Paper RW06

Deriving Line of Therapy: an algorithmic approach in a real-world oncology database

Matthew Thompson, IQVIA, Leeds, United Kingdom

ABSTRACT

In oncology, line of therapy is used to describe the regimen(s) of chemotherapy a patient receives to treat their cancer. A change in line of therapy is usually initiated when treatment fails to control the disease, is not tolerated by the patient or at the point of disease relapse. Not all hospital prescribing systems collect this data in an easily identifiable and extractable form. This paper describes the development of a SAS algorithm to determine changes in line of therapy using chemotherapy cycles data. The algorithm compares a single cycle of treatment with the previous and next cycle and generates flags to indicate changes in the drugs administered and gaps in between treatment cycles that differ from expectations. The algorithm then uses these indicators to assign a line of therapy to the patient's data, understanding that some patient's treatment profiles will be too complicated to assign via a logical, rule-based approach.

INTRODUCTION

Line of therapy is used to describe the regimen(s) of chemotherapy a patient receives to treat their cancer. The ability to identify the line of therapy for a cancer patient has great value in the clinical world. Based on experience of working with healthcare professionals in national healthcare services, there is a constant desire to assess how patients respond to treatment at various lines of therapy. Additionally, the pharmaceutical industry develops drugs that target patients at a particular line of therapy. Therefore, if the line of therapy for cancer patients can be identified, it will help determine information such as:

a) Numbers of patients who progress or are treated per line of therapy
b) Patient endpoints such as overall and progression-free survival per line of therapy
c) Identification of meaningful patient cohorts for real-world studies

A change in line of therapy is usually initiated when treatment fails to control the disease, is not tolerated by the patient or at the point of disease relapse/progression. These changes in line of therapy can be difficult to identify in a real-world database; management of cancer patients is often complex, individualized to the patient, and treatments may be changed to improve tolerability, even in the face of treatment responses.

In a hospital EMR, line of therapy will be captured in patient notes or clinical letters, but not always in a structured, easily extractable form. When working with retrospective data, centres may not have the resource to populate database fields for this information. Therefore, other indicators of change in line of therapy are required, derived using an algorithm approach.

Having worked with a hospital chemotherapy database and found that line of therapy, treatment response and dates of progression were not well captured in a structured or easily extractable form, another method to derive line of therapy via use of a SAS algorithm has been developed.

To validate the accuracy of the algorithm, it has been applied to a cohort of 226 urothelial cancer patients (cancers of bladder, ureter, renal pelvis and other, unspecified urinary organs) who have had their data clinically enhanced by a senior oncologist, reviewing patient the records and populating the database with this information, and the line of therapy derived using the algorithm has been compared with the clinically assessed line of therapy.

DATA SOURCE

The Leeds Cancer Centre (LCC) serves a metropolitan catchment area of over 850,000 people for secondary care at Leeds Teaching Hospitals NHS Trust (LTHT). ChemoCare is an expert chemotherapy electronic prescribing system[1] and is the LTHT hospital prescribing system in use since 2004.

Real-world Evidence Alliance Leeds (REAL) is a collaboration between LCC, Leeds Institute for Data Analytics (University of Leeds) and IQVIA Real World Insights. REAL accesses continually updated patient data stored in PPM, including patient demographics, cancer diagnoses, tumour staging and anti-cancer therapy. REAL studies are conducted on-site within LTHT under the strict legal framework governing access to and use of personal information

in the NHS. The work was completed with UK Heath Research Authority formal approval; the need for ethics approval was waived.

METHODS

DATA AVAILABILITY

The chemotherapy data in the database is imported from the ChemoCare system and structured in three tables: regimens, cycles and drugs. Data from all three tables are linked as part of the data preparation prior to running the algorithm.

The chemotherapy regimens data includes complete data on the regimen/treatment label and regimen start date, which is the date of the start of the first cycle for that regimen. Regimen end dates are not captured; this can be imputed by taking the start date of the last cycles in a regimen and adding the prescribed cycle duration. Intent of treatment (neoadjuvant, adjuvant, palliative etc.) is an available field, but this data isn't always well captured in structured form. Additionally, a single regimen may have multiple drugs and different dosing schedules. Repeated admissions of a regimen may constitute the same line of therapy for a patient who has undergone dose modifications. This can be a challenge when determining whether an apparent switch in a regimen is a genuine change in line of therapy.

The cycles data table details each cycle of chemotherapy, links it to the regimen via a unique identifier, and includes complete data on the cycle start date and the maximum number of days each cycle is treated for. This isn't necessarily a cycle duration but is the protocolled number of days the cycle is prescribed for. This can be used to infer a cycle end date. The number of cycles administered can be derived.

The drugs data can be linked back to the cycles and regimens data via unique identifiers, and details dosage information and drug administration dates. The regimen is broken down into the specific drugs that make up the regimen and this is used to help standardize the inconsistencies in the regimen labelling.

DATA CLEANING

According to senior clinicians, patients who have a single solid tumour typically have less complex chemotherapy regimens than, for example, haematological malignancies. Similarly, if patients have multiple malignancies it can be hard to determine which treatment is for which cancer, making identifying the line of therapy very challenging. For this reason, the developed approach focuses on those patients with a single, solid tumour. Any patients who have multiple primary cancers are excluded during the cohort selection.

Following the selection of the patient cohort, steps are taken to clean the data and make it more amenable to an algorithm. To more easily determine changes in regimen, inconsistencies in regimen labelling needed to be addressed. To do this, mapping tables to standardize regimens to SACT regimens were used where possible and combined with drug labels[2].

Database Regimen Label	SACT Standardised Regimen Label	drug1	drug2	drug3
CARBO/PAC 1W L3	Carboplatin + Paclitaxel (3 weekly)	CARBOPLATIN	PACLITAXEL	
PAC L3 ALL/CARBO	Carboplatin + Paclitaxel (3 weekly)	CARBOPLATIN	PACLITAXEL	
PACLITAX/CARBO	Carboplatin + Paclitaxel (3 weekly)	CARBOPLATIN	PACLITAXEL	
CARBO/PAC 1W	Carboplatin + Paclitaxel (weekly)	CARBOPLATIN	PACLITAXEL	
CARBO/PACLITAX1W	Carboplatin + Paclitaxel (weekly)	CARBOPLATIN	PACLITAXEL	

Drugs that are not considered chemotherapy (e.g. growth colony stimulating factors) are excluded at this point, as well as other regimens in the data that are not relevant to the line of therapy derivation. These are identified by clinicians on review of the data and excluded manually in the code.

Finally, treatment regimens starting on the same date are combined into a single regimen. The contents of dataset that is processed by the algorithm are detailed in the appendices at the end of this paper.

LINE OF THERAPY RULES

The algorithm compares the start date, regimen and drugs of a cycle with the previous and next cycles to identify indicators of increments in line of therapy based on the following rules, defined with the guidance of senior oncologists:

- If the patient progresses, where a progression date is available, the next cycle after the date of progression is the start of a new line of therapy

- If there is a gap between cycles of a pre-determined number of days (usually 70 days + the maximum duration of the cycle), the next cycle is a new line of therapy.
- If the regimen is changed and the drugs in the new regimen are entirely different to those in the previous regimen, this is a new line of therapy.
- If the regimen is changed, but all drugs given as part of the new regimen are contained in the regimen given as part of the previous cycle, this is not considered a change in line of therapy.
- If drugs are added, as part of a change in regimen, this is considered a new line of therapy.
- If only one cycle of a regimen is given before switching to a new regimen, and that switch occurs within 60 days of the start of the previous cycle, this isn't considered a new line of therapy

In determining gaps in treatment, a "reference date" is calculated, based on an imputed cycle end date. The cycle end date is imputed as the cycle start date, or last drug administration date if a cycle has multiple drug administrations, plus the cycle duration. The cycle duration is the maximum number of days a cycle of chemotherapy is prescribed for e.g. a regimen with a cycle duration of 21 days means a planned gap of 21 days between cycle start dates. A time window of 70 days is then added to the imputed cycle end date as the reference date to used to identify gaps in treatment.

The gap of 70 days has been chosen, following discussion with senior oncologists, as this would allow for a gap of approximately three months between cycle start dates when combined with a typical cycle duration of around 21-28 days. Using the maximum prescribed cycle duration allows for treatment such as maintenance therapy which is often prescribed in cycles three months apart.

VALIDATION

Clinical data was enhanced for a cohort of urothelial cancer patients. Enhancement was performed by a senior oncologist reviewing patient notes and clinical letters and populating the structured fields in the EMR, including the manual coding of line of therapy into a database field. This enhanced data was extracted and used as a validation dataset for the derived line of therapy method. In instances where clinically enhanced data isn't available a sample of patient level chemotherapy data from a cohort are reviewed by an oncologist and any misclassifications made using the algorithm identified and addressed.

RESULTS

A SAS algorithm to derive line of therapy in patients, receiving chemotherapy, with solid tumour cancers has been written to apply the rules defined above. This has been run on a cohort of urothelial cancer patients as an exemplar.

USING PROGRESSION DATES

Disease progression is the most common reason a patient begins a new line of therapy. Generally, if a patient receives chemotherapy and their disease progresses, the chemotherapy administered after progression would be considered a new line of therapy. Progression date is therefore the first thing used in identifying changes in line of therapy in the data. This is often not well captured outside of clinical letters or free-text notes, but these dates are available for the urothelial cancer cohort, as a result of the clinical data enhancement work, and have therefore been used in the algorithm to compare with the start date of a chemotherapy cycle and thus find the cycle that occurred following progression:

```
data [output];
    set [input];
    by patid eccstart regstart regcode;
    array pr[*] progdt1-progdt&maxprog.; /*progression dates are included as columns*/
    if not first.patid then do i=1 to &maxprog.;
        /*flag cycle if progression date is between cycle start and previous cycle
          start*/
        if eccstart ge pr[i] and prev_strt lt pr[i] then progress=1;
    end;
run;
```

IDENTIFYING GAPS IN TREATMENT AND SWITCHES IN REGIMENS

Gaps in treatment and switches in regimen, as specified in the agreed rules for a change in line of therapy, are identified by comparing one cycle with the previous cycle:

```
**Determine start date and regimen name of previous cycle**;
```

```
data prev_cycle;
    set [input];
    by patid eccstart regstart regcode;
    format prev_strt prev_date datetime20.;
    prev_strt=lag(eccstart);
    prev_reg=lag(regcode);
    prev_rlab=lag(reglabel);
    prev_date=lag(compare_date);
    prev_comb=lag(combined_reg);
    if first.patid then do;
        prev_strt=.;
        prev_reg=.;
        prev_rlab=.;
        prev_date=.;
        prev_comb=.;
    end;
run;

**Identify gaps between cycles and chronological switches in regimen**;
data cyclegap;
    set prev_cycle;
    by patid eccstart;
    if not first.patid then do;
        diff=eccstart-prev_date;
        if regcode ne prev_reg then switch=1;
        if diff ge 0 then gap=1;
        if switch=1 then do; /*Check if a switch in regimen is new line of therapy*/
            **Use SAS "sounds like" functions to determine switches between similar
              regimens**;
            if soundex(reglabel) = soundex(prev_rlab) then similar=1;
            **Flag if regimen is contained in previous regimen or previous combined
              regimens**;
            if index(compress(prev_rlab),compress(reglabel)) ne 0 then contain=1;
            if index(compress(prev_comb),compress(reglabel)) ne 0 then contain=1;
        end;
    end;
run;
```

Once the switches in regimen have been identified checking whether the regimen labels are similar, or whether one regimen label is “contained” within the previous, will help determine if the switch is an actual change in line of therapy or, per the rules defined above, one where the new regimen is a variation of the previous and not a new line of therapy. For example, a switch between a regimen labelled “Carboplatin + Paclitaxel (Weekly)” and “Carboplatin + Paclitaxel (3 Weekly)” wouldn’t be considered a new line of therapy, nor would a switch between “Carboplatin + Paclitaxel” and just “Carboplatin”. This is because the drugs given in the new regimen are all also given in the old regimen i.e. drugs have either been taken away or haven’t changed, but no new drugs have been added.

COMPARING DRUG COMBINATIONS

By comparing drug combinations between regimens, it is also possible to determine whether a change of regimen equals a change in line of therapy. If new drugs are added to the previous regimen, the new regimen will be considered a new line of therapy. If the drugs are taken away from the previous regimen, but all remaining drugs are a part of the previous regimen, this is not considered a new line of therapy.

```
**Compare each drug with the drugs in the previous cycle and count the number that are
found in both**;
data cycledrugs2;
    set cycledrugs1;
    by patid eccstart regstart regcode reglabel;
    length pdrug1-pdrug&maxdrugs $100;
    array pd[*] pdrug1-pdrug&maxdrugs;
    array cd[*] drug1-drug&maxdrugs;
    do i=1 to &maxdrugs;
        pd[i]=lag(cd[i]);
```

```
    end;
    do ii=1 to &maxdrugs;
        if first.patid then pd[ii]='';
    end;
    samedrugs=0;
    do j=1 to &maxdrugs;
        do k=1 to &maxdrugs;
            /*Switches between platinum therapies are not considered a change in line
              of therapy*/
            if not missing(cd[j]) and (
                index(left(trim(pd[k])),left(trim(cd[j]))) gt 0
                or (index(left(trim(pd[k])),'CARBO') gt 0 and
                        index(left(trim(cd[j])),'CIS') gt 0)
                or (index(left(trim(pd[k])),'CIS') gt 0 and
                        index(left(trim(cd[j])),'CARBO') gt 0)
                )
            then samedrugs=samedrugs+1;
        end;
    end;
run;

**If all drugs are in the previous cycle this means drugs are taken away and the line
of therapy will not increase**;
data samedrugs;
    set cycledrugs2;
    if switch=1 and samedrugs=num_drugs and num_drugs gt 0 then contain=1;
run;
```

The above steps create a series of flags in the SAS dataset and those flags are used to identify the cycles that constitute a change in line of therapy. Once those cycles have been determined, exceptions to the rules can be applied. Some of the exceptions will be general exceptions, for example, a switch between platinum therapies (Cisplatin and Carboplatin) would not be considered a change in line of therapy and that has been accounted for in the above steps. Other exceptions are manually hardcoded at the end of the algorithm and may be specific to the cancer type being analyzed or individual patients that have been identified as not possible to correctly assign line of therapy to algorithmically.

VALIDATION

Having run the algorithm on the cohort of 226 urothelial cancer patients, for which clinically enhanced data was available, the derived line of therapy was then compared against the enhanced line of therapy.

For 221 of these patients, the derived line of therapy matched the clinically enhanced line of therapy (97.8%). The remaining 5 patients were all cases where the disease progressed, but the decision was made to continue with the current course of treatment i.e. they were not advanced onto a further line of therapy. The algorithm calculated a change in line of therapy at the date of progression. For these 5 patients, they could be included in the “Exceptions” step and manually derived, returning a complete cohort of 226 patients with correctly a derived line of therapy.

By extending the cohort to include all urothelial cancer patients, that were treated with chemotherapy from 2006 onwards, in the EMR the proportion of patients treated at each line of therapy can be compared against the proportions of patients treated on each line of therapy in the enhanced cohort.

Of 369 patients overall, 29% were treated with neoadjuvant chemotherapy and 74.5% were treated with 1L chemotherapy (some patients can be treated with both) Of patients treated with 1L chemotherapy, 24.7% were treated with 2L, 5.8% with 3L and 2.5% with 4L or greater.

Compare that with the 226 enhanced patients: 40.7% were treated with neoadjuvant chemotherapy and 62.4% were treated with 1L chemotherapy. Of patients treated with 1L chemotherapy, 26.2% were treated with 2L, 5.6% with 3L and 2.8% with 4L or greater.

CONCLUSION

The objective was to create an algorithm that could accurately derive line of therapy. By comparing an algorithmically derived line of therapy versus a clinically determined and manually enhanced line of therapy for 226 urothelial cancer patients, the algorithm was correct for 97.8% of patients.

Generally, if a patient receives chemotherapy and their disease progresses, the chemotherapy administered after progression would be considered a new line of therapy. The 2.2% of the derived cohort that did not correlate with the clinical enhancement were again manually reviewed by a clinician. This small group of patients displayed progression in their disease but continued with their current line of therapy. These individualized patient decisions show the challenge in deriving line of therapy for 100% of patients and that there will always be exceptions to logic-based rules. However, these exceptions account for <3% of patients and demonstrate that the algorithm is accurate enough to be used to support research projects and determine valuable information for the healthcare and pharmaceutical industries.

The algorithm can be used on much larger cohorts than the urothelial cancer example illustrated here, deriving line of therapy for thousands of patients and saving hours of clinical resource that would be required to retrospectively enter this data into the EMR.

REFERENCES

[1] “CIS Healthcare - ChemoCare”, https://www.cis-healthcare.com/products-amp-services/chemocare/
[2] “About SACT”, http://www.chemodataset.nhs.uk/about_sact/

APPENDIX 1 - INPUT DATASET:
Below outlines the contents and specifications for the dataset that is input into the algorithm code:

Variable Name	Variable Type	Variable Description	Variable Length	Variable Format	Format Length
PATID	Num	Patient ID	8		11
FIRSTDIAGNOSIS	Num	Patient First Diagnosis Date	8	DATETIME	22
ICD10	Char	Patient ICD10 Diagnosis Code	4	$	4
REGID	Num	Regimen ID Number (Database ID)	8		11
REGSTART	Num	Regimen Start Date (First Cycle per Regimen)	8	DATETIME	22
REGCODE	Num	Regimen Numeric Code (Database Code)	8		11
REGLABEL	Char	Regimen Label	100	$	100
REGEND	Num	Regimen End Date (where available)	8	DATETIME	22
COMBINED_REG	Char	Combined regimen labels (for regimens starting on same date)	200	$	200
ECCSTART	Num	Cycle Start Date	8	DATETIME	22
ECCNUM	Num	Cycle Number	8		11
ECCDAYS	Num	Maximum Cycle Duration	8		11
LAST_DRUG	Num	Date of last drug administration (where available)	8	DATETIME	22
INTENT	Char	Regimen Intent	100	$	100
DRUG1-DRUG[N]	Char	Individual drug labels per regimen as separate columns	200	$	200
NUM_DRUGS	Num	Number of drugs per regimen	8		

APPENDIX 2 – ALGORITHM CODE:
The algorithm code is included as a SAS macro below:

```
%macro lineoftherapy(inputdata=,outdata=);

*************************************************************************;
**** Import data and set up                                          ****;
*************************************************************************;

**Create a 'compare date' to determine if there is a gap between cycles**;
data comparedate;
    set &inputdata.;
    format compare_date datetime20.;
    *Apply gap of 70 days to date of last drug for a cycle*;
    if not missing(last_drug) and last_drug gt eccstart and eccdays ge 0 then
compare_date=last_drug+((eccdays+70)*24*3600);
    *Apply gap of 70 days to date of last drug for a cycle*;
    else if not missing(last_drug) and last_drug gt eccstart then
compare_date=last_drug+(70*24*3600);
    *If drugs data unavailable, use max number of days the regimen is protocolled*;
    *for to determine an end date + 70 days*;
    else if eccdays ge 0 then compare_date=eccstart+((eccdays+70)*24*3600);
    *Neither available, apply gap of 90 days from cycle start date*;
    else compare_date=eccstart+(90*24*3600);
run;
```

```
**Read in progression dates for each patient - one row per progression date**;
data progress1;
    set progress_in;
    by patid ProgressionDate;
    if first.patid then progno=1;
    else progno+1;
run;

data _null_;
    set progress1;
    retain _prog;
    if progno gt _prog then _prog=progno;
    call symput('maxprog',compress(_prog));
run;

proc transpose data=progress1 out=progress2 prefix=progdt;
    by patid;
    id progno;
    var ProgressionDate;
run;

**Combine progression dates as columns with chemo data to determine cycles after
progression**;
data progress3;
    merge comparedate(in=a) progress2(drop=_:);
    by patid;
    if a;
run;

proc sort data=progress3; by patid eccstart regstart regcode; run;

*************************************************************************;
**** Derive flags to indicate line of therapy changes                ****;
*************************************************************************;

**Determine start date and regimen name of previous cycle**;
data prev_cycle;
    set progress3;
    by patid eccstart regstart regcode;
    format prev_strt prev_rgdt prev_date datetime20.;
    prev_strt=lag(eccstart);
    prev_reg=lag(regcode);
    prev_rgdt=lag(regstart);
    prev_rlab=lag(reglabel);
    prev_comb=lag(combined_reg);
    prev_date=lag(compare_date);
    if first.patid then do;
        prev_strt=.;
        prev_reg=.;
        prev_rgdt=.;
        prev_rlab=.;
        prev_comb=.;
        prev_date=.;
    end;
run;

**Determine start date and regimen name of next cycle**;
proc sort data = prev_cycle out= next1;
    by patid eccstart regstart regcode;
run;

**Create ordering date var**;
data next2;
```

```
    set next1;
    orddate=eccstart;
    laggrp = lag(patid);
run;

proc sort data = next2;
    by patid orddate;
run;

**Point to next observation**;
data next_cycle (drop=laggrp orddate);
    format next_date datetime20.;
    _N_ ++ 1;
    if _N_ <= N then do;

        set next2 POINT=_N_;

        if laggrp = patid then do;
            next_date = eccstart;
            next_reg = regcode;
        end;
    end;
    else do;
        next_date = .;
        next_reg = .;
    end;
    set next2 nobs = n;
    by patid orddate;
run;

proc sort data = next_cycle;
  by patid eccstart regstart regcode;
run;

**Flag next cycle that occurs following progression event**;
data progress4;
    set next_cycle;
    by patid eccstart regstart regcode;
    array pr[*] progdt1-progdt&maxprog.; /*progression dates are included as columns*/
    if not first.patid then do i=1 to &maxprog.;
    /*flag cycle if progression date is between cycle start and previous cycle start*/
        if eccstart ge pr[i] and prev_strt lt pr[i] then progress=1;
    end;
run;

**Identify gaps between cycles and chronological switches in regimen**;
data cyclegap;
    set progress4;
    by patid eccstart;
    if not first.patid then do;
        diff=eccstart-prev_date;
        if regcode ne prev_reg then switch=1;
        if diff ge 0 then gap=1;
        if switch=1 then do; /*Check if a switch in regimen is new line of therapy*/
            **Use SAS "sounds like" functions to determine switches between similar
regimens**;
            if soundex(reglabel) = soundex(prev_rlab) then similar=1;
            **Flag if regimen is contained in previous regimen or previous combined
regimens**;
            if index(compress(prev_rlab),compress(reglabel)) ne 0 then contain=1;
            if index(compress(prev_comb),compress(reglabel)) ne 0 then contain=1;
        end;
    end;
```

```
run;

**Compare each drug with the drugs in the previous cycle and count the number that are
found in both**;
data cycledrugs2;
    set cyclegap;
    by patid eccstart regstart regcode reglabel;
    length pdrug1-pdrug&maxdrugs $100;
    array pd[*] pdrug1-pdrug&maxdrugs;
    array cd[*] drug1-drug&maxdrugs;
    do i=1 to &maxdrugs;
        pd[i]=lag(cd[i]);
    end;
    do ii=1 to &maxdrugs;
        if first.patid then pd[ii]='';
    end;
    samedrugs=0;
    do j=1 to &maxdrugs;
        do k=1 to &maxdrugs;
        /*Switches between platinum therapies are not considered a change in line
              of therapy*/
            if not missing(cd[j]) and (
                index(left(trim(pd[k])),left(trim(cd[j]))) gt 0
                or (index(left(trim(pd[k])),'CARBO') gt 0 and
index(left(trim(cd[j])),'CIS') gt 0)
                or (index(left(trim(pd[k])),'CIS') gt 0 and
index(left(trim(cd[j])),'CARBO') gt 0)
                )
            then samedrugs=samedrugs+1;
        end;
    end;
run;

**If all drugs are in the previous cycle this means drugs are taken away and the line
of therapy will not increase**;
data samedrugs;
    set cycledrugs2;
    if switch=1 and samedrugs=num_drugs and num_drugs gt 0 then contain=1;
run;

proc sort data=samedrugs; by patid eccstart regstart regcode; run;

**Group all regimens between a treatment switch to count number of cycles per grouped
regimen**;
data cycleswitch1;
    set samedrugs;
    by patid eccstart regstart regcode;
    retain groupno;
    if first.patid then groupno=1;
    else if progress=1 then groupno+1;
    else if switch=1 and (similar ne 1 and contain ne 1) then groupno+1;
run;

**Count number of cycles**;
proc freq data=cycleswitch1 noprint;
    by patid;
    table groupno / out=cyclecount(keep=patid groupno count rename=(count=numcycles));
run;

**Get start date of each group**;
data groupdate;
    set cycleswitch1;
    by patid groupno;
```

```
    if first.groupno then output;
    rename eccstart=groupdate;
    keep patid groupno eccstart;
run;

**Get start date of each group**;
data groupend;
    set cycleswitch1;
    by patid groupno;
    if last.groupno then output;
    rename eccstart=groupend;
    keep patid groupno eccstart;
run;

data cycleswitch2;
    merge cycleswitch1 cyclecount groupdate groupend;
    by patid groupno;
run;

data cycleswitch3;
    set cycleswitch2;
    by patid groupno;

       **If regimen has one cycle and start of next cycle is within 60 days of start of
       previous flag as single cycle**;
    if numcycles eq 1 and (not last.patid) then do;
        singlediff=datepart(next_date)-datepart(prev_strt);
        if singlediff lt 60 and gap=. and progress=. then single=1;
    end;

    if (first.groupno and single=. and lag(single)=.) or gap=1 then new_line=1;
run;

*************************************************************************;
**** Exceptions                                                      ****;
*************************************************************************;

data except1;
    set cycleswitch3;

    ******************************************;
    ** ENTER EXCEPTIONS TO STANDARD RULES HERE **;
    ******************************************;

run;

*************************************************************************;
**** Count line of therapy numbers and output                        ****;
*************************************************************************;

data assign_lot;
    set except1;
    by patid eccstart;
    retain LOTnum;
    **For urothelial cancer - data ehancement on treatment intent done so can be
included in derivation**;
    if ICD10 in ('C65', 'C66', 'C67', 'C68') and first.patid and
substr(upcase(intentlbl),1,5) in ('NEOAD','ADJUV') then LOTnum=0;
    **For all other cancers - treatment intent not factored in as unenhanced data not
reliable**;
    else if first.patid then LOTnum=1;
    else if new_line=1 then LOTnum+1;
run;
```


```
proc sql noprint;
    create table &outdata. as
    select patid, eccstart as CycleStartDate, eccdays as CycleMaxDays,
    regid as RegimenID, regstart as RegimenStartDate, regend as RegimenEndDate,
    last_drug as LastDrugDate, combined_reg as StandardRegLabel,
    numcycles as NumOfCycles
        %do j=1 %to &maxdrugs;
           , DRUG&j
        %end;
    , LOTnum, FirstDiagnosis
    from assign_lot
    order by patid, eccstart, regstart, regcode;
quit;

%mend lineoftherapy;
```

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Matthew Thompson
IQVIA
Email: matthew.thompson@iqvia.com
Work Phone: +44 7899 067 185