Establishing a Condition Era Persistence Window for Active Surveillance

Patrick Ryan
Last revised: 12 January 2010

Background

As part of its common data model (http://omop.fnih.org/CDMandTerminologies), OMOP has proposed the creation of 'condition era' tables as a means to systematically apply consistent rules for all medical conditions to infer distinct episodes of care from available information, such as diagnosis codes and problem lists. One important decision to make when applying condition era logic to an observational database (administrative claims or electronic health record) is to define the persistence window, or the period of time when two related conditions would be considered part of the same episode of care. Using a 0-day persistence window ensures that all events on the same day are combined, but events that occur over time are each treated separately. Applying a 30-day window allows for multiple observations that occur within a month timeframe to be considered part of the same episode of care. This analysis evaluates the impact of the 0-day and 30-day persistence window on the construction of condition eras across the central databases within the OMOP data community. Based on the findings, the OMOP research team recommends a 30-day persistence window assumption be consistently applied as a shared standard across the active surveillance analyses.

Methods

The OMOP common data model was developed to accommodate observational data from disparate sources, including administrative claims and electronic health records. As part of its design, the common data model contains a CONDITION_OCCURRENCE table, which stores all verbatim records from the source database that could be potentially used to infer condition occurrences. Most source databases provide an identifier for the condition (such as ICD-9-CM diagnosis code) used and a diagnosis date. However, particularly in administrative claims systems, diagnoses may be recorded to facilitate reimbursement of a particular procedure, and may be recorded multiple times on the same or successive dates if more than one service is provided. The CONDITION_ERA table is intended to provide one common structure for aggregating distinct diagnosis records into episodes of care for a given condition. CONDITION_ERA is a derived table, based on the CONDITION_OCCURRENCE table, that pre-processes the data to make it more analysis-friendly and minimize the computational burden. The intent behind developing this framework is to establish one systematic, transparent process for building CONDITION_ERAs that can be consistently applied across all conditions in a database, and potentially across multiple databases.

OMOP has 5 central databases: 4 administrative claims databases from Thomson Reuters, and 1 electronic health record database from GE Centricity. CCAE is a large commercial claims database of a privately insured population covering 59m lives. MDCD is an 11m person database containing claims from Medicaid services. MDCR contains Medicare supplemental claims for 4.6m lives. MSLR is a 1.5m person subset of the broader population that contains medical and pharmacy claims, along with laboratory values. GE contains 11.2m lives. OMOP has developed standardized procedures for constructing condition eras that have been successfully applied against 5 central databases, and have informed the development within its distributed partners.

One parameter within the standardized procedure is the 'persistence window'. The CONDITION_ERA table was developed to accommodate any persistence window. In its current work, the OMOP research team has populated the central databases with two sets of condition eras under two different persistence window assumptions, '0-day' and '30-day'. Creating eras under both scenarios have enabled exploration of the impact of the persistence window on the number of condition occurrences and its downstream effects on active surveillance analyses.

Similar to Drug Eras, Condition Eras are chronological periods of Condition Occurrence. Combining individual Condition Occurrences into a single Condition Era serves at least two purposes:

- It allows aggregation of chronic conditions that require frequent ongoing care, instead of treating each Condition Occurrence as an independent event.
- It allows aggregation of multiple, closely-timed doctor visits for the same condition to avoid double-counting the Condition Occurrences.

For example, consider a Person who visits his Primary Care Physician (PCP), who diagnoses the Person with a specific condition and refers the Person to a Specialist. One week later, the Person visits the Specialist, who confirms the PCP's diagnosis and provides the appropriate treatment to resolve the condition with no further care required. These two independent doctor visits should be aggregated into one Condition Era.

This model generally fits well for acute conditions, but may be less robust for chronic conditions. For example, chronic conditions that do not require regular follow-up may be recorded as multiple Condition Eras because the absence of data in the periods between visits does not justify the aggregation of the eras. Because the persistence window is small, it is likely that multiple visits will be captured in rapid succession for the same condition; however, it is unlikely that infrequent visits for chronic conditions (e.g. a Person with Rheumatoid Arthritis who visits his rheumatologist every three months) will be captured. However, the small window also reduces the likelihood that independent events will be falsely classified as the same Condition Era.

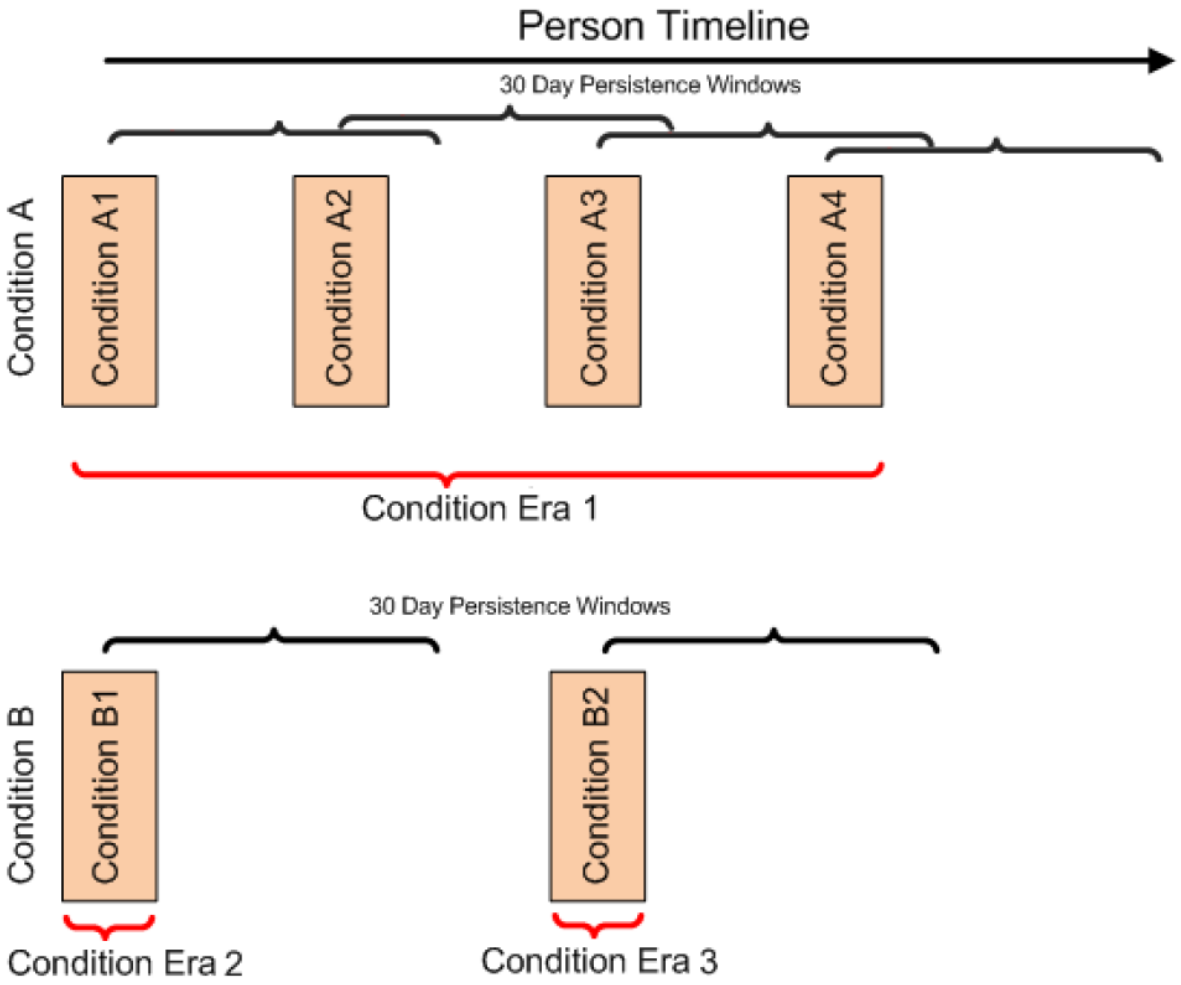

Figure 1: Example of constructing condition eras from condition occurrence records

Figure 1 provides an illustration of how the logic for condition eras is applied to diagnosis codes. Imagine a Person who has been diagnosed with two conditions during his insurance coverage period: Condition A and Condition B. The Person has been diagnosed with Condition A four times (A1, A2, A3, A4), and has been diagnosed with Condition B twice (B1, B2).

To define condition persistence for Condition A, the timing of successive diagnoses is considered. Here, A2 is within the persistence window of A1. Similarly, A3 is within the persistence window of A2, and A4 is within the persistence window of A3. Thus, the four diagnoses of Condition A should be consolidated into Condition Era1, with the start date equal to the diagnosis date for A1, and the end date equal to the diagnosis date for A4. With Condition B, significant time has elapsed between diagnoses B1 and B2. Therefore, it cannot be assumed that there is dependence between the diagnoses as the time exceeded the persistence window for B1. Therefore two distinct Condition Eras are defined, one each that corresponds to B1 and B2.

Note, that for Eras built using 30 day-persistence windows no additional 30 days is being added at the end of the last Condition Occurrence. That means, that Condition-free times within an Era is treated as continual Condition, while in the time following the Era no Condition is assumed.

Results

Table 1 shows the number of records created for each database using the two persistence window assumptions. CCAE is the largest database and observed 1.2b condition era records when using a 0-day persistence window, and 833m eras when using the 30-day persistence window. Thus, the consolidation in records was 33.7%. In other words, applying a 30-day persistence window, rather than the 0d assumption yielded 34% fewer condition episodes. This general phenomenon is seen fairly consistently between the claims databases (MDCD, MDCR, and MSLR), with the Medicaid database observing the largest reduction in eras, at 48.7%.

In contrast, the impact of the persistence window on GE conditions is much less significant. This should be expected due to the nature of condition capture in these systems. In GE, conditions are inferred from a problem list. Providers can enter a diagnosis with a start date and end date. As a result, it is much less likely to observe the same patient with the same condition at the same period of time. Across the entire database, GE has 0.9% consolidation of records when comparing the 0-day and 30-day persistence window assumptions.

Table 1: Condition eras by persistence window assumption

	Era Count	% reduction
CCAE		
0d	1,256,283,225	
30d	832,681,743	33.7%
MDCD		
0d	506,851,825	
30d	259,981,710	48.7%
MDCR		
0d	325,847,987	
30d	198,285,908	39.1%
MSLR		
0d	55,441,300	
30d	35,520,669	35.9%
GE		
0d	65,349,311	
30d	64,742,733	0.9%

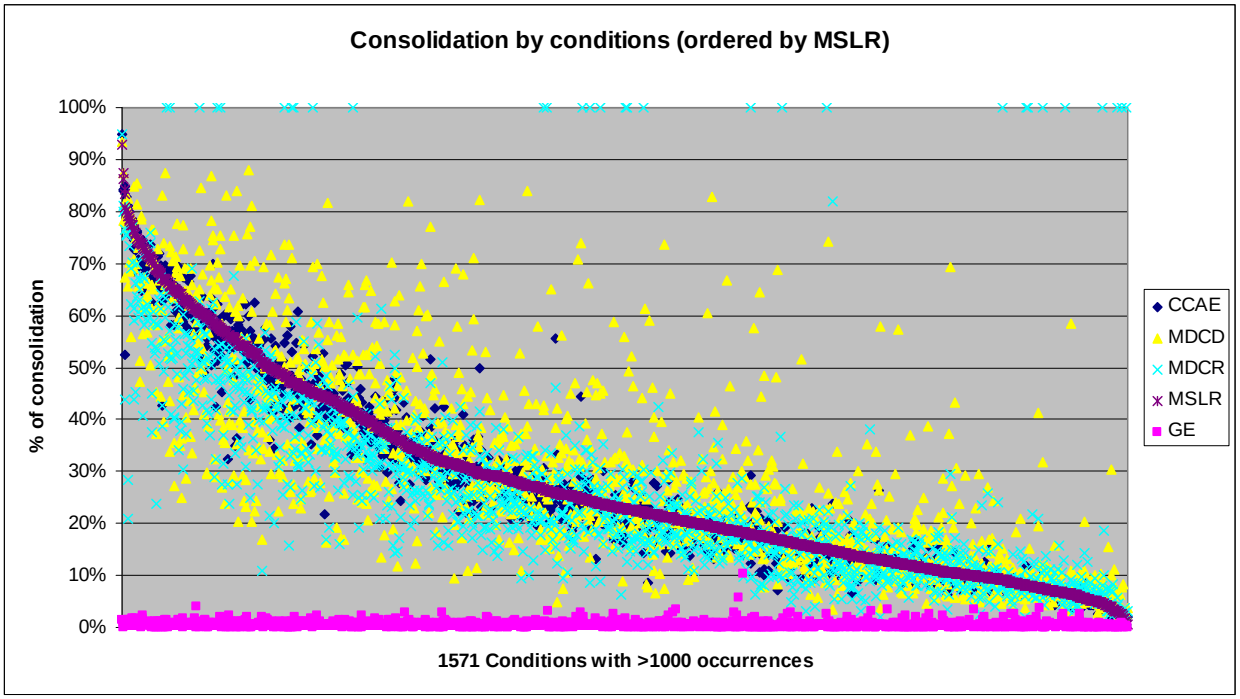

Figure 2: Era consolidation across conditions

Figure 2 shows the consolidation by conditions across the 5 central databases. Here, conditions are distinct SNOMED concepts, and are restricted to the 1571 distinct conditions with at least 1000 exposures (using 0d persistence window) in each database. The four claims databases (CCAE, MDCD, MDCR, MSLR) follow a consistent pattern of consolidation by condition. In CCAE and MSLR, 9 conditions have >80% consolidation (including 'End stage renal disease', 'Multiple myeloma', 'Anemia in chronic kidney disease', 'Chronic renal failure syndrome', and 'Primary malignant neoplasm of pancreas'. All of these conditions represent serious, highly morbid diseases that are likely to require intensive care during a short time duration; as such, observing most records combining into a small number of distinct episodes of care seems reasonable. In CCAE, 238 conditions have >50% consolidation. In general, this relative consistency makes sense since diagnoses are recorded as part of the administrative claims process, and variance (particularly within the Medicare database) likely reflects the underlying health of the source populations. In GE, only one condition ('Acute suppurative otitis media without spontaneous rupture of ear drum') has consolidation over 10% (10.3%).

Table 2 highlights the impact of the persistence window assumption within MDCD for conditions that are subsumed within eight of the OMOP Health Outcomes of Interest (HOI). A consistent pattern is observed across the HOIs that applying a 0-day persistence window to the source condition occurrences results in a 20-40% reduction in records. Using a 30-day persistence window further aggregates the records by at least 28%. The 30-day assumption had the largest impact on acute renal failure conditions, where 82% of the source diagnoses and 75% of the 0-day eras are consolidated into 205,318 distinct episodes.

Table 2: Impact of persistence windows across OMOP Health Outcomes of Interest

MDCD	CONDITION_OCCURRENCE		CONDITION ERAS				
Conditions included within HOI definitions:	PERSON_COUNT	OCCUR_COUNT	0d PW	% reduction from OCCURRENCE	30d PW	% reduction from OCCURRENCE	% reduction from 0d PW
Angioedema	19,668	52,079	31,394	39.7%	22,440	56.9%	28.5%
Aplastic Anemia	23,022	169,324	111,495	34.2%	38,052	77.5%	65.9%
Acute Liver Failure	278,711	1,752,630	1,146,140	34.6%	651,229	62.8%	43.2%
Acute Renal Failure	133,466	1,138,148	825,258	27.5%	205,318	82.0%	75.1%
Bleeding	1,328,161	6,430,307	4,480,749	30.3%	2,581,786	59.8%	42.4%
Hip Fracture	56,047	855,962	682,417	20.3%	175,948	79.4%	74.2%
Acute myocardial Infarction	214,954	1,814,000	1,280,730	29.4%	466,607	74.3%	63.6%
Upper GI Ulcer Hospitalization	274,123	1,193,468	861,778	27.8%	421,083	64.7%	51.1%

Error! Reference source not found. Table 3 provides a real de-identified person-level example from the database. Condition eras were constructed for 'Acute peptic ulcer with hemorrhage AND obstruction', one of the conditions that is part of the Upper GI Ulcer Hospitalization definition. Using the 0d persistence window, one person had 248 condition eras, with the earliest occurring in Mar2003, and the latest ending in Nov2007. As the pattern suggests, the patient appears to be receiving continuous care for the condition, every one or two weeks. Only once, from Jun2004-Aug2004 (eras 65-66) did the patient not have data that suggested care for over a month. Using the 30d persistence window assumption, these condition occurrence records were consolidated into two condition eras: 1 from Mar2003 through Jun2004, and another from Aug2004 through Nov2007.

The overall consequence of selecting the persistence window assumption is that using '0d' yields 248 distinct episodes, while the'30d' persistence window produces 2 distinct episodes. If an analysis were to consider acute peptic ulcer to be a chronic condition or to be focused on incident events, then selection of the first occurrence could provide a proxy for incidence and both assumptions would provide consistent results. However, if an analysis intended to assess each occurrence as a potential event to monitor in active surveillance, it is easy to see the potential consequence in estimating the number of observed cases.

In this case, one could argue the 30-day persistence window may have been too stringent, in that the data may suggest that this individual has one continuous episode of care for 5-year window. In the extreme to accommodate chronic conditions, one could imagine using an infinite persistence window to combine occurrence records into single periods of conditions for each individual. However, there may be cases when the start of a new episode of care is significant. For example, to the previous example, the patients with rheumatoid arthritis may have only one extended diagnosis history for the chronic condition, but the observation of a new episode of care

could be an indicator for change in disease severity (such as joint flare or treatment ineffectiveness) that could be used as a potential marker for observing the effects of medicines.

Table 3: Example from MDCD of one person with multiple records for "Acute peptic ulcer with hemorrhage AND obstruction"

Example from MDCD: One person with multiple records for "Acute peptic ulcer with hemorrhage AND obstruction"

Eras using 0-day persistence window

Era #	Era Start Date	Era End Date	Era Gap
1	25-Mar-03	25-Mar-03	
2	1-Apr-03	1-Apr-03	7
3	8-Apr-03	8-Apr-03	7
4	15-Apr-03	15-Apr-03	7
5	22-Apr-03	22-Apr-03	7
6	1-May-03	1-May-03	9
7	7-May-03	7-May-03	6
8	14-May-03	14-May-03	7
9	21-May-03	21-May-03	7
10	28-May-03	28-May-03	7
11	1-Jun-03	1-Jun-03	4
12	12-Jun-03	12-Jun-03	11
13	19-Jun-03	19-Jun-03	7
14	1-Jul-03	1-Jul-03	12
15	9-Jul-03	9-Jul-03	8
16	15-Jul-03	15-Jul-03	6
17	22-Jul-03	22-Jul-03	7
18	29-Jul-03	29-Jul-03	7
19	1-Aug-03	1-Aug-03	3
20	12-Aug-03	12-Aug-03	11
21	19-Aug-03	19-Aug-03	7
22	26-Aug-03	26-Aug-03	7
23	1-Sep-03	1-Sep-03	6
24	9-Sep-03	9-Sep-03	8
25	16-Sep-03	16-Sep-03	7
26	23-Sep-03	23-Sep-03	7
27	1-Oct-03	1-Oct-03	8
28	8-Oct-03	8-Oct-03	7
29	14-Oct-03	14-Oct-03	6
30	28-Oct-03	28-Oct-03	14
31	1-Nov-03	1-Nov-03	4
32	7-Nov-03	7-Nov-03	6
33	13-Nov-03	13-Nov-03	6
34	18-Nov-03	18-Nov-03	5
35	25-Nov-03	25-Nov-03	7
36	1-Dec-03	1-Dec-03	6
37	9-Dec-03	9-Dec-03	8
38	17-Dec-03	17-Dec-03	8
39	22-Dec-03	22-Dec-03	5
40	29-Dec-03	29-Dec-03	7
41	1-Jan-04	1-Jan-04	3
42	7-Jan-04	7-Jan-04	6
43	14-Jan-04	14-Jan-04	7
44	21-Jan-04	21-Jan-04	7
45	27-Jan-04	27-Jan-04	6
46	1-Feb-04	1-Feb-04	5
47	10-Feb-04	10-Feb-04	9
48	17-Feb-04	17-Feb-04	7
49	24-Feb-04	24-Feb-04	7
50	1-Mar-04	1-Mar-04	6
51	11-Mar-04	11-Mar-04	10
52	18-Mar-04	18-Mar-04	7
53	25-Mar-04	25-Mar-04	7
54	30-Mar-04	30-Mar-04	5
55	1-Apr-04	1-Apr-04	2
56	8-Apr-04	8-Apr-04	7
57	13-Apr-04	13-Apr-04	5
58	16-Apr-04	16-Apr-04	3
59	20-Apr-04	20-Apr-04	4
60	27-Apr-04	27-Apr-04	7
61	1-May-04	1-May-04	4
62	12-May-04	12-May-04	11
63	19-May-04	19-May-04	7
64	16-Jun-04	16-Jun-04	28
65	18-Jun-04	18-Jun-04	2
66	1-Aug-04	1-Aug-04	44
67	9-Aug-04	9-Aug-04	8
68	17-Aug-04	17-Aug-04	8
69	24-Aug-04	24-Aug-04	7
70	1-Sep-04	1-Sep-04	8
71	7-Sep-04	7-Sep-04	6
72	13-Sep-04	13-Sep-04	6
73	20-Sep-04	20-Sep-04	7
74	27-Sep-04	27-Sep-04	7
75	1-Oct-04	1-Oct-04	4
76	11-Oct-04	11-Oct-04	10
77	18-Oct-04	18-Oct-04	7
78	25-Oct-04	25-Oct-04	7
79	1-Nov-04	1-Nov-04	7
80	8-Nov-04	8-Nov-04	7
81	15-Nov-04	15-Nov-04	7
82	22-Nov-04	22-Nov-04	7
83	1-Dec-04	1-Dec-04	9
84	8-Dec-04	8-Dec-04	7
85	13-Dec-04	13-Dec-04	5
86	21-Dec-04	21-Dec-04	8
87	27-Dec-04	27-Dec-04	6
88	1-Jan-05	1-Jan-05	5
89	10-Jan-05	10-Jan-05	9
90	17-Jan-05	17-Jan-05	7
91	25-Jan-05	25-Jan-05	8
92	1-Feb-05	1-Feb-05	7
93	7-Feb-05	7-Feb-05	6
94	15-Feb-05	15-Feb-05	8
95	22-Feb-05	22-Feb-05	7
96	1-Mar-05	1-Mar-05	7
97	8-Mar-05	8-Mar-05	7
98	15-Mar-05	15-Mar-05	7
99	22-Mar-05	22-Mar-05	7
100	29-Mar-05	29-Mar-05	7
101	1-Apr-05	1-Apr-05	3
102	5-Apr-05	5-Apr-05	4
103	12-Apr-05	12-Apr-05	7
104	19-Apr-05	19-Apr-05	7
105	25-Apr-05	25-Apr-05	6
106	1-May-05	1-May-05	6
107	10-May-05	10-May-05	9
108	17-May-05	17-May-05	7
109	24-May-05	24-May-05	7
110	1-Jun-05	1-Jun-05	8
111	7-Jun-05	7-Jun-05	6
112	14-Jun-05	14-Jun-05	7
113	21-Jun-05	21-Jun-05	7
114	28-Jun-05	28-Jun-05	7
115	1-Jul-05	1-Jul-05	3
116	12-Jul-05	12-Jul-05	11
117	18-Jul-05	18-Jul-05	6
118	27-Jul-05	27-Jul-05	9
119	1-Aug-05	1-Aug-05	5
120	8-Aug-05	8-Aug-05	7
121	15-Aug-05	15-Aug-05	7
122	22-Aug-05	22-Aug-05	7
123	29-Aug-05	29-Aug-05	7
124	1-Sep-05	1-Sep-05	3
125	6-Sep-05	6-Sep-05	5
126	12-Sep-05	12-Sep-05	6
127	20-Sep-05	20-Sep-05	8
128	26-Sep-05	26-Sep-05	6
129	1-Oct-05	1-Oct-05	5
130	7-Oct-05	7-Oct-05	6
131	12-Oct-05	12-Oct-05	5
132	17-Oct-05	17-Oct-05	5
133	21-Oct-05	21-Oct-05	4
134	25-Oct-05	25-Oct-05	4
135	1-Nov-05	1-Nov-05	7
136	4-Nov-05	4-Nov-05	3
137	9-Nov-05	9-Nov-05	5
138	16-Nov-05	16-Nov-05	7
139	22-Nov-05	22-Nov-05	6
140	1-Dec-05	1-Dec-05	9
141	7-Dec-05	7-Dec-05	6
142	14-Dec-05	14-Dec-05	7
143	21-Dec-05	21-Dec-05	7
144	26-Dec-05	26-Dec-05	5
145	1-Jan-06	1-Jan-06	6
146	10-Jan-06	10-Jan-06	9
147	18-Jan-06	18-Jan-06	8
148	24-Jan-06	24-Jan-06	6
149	1-Feb-06	1-Feb-06	8
150	8-Feb-06	8-Feb-06	7
151	15-Feb-06	15-Feb-06	7
152	21-Feb-06	21-Feb-06	6
153	1-Mar-06	1-Mar-06	8
154	8-Mar-06	8-Mar-06	7
155	15-Mar-06	15-Mar-06	7
156	22-Mar-06	22-Mar-06	7
157	29-Mar-06	29-Mar-06	7
158	1-Apr-06	1-Apr-06	3
159	5-Apr-06	5-Apr-06	4
160	12-Apr-06	12-Apr-06	7
161	19-Apr-06	19-Apr-06	7
162	26-Apr-06	26-Apr-06	7
163	1-May-06	1-May-06	5
164	3-May-06	3-May-06	2
165	10-May-06	10-May-06	7
166	17-May-06	17-May-06	7
167	24-May-06	24-May-06	7
168	1-Jun-06	1-Jun-06	8
169	7-Jun-06	7-Jun-06	6
170	14-Jun-06	14-Jun-06	7
171	21-Jun-06	21-Jun-06	7
172	28-Jun-06	28-Jun-06	7
173	1-Jul-06	1-Jul-06	3
174	5-Jul-06	5-Jul-06	4
175	12-Jul-06	12-Jul-06	7
176	18-Jul-06	18-Jul-06	6
177	26-Jul-06	26-Jul-06	8
178	1-Aug-06	1-Aug-06	6
179	9-Aug-06	9-Aug-06	8
180	16-Aug-06	16-Aug-06	7
181	23-Aug-06	23-Aug-06	7
182	29-Aug-06	29-Aug-06	6
183	1-Sep-06	1-Sep-06	3
184	6-Sep-06	6-Sep-06	5
185	13-Sep-06	13-Sep-06	7
186	20-Sep-06	20-Sep-06	7
187	27-Sep-06	27-Sep-06	7
188	1-Oct-06	1-Oct-06	4
189	5-Oct-06	5-Oct-06	4
190	11-Oct-06	11-Oct-06	6
191	1-Nov-06	1-Nov-06	21
192	1-Dec-06	1-Dec-06	30
193	16-Dec-06	16-Dec-06	15
194	20-Dec-06	20-Dec-06	4
195	27-Dec-06	27-Dec-06	7
196	1-Jan-07	1-Jan-07	5
197	3-Jan-07	3-Jan-07	2
198	10-Jan-07	10-Jan-07	7
199	17-Jan-07	17-Jan-07	7
200	23-Jan-07	23-Jan-07	6
201	1-Feb-07	1-Feb-07	9
202	7-Feb-07	7-Feb-07	6
203	11-Feb-07	11-Feb-07	4
204	14-Feb-07	14-Feb-07	3
205	21-Feb-07	21-Feb-07	7
206	1-Mar-07	1-Mar-07	8
207	7-Mar-07	7-Mar-07	6
208	14-Mar-07	14-Mar-07	7
209	21-Mar-07	21-Mar-07	7
210	28-Mar-07	28-Mar-07	7
211	1-Apr-07	1-Apr-07	4
212	4-Apr-07	4-Apr-07	3
213	11-Apr-07	11-Apr-07	7
214	18-Apr-07	18-Apr-07	7
215	25-Apr-07	25-Apr-07	7
216	1-May-07	1-May-07	6
217	9-May-07	9-May-07	8
218	16-May-07	16-May-07	7
219	23-May-07	23-May-07	7
220	30-May-07	30-May-07	7
221	1-Jun-07	1-Jun-07	2
222	6-Jun-07	6-Jun-07	5
223	13-Jun-07	13-Jun-07	7
224	20-Jun-07	20-Jun-07	7
225	27-Jun-07	27-Jun-07	7
226	1-Jul-07	1-Jul-07	4
227	11-Jul-07	11-Jul-07	10
228	18-Jul-07	18-Jul-07	7
229	25-Jul-07	25-Jul-07	7
230	1-Aug-07	1-Aug-07	7
231	8-Aug-07	8-Aug-07	7
232	16-Aug-07	16-Aug-07	8
233	23-Aug-07	23-Aug-07	7
234	1-Sep-07	1-Sep-07	9
235	5-Sep-07	5-Sep-07	4
236	12-Sep-07	12-Sep-07	7
237	19-Sep-07	19-Sep-07	7
238	26-Sep-07	26-Sep-07	7
239	1-Oct-07	1-Oct-07	5
240	3-Oct-07	3-Oct-07	2
241	10-Oct-07	10-Oct-07	7
242	17-Oct-07	17-Oct-07	7
243	24-Oct-07	24-Oct-07	7
244	1-Nov-07	1-Nov-07	8
245	7-Nov-07	7-Nov-07	6
246	14-Nov-07	14-Nov-07	7
247	21-Nov-07	21-Nov-07	7
248	28-Nov-07	28-Nov-07	7

Eras using 30-day peristence window

Era	Era Start	Era End	Era
1	25-Mar-03	18-Jun-04	
2	1-Aug-04	28-Nov-07	44

Figure 3 shows the distribution of era gaps that exist when applying the 0d persistence window, as observed against the conditions within eight of the OMOP Health Outcomes of Interest. Gaps were calculated for each person and each distinct SNOMED concept within these HOIs as the length of time from the era end in the preceding era to the era start in the subsequent era. All HOIs exhibit a similar pattern with >70% of gaps are <30 days; these gaps represent the eras which would be combined by applying the 30d persistence window assumption. All eras with gap >30d would remain distinct.

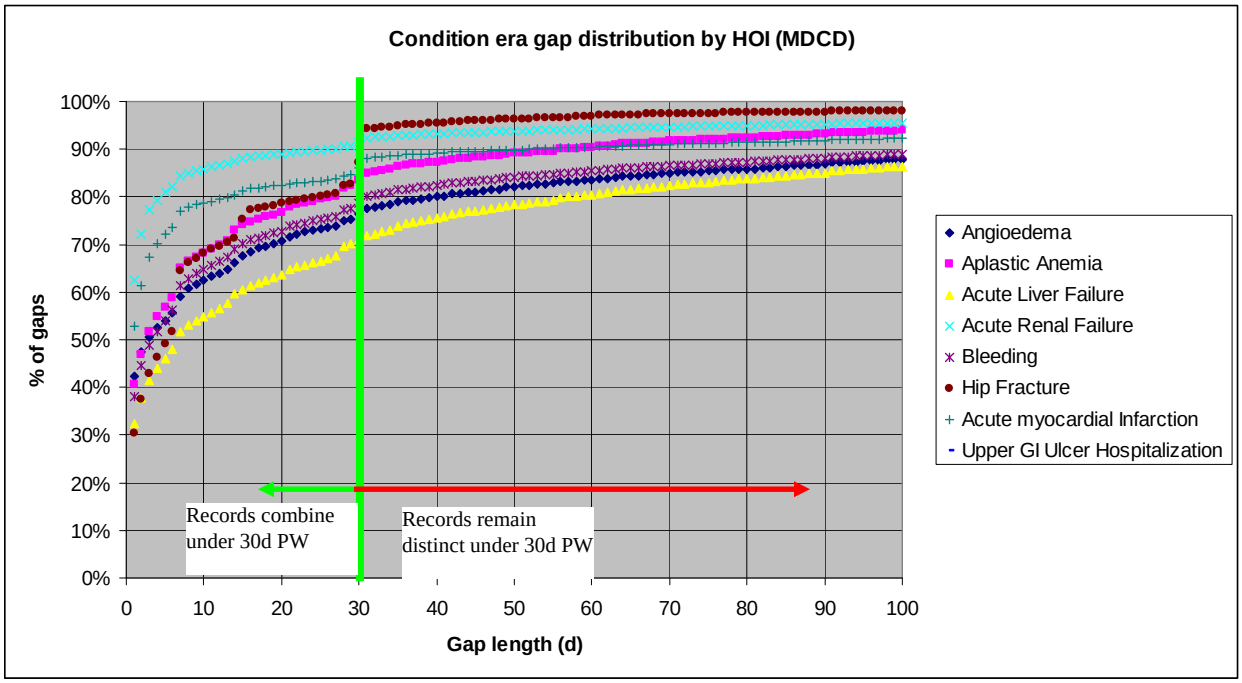

Figure 3: Era gap distribution for OMOP Health Outcomes of interest, in MDCD

As Table 4 details, among the gaps <30d (those to be aggregated under the 30d persistence window), >77% of gaps were <10d, and >89% were <20d for all conditions.

Table 4: Era gap analysis for 30d persistence window in MDCD

MDCD		Among gaps <30d	
Among conditions within HOI definitions:	% of era gaps <30d	% <10d	% <20d
Angioedema	76.36%	81.91%	92.62%
Aplastic Anemia	83.57%	81.69%	91.87%
Acute Liver Failure	70.92%	77.50%	89.63%
Acute Renal Failure	91.39%	94.00%	97.44%
Bleeding	78.67%	82.32%	92.63%
Hip Fracture	87.24%	78.08%	90.08%
Acute myocardial Infarction	86.00%	91.54%	95.91%
Upper GI Ulcer Hospitalization	85.16%	88.15%	95.22%

Figure 4 and Table 5 show the impact of the persistence window on alternative analysis methods. There is fair agreement between two disproportionality analysis metrics: PRR and IC, when applying the two drug era persistence window assumptions.

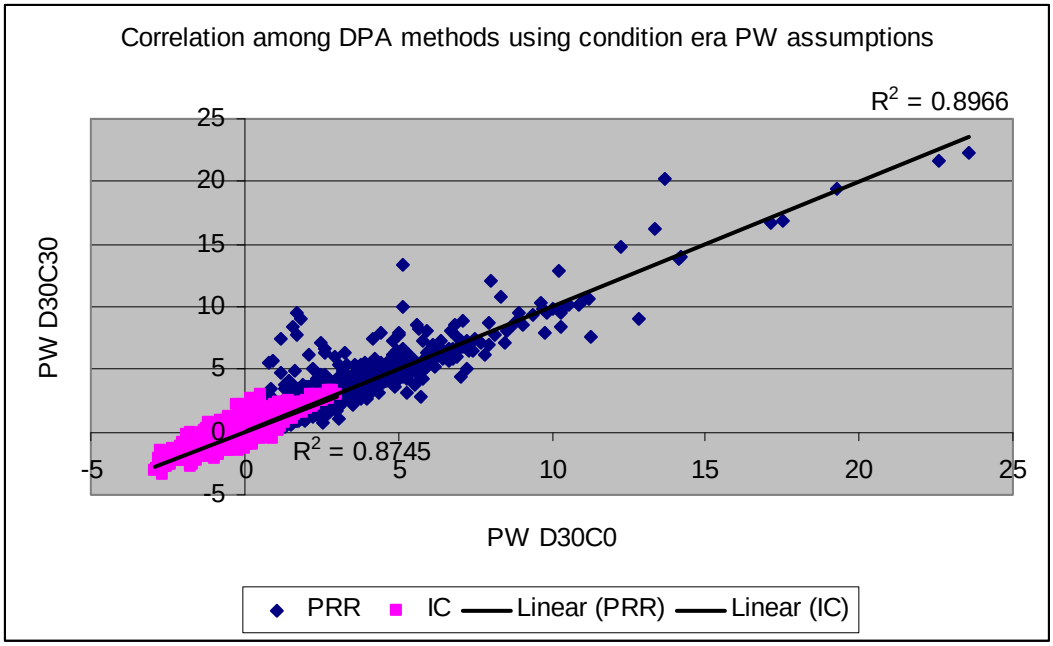

Figure 4: Correlation between disproportionality analysis methods under different condition era persistence windows

Table 5: Condition era persistence window impact

Condition era persistence window impact

		Difference from 0d - 30d		
	R^2	MIN	AVG	MAX
PRR	0.8966	0.00	0.15	8.20
IC	0.8745	0.00	0.15	2.42

Discussion

Deriving condition eras from disparate diagnosis records can be a valuable component of a common data model for use in active drug safety surveillance. It creates standardized, transparent approach that can be applied consistently across any analysis. To date, all methods developed within OMOP have used condition eras, even though common data model can accommodate using raw condition occurrence records. An open question is what persistence window to use: during the initial design, the OMOP advisory board requested studies using both the 0-day and 30-day assumptions.

The potential concern with applying any persistence window when defining episodes of care is misclassification. A longer persistence window risks treating diagnoses that reflected independent conditions are part of the same continuum of care, while shorter persistence window assumptions may falsely separate the records from the same episode of care and observe them as distinct occurrences. In the context of active surveillance, where condition occurrences may be used as proxies for potential observations of adverse events, both forms of misclassification bias require careful consideration. As Table 4 reveals, even when using a 30d persistence window assumption, the large majority of aggregated eras come from the same diagnosis occurring less than 10 days from one another. In these cases, it seems more unlikely these conditions represent independent events than it does that the gaps coincide with a common episode of care.

If method only uses the first occurrence of cases as a proxy for incident events, then consolidation of eras does not matter (since both have the same first start date). However, if a method attempts to use prevalent cases, as measured by each distinct era occurrence, the selection of the persistence window can be significant. To reiterate, multiple eras for the same condition does not necessarily indicate distinct occurrences of the condition, but instead represent independent periods of time where the data suggests the condition may have occurred. That is, chronic conditions, such as diabetes, are likely to be considered to persist following the first occurrence, but a person may have multiple eras for diabetes because they do not receive care of the disease on a regular basis.

Each analysis method may be parameterized to tailor the persistence window to the specific drug-condition relationship of interest. For example, the multi-set case control estimation program (http://omop.fnih.org/MethodsLibrary) allows the user to select the persistence window assumption (0d or 30d), but only uses first occurrence of conditions. The disproportionality analysis program allows the user to specify both the condition era persistence window and whether to count ‘first occurrence’ or ‘all occurrences’ of conditions as potential events.

As with selecting best practice rules for drug era persistence window, it seems wherever possible when conducting first-pass analyses for active surveillance, it would be preferable to minimize the number of subjective decisions needed and to establish best practices that can be applied in many contexts, across multiple databases for different drugs and outcomes. If potential relationships are uncovered through this process, further tuning of the analysis-level parameters can be made to refine the estimate of the drug-condition association. In the extreme case, if reviewers were concerned with the potential impact of the logic used to infer condition eras (such as wanting to focus only on primary inpatient diagnoses), the analysis could be manually

reproduced by directly using the source records within the CONDITION_OCCURRENCE table. In this regard, the common data model should be able to accommodate any level of analysis, from initial stages using scalable, automated exploratory processes to later stage evaluation studies.

Based on this study's findings, it is clear that selection of a persistence window can have a significant effect on the observed number of condition episodes. The interpretation of all analyses using condition eras should consider the potential impact of the assumption when drawing conclusions. As an initial assumption, it appears using the 30-day persistence window is sufficient for active surveillance. Condition eras that are aggregated under this 30-day rule appear clinically meaningful, and seem more likely to represent a continuum of care than independent events. Most aggregation comes from diagnoses whose gaps are less than 10 days.

Further study can evaluate the consistency in active surveillance analysis results based on the persistence window assumption. For example, a screening method could be run against a database using both 0d and 30d assumption, and the correlation of the results could be assessed, as well was exploration of which drug-condition pairs showed the largest differences.

