Appendix 1: Main rules

The main rules of the V*ae*TM grammar are described below:

1. **'TIME_TO_ONSET'** rule that incorporated the 'NUMBER', the 'TimeUnit' and 'TimeModifier' tags along with the 'Vaccination' or 'Vaccine' tags that denoted the event of vaccination.

2. **'RULE_OUT_DIAGNOSIS'** rule that incorporated tags to represent all variations of the key phrases 'rule out' or the key abbreviation 'ro'. It also included the two basic rules as well as the 'NEGATION' tag.

3. **'CAUSE_OF_DEATH'** rule that incorporated tags to represent 'cod' (stands for cause of death), 'autopsy' and variations of the key phrases 'cause of death' and 'manner of death'. It also included the two basic rules as well as the 'NEGATION' and 'POSSIBILITY' tags.

4. **'DIAGNOSIS'** rule that incorporated a tag to represent all the possible inflections of the 'dx' and 'diagnosis' triggers, the 'NEGATION' tag and the two basic rules.

5. **'SECOND_LEVEL_DIAGNOSIS'** rule that incorporated a tag to represent the corresponding triggers (see '2nd level diagnosis' triggers in Supplementary Table 1). It also included the two basic rules as well as the 'NEGATION' tag.

6. **'FAMILY_HISTORY'** rule that incorporated tags to represent all variations of the key phrase 'family history' and the key abbreviation 'fhx'. It also included the two basic rules as well as the 'NEGATION' tag.

7. **'MEDICAL_HISTORY'** rule that incorporated tags to represent all variations of the key phrase 'medical history' and the key abbreviations 'mhx' and 'pmhx'. It also included the two basic rules as well as the 'NEGATION' tag.

8. **'VACCINE'** rule that incorporated the 'Vaccine' tag and was executed only if the term tagged as 'Vaccine' had not been used by the 'TIME_TO_ONSET' rule.

9. **'DRUG'** rule that incorporated the 'Drug' tag.

10. **'SYMPTOM'** rule that incorporated the two basic rules as well as the 'NEGATION' tag. Given that the main rules were executed in the presented order, a 'SYMPTOM' feature was extracted only if any of the above main rules was not satisfied.

11. **'LOT'** rule that incorporated a tag to represent the trigger 'lot' and the 'LotNumber' tag.