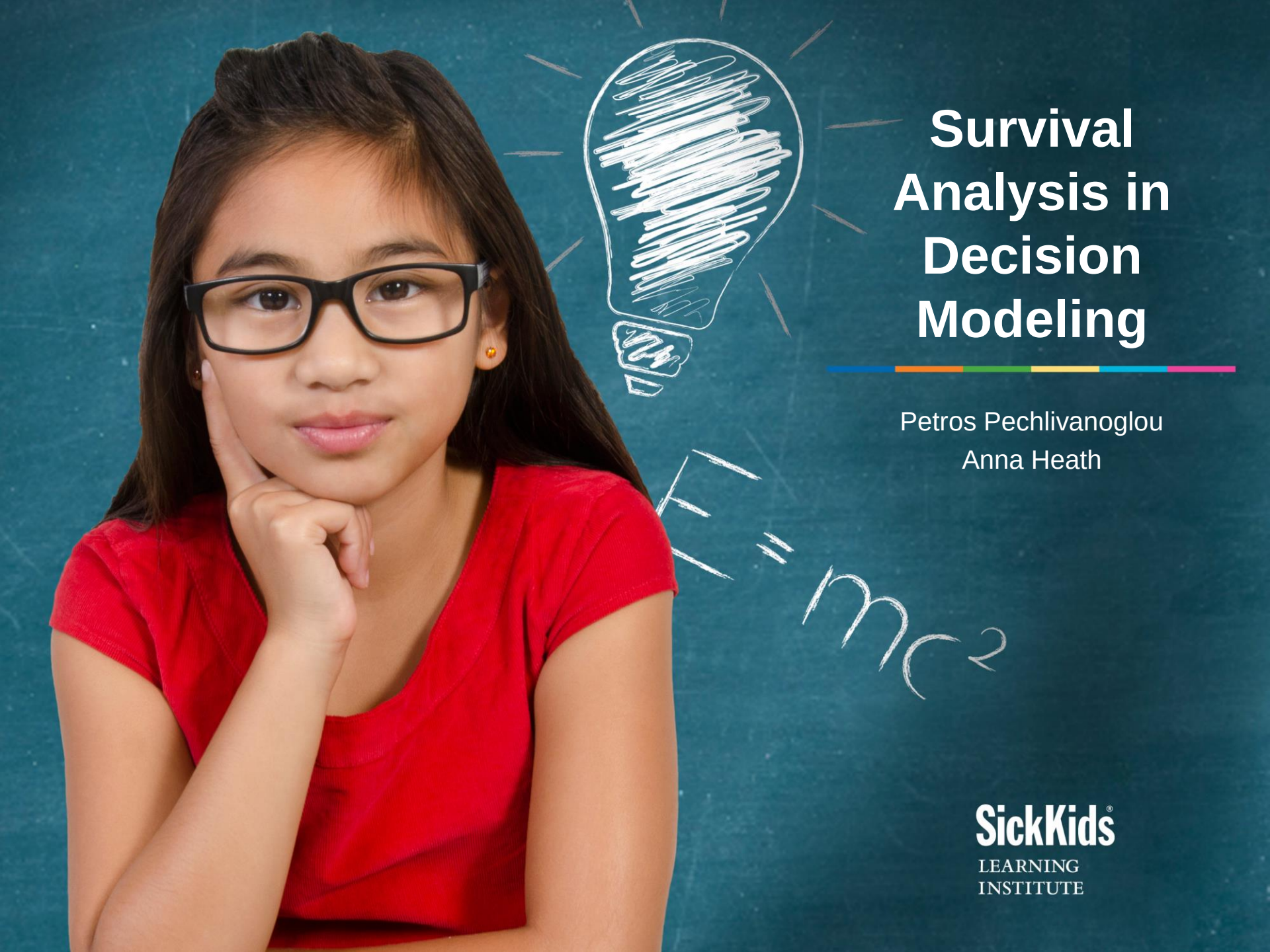


Survival Analysis in Decision Modeling

Petros Pechlivanoglou
Anna Heath

SickKids®
LEARNING INSTITUTE

Anna Heath

RECONSTRUCTION OF IPD

Background

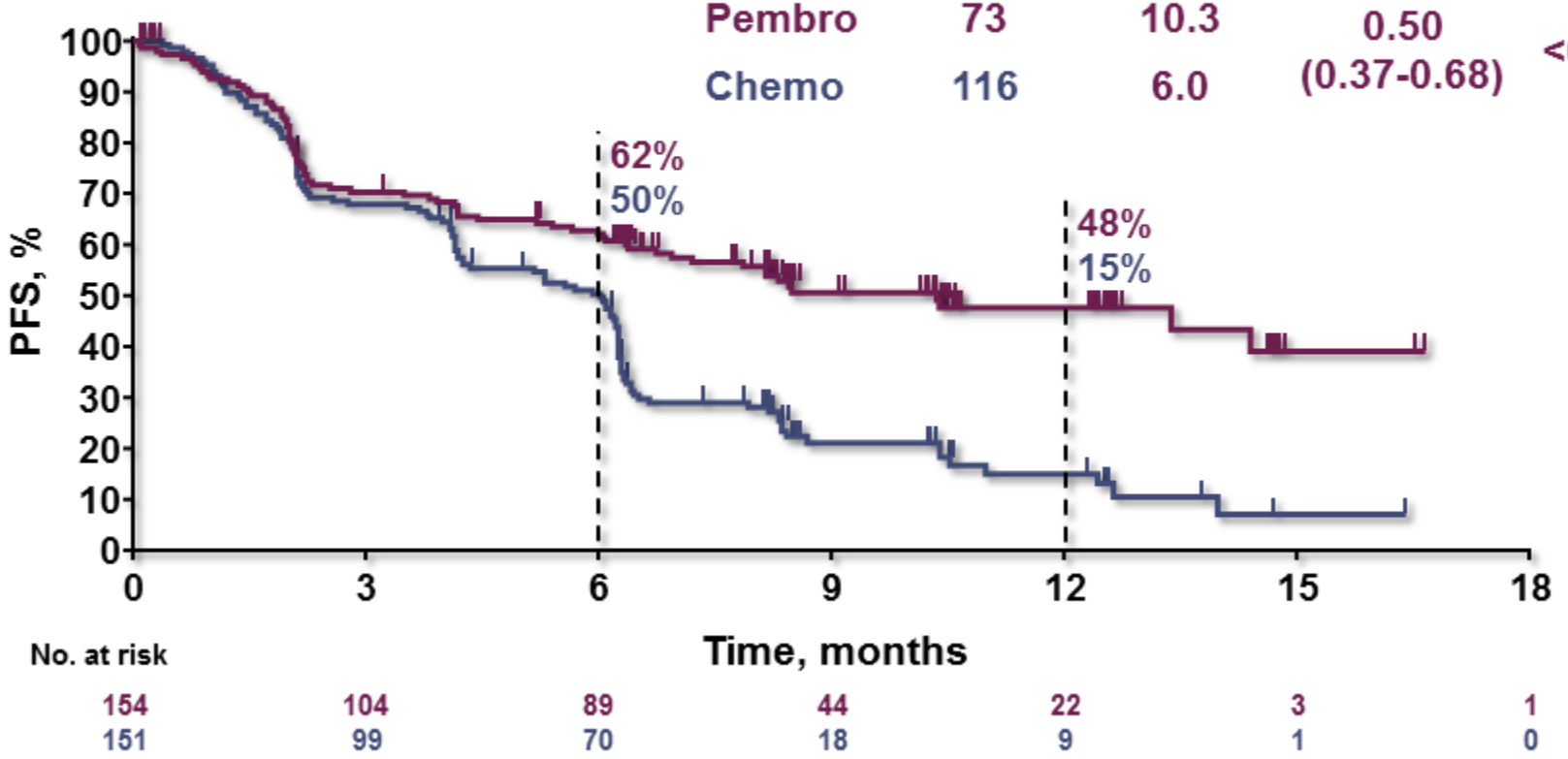
- Decision modelling is often (at least in part) based on previously published data.
- In survival analysis, published data for outcomes is frequently insufficient to perform cost-effectiveness analysis
- Extrapolation of survival curves past the trial follow-up time requires individual patient data (IPD)
- We aim to use published graphics to reconstruct IPD so published data can be used in a CEA.

Published Kaplan-Meier Data

M Reck. ESMO 2016.

Progression-Free Survival

	Events, n	Median, mo	HR (95% CI)	*P*
Pembro	73	10.3	0.50 (0.37-0.68)	<0.001
Chemo	116	6.0		



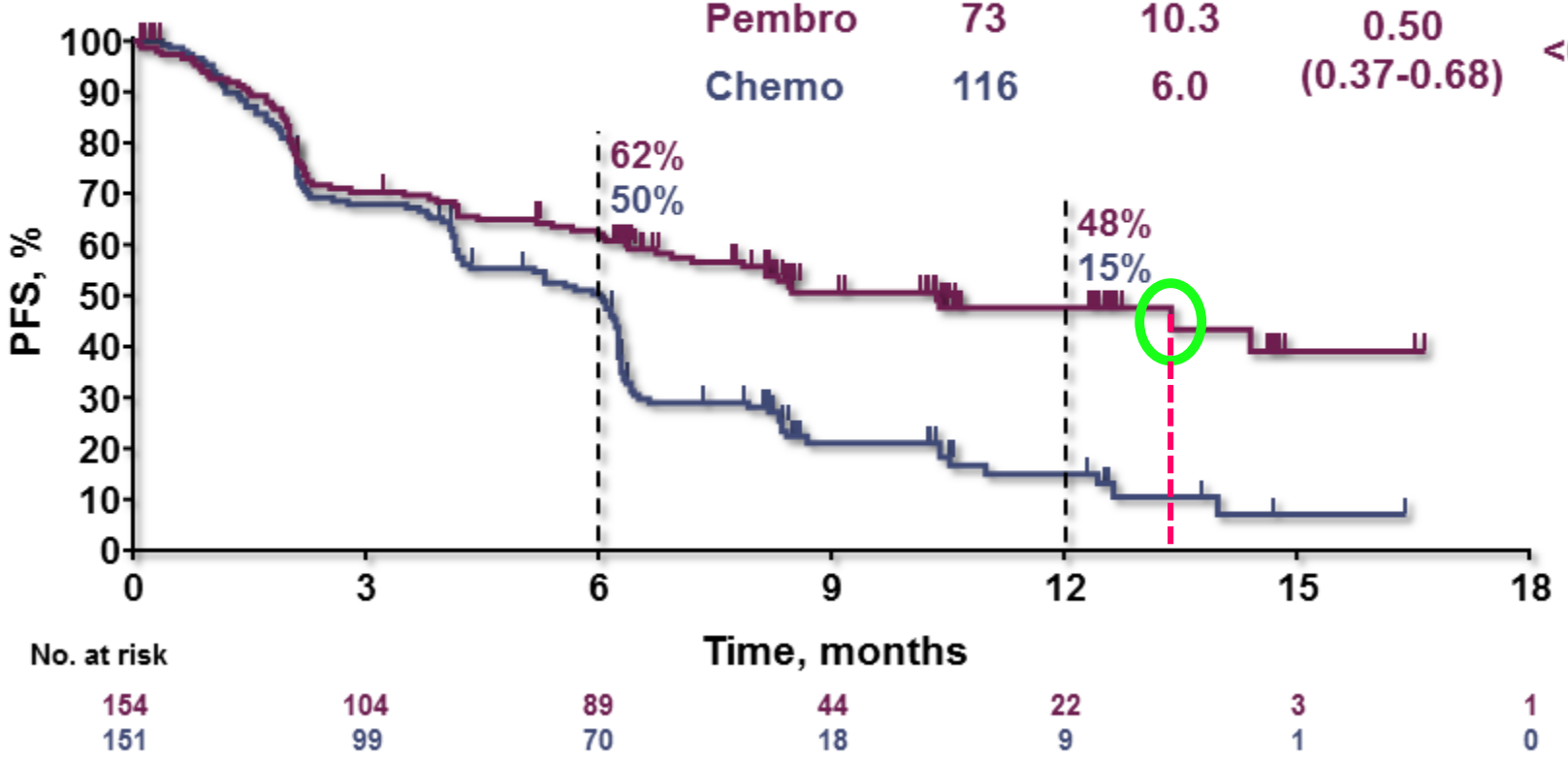
Assessed per RECIST v1.1 by blinded, independent central review.
Data cut-off: May 9, 2016.

Published Kaplan-Meier Data

M Reck. ESMO 2016.

Progression-Free Survival

	Events, n	Median, mo	HR (95% CI)	*P*
Pembro	73	10.3	0.50 (0.37-0.68)	<0.001
Chemo	116	6.0		



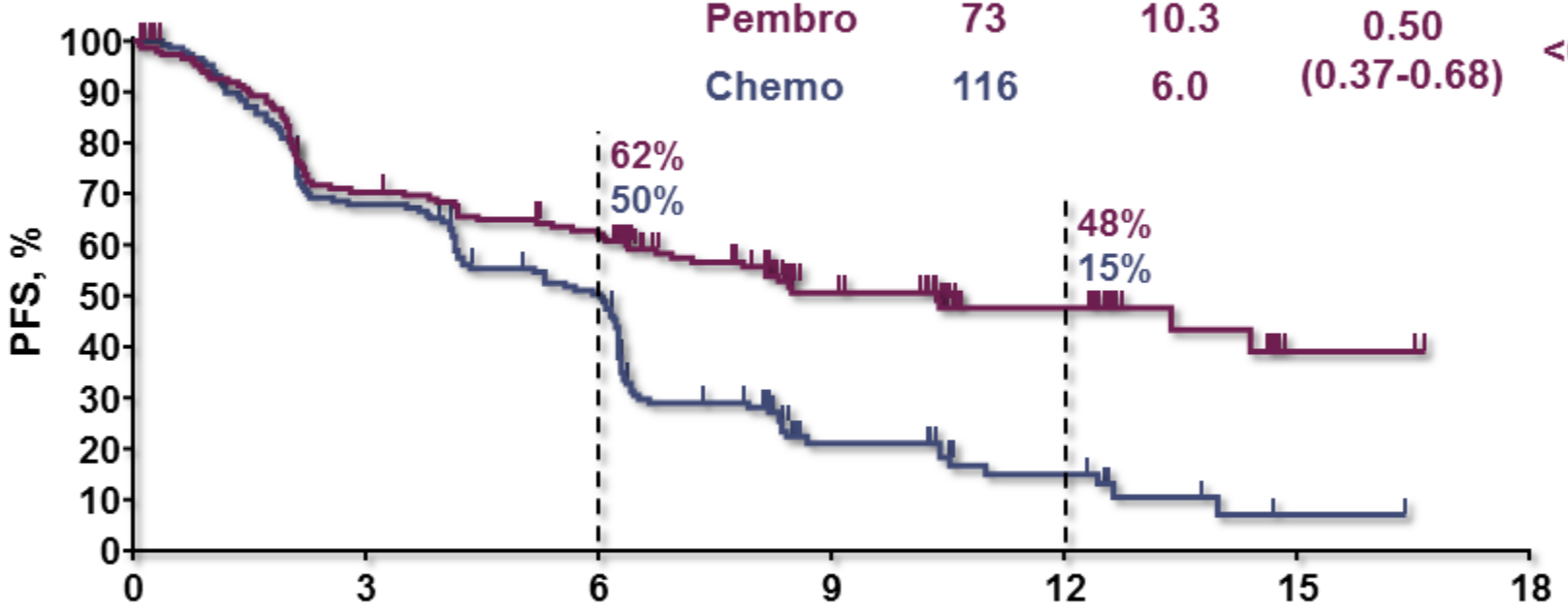
Assessed per RECIST v1.1 by blinded, independent central review.
Data cut-off: May 9, 2016.

Published Kaplan-Meier Data

M Reck. ESMO 2016.

Progression-Free Survival

	Events, n	Median, mo	HR (95% CI)	*P*
Pembro	73	10.3	0.50 (0.37-0.68)	<0.001
Chemo	116	6.0		



No. at risk

154	104	89	44	22	3	1
151	99	70	18	9	1	0

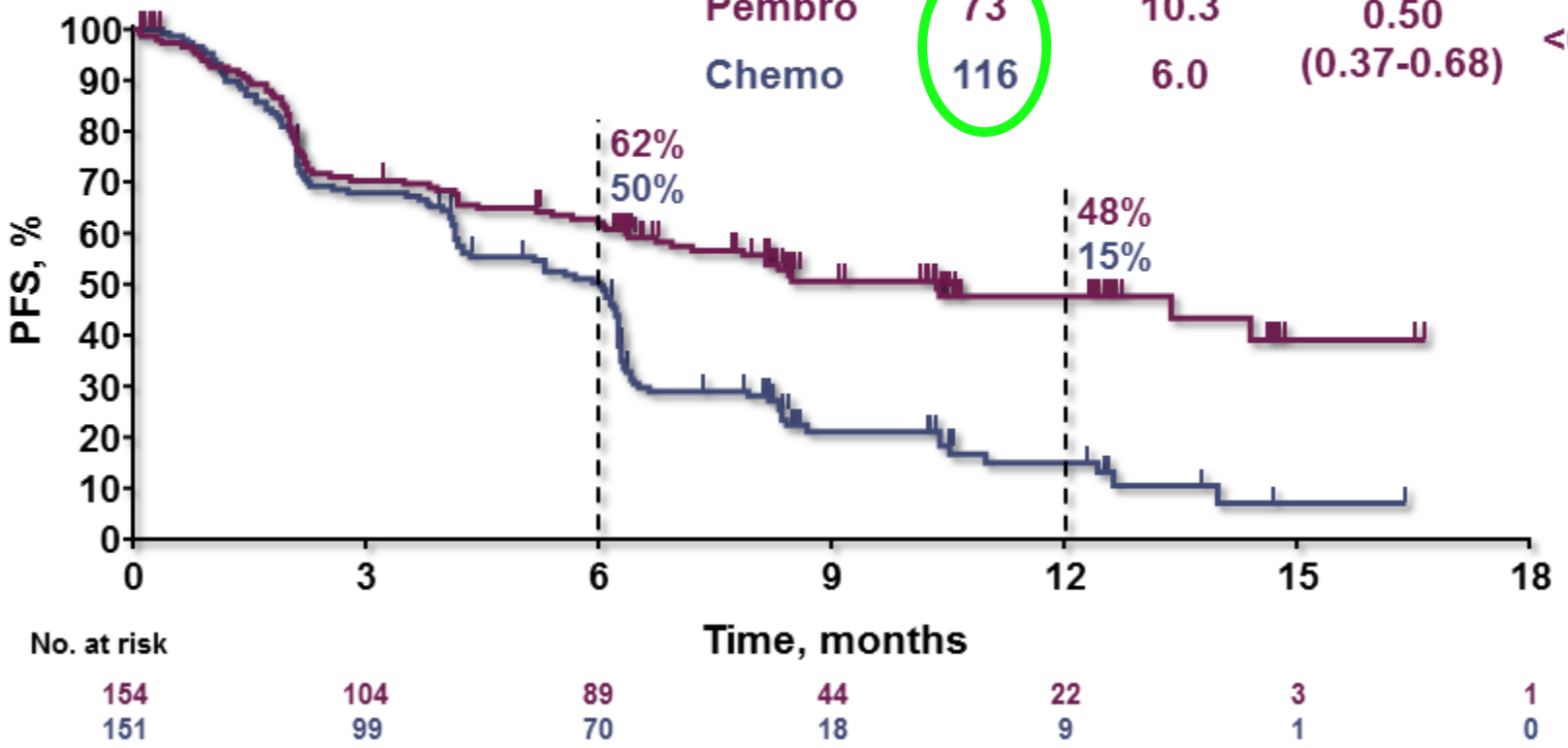
Assessed per RECIST v1.1 by blinded, independent central review.
Data cut-off: May 9, 2016.

Published Kaplan-Meier Data

M Reck. ESMO 2016.

Progression-Free Survival

	Events, n	Median, mo	HR (95% CI)	*P*
Pembro	73	10.3	0.50 (0.37-0.68)	<0.001
Chemo	116	6.0		



Assessed per RECIST v1.1 by blinded, independent central review.
Data cut-off: May 9, 2016.

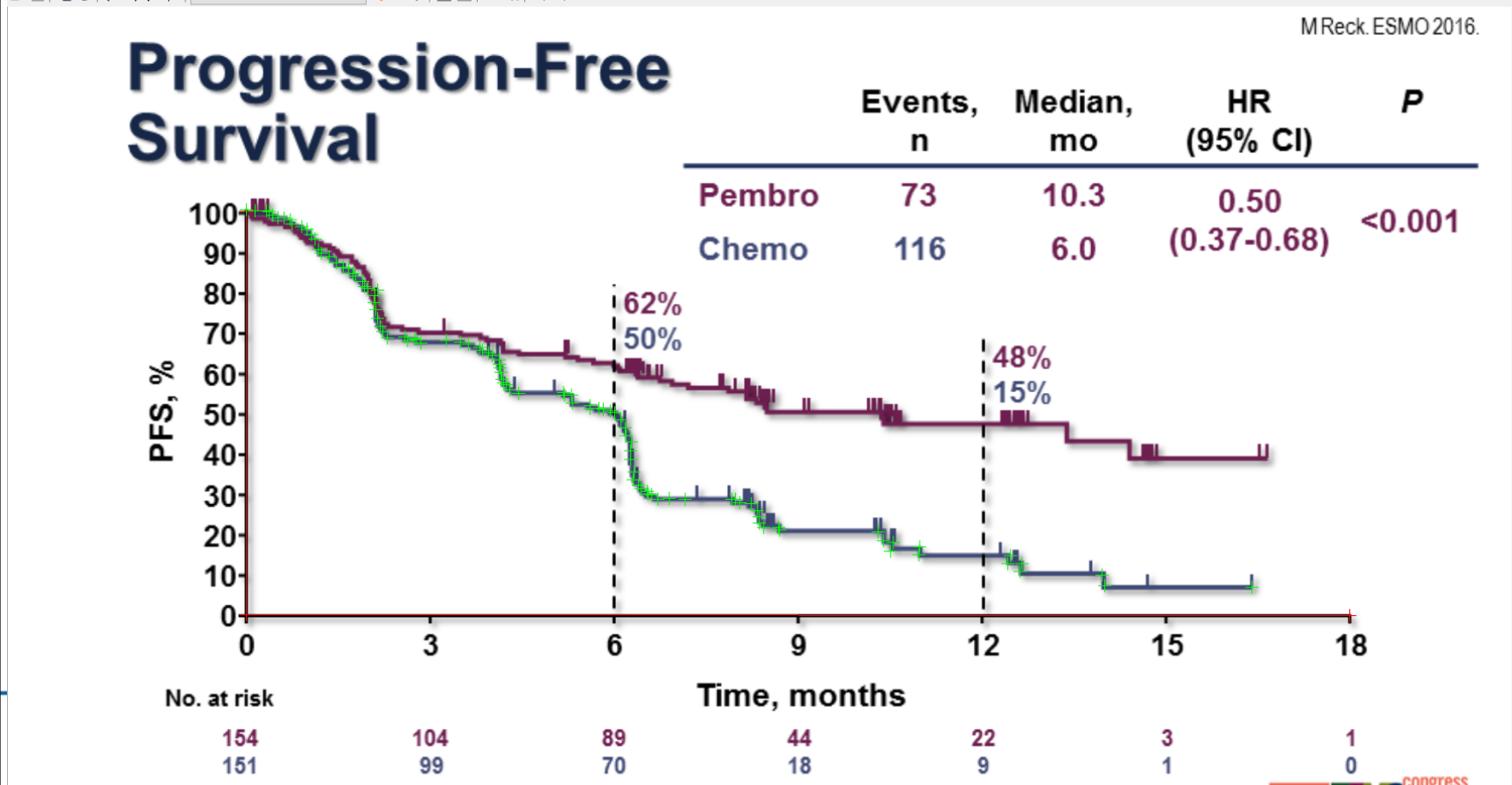
“Digitizing” the curves

https://www.digitizeit.de/

DigitizeIt - PFS.dig

File Edit View Axis Dataset Data Auto Graph Help

Dataset



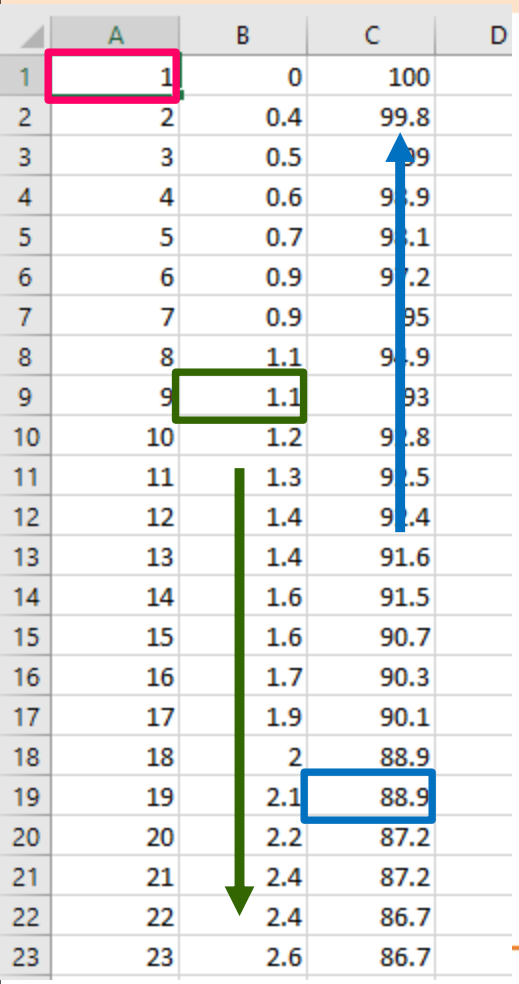

```
x-range:
     0.00;       18.0
scale: linear

y-range:
      100;       0.00
scale: linear

points: 103
sorting x ascending
     0.00         100
    0.144         100
    0.331         100
    0.374         100
    0.431        99.1
    0.518        98.5
    0.618        98.5
    0.719        98.0
    0.748        96.9
    0.906        96.3
    0.992        95.6
     1.06        94.3
     1.12        93.2
     1.18        90.6
     1.22        89.7
     1.35        88.9
     1.47        87.3
     1.58        86.2
     1.67        86.0
     1.74        84.7
     1.78        84.0
     1.90        82.7
     1.93        81.4
     1.98        81.0
     2.08        77.5
     2.10        79.2
     2.11        75.7
     2.14        80.6
     2.14        73.5
     2.19        71.6
     2.23        70.3
     2.30        68.7
     2.59        69.0
     2.63        68.1
     2.75        68.5
     2.77        68.3
     2.85        67.4
     3.26        68.1
     3.49        67.7
     3.62        67.2
     3.81        66.3
     3.85        65.7
     3.95        65.0
     4.01        64.4
     4.13        63.3
     4.13        62.2
     4.14        61.3
     4.15        60.2
     4.17        58.5
     4.20        57.6
     4.24        57.2
     4.30        55.9
     4.44        55.0
     5.18        55.2
     5.19        54.5
     5.31        54.6
     5.32        52.6
```

“Digitizing” the curves

	A	B	C	D
1	1	0	100	
2	2	0.4	99.8	
3	3	0.5	99	
4	4	0.6	98.9	
5	5	0.7	98.1	
6	6	0.9	97.2	
7	7	0.9	95	
8	8	1.1	94.9	
9	9	1.1	93	
10	10	1.2	92.8	
11	11	1.3	92.5	
12	12	1.4	92.4	
13	13	1.4	91.6	
14	14	1.6	91.5	
15	15	1.6	90.7	
16	16	1.7	90.3	
17	17	1.9	90.1	
18	18	2	88.9	
19	19	2.1	88.9	
20	20	2.2	87.2	
21	21	2.4	87.2	
22	22	2.4	86.7	
23	23	2.6	86.7	

1. The index number of click
2. The value on the x-axis – the time.
3. The value on the y-axis – the proportion of patients “alive” at that specific time point.

“Digitizing” the curves



	A	B	C	D
1	1	0	100	
2	2	0.4	99.8	
3	3	0.5	99	
4	4	0.6	9[illegible].9	
5	5	0.7	9[illegible].1	
6	6	0.9	9[illegible].2	
7	7	0.9	95	
8	8	1.1	9[illegible].9	
9	9	1.1	93	
10	10	1.2	9[illegible].8	
11	11	1.3	9[illegible].5	
12	12	1.4	9[illegible].4	
13	13	1.4	91.6	
14	14	1.6	91.5	
15	15	1.6	90.7	
16	16	1.7	90.3	
17	17	1.9	90.1	
18	18	2	88.9	
19	19	2.1	88.9	
20	20	2.2	87.2	
21	21	2.4	87.2	
22	22	2.4	86.7	
23	23	2.6	86.7	

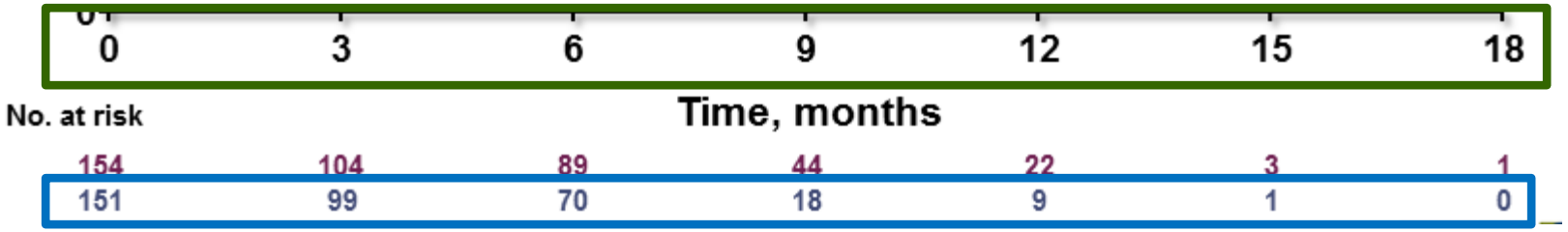
1. The index number of click
2. The value on the x-axis – the time.
3. The value on the y-axis – the proportion of patients “alive” at that specific time point.

- The times must be increasing
- The survival proportions must be decreasing.
- This needs to be checked manually because digitization can be imprecise

Tabulating the number “at risk”

	A	B	C	D	E
1	1	0	1	37	151
2	2	3	38	62	99
3	3	6	63	90	70
4	4	9	91	96	18
5	5	12	97	102	9
6	6	15	103	103	1
7	7	18	103	103	0

1. The index number for the interval
2. The time at the beginning of the interval
3. Creates a link between the digitized curves and the “At Risk” table – this is index of the first click that is within the interval
4. This is the index of the last click that is within the interval
5. The number of patients at risk at the beginning of the interval



Creating the “pseudo” IPD

- Guyot et al. developed an algorithm to reconstruct the IPD from these two files.
- It is included in R in the package survHE.

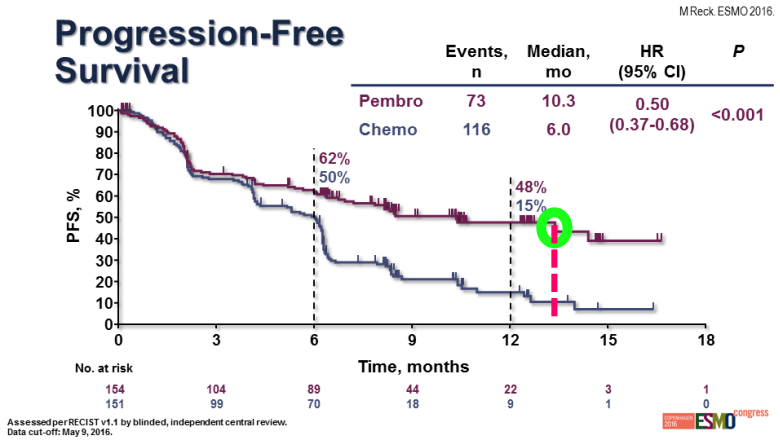
```
> library(survHE)
> digitise("DIGITIZED.txt","ATRISK.txt",
      km_output="KMdata.txt", ipd_output="IPDdata.txt")
```

- This creates two datasets:

	A	B	C	D
1	time	n.risk	n.event	n.censored
2	0	672	0	3
3	72	669	1	3
4	140	665	1	1
5	177	663	1	1
6	210	661	1	4
7	275	656	1	1
8	294	654	1	1
9	311	652	1	0

	A	B	C
1	time	event	arm
2	72	1	1
3	140	1	1
4	177	1	1
5	210	1	1
6	275	1	1
7	294	1	1
8	311	1	1
9	323	1	1

Creating the "pseudo" IPD



	A	B	C	D
1	time	n.risk	n.event	n.censored
2	0	672	0	3
3	72	669	1	3
4	140	665	1	1
5	177	663	1	1
6	210	661	1	4
7	275	656	1	1
8	294	654	1	1
9	311	652	1	0

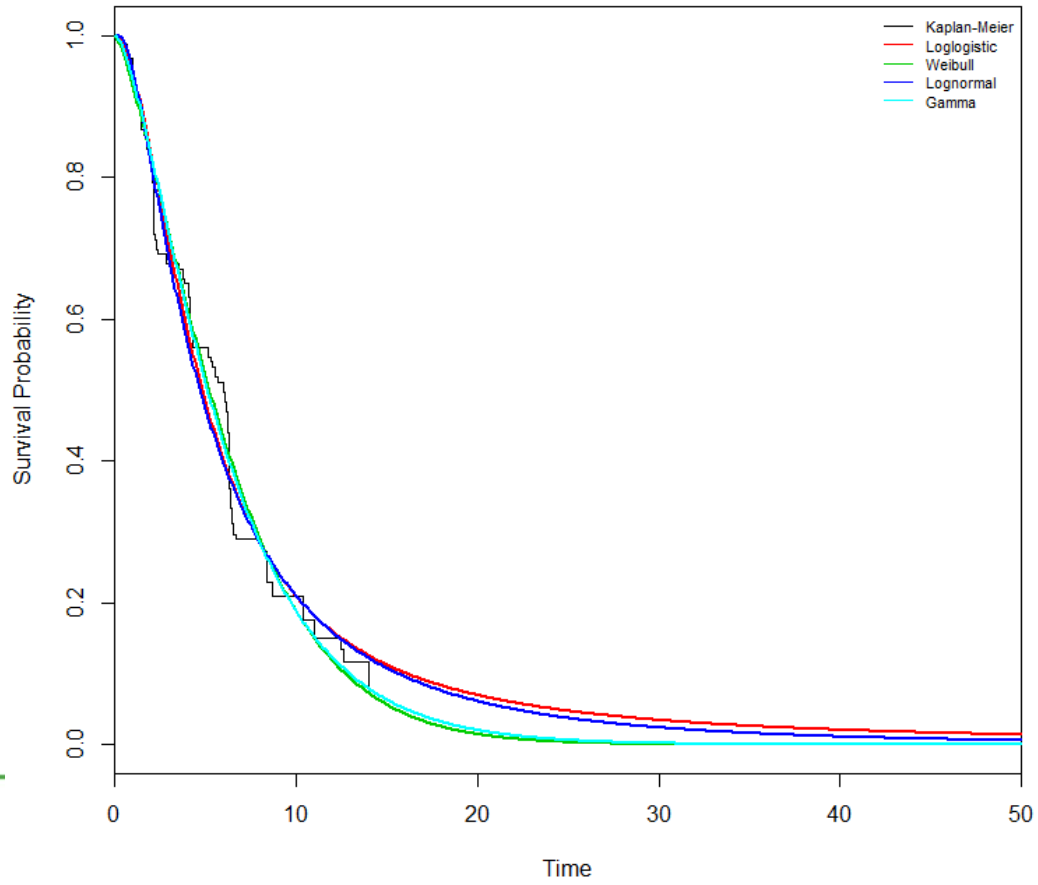
- Drops correspond to events.
- The digitized data highlights these drops.
- The *size* of the drop relates to $\frac{d_t}{n_t}$
  - n_t: patients at risk just before time t
  - d_t: patients with event at time t
- The algorithm randomly spaces censored values and then iteratively matches the number at risk.
- We are solving the KM equation:

$$S(t) = S(t - \delta_t)\frac{n_t - d_t}{n_t}$$

- δ_t: The time since last event

Model Fitting using pseudo IPD

- Parametric survival models can be fitted as normal to the reconstructed IPD.
- The best fitting model can be selected using standard model fitting criterion.
- Extrapolation to the full model horizon can be achieved.
- Partitioned survival analysis can be used to value health states.



Limitations

- There are no links between the curves:
  - Relationships between linked outcomes (e.g. PFS and OS) are not preserved.
- Data cannot be used to inform a multi-state model
- Cannot perform any subgroup analysis, e.g. by age, tumour category unless these curves are available and have been digitized