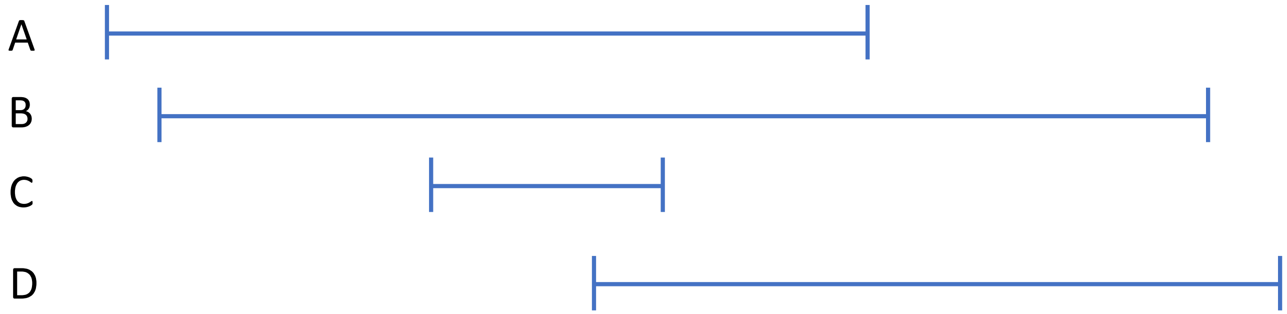
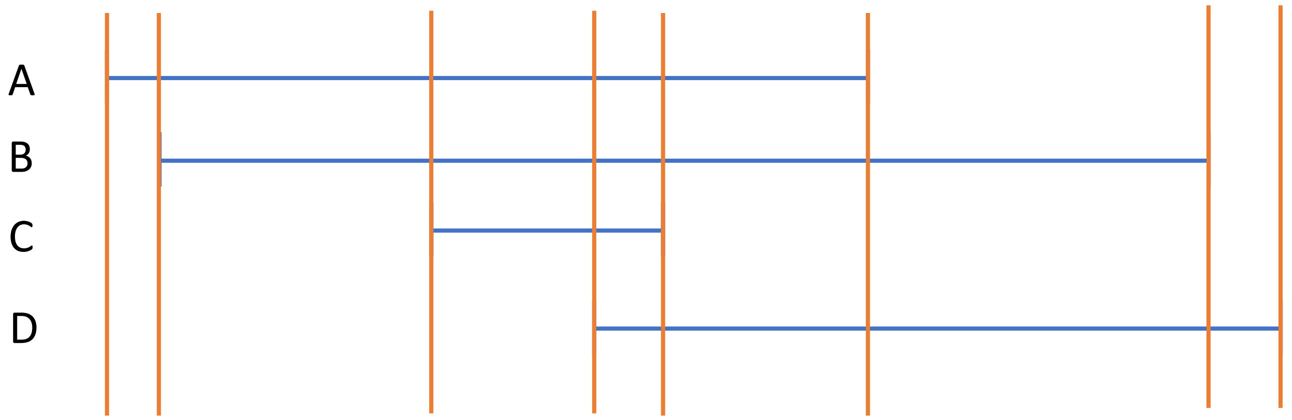
Case 1: multiple overlaps, no collapse window

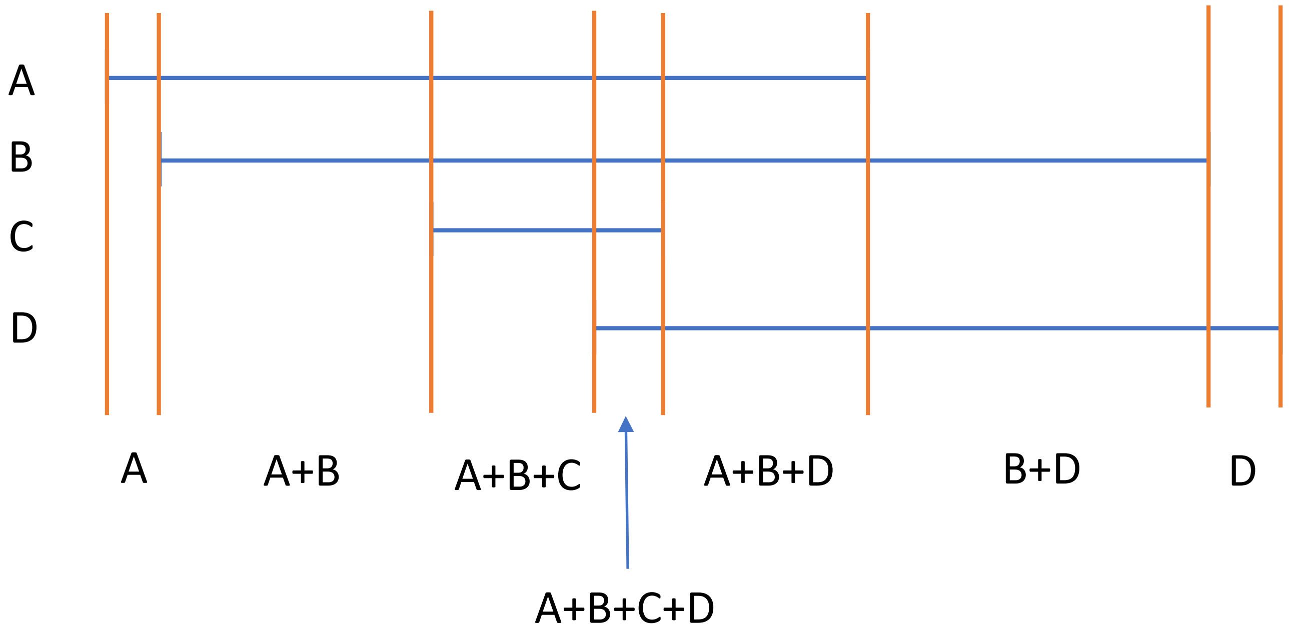


Case 1: multiple overlaps, no collapse window



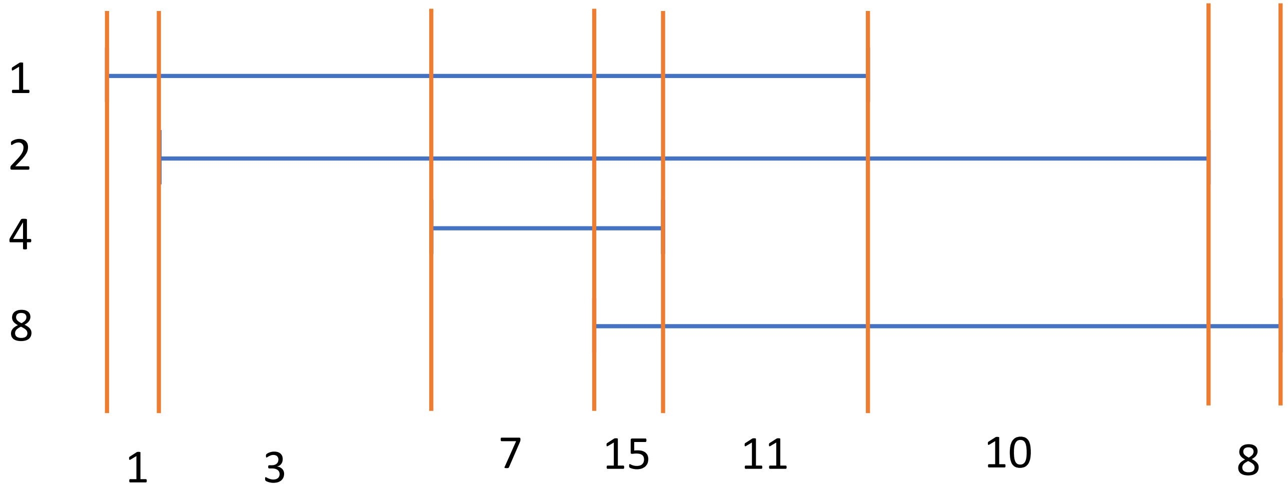
The red lines indicate the distinct dates (start or end) that appear in the patient’s event cohorts that are used to split up each event cohort.

Case 1: multiple overlaps, no collapse window



Using the start/ends that match across event cohorts, we can identify overlaps.

Case 1: multiple overlaps, no collapse window

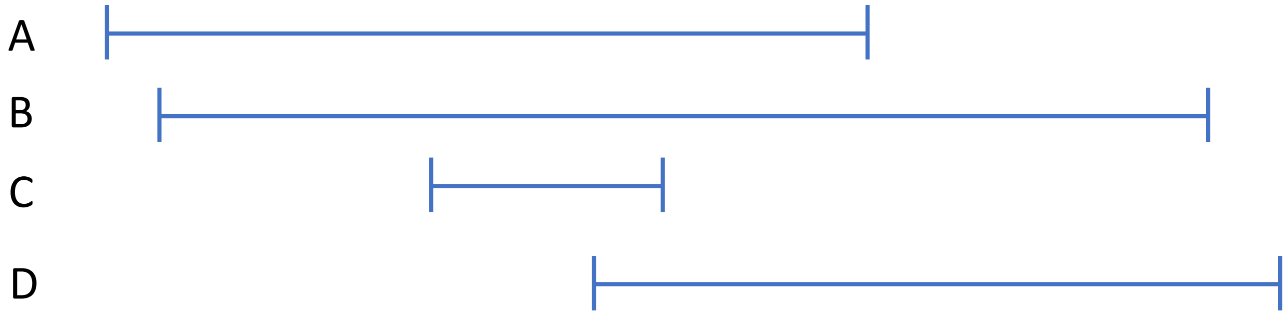


In SQL, we use add bit-wise distinct binary numbers together to find the combinations:

SUM(EventCohortBit) GROUP BY START_DATE, END_DATE

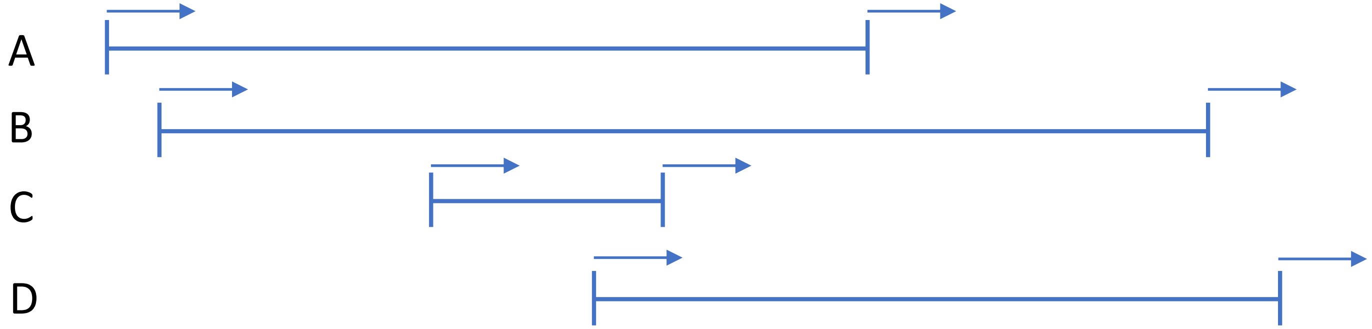
Case 2: multiple overlaps, 10 day collapse window

→ = 10 days



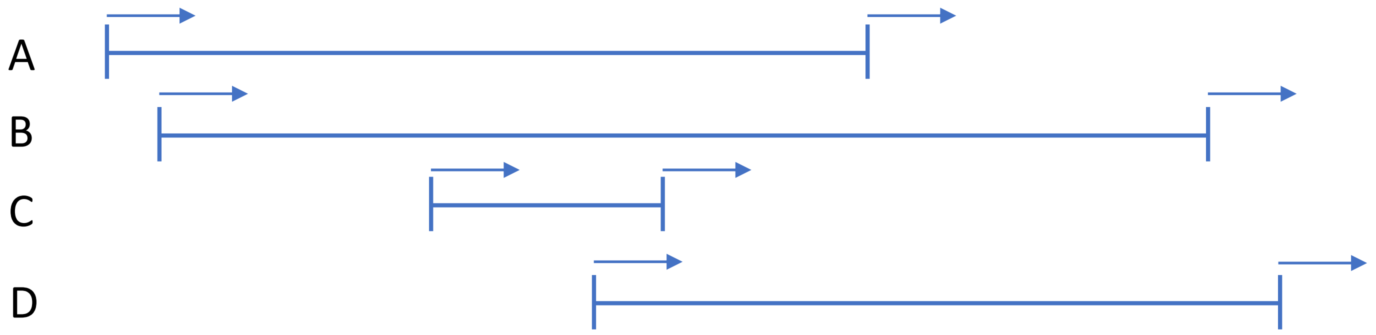
Case 2: multiple overlaps, 10 day collapse window

→ = 10 days



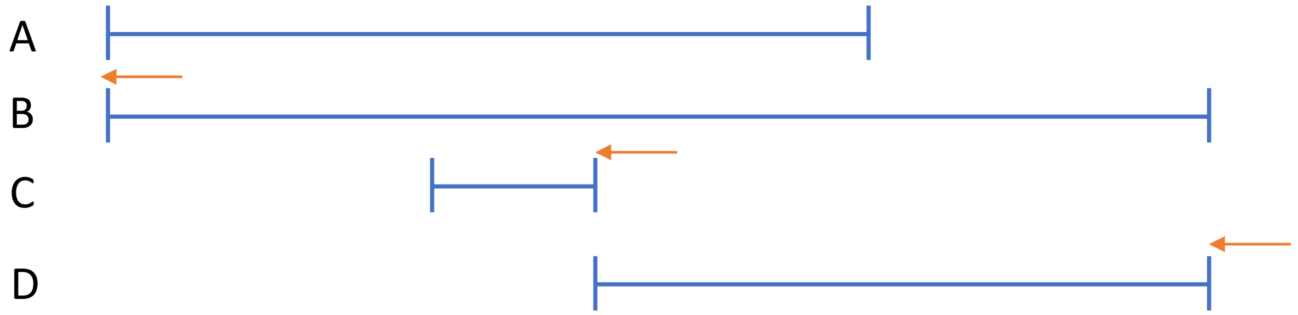
Case 2: multiple overlaps, 10 day collapse window

→ = 10 days

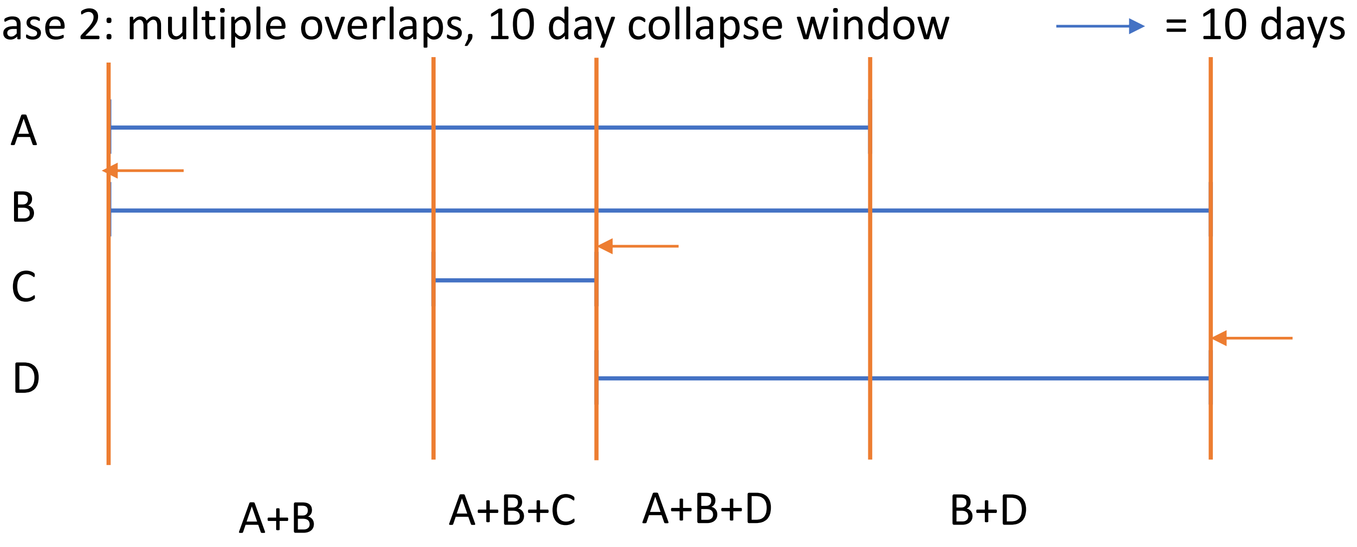


Each start/end looks forward the length of collapse window, and the date is re-assigned to the lowest date found. In SQL, this may be actually a ‘look backwards’, but the logic is the same.

This results in:



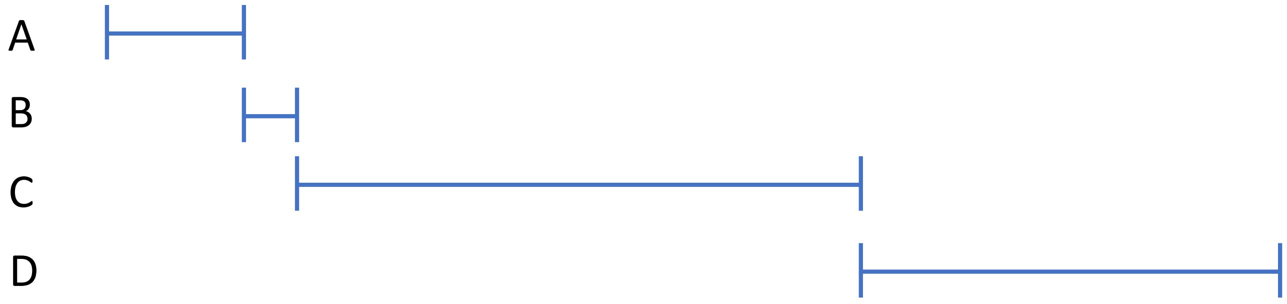
Case 2: multiple overlaps, 10 day collapse window



With the start/ends ‘paired up’ based on the collapse window, we will have less split intervals, hence reducing some ‘noise’. The resulting path using a collapse window is shorter, and could be seen as ‘more reasonable’ as to the progression between events. In addition, after 5 levels of ‘path-depth’, the data becomes harder to interpret. Also, with more distinct paths, you have a higher chance for people to be placed into their own distinct groups, making it harder to see commonality. Ultimately, the researcher needs to decide how to adjust this.

Without Collapse window:
A -> A+B -> A+B+C -> A+B+C+D -> A+B+D -> B+D -> D
With Collapse window:
A+B -> A+B+C -> A+B+D -> B+D

Question 1: should an event that ends when another event starts be considered an overlap?



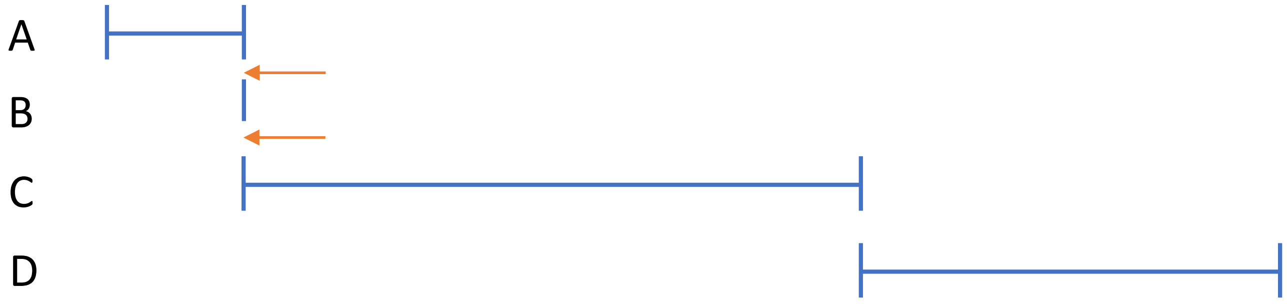
Is the correct result:

A -> A+B -> B -> B+C -> C -> C+D -> D

Or:

A->B->C->D

Question 2: What if we apply a collapse window?



Is the correct result:

A -> A+B+C -> C -> C+D -> D

Or:

A-> A+B+C->C->D

Or:

A->B->C->D

Other Questions

- Is collapse window the right approach or do we want to be able to ignore intervals that are lower than a threshold? Or do we need both?
- Does the algorithm described have problems with events where start=end? Does that confuse the logic?
- Is there a broader ‘researcher perspective’ that we did not account for that might make the described approach counter-intuitive or against ‘best practices’ of describing pathway progression?