Grouping patients for the High Complexity Model

Contents

Introduction

The HCM is suitable for services that provides treatment and care for patients with a long-term condition. Patients in these services will require packages of care that may vary in terms of clinical resources and frequency of contact, depending on their condition. These packages of care may also change over time, as the patient's condition changes.

To predict the capacity requirements for such services, we will need to understand how patients typically move between different types of care within a service. A critical part of this analysis is to first understand how to categorise similar types of care packages, so that we can measure how patients move between these types of intervention.

How to use this guide

This document provides guidance on how to transform an activity dataset for a high complexity service into information that can be used to predict patient group behaviour.

The guidance is aimed at information analysts and information leaders within provider organisations. We assume that users will already have some experience in the collection of data and management of information in the NHS. This includes as a minimum:

- Understanding of the NHS Data Model and Dictionary

- Experience in generating activity and service level reports in a provider setting

- Some expertise setting rules in a standard language for relational database management systems like SQL.

This guidance is made up of the following sections:

- Required data items

- Establishing your patient population

- Stratifying patients into levels of intensity

- Using frequency of contact to group patients

- Counting the number of patients per level of intensity

- Demand data outputs

Required data items

The dataset you extract will need to include the following data items. The outputs are generally expected to be in line with NHS data dictionary standards:

1. Pathway related items:

- Service pathway ID: care pathway identifier

- Local patient ID: patient identifier

- Service Pathway start date (ie referral received date)

- Service discharge date (ie discharge date)

Activity related items:

- Appointment/contact ID: identifier for specific appointment or contact

- Appt / contact date: date at which patient should have been seen

- Outcome of the appointment: appointment status (eg attended, did not attend, cancelled by professional / patient)

The following data items are not required, unless you are using them to assist in categorising patients into intensity levels:

- Clinical contact duration

- Activity description

- Clinic/staff type identifier (eg. clinic code, staff type etc.):

Establishing the patient population for your service

To populate the High Complexity Model, you will need patient level data for appointment bookings, covering a period of one to three years.

The patients that you will be capturing in your analysis should be patients in the service of interest, who are:

- treated using a shared group of resources (eg clinicians, multi-disciplinary teams, equipment or facilities) and may extend across multiple services

- diagnosed with the same condition (eg diabetes, first episode of psychosis, autistic spectrum disorder service) or a group of conditions that may be analysed together (eg. mental health clusters)

The key identifier to establishing the pool of patients should be that the same set of resources is deployed consistently within this cohort of patients.

If the HCM is being used to analyse one part of your service, you will need to ensure that your patient population is restricted to only those patients who interact with the process of interest.

Intensity levels: stratifying patients into groups

You will also need to obtain rules that will allow yourself to categorise the patients into separate groups.

Key identifiers that can be used to categorise the patients are:

- Frequency of contact (eg. daily, weekly, monthly, quarterly, yearly)

- Skills mix / clinic type (eg. nurse [band 5], consultant)

- Type of contact (eg. domiciliary visit, outpatient appointment)

- Clinical criteria

Rules for the above should be developed by your organisation in accordance with local practice – the best source of information on this would be operational colleagues involved in the running of the service.

Ideally, if clinical services are already delivered in a way that naturally groups patients (eg. a local definition of 'complex' vs. 'simple' patients), and this is already recorded on your electronic record system, then the criteria for these groupings should be used.

Please note that the High Complexity Model is limited to a maximum of 5 intensity levels – if your service involves more groupings, it may be worth looking at either modelling at a sub-service level or aggregating similar groups together.

A practical example of this type of grouping could be:

- A district nursing service – based on length and frequency of appointment:

  o 1 hr visit, daily

  o ½ hr visit, daily to every 4 days

  o ½ hr visit, weekly

If a natural grouping does not currently exist, and there are no clinical criteria that may help with this, you will need to establish your own groupings.

A good starting point is grouping by the frequency of contact – there may be other factors such as appointment length which may be more useful. Once again, we would strongly recommend a discussion with colleagues involved in running the service to help determine a relevant metric.

Using frequency of contact to group patients

If you do not have any existing criteria to group your patients, you can use the frequency of contact as a starting point for grouping.

To do this, you will need to:

- Estimate the intended interval between contacts. The intended booking interval may differ from actual contacts due to non-attendances or rescheduling (eg. a 1 month follow up who DNAs, and is subsequently seen after an additional week), so you may need to account for this when looking at your data.

- Count the number of contacts by intended interval, and plot this data in a frequency chart as shown below – note that in this example the booking interval is in days:
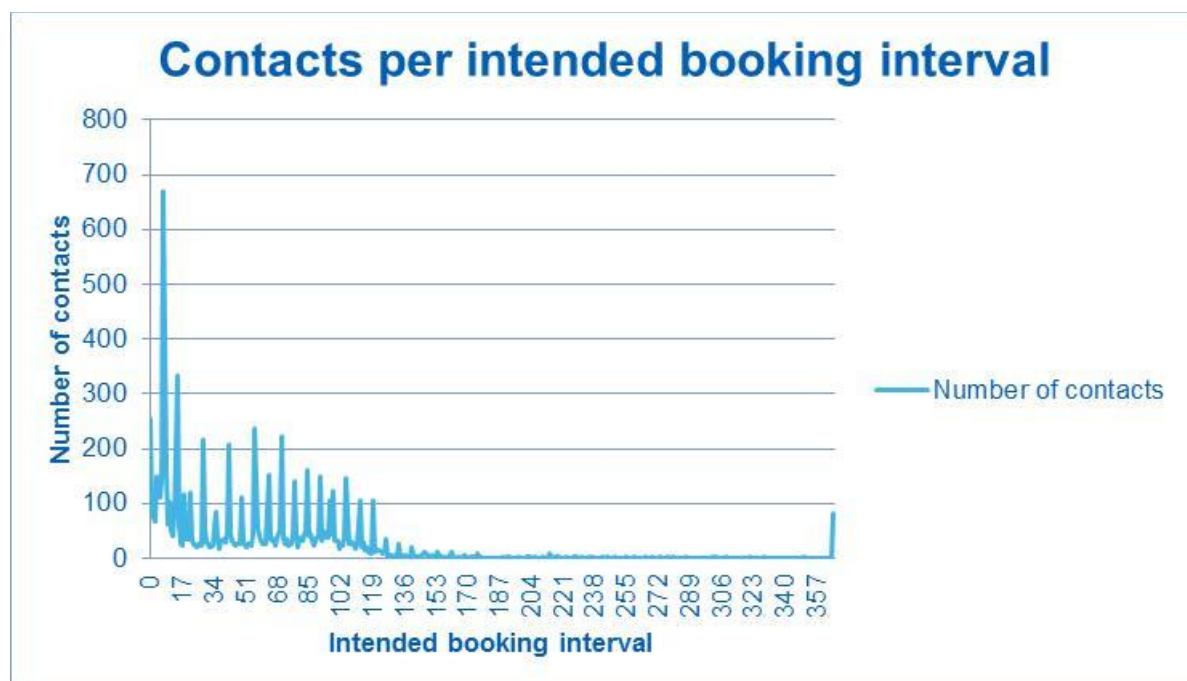
- On a second axis within the same graph, plot the cumulative frequency (percentage) distribution of intended appointments.

To calculate this distribution, take the running total number of contacts, and divide this by the net total contacts across the whole data sample period.

Frequency distribution of intended intervals
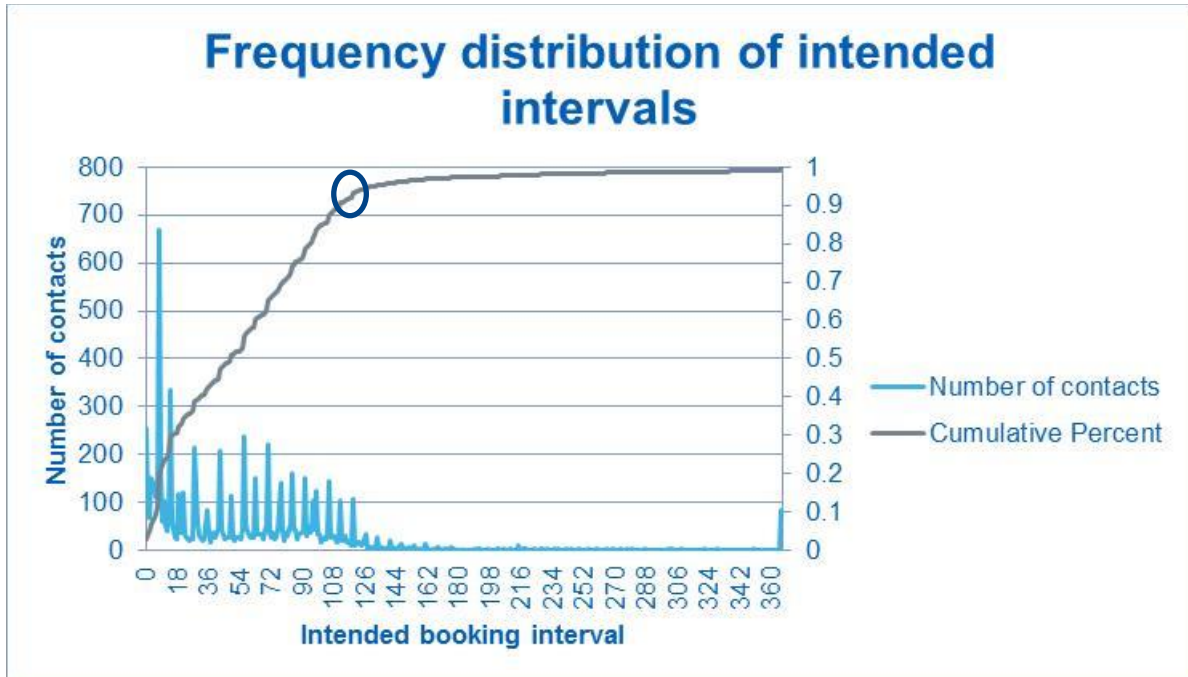
Reading the cumulative percentage scale on the right side of the chart with the intended booking interval scale on the x-axis will give you a sense of what proportion of your patient contacts are seen within a given period.

In the example above, 95% of the patient contacts (the circled point) occur within 120 days.

- We can use this information to determine cut-off points for intervals. In this case we can see a long 'tail' of intervals over 120 days which constitute a small (< 5%) proportion of our total contacts, which will factor into our decision on how to apply our groupings.

- Now we will inspect the chart and determine some initial groupings.

Equal interval

This is more appropriate for a homogenous spread, where a similar number of contacts take place at any frequency of contact.

Put more simply, if the cumulative percentage chart goes up in a straight line, then we can split the line into equal parts as a starting point.

The length of the interval can be calculated through the following formula

$$Length = \frac{Max - Min}{\# \ of \ intervals}$$

A more visual representation is shown below:
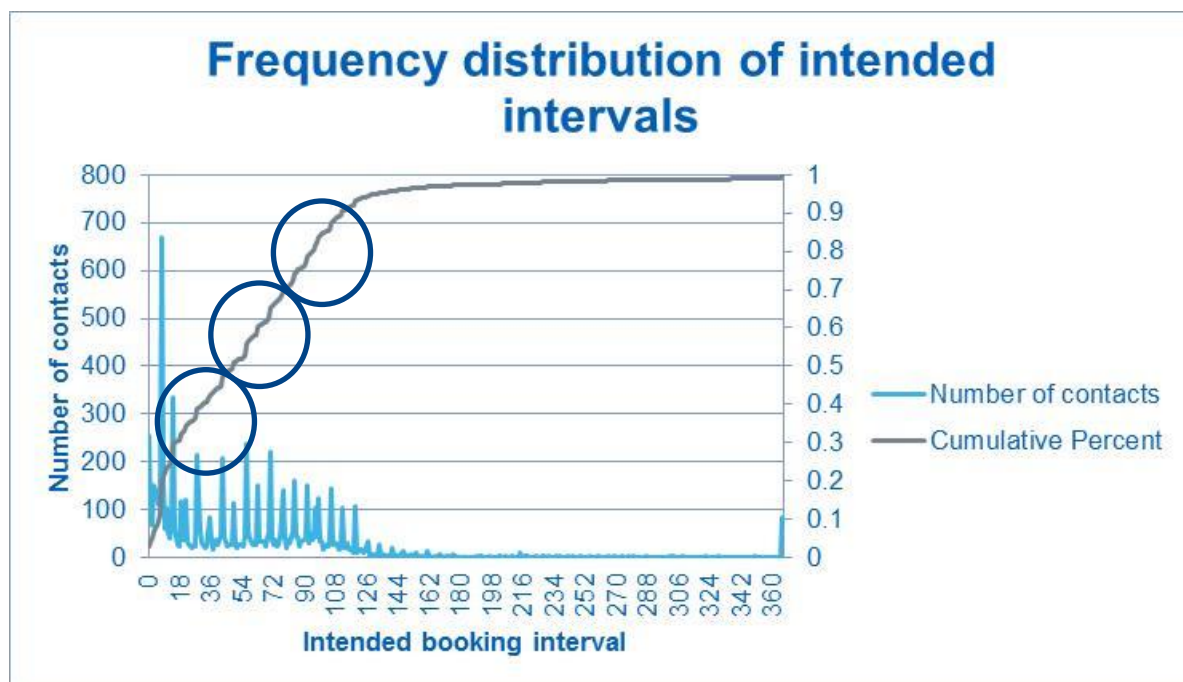
Dividing the spread of data into equal intervals

Other grouping methods

Other ways of grouping datasets are available. Some of methods are explained below, and Appendix 1 explains how the groups are calculated in more detail. The [Grouping worksheet] Excel tool uses some of methods. You can download it from our website alongside this guidance.

Grouping by visual inspection

You may be able to see groups by simply looking at the cumulative percentage chart. Groups can be identified by regions where the gradient of the cumulative percentage line changes suddenly. An example of this is shown below:
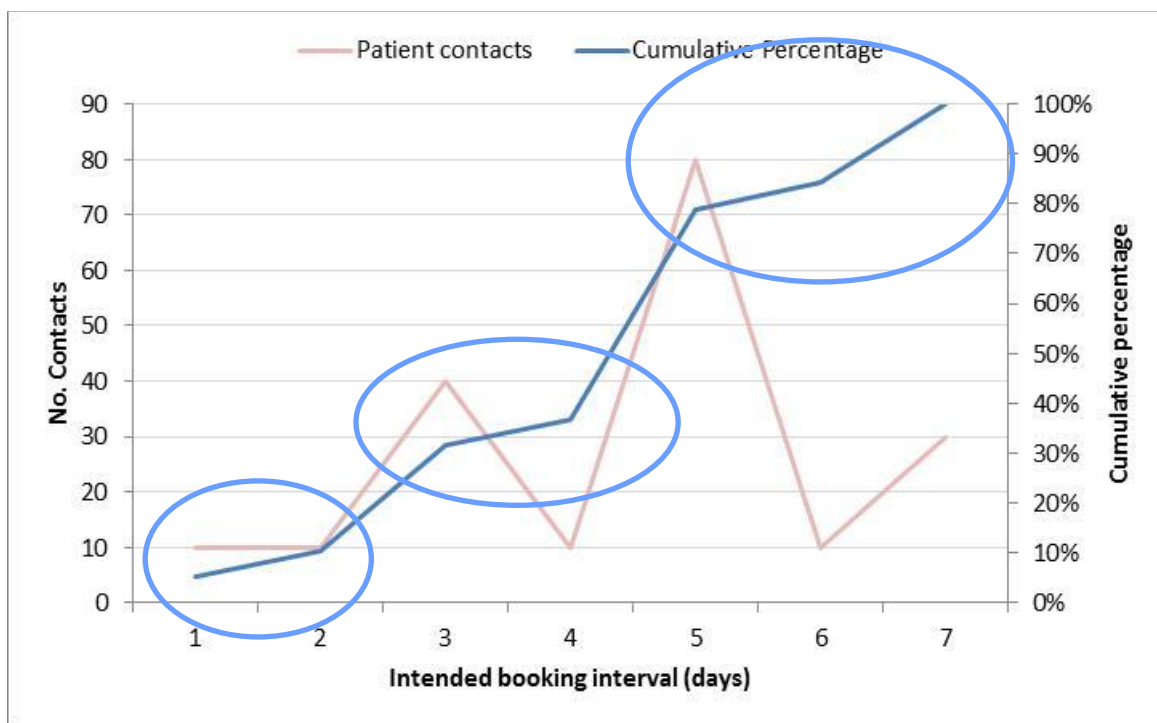
In this instance we have identified groups by looking for 'plateaus', where the gradient of the cumulative percentage chart has flattened out in between relatively steep rises.

Quantitative methods

Standard Deviation

This is useful when the dataset shows a normal distribution (also called a 'bell curve' - an SPC chart can be used to validate this assumption). The limits of the groups are established based on the standard deviation and the mean. More details for setting up to 5 groups can be found in Appendix 1.

Quantile

This assigns equal number of records/observations into each group. It is more suitable for datasets which do not show a normal distribution, but where the 'bell curve' is slanted either to the left or right – this type of dataset is normally described as 'skewed'. The limits of the groups are based on percentiles, e.g. interquartile range. More details for setting up to 5 groups can be found in Appendix 1.

Natural breaks / Jenks

Jenks Natural Breaks Optimisation is a form of clustering that creates groups which are as similar as possible within a group, while being as different as possible between groups.

This form of grouping is only carried out using a program – the Excel grouping worksheet which can be downloaded from our website includes an implementation of this.

Please note that large datasets (eg. more than 1000 entries) will result in slow run times or may crash the programme altogether. If you do have a large dataset we recommend you sample your dataset and use no more than 300 data points.

Combining quantitative and operational information

If you use any of the methods described in this section, we strongly recommend that you cross reference the outputs with operational colleagues, using the results as a starting point for further refinement.

As an example, a possible set of refined grouping definitions using the original dataset is shown below.

Level of intensity	Intended booking interval
Level of intensity 1	1 – 15 days
Level of intensity 2	16 – 30 days
Level of intensity 3	31 - 60 days
Level of intensity 4	60 – 120 days
Level of intensity 5	Above 120 days

Note that the final definition shown above is more heavily influenced by operational considerations (eg. the groupings above are based on fortnightly, monthly and quarterly appointment groupings), but the quantitative methods are a useful way of generating initial discussion points in the absence of any existing information.

This form of analysis is a mixture of judgement, data exploration and collaboration with operational staff, and you can expect to go through several iterations before reaching a result that is agreed on.

Counting the numbers of patients per level of intensity

One you have established a rule-set for determining the intensity level of a patient, you will need to apply this to your patient dataset.

In order to track the movement of patients between intensity levels, you will need to count the patients, by level of intensity, on a per-week basis. This will require the following analysis to take place:

- On a per-week basis (eg. start of every Monday), track all active patients within the service, and note which intensity group they belong to.

This grouping will be based on the patient's current intended booking interval – either based on their next booked appointment, or if that data is unavailable, the grouping based on the most recently attended appointment. An example output of the expected output is provided below (changes in intensity level are highlighted in red):

PatientID	Week 1	Week 2	Week 3	Week 4	Week 5
A	Level 1	Level 1	Level 1	Level 1	Level 2
B	Level 2	Level 2	Level 2	Level 2	Level 2
C	Level 2	Level 2	Level 1	Level 1	Level 1
D	Level 3	Level 3	Level 3	Level 3	Level 3

- This table can then be used to count the number of patients per level of intensity:

Week	Level 1	Level 2	Level 3
Week 1	1	2	1
Week 2	1	2	1
Week 3	2	1	1
Week 4	2	1	1
Week 5	1	2	1

- You will also need to highlight changes in intensity levels for individual patients, and when this has occurred. These events are highlighted in red in the first table in this section. You will need to ensure that you can describe:

  o When patients move between intensity levels

  o The new intensity level

  o The previous intensity level

All of the above data items will be used to describe:

- The size of the patient caseload per intensity level at a given point in time

- The number of patients who flow between different levels of intensity over time.

The flow diagram on the following page shows a set of suggested logical steps which can be used to generate an appropriate set of grouping tables – this is best implemented in a data warehouse.
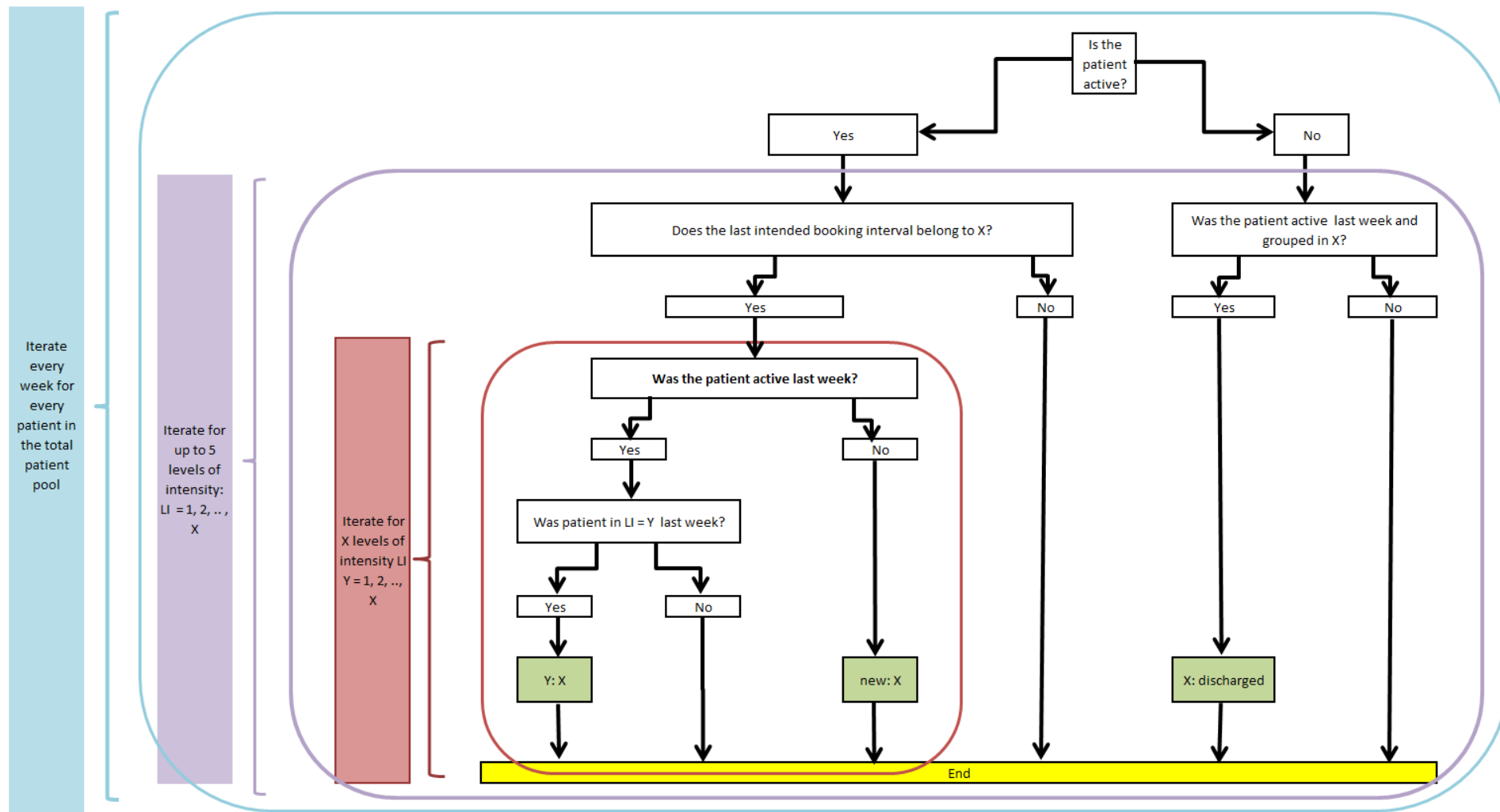
The final set of data that is required is the number of external additions (ie. new patients) and removals (ie. discharges) from your patient population.

These will need to be linked to your intensity levels, so patient tracking will be needed once again.

If you are finding difficulties in assigning intensity levels to new patients, a good starting point may be to do a clinical audit of new patients, to see if there is a consistent pattern to the caseload mix.

A revised data table to take account of these additions could therefore look like the following:

PatientID	Week 1	Week 2	Week 3	Week 4	Week 5
A	L1	L1	L1	L1	L2 (from L1)
B	L2	L2	L2	L2	L2
C	L2	L2	L1 (from L2)	L1	L1
D	L3	L3	L3	L3	L3
E	-	-	New - L1	L1	L1
F	L1	L1	Discharge (L1)		

Final outputs

The data table you have constructed should include the following:

- A per-week count of patients showing

    o Current intensity level

    o Previous intensity level (if different to current)

    o New additions (by intensity level)

    o Discharges (by intensity level)

    o ?

The last step is to aggregate this data in a form that fits the tables below:

- Patients on the caseload

Week number / week starting	Patients on the caseload				
	L 1	L 2	L 3	L4	L5
1					
2					
3					
4					
.					
.					
.					

- Flow of patients between levels of intensity

Week number / week starting	Internal transfers (as many as required, this patient were already in a different level of intensity)				
	From L1 to L2	From L1 to L3	From L3 to L2	etc	etc
1					
2					
3					
4					
.					
.					
.					

- New patients added to the caseloads

Week number / week starting	External additions (patients who weren't in either level of complexity before)				
	L1	L2	L3	L4	L5
1					
2					
3					
4					
5					
6					
7					

- Patients discharged from the caseloads

Week number / week starting	Discharges				
	L1	L2	L3	L4	L5
1					
2					
3					
4					
5					
6					
7					

If your data is tabulated as shown above, then the information can then be directly entered into the High Complexity Model.

Appendix

The following tables show how calculations for the standard deviation, equal interval and quantiles methods are applied at varying levels of grouping (from 2 to 5 groups) across different combinations of groups:

Equal Interval								
Level of Intensity	2		3		4		5	
	Min	Max	Min	Max	Min	Max	Min	Max
A Size	Min	Min + (max - min)/2	Min	Min + (Max - Min)/3	Min	Min + (Max - Min)/4	Min	Min + (Max - Min)/5
B Size	Min + (max - min)/2 + 1	Max	Min + (Max - Min)/3 + 1	Min + 2*(max - min)/3	Min + (Max - Min)/4 + 1	Min + (Max - Min)/2	Min + (Max - Min)/5 + 1	Min + 2*(Max - Min)/5
C Size			Min + 2*(Max - Min)/3 + 1	Max	Min + (Max - Min)/2 + 1	Min + 3*(Max - Min)/4	Min + 2*(Max - Min)/5 + 1	Min + 3*(Max - Min)/5
D Size					Min + 3*(Max - Min)/4 + 1	Max	Min + 3*(Max - Min)/5 + 1	Min + 4*(Max - Min)/5
E Size							Min + 4*(Max - Min)/5 + 1	Max

Standard Deviation								
Level of Intensity	2		3		4		5	
	Min	Max	Min	Max	Min	Max	Min	Max
A Size	Min	Mean	Min	Mean - 1SD	Min	Mean - 1SD	Min	Mean - 1SD
B Size	Mean + 1	Max	Mean - 1SD + 1	Mean + 1SD	Mean - 1SD + 1	Mean	Mean - 1SD + 1	Mean - 0.5*SD
C Size			Mean + 1SD + 1	Max	Mean + 1	Mean + 1SD	Mean - 0.5*SD + 1	Mean + 0.5*SD
D Size					Mean + 1SD + 1	Max	Mean + 0.5*SD + 1	Mean + 1SD
E Size							Mean + 1SD + 1	Max

Note that the standard deviation (SD) for the sample group can be calculated using the Excel formula =*STDEV.S([data range])*

Level of Intensity	Quantiles							
	2		3		4		5	
	Min	Max	Min	Max	Min	Max	Min	Max
A Size	Min	Median	Min	33rd Percentil	Min	25th Percentil	Min	20th Percentile
B Size	Median + 1	Max	33rd Percentile +	66th Percentil	25th Percentile + 1	Median	20th Percentile + 1	40th Percentile
C Size			66th Percentile +	Max	Median + 1	75th Percentil	40th Percentile + 1	60th Percentile
D Size					75th Percentile + 1	Max	60th Percentile + 1	80th Percentile
E Size							80th Percentile + 1	Max

As with the standard deviation calculation, the relevant percentile for a dataset can be calculated using the formula *=PERCENTILE([data range], Percentile)*, where *Percentile* is a value between 0 and 1.

The 33rd percentile would therefore be calculated as

=PERCENTILE([data range], 0.33)

The grouping worksheet automatically calculates these values for

Enter the data you want to group in column C under the shaded area, then select the number of groups you would like to assign in cell 'F2' ('Levels of intensity').

Levels of Intensity: 3

Data

3	
411	
4	
2	
339	
45	
310	
55	
14	
441	
54	
490	
416	
297	
29	
102	
395	
30	
256	
167	
104	
15	
347	
353	
382	
168	
170	
129	
329	
405	
362	
156	
39	

Grouping by Standard Deviation

Group	Min Value	Max Value
a	2	102
b	103	382
c	383	498
d		
e		

Grouping by Quantiles

Group	Min Value	Max Value
a	2	158
b	159	320
c	321	498
d		
e		

Grouping by Equal Interval

Group	Min Value	Max Value
a	2	167
b	168	333
c	334	498
d		
e		

Grouping with Jenks

Group	Min Value	Max Value
a	2	161
b	165	321
c	324	498
d		
e		

Estimated run time

0.12 minutes

Refresh

1. Enter your dataset here

2. Select number of groups

3. Refresh the Jenks optimisation calculation.

This section will turn red if the expected run time is more than 5 minutes.