculate the sampling fraction

Checking for bias: is sequenced sample representative?

- VE against all viruses should be the same as VE against sequenced viruses

Reasons for bias

- Samples difficult to sequence with high CT value (low viral load)
  - Viral load could vary by variant → specific variants easier/harder to sequence
    - Bias in description of variants, no bias for VE (lower precision)
  - Viral load may be lower among vaccinated
    - Bias in VE

Variant-specific VE

- Difficult to have high precision
  - Many variants circulating
  - Not all viruses sequenced
- Other methods (e.g., case-only approaches) may provide a signal that variants have lower VE
  - Signal to measure variant-specific VE
- Variant-specific VE crucial to help influenza and SARS-CoV-2 control measures

Any questions?