COVID-19に関する頻出医療指示パターンの解析と可視化*

お茶の水女子大学

Le Hieu Hanh　准教授

le@is.ocha.ac.jp

*東工大情報理工学院横田研究室の研究成果

2 電子カルテデータベース

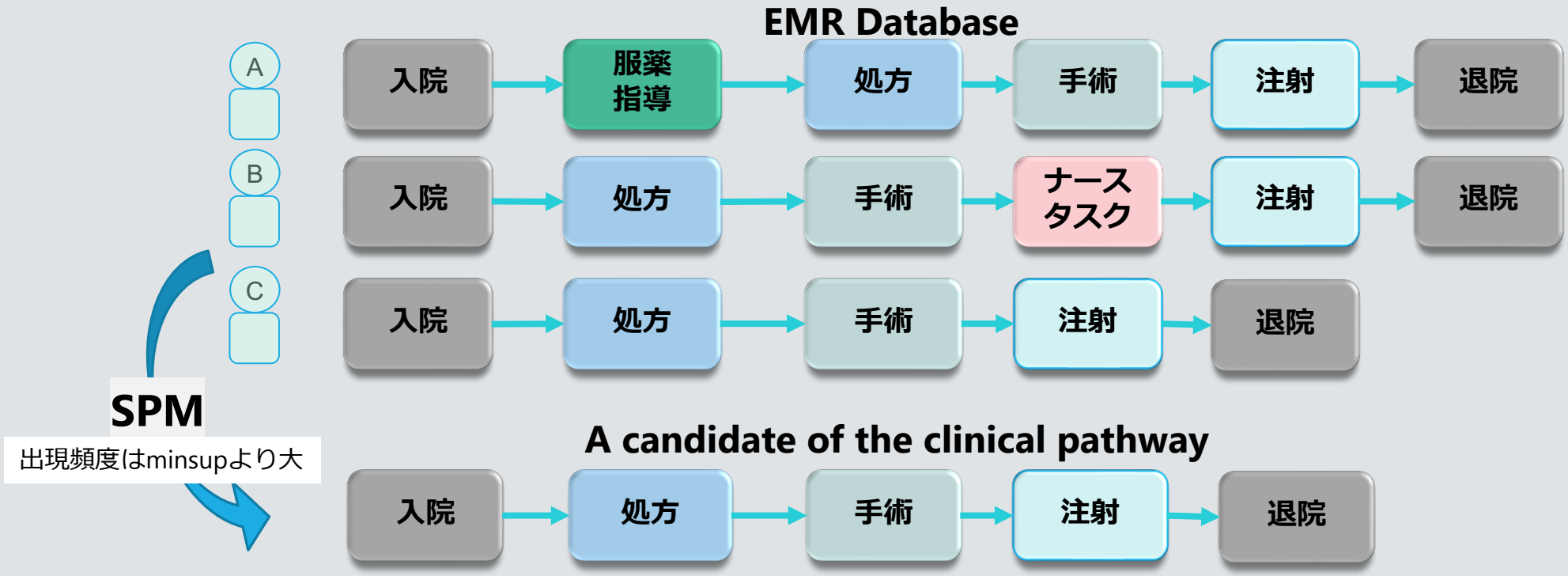
- 紙のカルテに変わり，電子カルテが大規模病院を中心に普及している
- 電子カルテはコスト削減が期待できる（一次利用）
  - 対象となる患者の検索や医療指示の履歴閲覧の高速化を可能に
- さらに，複数の医療機関による集約したデータの分析による医療支援も期待大（二次利用）
  - 千年カルテプロジェクト

3 データ解析の対象

- テキスト（自然言語）
  - テキストマイニング技術を用いて電子カルテデータを解析
    - 例：疾患の症状発見など
- イメージ
  - コンピュータ断層撮影（CT），磁気共鳴画像（MRI）検査結果などに基づいて病気診断を行う
- **医療指示**
  - **疾患の治療のパターン発見**

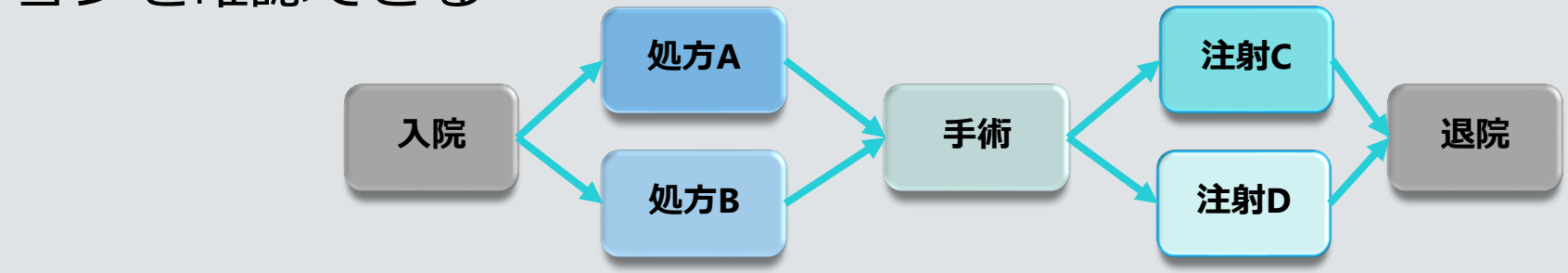
4 コンピュータベースクリニカルパス抽出

- 電子カルテデータベースからSequential Pattern Mining (SPM)を適用し候補なクリニカルパスを抽出可能



5 シーケンシャルバリアント（SV）

- 実験結果から頻出シーケンシャルパターンは共通なパータンを含めることが多い
- このようなグループをSVという
- SVは頻出シーケンシャルパターンにおける医療指示の分岐を示す
  - 医療オプションを確認できる



6 SVの要因推定

- 非ホジキンリンパ治療で**年齢・入院歴**と**検査値**が分岐要因

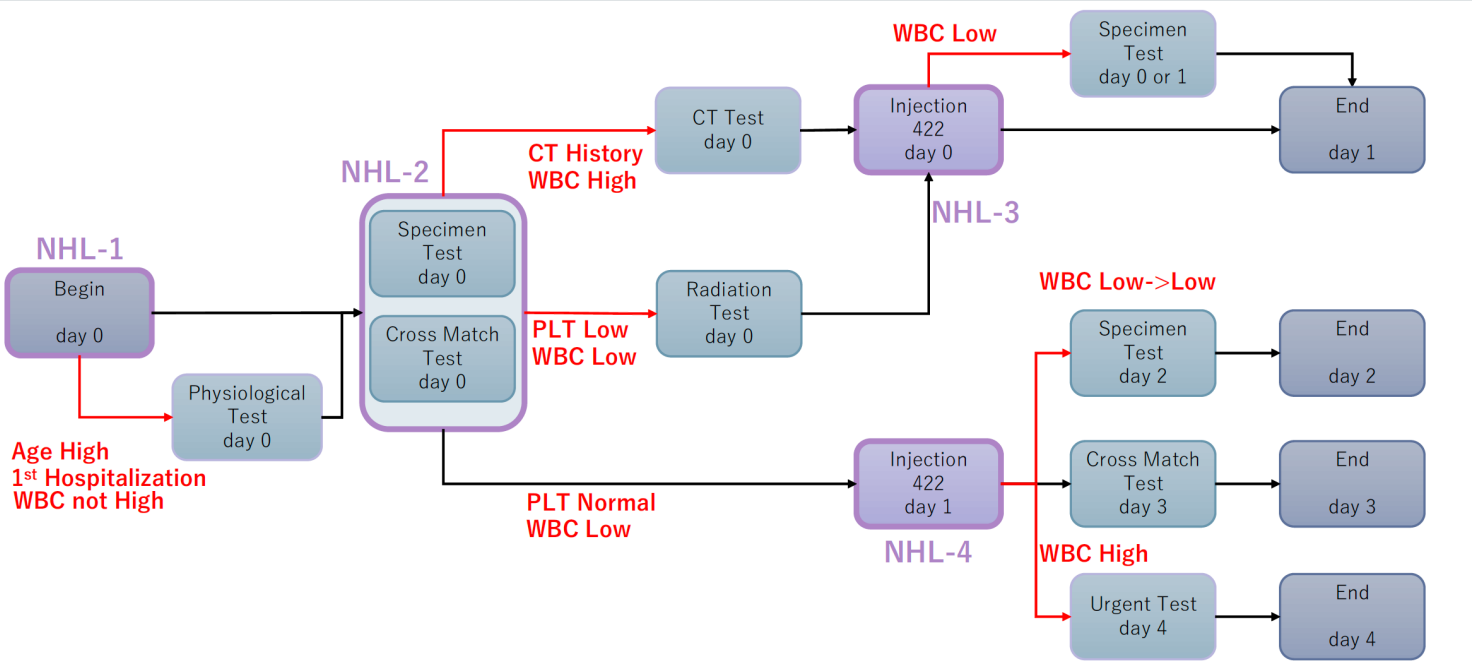


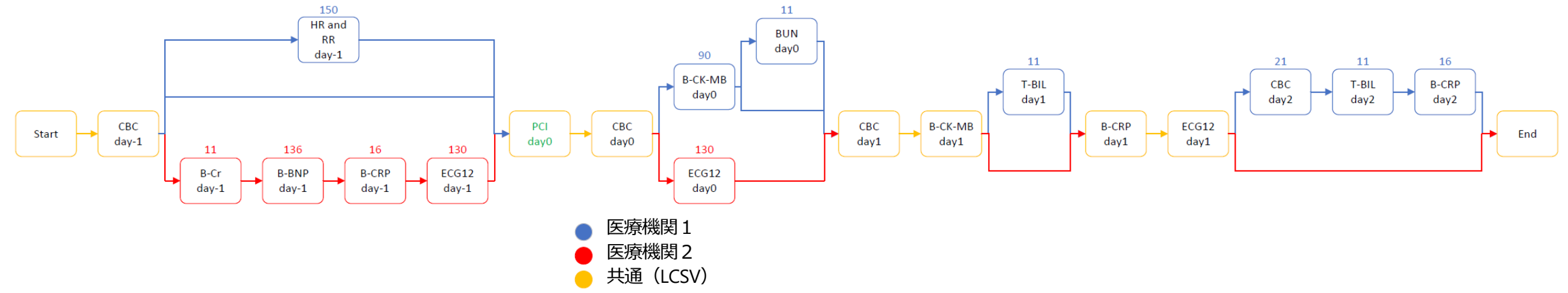
Fig. 12. Part of the extracted frequent paths for NHL (422 is the efficacy code of antimetabolite medicines).

WBC:白血球数
PLT:血小板数

Le et al., ACM Transaction on Computing for Healthcare, 2023

7 ２医療機関におけるSVの比較

- ２SVを抽出し，共通シーケンスバリアント（LCSV）と併合シーケンスバリアント（MSV）を生成し，共通点と差異点を可視化



医療機関１ (minSup=0.15) and **医療機関２** (minSup=0.25)

Li et al., DEXA 2022

8 既存研究の課題

- **3つ以上の医療機関におけるSV比較**が未実施
  - SPMで抽出されたSVが頻出な検査項目のみを含み，**重要な投薬などの処置項目が抽出できない**
  - 多くのSVを一度に比較しようとすると，**共通部分がなくなる**

- **異なる時期の違いに関する分析**が未実施
  - COVID-19治療のように治療中に**投薬等の変更が生じた可能性有**

9 本研究の貢献

- 貢献1
  - 研究1:**3つ以上に医療機関での頻出医療指示パターンの比較**手法の提案
  - 研究2:異なる時期の疾病に対する**頻出医療指示パターンの違いの発生時期を発見**する手法の提案
- 貢献2
  - COVID-19のデータに提案手法を適用し，その効果を確認

研究1:3つ以上の医療機関のでの頻出医療指示パターンの比較

① 頻出な検査項目の除外，各医療機関のSV抽出

② 医療機関群のクラスタリング

③ クラスタ内でMSVを作成

① 頻出な検査項目の除外(1/3)

- 多くのシーケンスに出現しシーケンス中の出現回数が多い検査項目があると**投薬等の処置項目が抽出できない**
- 医療指示がシーケンスに出現する平均回数を新たに定義

- $TO/S\,(\mathrm{order}) = \dfrac{\mathrm{order}\text{が全シーケンス中に出現する回数}}{\text{全シーケンス数}}$

- TO/S が高い検査項目を除外

① 頻出な検査項目の除外(2/3)

- 複数の医療機関で除外する医療指示を一元的に定める
- それぞれの医療指示について，TO/S が最大となる医療機関における TO/S の値，$\mathrm{Max}(TO/S)$ を算出

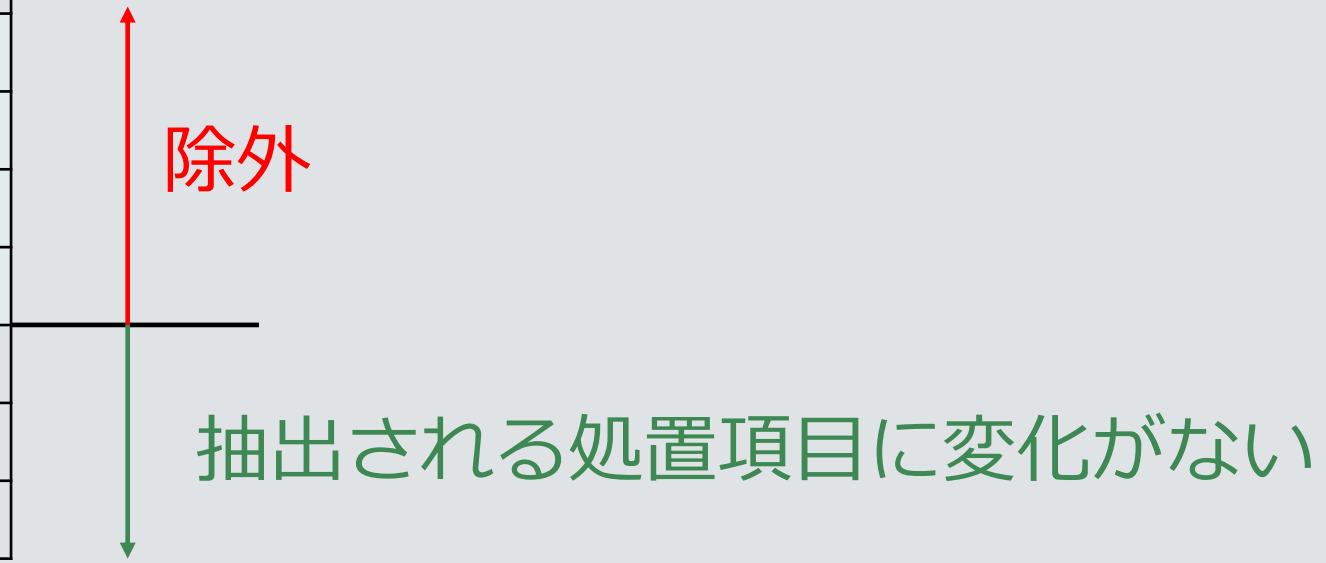
	医療機関1	医療機関2	医療機関3	医療機関4
医療指示a	0.3	1.8	0.5	1.2

$\mathrm{Max}(TO/S)$

13 ① 頻出な検査項目の除外(3/3)

- $\mathrm{Max}(TO/S)$ が高い検査項目から順に除外し，抽出される処置項目に変化がない医療指示の $\mathrm{Max}(TO/S)$ の最大値をまとめ，それよりも $\mathrm{Max}(TO/S)$ が高い検査項目を除外

	$\mathrm{Max}(TO/S)$
検査e	10.0
検査a	7.8
検査c	4.6
検査d	2.2
検査g	1.4
検査b	0.5
検査f	0.1

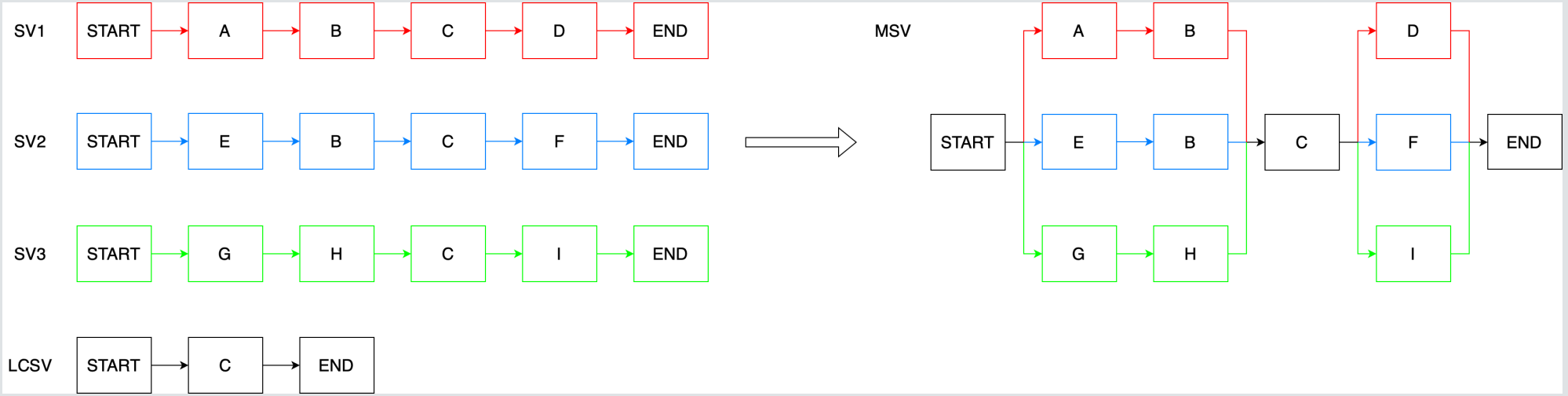


14 ② 医療機関群のクラスタリング

- 共通部分を持ついくつかのクラスタに分類
  - 多くのSVを一度に比較しようとすると，共通部分であるLCSVが存在しなくなる
- 階層的クラスタリングを用いる
  - SV間の距離を算出
  - 距離の近いクラスタを順に併合し，デンドログラムを作成

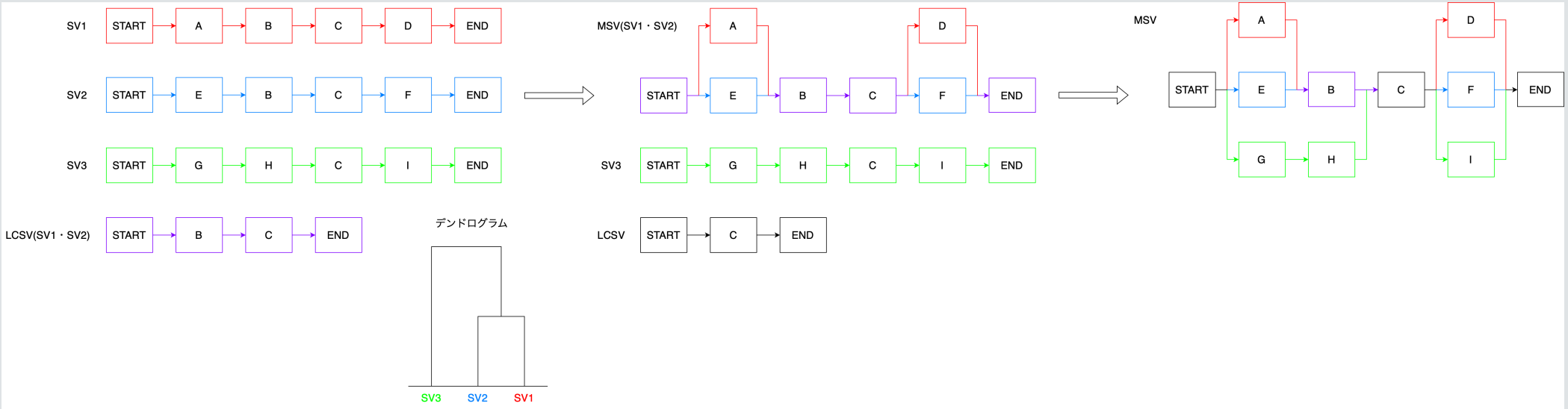
15 ③ 3つ以上の医療機関のMSVを導出(1/2) 直接併合法

- クラスタ内でのLCSVにそれぞれのSVを直接併合
  - **個々のSVがわかりやすい**
  - 全体の共通部分しか表せない



16 ③ 3つ以上の医療機関のMSVを導出(2/2) 距離ベース併合法

- クラスタリングで得られたデンドログラムを基に距離が近いSVから順に併合
  - **部分的な共通部分も表せる**
  - ラベルの種類が増え，複雑



17 評価実験に用いたデータセット

- 関西経済連合会の支援による「COVID-19感染の診療への影響調査と予測モデル開発」に関する研究の研究協力機関から提供されたデータに適用
  - 23医療機関とCOVID-19の第5波

医療機関	A	B	C	D	E	F	G	H	I	J	K	L
シーケンス数	130	23	11	102	89	79	126	36	29	104	104	30
平均シーケンス長(除外前)	181.7	256.0	45.3	153.7	242.6	284.3	90.2	127.4	211.4	317.5	112.2	275.7
平均シーケンス長(除外後)	18.7	56.6	5.3	22.0	47.2	31.8	23.2	15.6	37.2	51.3	35.3	57.8
医療機関	M	N	O	P	Q	R	S	T	U	V	W	
シーケンス数	179	228	182	29	57	283	55	69	22	28	22	
平均シーケンス長(除外前)	282.4	272.5	181.0	73.7	310.0	139.2	174.5	189.4	340.0	220.3	137.1	
平均シーケンス長(除外後)	40.6	46.2	23.0	16.4	66.0	13.3	21.3	36.2	51.1	38.6	18.4	

18 実験結果　医療指示の略称

検査項目

略号	医療指示名	略号	医療指示名
CT	CT撮影	CRP	C反応性蛋白
BUN	尿素窒素	DD	Dダイマー
AST	AST	FRTN	フェリチン
GB	グロブリンクラス別ウイルス抗体	CBC	末梢血液
QFT	結核菌特異的インターフェロン-γ産生能	XP	X線検査
		HV	肝炎ウイルス検査

処置項目

略号	医療指示名	概要
DM	デキサメタゾン	ステロイド剤
PD	プレドニゾロン	ステロイド剤
HP	ヘパリン	血液凝固防止剤
AA	アセトアミノフェン	解熱鎮痛剤
IN	インスリン	糖尿病治療薬
OT	酸素吸入	酸素の投与

19 医療指示の除外(1/2)

- $Max(TO/S)$が肝炎ウイルス検査以上の検査項目を除外

医療指示名(検査項目)	Max(TO/S)
経皮的血酸素飽和度測定	10.955
血糖値測定	10.222
呼吸心拍監視	9.528
X線検査	9.467
液体酸素定置式液化酸素貯槽	9.318
生食注シリンジ	8.267
末梢血液	7.455
グルコース	6.828
総ビリルビン	6.545
CRP	6.364
点滴注射	6.203
フェリチン	5.611
血液ガス分析	5.227
Dダイマー	5.091
PT	4.043
ナトリウム及びクロール	3.722
観血的動脈圧測定	3.633
動脈血採取	3.333
APTT	3.31
乳酸	3.202
総蛋白	3.152
中心静脈注射	3.045
中心静脈圧測定	3
麻薬等の鎮痛	2.87
細菌培養同定検査	2.652
終末呼気炭酸ガス濃度測定	2.6
R - 吸気分布	2.433
細菌顕微鏡検査	2.364

医療指示名(検査項目)	Max(TO/S)
SARS-CoV-2	2.304
VWF抗原	2.232
肝炎ウイルス検査	2.103
尿一般	1.867
梅毒検査	1.646
心電図	1.591
CT撮影	1.467
PCT	1.46
BUN	1.414
クレアチニン	1.379
AST	1.353
鼻腔咽頭拭い液採取	1.342
KL6	0.962
BNP	0.933
インターロイキン	0.904
ABO血液型	0.888
HbA1c	0.873
FT4	0.768
TSH	0.768
TARC	0.745
LDLコレステロール	0.621
CK-MB	0.607
グロブリンクラス別ウイルス抗体	0.552
結核菌特異的インターフェロン-γ産生	0.517
肺サーファクタント蛋白	0.517
留置カテーテル	0.478
SiL2	0.455
フィブリノゲン	0.217

医療指示名(処置項目)	Max(TO/S)
ヘパリン	8
酸素吸入	7
ネキシウムカプセル	6.686
ランソプラゾール	5.773
プレドニゾロン	5.406
インスリン	4.667
ミヤBM細粒	4.5
メジコン	3.462
デキサメタゾン	3.427
ロゼレム	3.367
人工呼吸	2.636
コルヒチン	2.615
カルボシステイン	2.145
アセトアミノフェン	2.135
ベルソムラ	2.026
オルミエント	1.955
センノシド	1.921
ツムラ補中益気湯エキス顆粒	1.733
リスペリドン	1.072
ルネスタ	0.928
メトクロプラミド	0.754

Max(TO/S)で除外

Covid-19の治療と関係が薄いので除外

20 医療指示の除外(2/2)

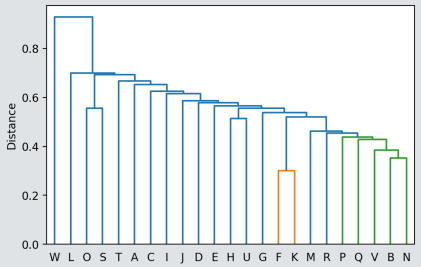
- 除外した検査項目には，**CRP・Dダイマー・末梢血液・単純撮影・経皮的血酸素飽和度測定・呼吸心拍監視・総ビリルビン・フェリチン**などCOVID-19の治療においてよく出現する検査項目が含まれている

- このような検査項目がSPMで抽出されても有用性は低い

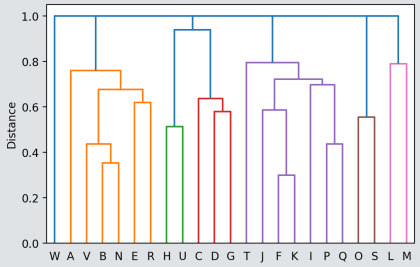
21 クラスタリング

- ウォード法・群平均法を採用
  - 3つ以上の医療機関でMSVを作成する提案手法を多くの医療機関に適用したい
    - 要素数が3以上のクラスタに含まれる医療機関が最多となる手法

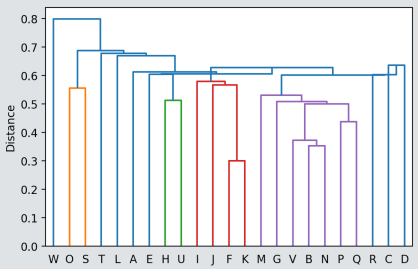
最短距離法



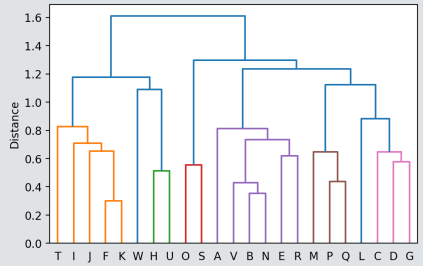
最長距離法



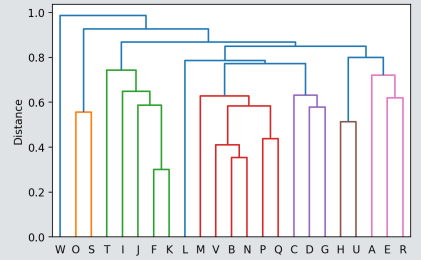
重心法



ウォード法

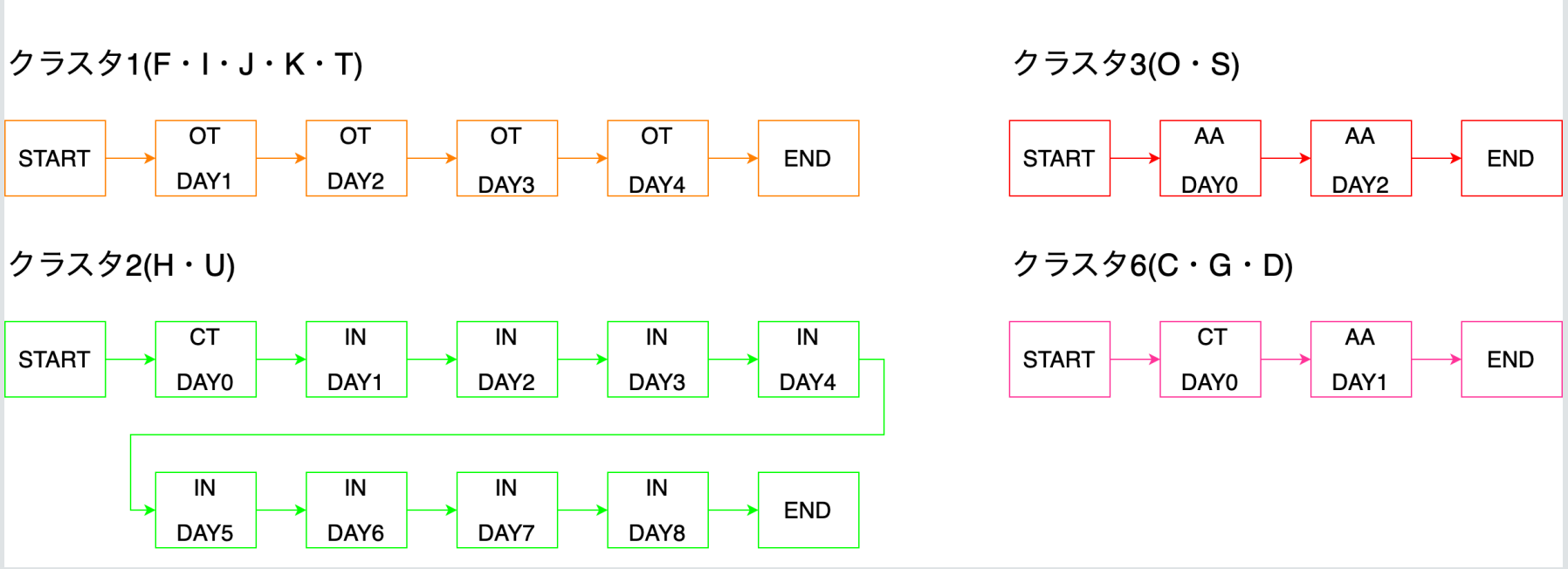


群平均法



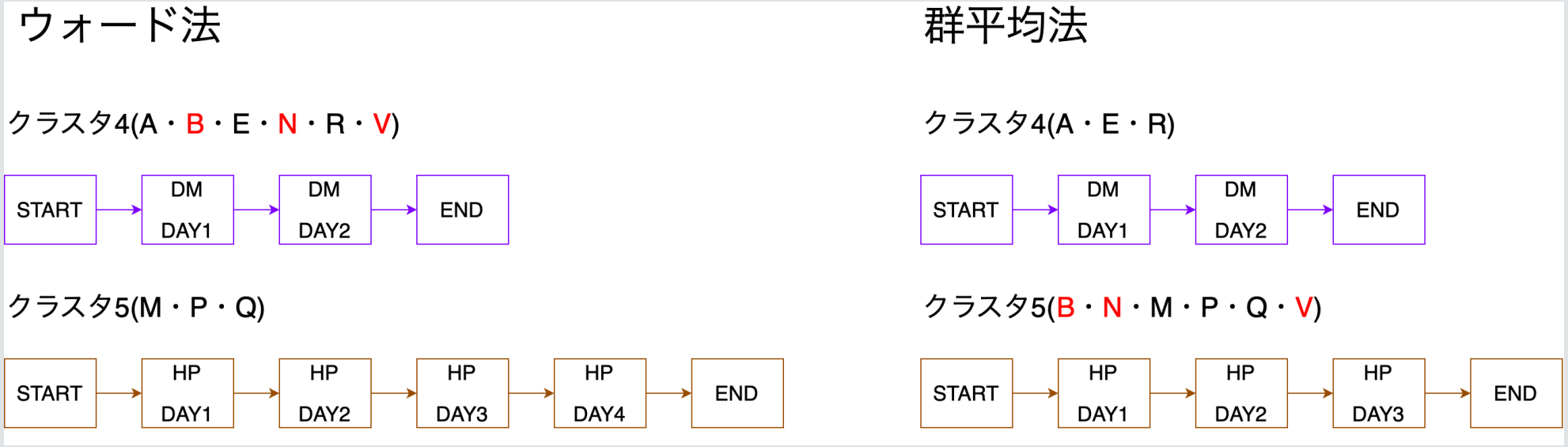
22 クラスタ内のLCSV（1/2）ウォード法・群平均法の共通

- クラスタ1は中等症の患者に酸素の投与
- クラスタ２は糖尿病の患者に糖尿病の治療薬
- クラスタ３と６は軽症者に解熱剤を投与



23 クラスタ内のLCSV（2/2）ウォード法・群平均法の差異

- クラスタ４は中等症患者に使用される炎症を抑えるステロイド剤のデキサメタゾンが共通
- クラスタ５は中等症患者に血栓症を予防する血液凝固防止剤のヘパリンが共通
- B, N, VのSVがデキサメタゾンとヘパリンの両方を含む

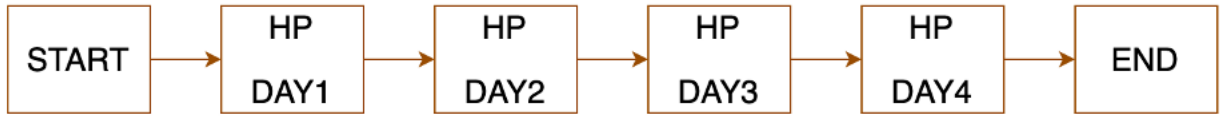


24 実験結果 共通点・差異点の可視化（1/3）

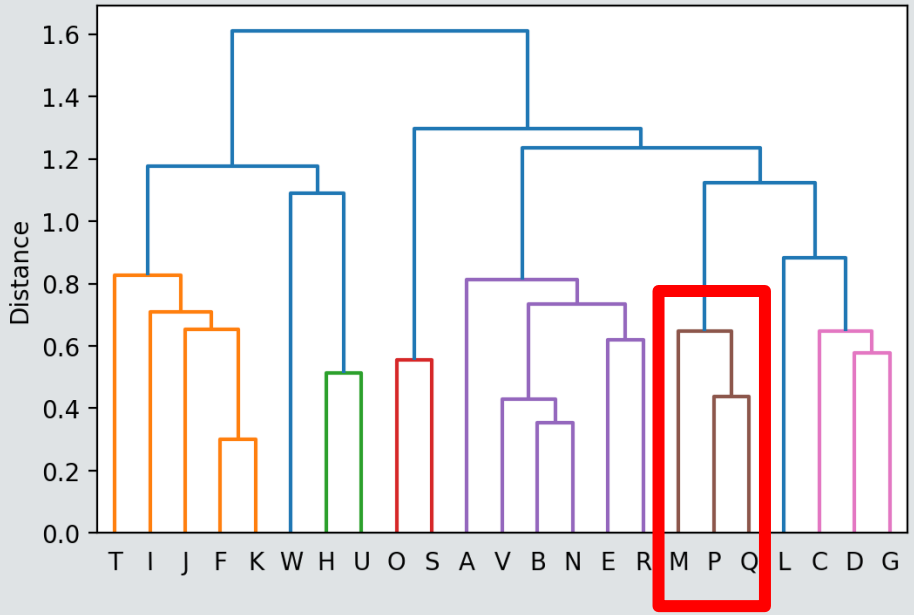
- ウォード法で得られたクラスタ5についてMSVを求めた

LCSV

クラスタ5(M・P・Q)

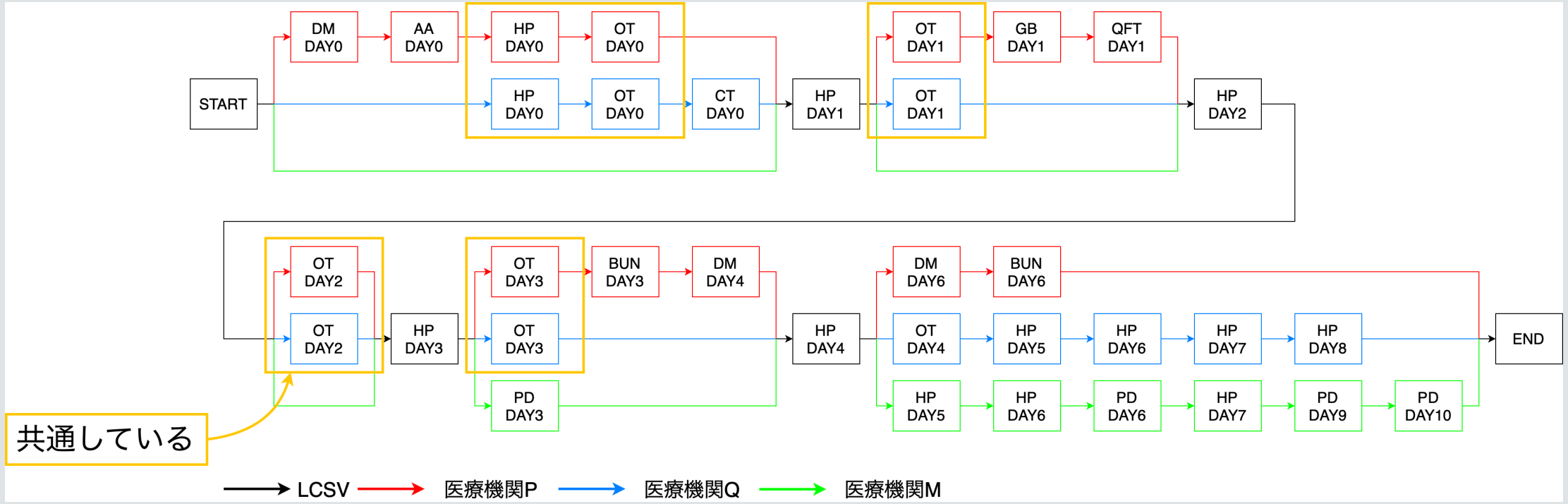


デンドログラム



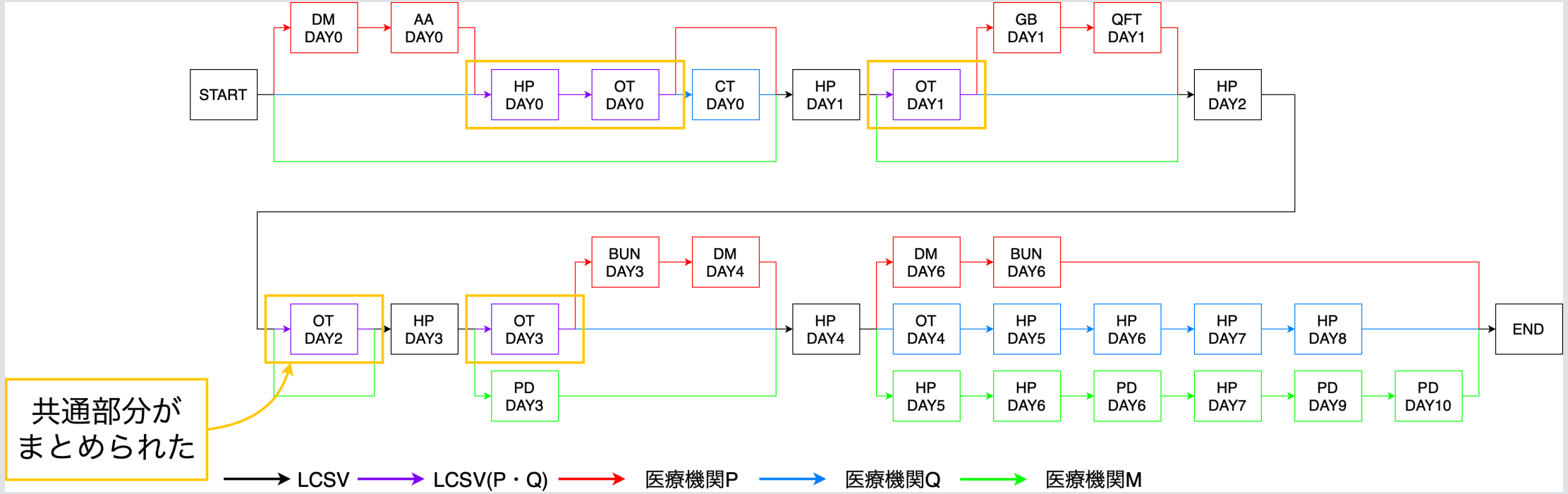
25 共通点・差異点の可視化（2/3）直接併合法

- 全体の共通部分しか表せない



26 共通点・差異点の可視化（3/3）距離ベース併合法

- 部分的な共通部分も表せる



27 考察

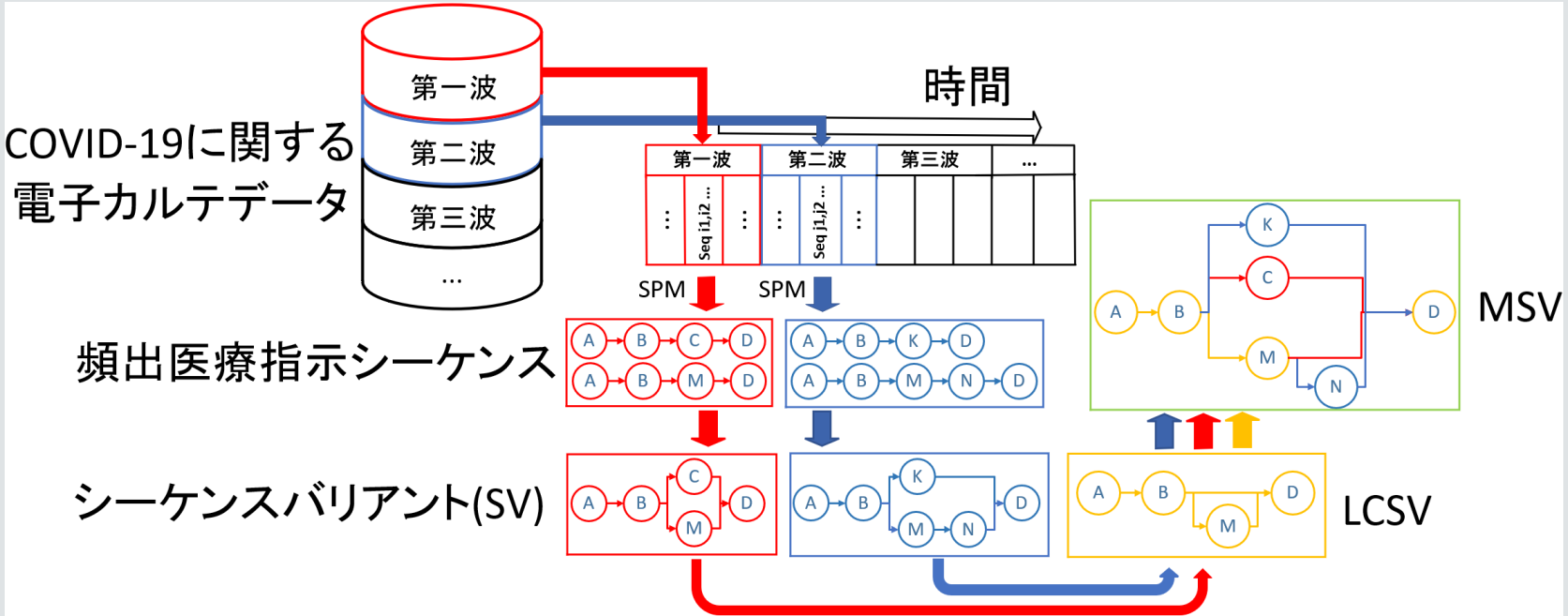
- 直接併合法の方が個々のSVがわかりやすい
- 距離ベース併合法の方が部分的な共通部分についても示すことができる
- 見やすさと冗長性どちらを優先するかによって選択可能

研究２:頻出医療指示パターンの違いの発生時期を発見する手法

スライディングウィンドウの導入

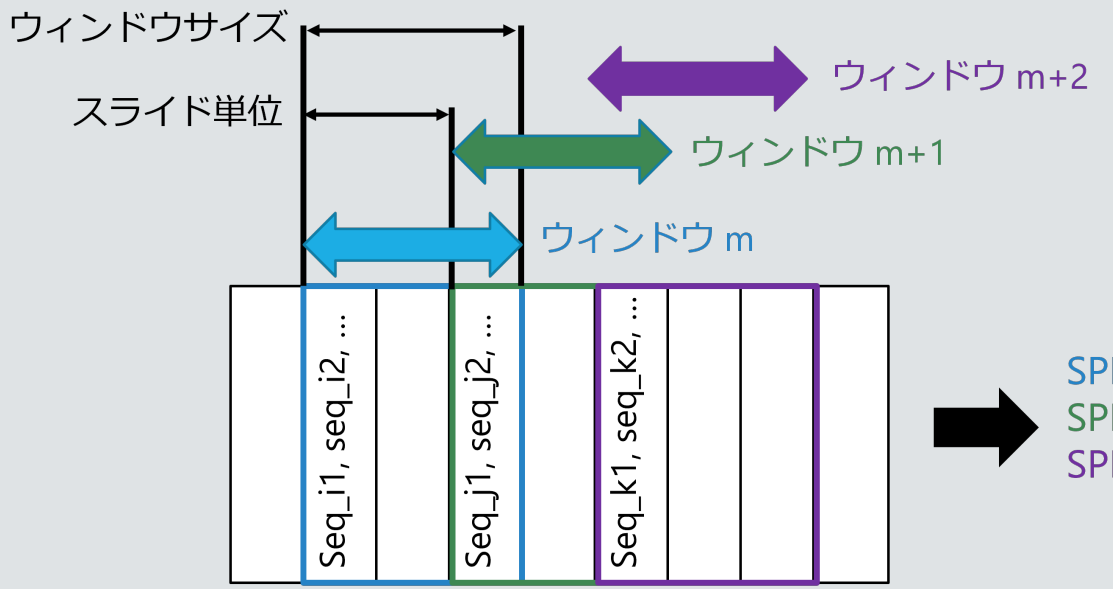
29 時期毎に頻出SVの抽出

- 時期毎の医療指示のシーケンスDBにSPMを適用し，頻出シーケンスを抽出し，分岐が発生したらSVに併合
- LCSVとMSVを隣接時期の頻出SV間に適用し違いを示す



30 違いの発生時期を示す

- スライディングウィンドウを導入しウィンドウ毎のデータにSPMを適用し，頻出パターンを抽出
  - ウィンドウをスライド単位移動させる



SPM
SPM
SPM

ウィンドウ間の類似度を算出

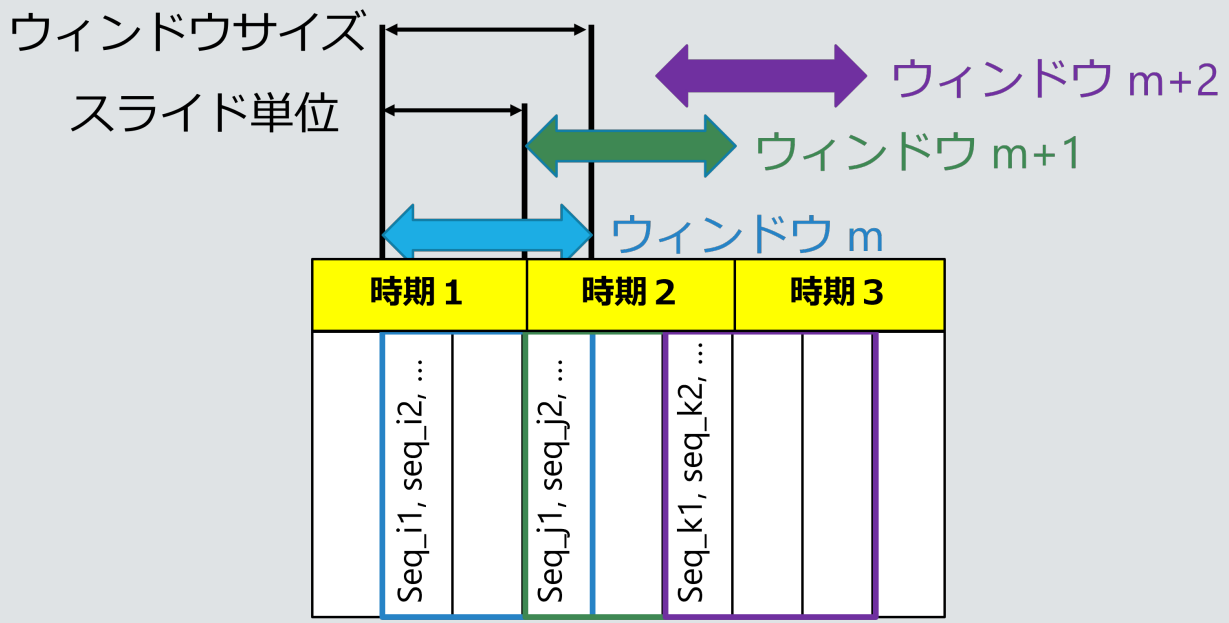
類似度が低い場合，LCSVとMSVを生成し，差異を可視化

要検討

・ウィンドウの種類
・ウィンドウサイズ
・ウィンドウ単位

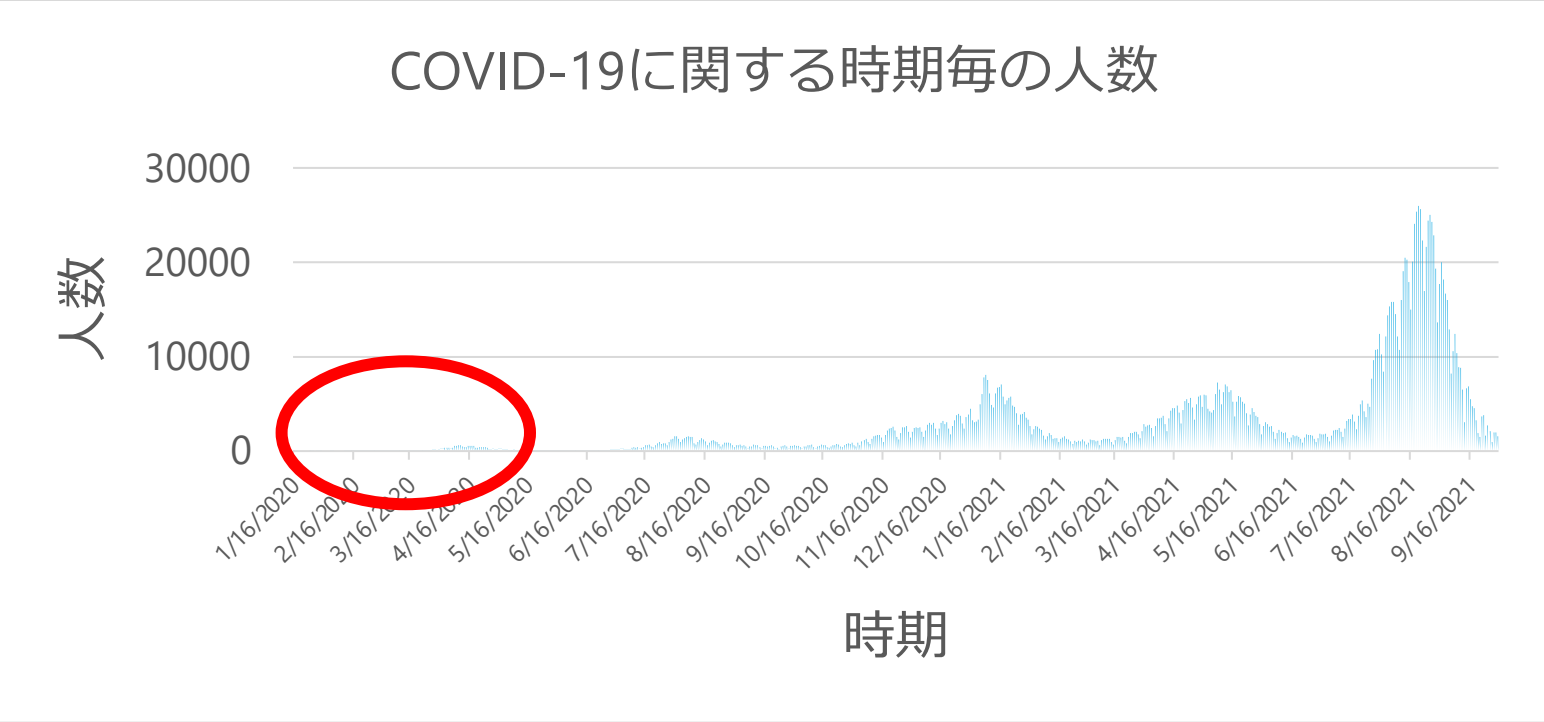
31 検討１:時間間隔ウィンドウ

- ある時期間のシーケンスデータにSPMを適用し**時間単位**でウィンドウをずらす
  - ウィンドウサイズとスライド単位を時間間隔で決める
    - 例：4週間と1週間



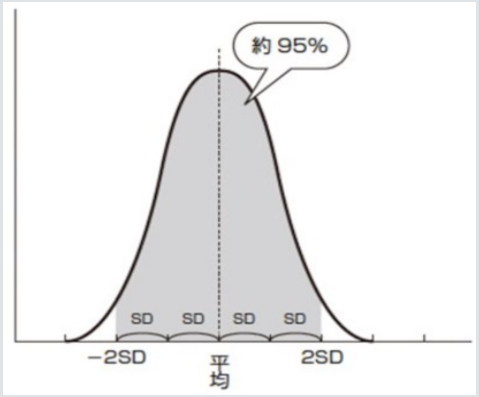
32 検討１:時間間隔ウィンドウの課題点

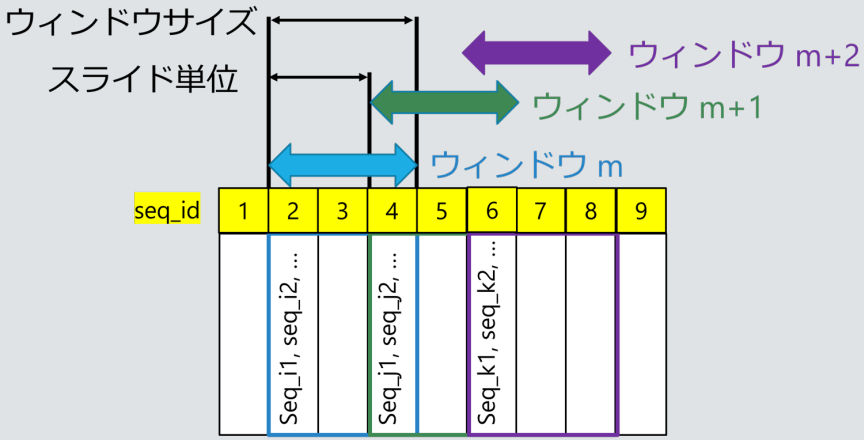
- シーケンスが少ない時期に対しSPMを適用しても抽出できない可能性あり



33 検討２:シーケンス数ウィンドウ

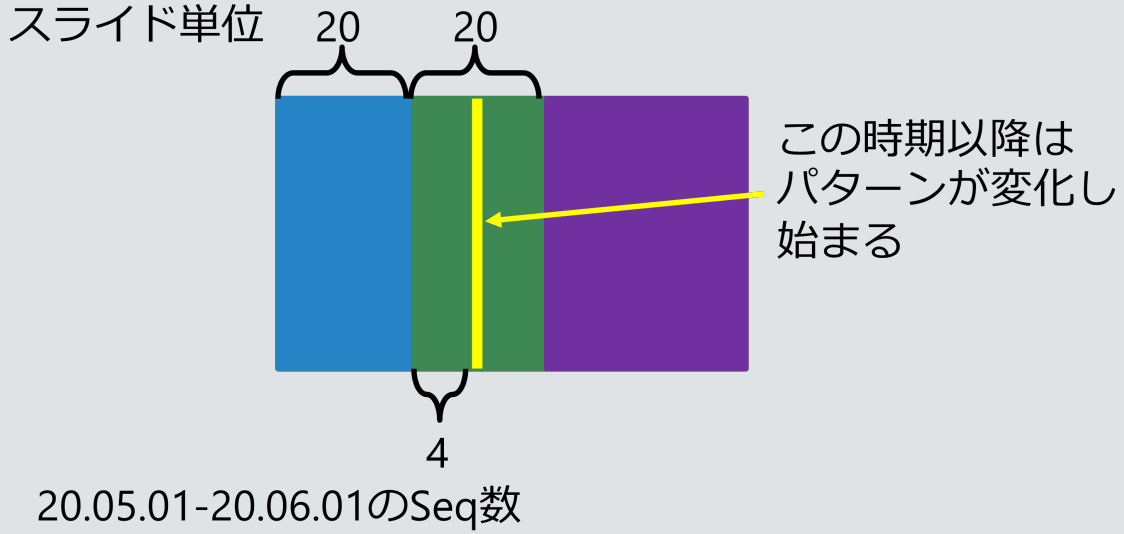
- シーケンスDBのシーケンスを時間順でソートし**シーケンス数単位**でウィンドウをずらす
  - ウィンドウサイズとスライド単位をシーケン数で決める
  - 時期的に特徴を持つ頻出パターンを抽出ように，ピークの周りのシーケンスを対象とする
    - ウィンドサイズを4SD
    - スライド単位を
      ウィンドウサイズの
      半分





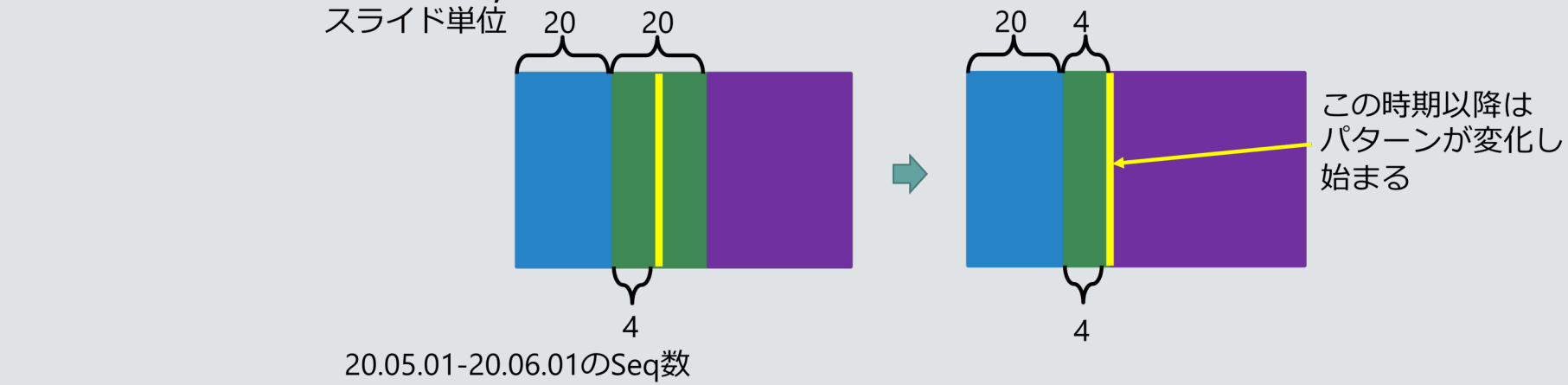
34 検討２:シーケンス数ウィンド課題点

- スライド単位が一定であれば，医療指示が切り替わるタイミングがスキップされる可能性あり
  - 時間的な頻出シーケンスが抽出できない
  - 後続き間のシーケンス数を考慮しスライド単位を動的に変更



35 検討３:時間・シーケンス数ウィンドウ

- シーケンス数ウィンドウをベースに**スライド単位を変更**
  - デフォルトスライド単位をウィンドウサイズの半分
  - ウィンドウ先頭から設定された時期範囲分のシーケンス数を調べる
    - シーケンス数がスライド単位より少ない場合，スライド単位をその時期のシーケンス数
    - そうでなければ，スライド単位をデフォルト値に



36 ウィンドウズ感のSVの類似度分析

- LCSVノード数の変化で類似度を評価
  - SVS (Sequence Variant Similarity)
    - $SVS(SV_1, SV_2) = \frac{2|LCSV|}{|SV_1|+|SV_2|}, SVS(SV_1, SV_2) \in [0,1]$
  - $|LCSV|$:LCSVのノード数
  - $|SV_1|$:SV1のノード数
  - $|SV_2|$:SV2のノード数

37 評価実験

- 「COVID-19感染の診療への影響調査と予測モデル開発」に関する研究の研究協力機関から提供されたデータ
- 医療機関Xに対するCOVID-19の第1波～第5波

COVID-19患者情報

患者数	329
最大医療指示数	2,319
最長在院日数	190

COVID-19の波毎の患者数とシーケンス数

	患者	Seq
第一波(2020.04.01-2020.06.30)	42	49
第二波(2020.07.01-2020.10.31)	52	63
第三波(2020.11.01-2021.02.28)	96	108
第四波(2021.03.01-2021.06.30)	89	99
第五波(2021.07.01-2021.09.30)	98	104

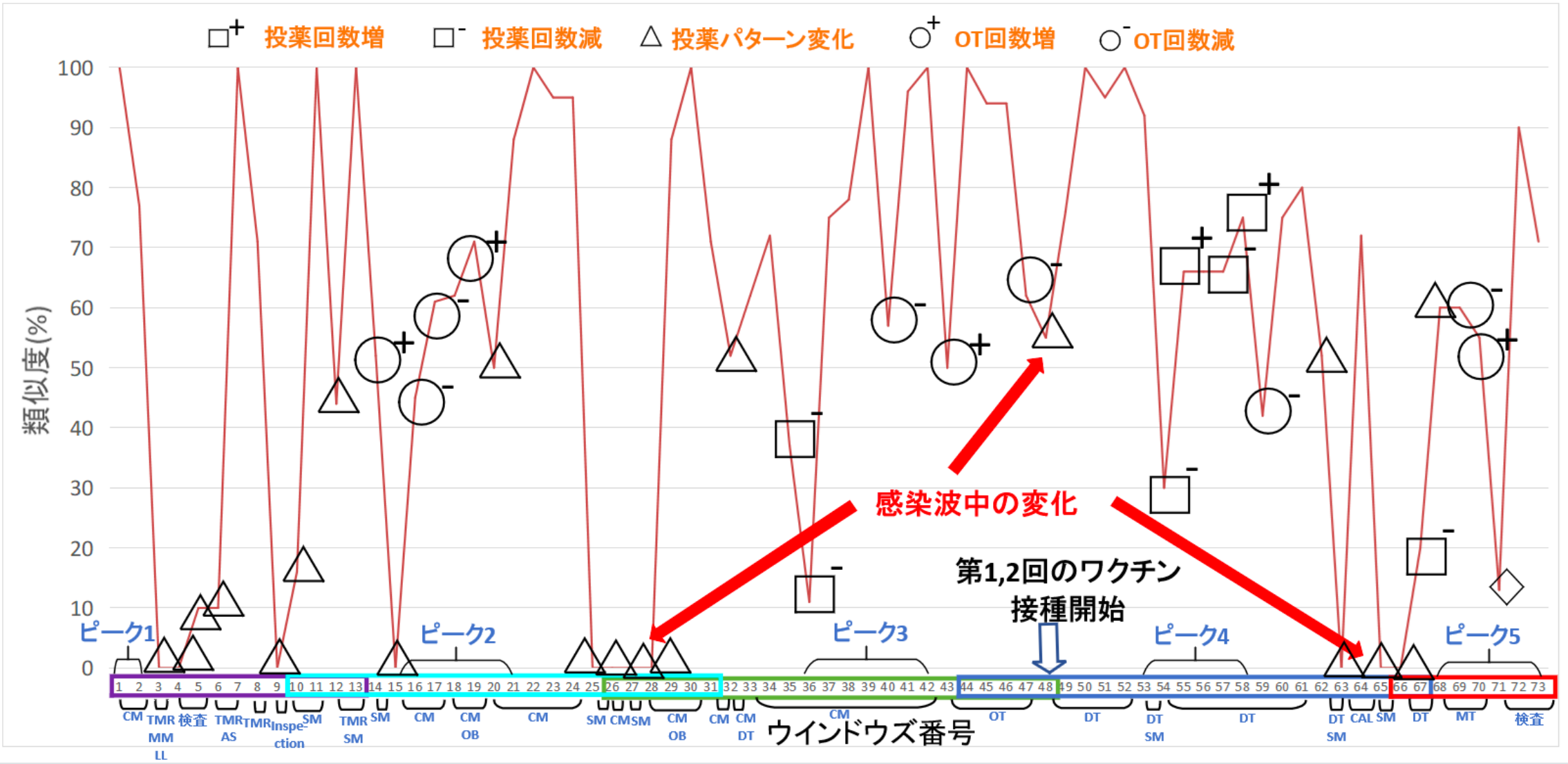
38 実験内容

- 第1波~第5波までの医療指示パターンの遷移状況を示す
- 適用したウィンドウ種類
  - 時間間隔ウィンドウ
  - シーケンス数ウィンドウ（時期的に特徴的な投薬を十分に抽出不可）
  - 時間・シーケンス数ウィンドウ

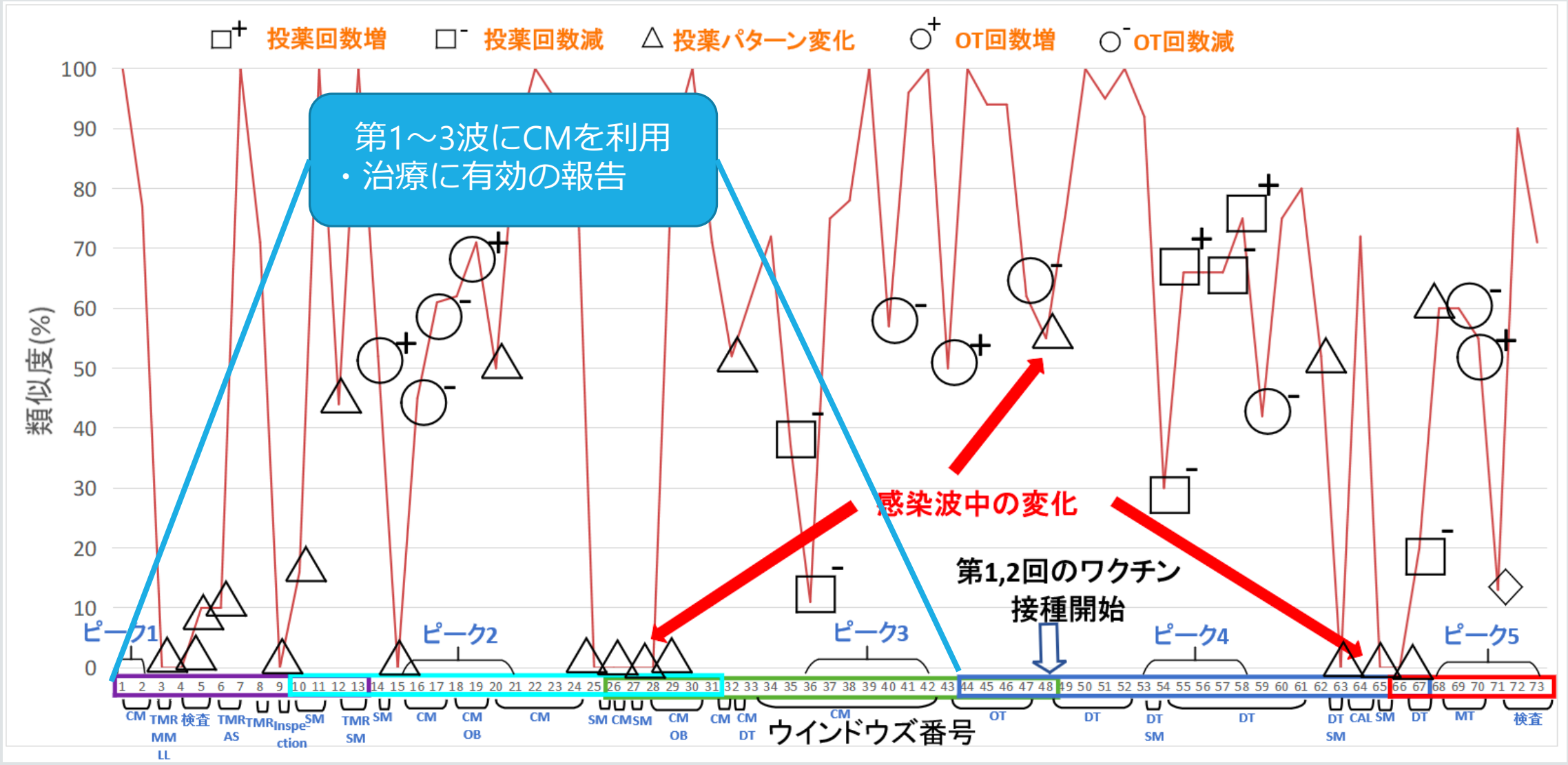
39 医療指示の略称

略号	医療指示名称	略号	医療指示名称
CM	カモスタットメシル酸塩錠	DT	デカドロン
TMR	ツムラ大建中湯エキス	OB	オルベスコ
MM	マグミット錠	LL	リリカOD錠
LS	ロキソプロフェンナトリウム	CAL	カロナール
ST	ダイフェン	MT	メジコン
AS	バイアスピリン	SM	酸化マグネシウム錠
OT	酸素吸入		

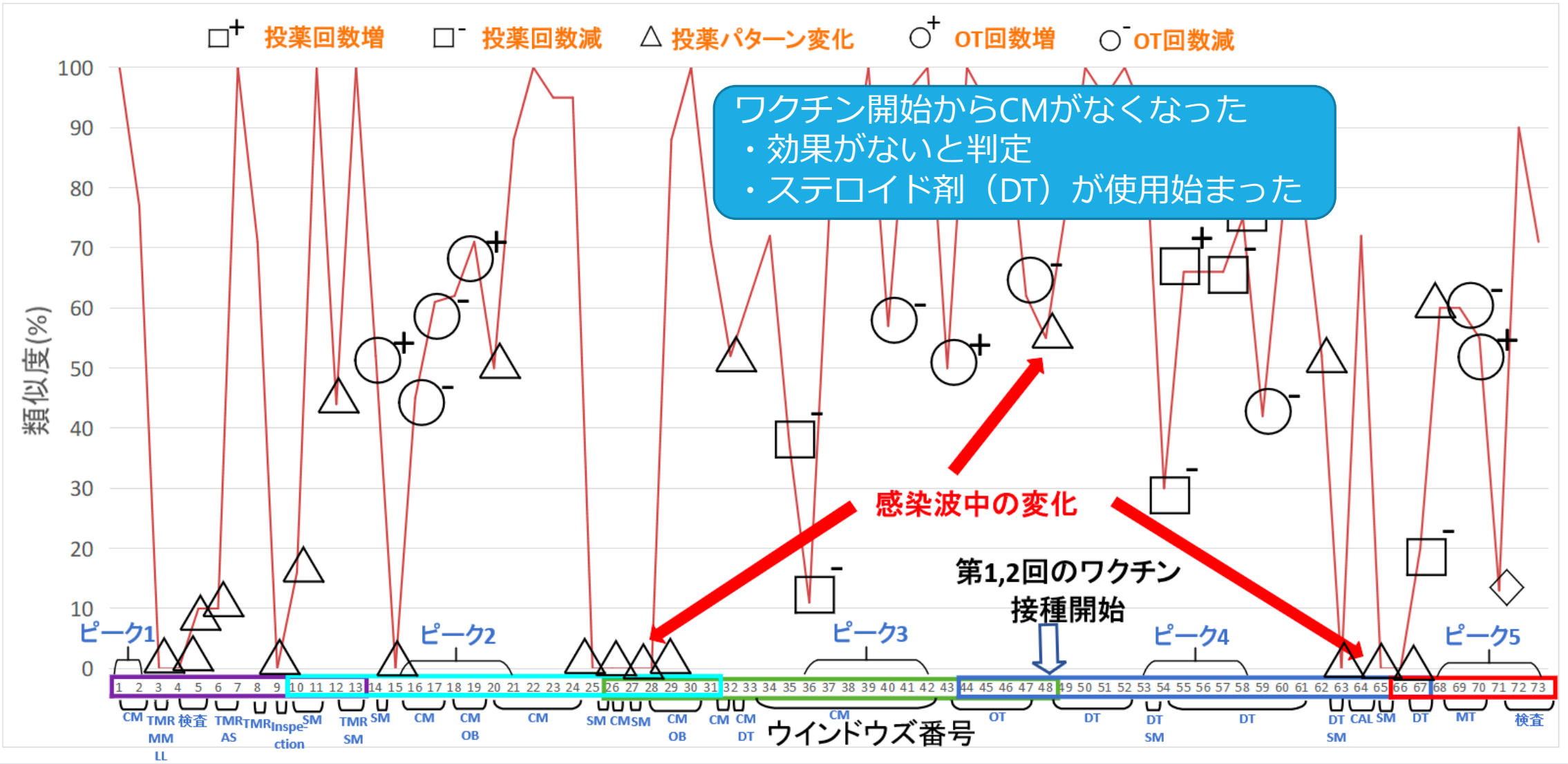
40 時間間隔ウィンドウ（1/4）



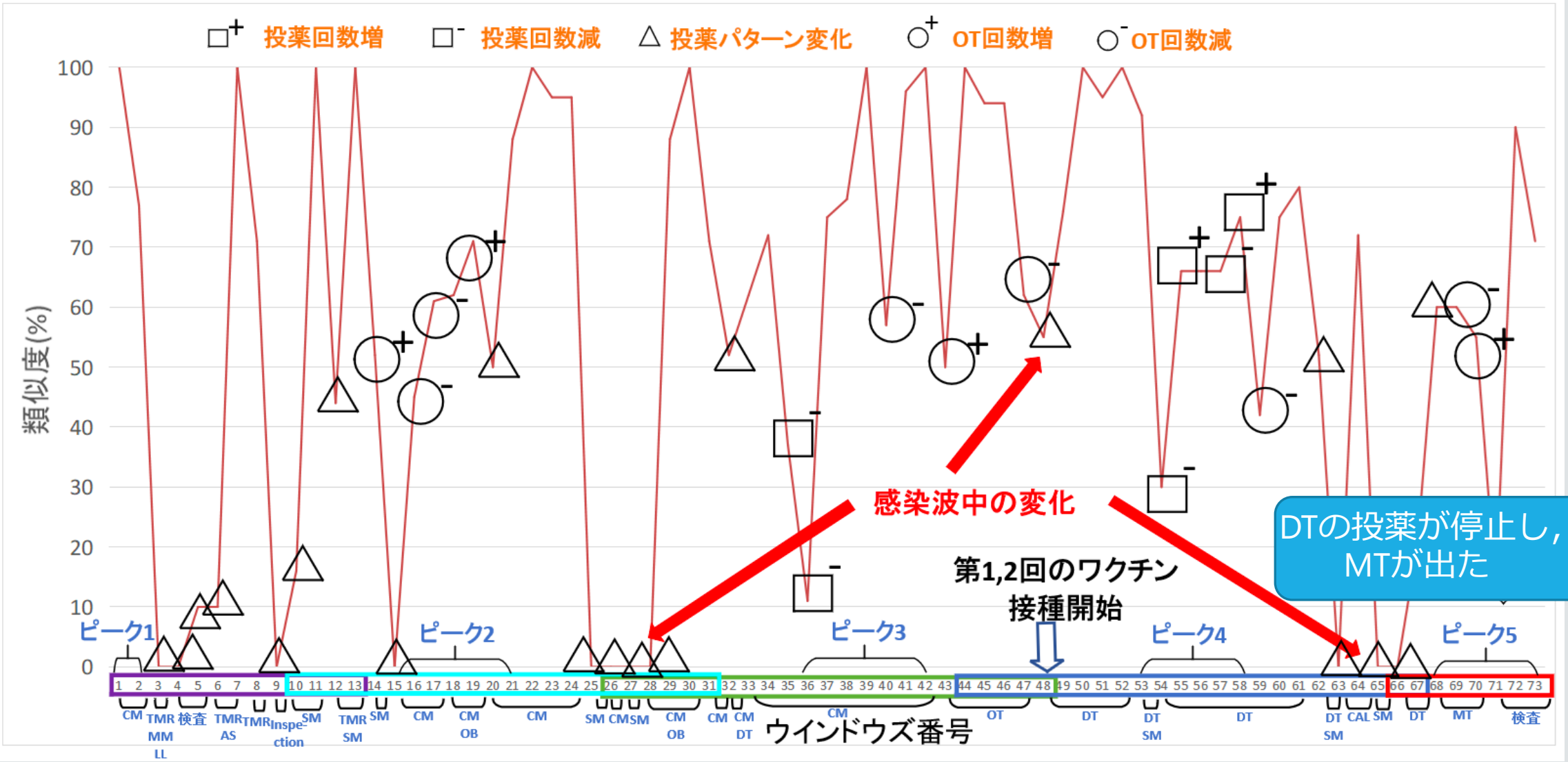
41 時間間隔ウィンドウ（2/4）



42 時間間隔ウィンドウ（3/4）

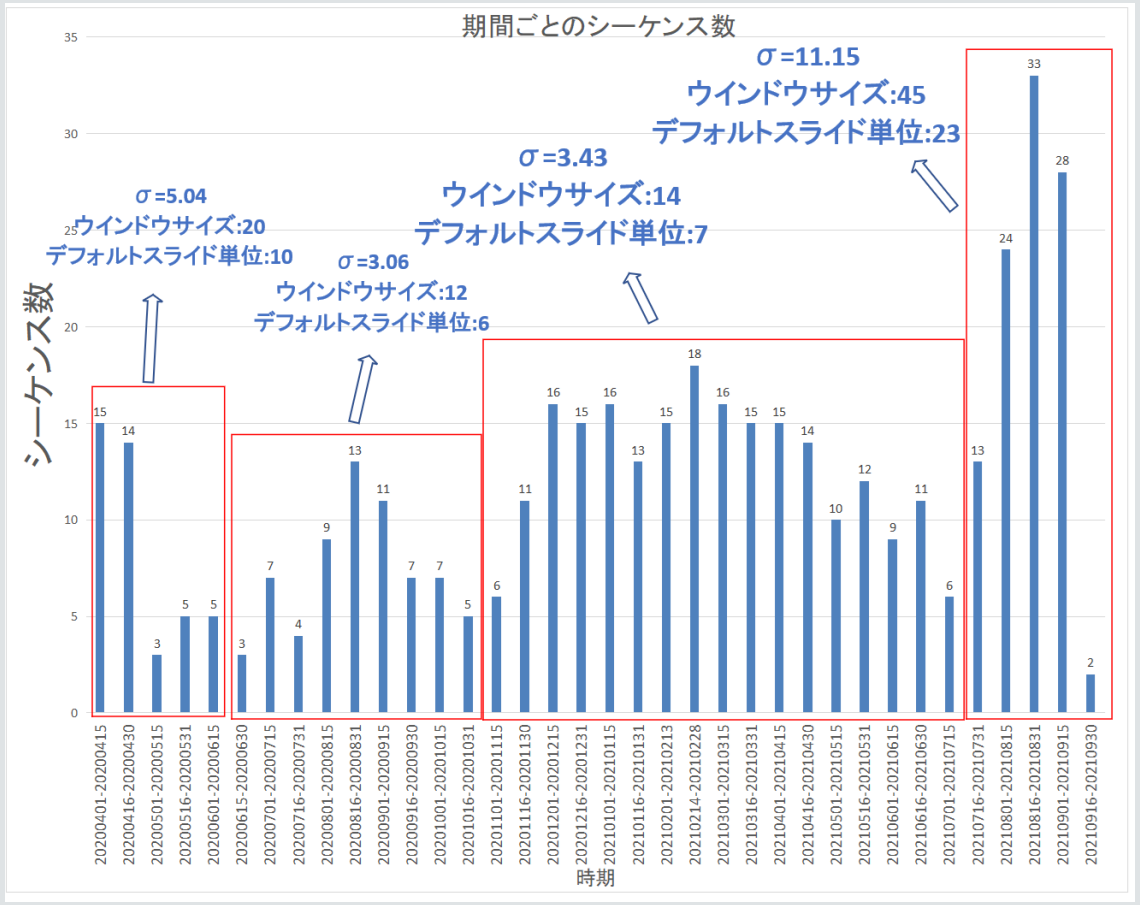


43 時間間隔ウィンドウ（4/4）

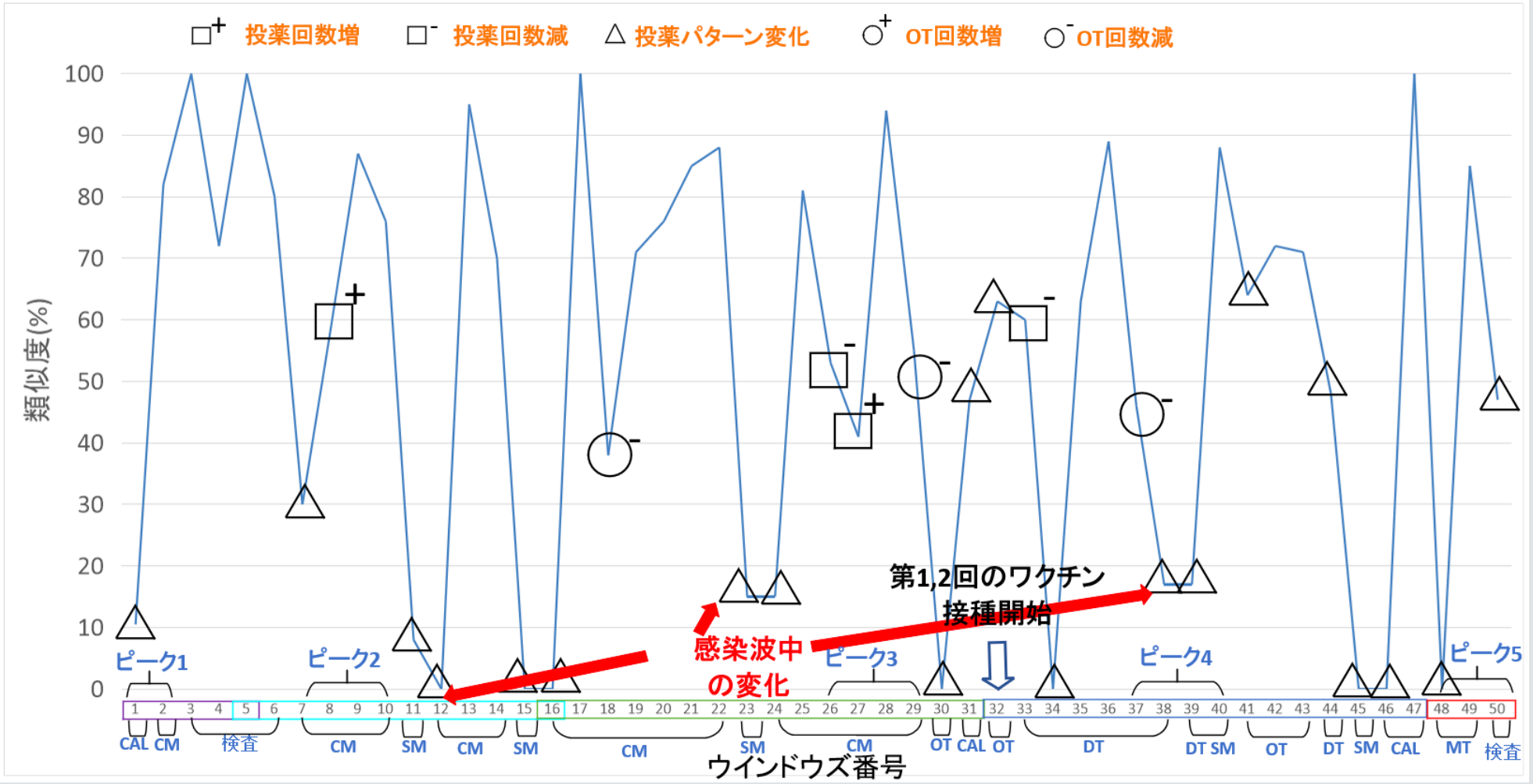


44 時間・シーケンス数ウィンドウ（1/3）

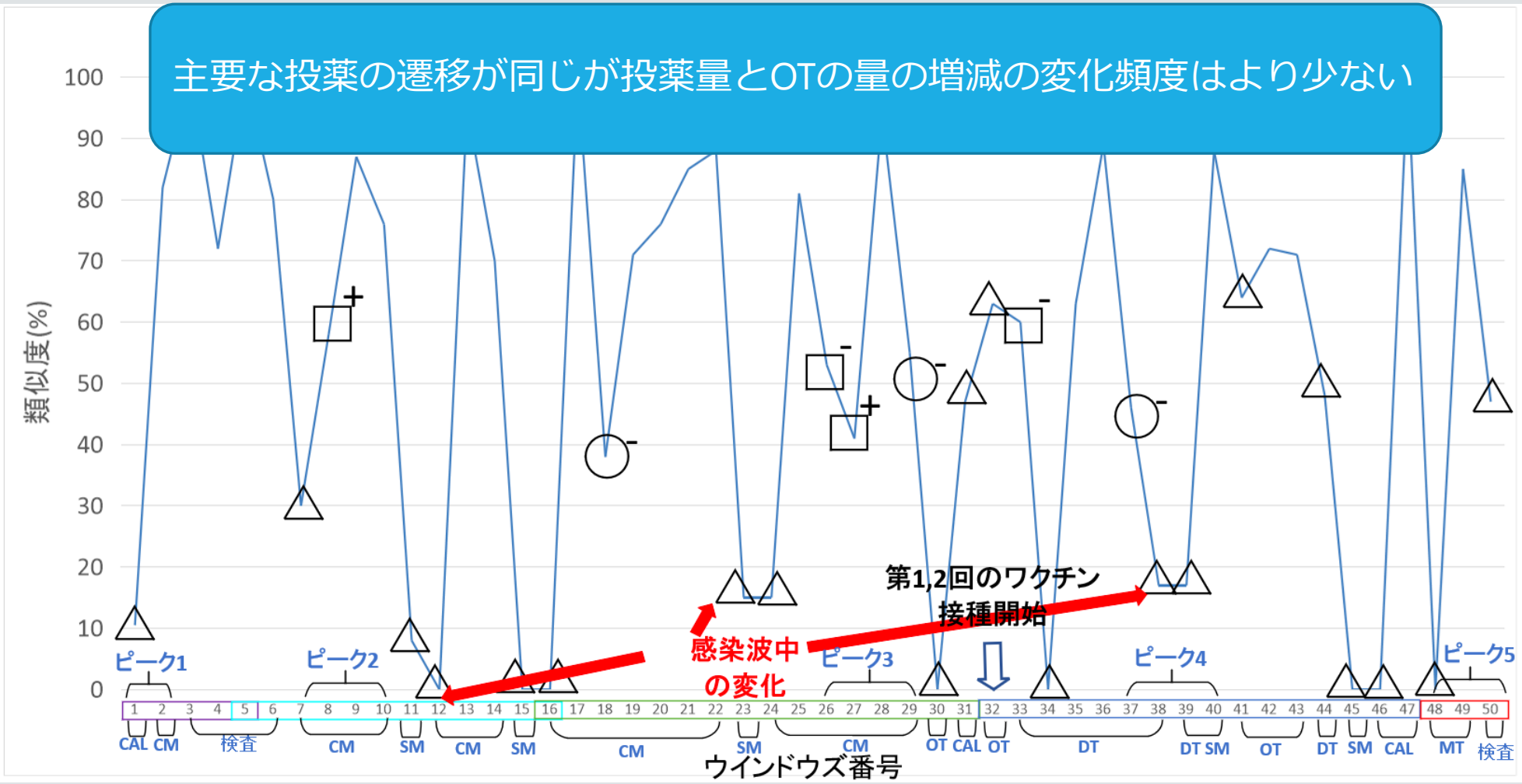
- ピーク前後のデータにより，ウインドウサイズとスライド単位を計算



45 時間・シーケンス数ウィンドウ（2/3）



46 時間・シーケンス数ウィンドウ（3/3）



47 異なるウィンドウ種類の違い

- 時間間隔ウィンドウ
  - パターンの投薬量がより頻繁に変更されて，より細かい投薬が出た
  - **時期毎の投薬量の変化と細かい投薬が把握できる**
- 時間・シーケンス数ウィドウ
  - **各時期の主要な投薬の使用状況の推移がより読み取りやすい**

48 まとめ

- クラスタリング等に基づいて**3つ以上の医療機関**におけるSV比較を可視化
  - 投薬等の処置項目を優先的に抽出
- 医療機関のSVにおいて**異なる時期の違い**を可視化
  - 医療指示パターンの遷移状況を時間間隔ウィンドウと時間・シーケンス数ウィンドウでしますことができた

49 関連外部発表

- **国際会議（査読付き）**
  - Analysis of Transitions in Differences between Frequent Medical-order Sequences for COVID-19 (IEEE CBMS'23), June 2023
- **国内会議（査読付き）**
  - COVID-19の電子カルテ履歴からの医療指示シーケンスパターン変化時期の抽出， 第27回日本医療情報学会春季学術大会予稿集，2023年6月
  - COVID-19の異なる医療機関と時期における頻出治療パターンの比較，第42回医療情報学連合大会，2022年11月
- **国内会議（査読なし）**
  - クラスタリングを用いた多病院間の頻出医療指示パターン比較, DEIM2023, 2023年3月
  - COVID-19に関する頻出医療指示パターンの時期による差異と差異発生時期の可視化，DEIM2023, 2023年3月
  - 動的に医療指示種類を変化したシーケンス間席における特徴的な治療パターン抽出，DEIM2023, 2023年3月