

S4. USE OF DATE INFORMATION IN IDENTIFYING 'PRIOR' ACTIVITY NODES

Offenses and incidents

Offenses/incidents recorded as having occurred between a start and end time-range were deemed to be 'prior' to the reference offense when the 'end date' (the latest point in time that the prior offense/incident could have occurred) pre-dated the reference offense 'start date' (the earliest date on which the reference offense could have occurred). This approach ensured that only offenses/incidents that definitely preceded the reference offense counted as prior activity nodes.

Address, family address and education records

The address, family address and education records included dates indicating the timeframe for which the information was current. Records with 'start dates' on or after the reference offense 'start date' were excluded as not preceding the reference offense. Issues with the way these date fields are used were managed as follows.

First, the start date often reflects the date that Police became aware of the address, school, etcetera rather than the true start of the individual's association with that node. To the extent of this error, it was accepted that the duration of a person's association with that node would be underestimated, or that some nodes would be missed as the Police did not find out about them until after the reference offense.

Second, some home, family home, work and 'unspecified' node records were not updated with end dates and so would still appear current after the association with the node has ended. To the extent of this error, the recency of the node would be overestimated. On inspection, the 'unspecified' node records tended to reflect one-off events such as arresting or noting the presence of the individual at a given location. Unspecified node records with no end date were therefore allocated end dates the same as the record's start date to reflect their nature as singular events rather than enduring associations.

Third, with education records the opposite error appeared: start and end dates were frequently the same date. To the extent of this error, the recency of the node would be underestimated. This was addressed by 1) end-dating education records with the start of any subsequent education record for the same individual, then 2) end-dating the most recent school record whichever was the latter of the record's end date or the date on which the offender turned 18, the typical school leaving age in New Zealand, and tertiary education records to the date on which the offender turned 22 (based on a 3 year bachelor degree, typical of NZ tertiary institutions). School and tertiary education records were distinguished using key words in the name of the institution. To the extent that the offender cohort are more likely than the general population to leave school or tertiary study at an earlier age, education node recency and duration are likely to be overestimated, though this error is believed to be less than the underestimation error from using the original end dates.

Lastly, outlier record start dates were detected in the offender address and family address datasets. Records with start dates that were blank (n=12/169), prior to 1950 (n= 6/8) or greater than the end date (n=10/5) were allocated the record end date and vice versa (i.e., the dates of records with start and end dates apparently the wrong way around were switched).