

**Appendix 2 (as supplied by the authors): Data sets used, relevant time periods and dates**

---

Dataset	Number of patients	Study start	Index date	Study end	Duration of follow up
Development dataset	300,000	01/01/2014	01/01/2015	31/12/2015	1 year
Validation dataset 1 (synchronous)	150,000	01/01/2014	01/01/2015	31/12/2015	1 year
Validation dataset 2 (asynchronous)	150,000	01/01/2010	01/01/2011	31/12/2011	1 year
		01/01/2010	01/01/2011	31/12/2015	5 years

Note that validation dataset 2 is used to calculate outcomes for both 1 year and 5 years follow up.