

eAppendix

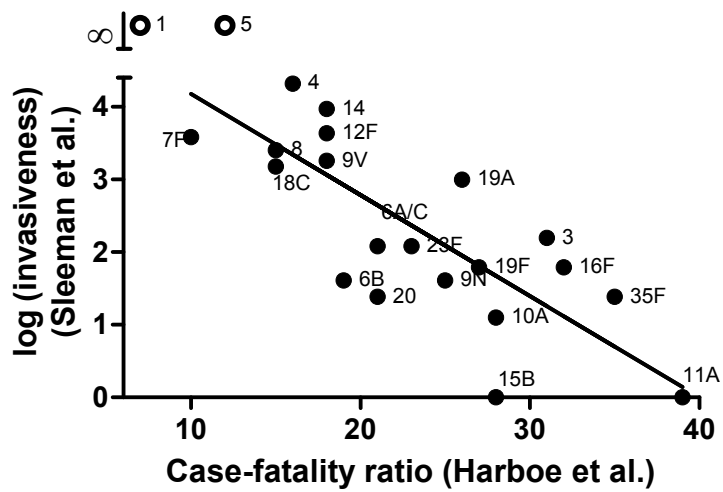
eTable 1. Estimated invasiveness based on linear regression

Serotype	Predicted invasiveness
1	98.8
2	35.8
3	3.5
4	28.3
5	49.3
6A/C	14.1
6B	18.6
7F	65.1
8	32.5
9A	55.9
9N	8.1
9V	21.4
10A	5.3
11A	1.2
12A	7.3
12F	21.4
13	14.7
14	21.4
15A	5.4
15B	5.3
15C	24.6
16F	3.1
17F	2.3
18C	32.5
19A	7.0
19F	6.1
20	14.1
22F	16.2
23A	5.3
23F	10.7
24F	10.7
31	0.6
33F	24.6
34	2.8
35B	14.5
35F	2.0
38	21.4

The regression used to derive these estimates was fit to “attack rate” data from Sleeman et al., where invasiveness is defined as the number of disease cases/100,000 acquisitions

eTable 2. Expected number of cases of invasive pneumococcal disease per serotype calculated from model 1 (using carriage prevalence and invasiveness as predictors) or model 2 (using polysaccharide structure and invasiveness as predictors). Fit to disease data from England and Wales.

Serotype	Expected (Model 1)	Expected (Model 2)	Observed
1	15	378	45
2	2	16	0
3	7	38	15
4	4	19	18
5	4	22	0
6A/C	38	36	27
6B	158	48	77
7F	9	6	22
8	22	74	12
9A	4	44	0
9N	6	4	2
9V	32	12	26
10A	3	1	1
11A	1	1	0
12A	0	1	0
12F	6	4	1
13	2	9	0
14	60	36	156
15A	2	3	0
15B	5	3	2
15C	27	19	3
16F	2	1	0
17F	2	0	1
18C	40	30	49
19A	10	26	26
19F	37	22	68
20	6	3	3
22F	11	6	6
23A	3	7	3
23F	56	15	45
24F	3	-	1
31	0	0	1
33F	17	10	5
34	0	3	1
35B	1	10	0
35F	1	2	2
38	11	-	1



eFigure. Correlation between case-fatality ratio and log(invasiveness).  $r^2=0.58$ . Serotypes 1 and 5 were not used in the calculation of the regression.