

Extracting systemic anticancer treatment lines from the Danish National Patient Registry for solid tumour patients treated in the North Denmark Region between 2009 and 2019

Supplementary material

Mapping SKS procedure codes to ATC

The procedure codes for treatment used in this study were mapped to the Anatomical Therapeutic Chemical (ATC) classification (see Supplementary Figure 1). SKS codes that were not directly related to one or a set of ATC codes were considered generic. For generic SKS codes (for example “BOHJ19B”), subcodes (for example “BOHJ19B1” and “BOHJ19B2”) were used to infer a generic ATC code (in this example “L01FG”). Subcodes were considered directly related to their parent code (for example “BOHJ19B1” is directly related to “BOHJ19B” and “BOHJ19”, see Supplementary Figure 1). The SKS codes were kept to group registrations into treatment lines as detailed in the main document.

Supplementary tables

Supplementary Table 1 - Primary tumour location grouping, inclusion status, and exclusion criteria.

ICD10	Primary tumour location	Specific group	Reason
C00-C14, C30-C33	Head and neck	No	Relatively rare and not primarily treated with systemic anticancer treatment
C15-C16	Gastro-oesophageal	Yes	
C17	Intestine	No	Rare
C18-C20	Colorectal	Yes	
C21	Anal	No	Rare
C22-C24	Hepato-biliary	No	Referred to other hospitals
C25	Pancreatic	Yes	
C26, C39, C55, C57, C76, C80	Ill-defined	No	Ill-defined
C34	Lung	Yes	
C37-C38	Thoracic other than lung	No	Rare
C40-C41	Bone and articular cartilage	No	Rare
C43	Melanoma	No	Referred to other hospitals
C44	Skin other than melanoma	No	Not treated with systemic anticancer treatment
C45-C49	Connective and soft tissue	No	Rare
C50	Breast	Yes	
C51-C52	Vulva and vagina	No	Rare
C53	Cervical	No	Referred to other hospitals
C54	Endometrial	Yes	
C56	Ovarian	Yes	
C58	Placenta	No	Rare
C61	Prostate	Yes	
C64-C68	Urinary	Yes	
C69-C70	Eye and meninges	No	Rare
C71	Brain	Yes	
C72	Nervous system	No	Rare
C73-C75	Endocrine	No	Rare
C77-C79	Secondary	No	Not primary

Supplementary Table 2 – Thresholds used according to the SKS codes of two consecutive prescriptions.

Prescription 2 Prescription 1	Same spec. SKS code (e.g. BWHA109)	Rel. gen. SKS code (e.g. BWHA1)	Diff. spec. SKS code (e.g. BOHJ19J3)	Diff. gen. SKS code (e.g. BOHJ1)
Spec. SKS code (e.g. BWHA109)	Same drug	Same drug class	Different drugs	Different drug class
Gen. SKS code (e.g. BWHA1)	Same drug class	Same drug class	Different drug classes	Different drug classes

Spec.: Specific; Gen.: Generic; Rel.: Related; Diff.: Different

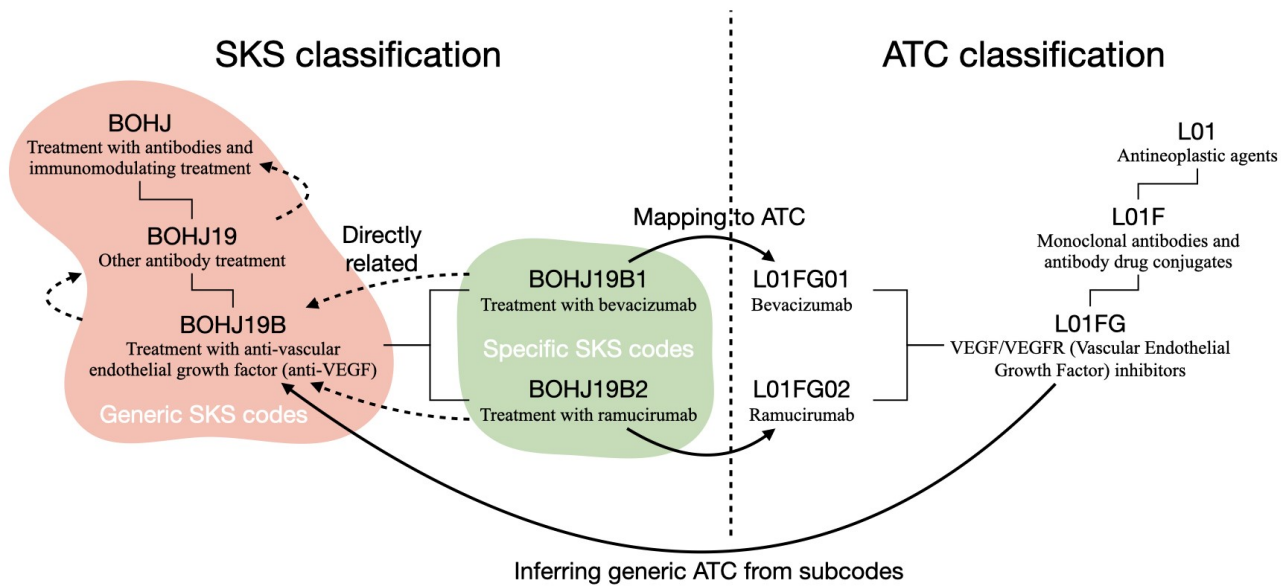
Supplementary Table 3 – Thresholds, systemic anticancer treatment lines or patients, and accuracy in terms of sensitivity, PPV and F1 score for the full dataset for the 3 matching types.

Method	TT	MedOnc	DNPR	Matching	Sensitivity	PPV	F1 score
Stringent matching							
Time-based	49	15,433	13,912	6,473	0.419	0.465	0.441
Mixed	33	15,433	12,942	6,400	0.415	0.495	0.451
Grid search	(33, 175)	15,433	13,452	6,727	0.436	0.500	0.466
Loose matching							
Time-based	45	15,433	14,231	12,134	0.786	0.853	0.818
Mixed	29	15,433	13,272	11,873	0.769	0.895	0.827
Grid search	(29, 175)	15,433	13,782	12,240	0.793	0.888	0.838
Line number matching							
Time-based	45	8,738	8,867	6,299	0.721	0.710	0.716
Mixed	31	8,738	8,867	6,170	0.706	0.696	0.701
Grid search	(31, 65)	8,738	8,867	6,395	0.732	0.721	0.726

TT: Time threshold(s) in days between two consecutive drug prescriptions. The values in parentheses for thresholds for the grid search are the “different drugs” and “same drug” thresholds in that order. The MedOnc, DNPR, and Matching columns show the number of lines found in MedOnc, DNPR, and matching in both, respectively. For the line number matching, the numbers shown are for the number of patients.

Supplementary figures

Supplementary Figure 1 – Example of mapping from SKS codes to ATC codes.



In this example, the focus is on treatment with anti-vascular endothelial growth factor (SKS code BOHJ19B) and the mapping of the specific codes BOHJ19B1 and BOHJ19B2 to ATC. Direct relation is illustrated by a dashed arrow. The green area includes the specific codes while the red area includes the generic codes.

Supplementary Figure 2 – Pseudocode for the 4 different methods used.

Let p_1 and p_2 be two consecutive drug prescriptions, at times t_1 and t_2 and for drugs with SKS code c_1 and c_2 , respectively. The interval between p_1 and p_2 is noted $\Delta t = t_2 - t_1$. $threshold$ is the time threshold in days for each method. For the grid search method, two time thresholds are used, $threshold_{same}$ and $threshold_{diff}$.

Time-based method:

```

if  $\Delta t \leq threshold$  then
   $p_1$  and  $p_2$  belong to the same line
else
   $p_1$  and  $p_2$  belong to different lines
  
```

Drug-based method:

```

if  $c_1$  is generic or  $c_2$  is generic then
  if  $\Delta t \leq threshold$  then
     $p_1$  and  $p_2$  belong to the same line
  else
     $p_1$  and  $p_2$  belong to different lines
else
  if  $c_1 = c_2$  then
     $p_1$  and  $p_2$  belong to the same line
  else
     $p_1$  and  $p_2$  belong to different lines
  
```

Mixed method:

```

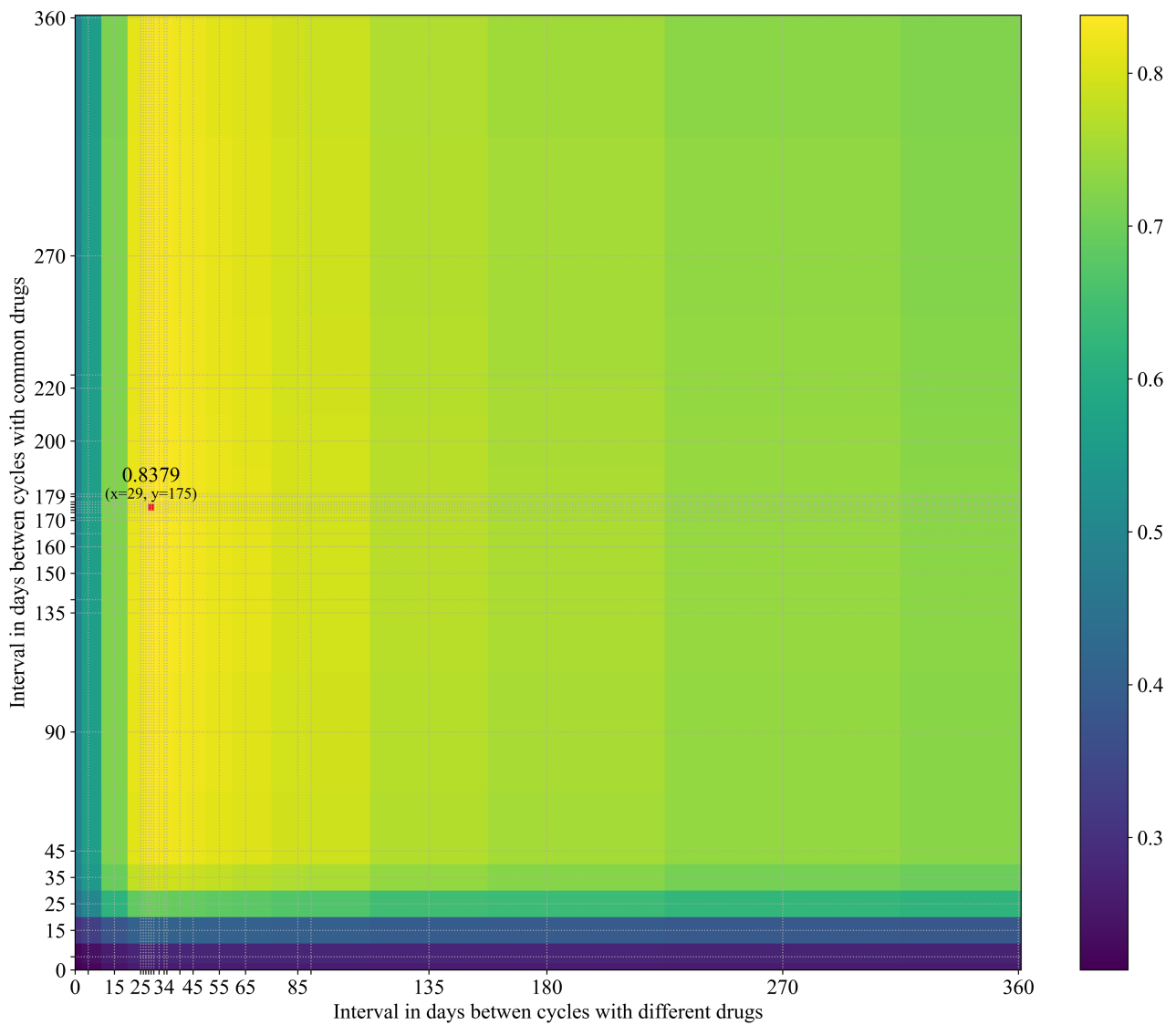
if  $c_1$  related to  $c_2$  or  $c_1 = c_2$  then
  if  $\Delta t \leq threshold$  then
     $p_1$  and  $p_2$  belong to the same line
  else
     $p_1$  and  $p_2$  belong to different lines
else
   $p_1$  and  $p_2$  belong to different lines
  
```

Grid search method:

```

if  $c_1$  related to  $c_2$  or  $c_1 = c_2$  then
  if  $\Delta t \leq threshold_{same}$  then
     $p_1$  and  $p_2$  belong to the same line
  else
     $p_1$  and  $p_2$  belong to different lines
else
  if  $\Delta t \leq threshold_{diff}$  then
     $p_1$  and  $p_2$  belong to the same line
  else
     $p_1$  and  $p_2$  belong to different lines
  
```

Supplementary Figure 3 – Grid search on F_1 score for the loose matching method.



The red dot corresponds to the maximum F_1 score, 0.8379 in that case. The corresponding coordinates are indicated, x being the “different drugs” threshold in days and y the “same drug” threshold.