Methods: Generating trust/specialty-level waitlist projections

1. Background

With future projections of referrals and capacity, forecasting waiting list size has typically been straightforward. What has been more difficult is forecasting waiting times. Through development of robust simulation techniques, *BNSSG Modelling and Analytics* have been able to confront such challenges[1,2].

However, with COVID-19, forecasting both waiting list size and waiting times is complicated by the nature and magnitude of ROTT[3], given the unprecedented durations that patients must wait for treatment. With longer waits, patients are at greater risk of going private, becoming inoperable, or dying before treatment.

Any credible forecasting approach must therefore take into account the impact of ROTT on early pre-treatment departures from the waiting list, in response to longer waiting times.

BNSSG Modelling and Analytics revised their simulation technique to account for ROTT. For demonstration, this has been applied at a national, all-England level[4]. However, given long computation time and unrealistic data quality requirements, the model could not be fit at a local trust/specialty level.

Consideration of other quantitative methods is required.

2. Objective

To forecast waiting list size and waiting times at a trust-specialty level; updatable every month with new data; and configurable with system-supplied scenarios concerning referrals and capacity (e.g. 2% annual referral growth; 5% annual capacity growth).

3. Mathematical model

Some of the main properties and characteristics of the waiting list problem can be represented through a differential equation[5] based model (Figure 1).

---

[1] Wood, RM (2019). Unravelling the dynamics of referral-to-treatment in the NHS. Health Systems.
[2] Wood, RM (2020). Modelling the impact of COVID-19 on elective waiting times. Journal of Simulation.
[3] Removals from the waiting list other than treatment.
[4] Howlett, NC & Wood, RM (2021). Modelling the recovery of elective waiting lists following COVID-19: scenario projections for England. Medrxiv pre-print.
[5] https://en.wikipedia.org/wiki/Ordinary_differential_equation.

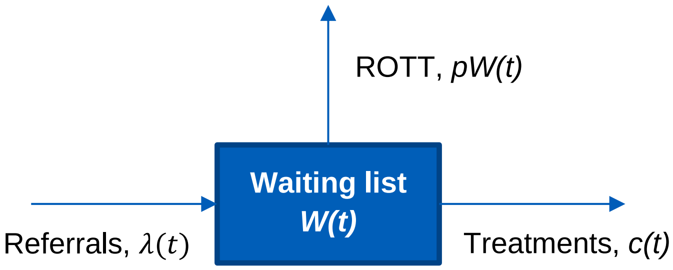

Figure 1. High-level dynamical behaviour captured by the model.

The model makes the following assumptions:
- The future *Referrals* rate ($\lambda(t)$) initially equals the mean over the last *n* months, and thereafter increases linearly according to the assumed annual growth rate (e.g. 2%, 5%).
- The *Treatments* rate is bounded above by the assumed capacity (*c(t)*).
- Future capacity (*c(t)*) initially equals the mean over the last *n* months, and thereafter increases linearly according to the assumed annual growth rate (e.g. 2%, 5%).
- The initial value of *ROTT* is set equal to the mean over the last *n* months, with changes thereafter in response to the waiting list dynamics (i.e. increases with longer waits).
- The value of *n* is set by the user.

The model formula for the waiting list size is given by:

$$W(t) = \frac{1}{p}\left[(\lambda_0 + \lambda_1 t) - (c_0 + c_1 t) - \frac{\lambda_1}{p} + \frac{c_1}{p}\right] + \left[W_0 - \frac{1}{p}\left(\lambda_0 - c_0 - \frac{\lambda_1}{p} + \frac{c_1}{p}\right)\right] e^{-pt} \qquad \text{Eqn. (1)}$$

where:
- $W(t)$ is the size of the waiting list at future time *t*
- λ_0 is the initial referral rate (i.e. that in the most recent data)
- λ_1 is the rate of increase in referral rate
- c_0 is the initial capacity level (i.e. that in the most recent data)
- c_1 is the rate of increase in capacity
- W_0 is the initial waiting list size (i.e. that in the most recent data)
- p is the ROTT-related parameter fitted from the data (larger p, more ROTT)

Note that:
- *Referrals* at future time *t*, $\lambda(t)$, is given by $\lambda_0 + \lambda_1 t$
- *Treatments* at future time *t*, *c(t)*, is given by $c_0 + c_1 t$

Other metrics are also calculated from Eqn. (1), including:
- *Mean wait for treatment* at future time *t*, $W(t)/(c_0 + c_1 t)$ (through Little's Law[6])
- *ROTT*, i.e. proportion of clockstops due to reneging at future time *t*, $pW(t)/\left(c_0 + c_1 t + pW(t)\right)$
- *Mean pathway duration* at future time *t*, $W(t)/\left(c_0 + c_1 t + pW(t)\right)$ (through Little's Law)

The proof of Eqn. (1) is provided in the Appendix, including more detail on the parameter p.

[6] https://en.wikipedia.org/wiki/Little%27s_law.

4. Model calibration

The overall process to model calibration and application is described in Figure 2.

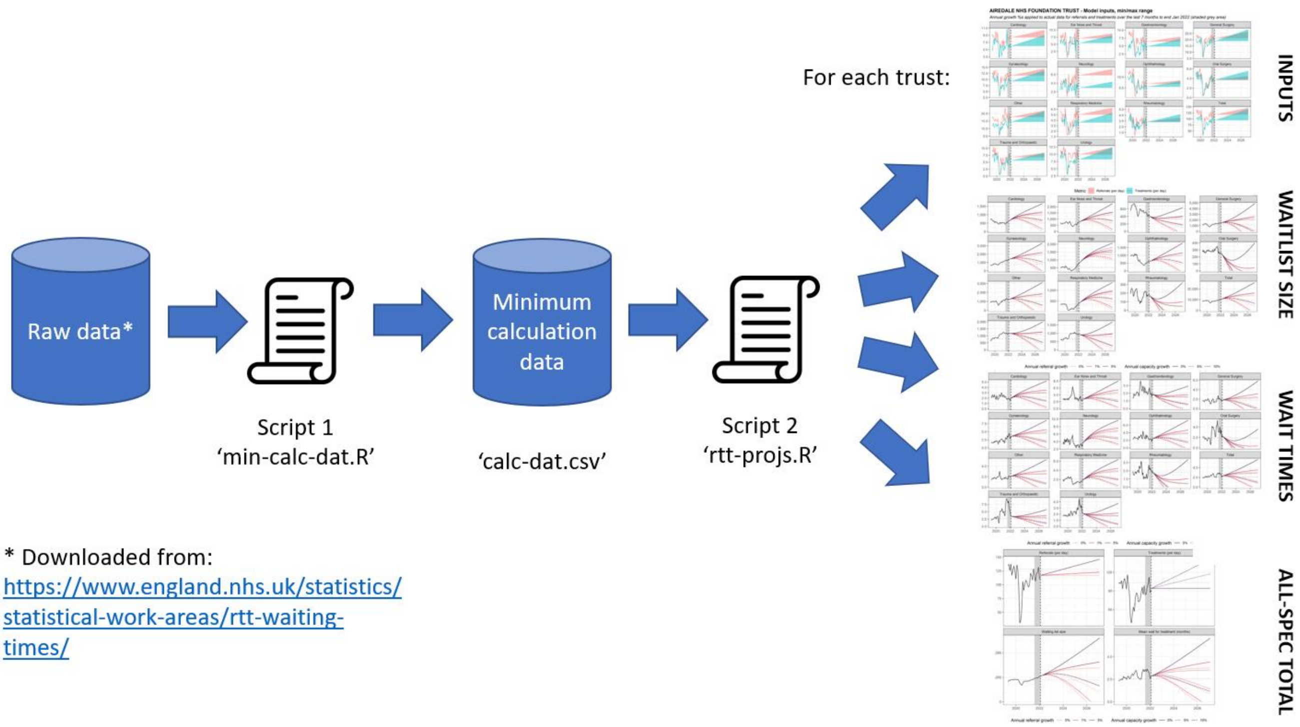

Figure 2. Dataflow, model calibration and result generation.

The process starts with obtaining the raw data from NHS-England[7], which contains monthly and trust/specialty-level summaries of referrals, treatments and waiting list size. ROTT is assumed to be the remainder of the difference between referrals and treatments when deducted from the waiting list size change from the previous month. That is, if there is no ROTT, then the difference in waiting list size from one month to the next should equal the difference in referrals and treatments in that month.

The model outputs are only as good as their inputs, and it has been noticed that there are some inconsistencies in the source data. For reasons including the lack of possibility to update past-submitted data, there appears to be a systematic under-reporting of referrals (clock-starts). This can lead, according to the abovementioned calculation approach, to negative ROTT (a clear impossibility). This is addressed, in Script 1 (Figure 2), by flooring ROTT at zero. This can lead to either spurious outputs or an inability to obtain a model fit.

Script 1 generates a minimum calculation dataset, containing the fields: trust, specialty, month, metric (wait list size, number of referrals, number of treatments, number of reneges, mean wait time for completed pathways), and value.

These are the inputs to Script 2, which also requires some other meta-parameters, including specification of the time period for calibration data (the *n* above), the forecast horizon (e.g. 5 years), and the different % annual growth to use for referrals and treatment capacity.

[7] https://www.england.nhs.uk/statistics/statistical-work-areas/rtt-waiting-times/.

Appendix: Proof of Eqn. (1) and detail on model calibration

Expression for numbers reneging

First, with referrals and treatment capacity known, consider the derivation of reneging (ROTT).

Let $d \in \mathbb{N}$ be the amount of time that a patient has been waiting for treatment since the point of referral, with each unit of time representing an arbitrarily small duration (e.g. a day).

Assume, if not treated, the probability of a waiting patient reneging at time *d* follows a series of *d* Bernoulli trials, each with a probability of success (reneging) of *p*. Thus, the patient's probability of reneging on the first day waiting is *p*. The probability of reneging on the second day waiting is the product of the $1-p$ probability of having not reneged on the first day and the *p* probability of reneging on the second day. Such that the probability of reneging on day *d* is

$$P(R = d) = (1-p)^{d-1}p \qquad \text{Eqn. (A.1)}$$

This is the probability mass function of the geometric distribution, which has a cumulative distribution function given by

$$P(R \leq d) = 1-(1-p)^{d} \qquad \text{Eqn. (A.2)}$$

These are plotted in Figure A.1 for example *p*=0.0025.

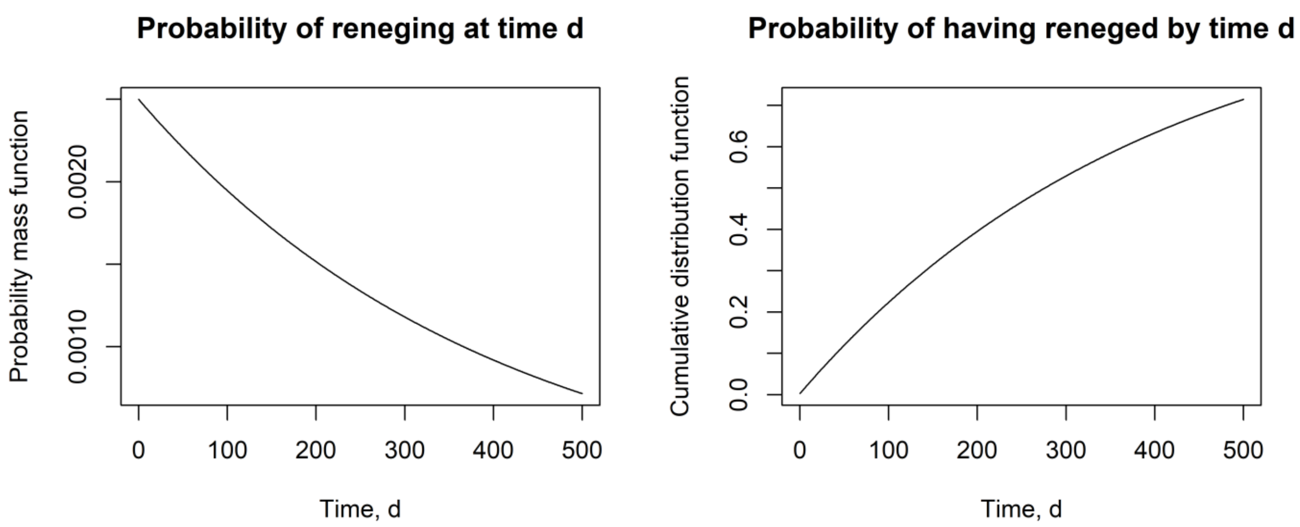

Figure A.1. Examples of reneging probability mass and cumulative distribution functions.

If the patient has not reneged until time *d*, then their probability of reneging on day *d* is therefore

$$P(R = d \mid R \geq d) = \frac{P(R=d)}{P(R \geq d)} = \frac{P(R=d)}{1-P(R<d)} = \frac{(1-p)^{d-1}p}{1-\left(1-(1-p)^{d-1}\right)} = p \qquad \text{Eqn. (A.3)}$$

Note that this exhibits the memorylessness property, meaning the probability of an event (patient renege) occurring at time *d* is independent of the number of times they have not reneged up to time *d*. This property is shared by the exponential distribution – the continuous equivalent of the (discrete) geometric distribution.

If there are $W(d)$ waiting patients at time *d*, each with expected probability of reneging equalling *p* within that arbitrarily small period of time, then it follows that

$$R(d) = pW(d) \qquad \text{Eqn. (A.4)}$$

Solution for W(t), i.e. proof of Eqn. (1)

The solution is sought for the dynamics described in Figure A.2.

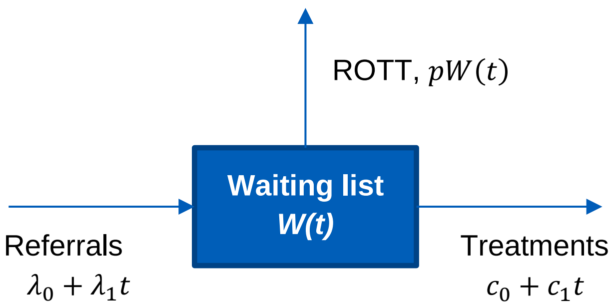

Figure A.2. Description of dynamics to be modelled.

This can be expressed as the first order differential equation

$$\frac{dW}{dt} = (\lambda_0 + \lambda_1 t) - (c_0 + c_1 t) - pW(t) \qquad \text{Eqn. (A.5)}$$

$$\frac{dW}{dt} + pW = (\lambda_0 + \lambda_1 t) - (c_0 + c_1 t) \qquad \text{Eqn. (A.6)}$$

Substituting $W = uv$ and using the *product rule* $\frac{dW}{dt} = u\frac{dv}{dt} + v\frac{du}{dt}$ gives

$$u\frac{dv}{dt} + v\frac{du}{dt} + puv = (\lambda_0 + \lambda_1 t) - (c_0 + c_1 t)$$

$$u\frac{dv}{dt} + v\left(\frac{du}{dt} + pu\right) = (\lambda_0 + \lambda_1 t) - (c_0 + c_1 t) \qquad \text{Eqn. (A.7)}$$

Setting equal to zero the *v* term in Eqn. (A.7) gives

$$\frac{du}{dt} = -up$$

$$\frac{du}{u} = -pdt$$

$$\int \frac{du}{u} = \int -pdt$$

$$\log u = -pt + k$$

$$u = Ae^{-pt} \qquad \text{Eqn. (A.8)}$$

where *A* is an unknown constant.

Putting Eqn. (A.8) into Eqn. (A.7) gives

$$Ae^{-pt}\frac{dv}{dt} = (\lambda_0 + \lambda_1 t) - (c_0 + c_1 t)$$

$$\int dv = \int \frac{e^{pt}}{A}(\lambda_0 + \lambda_1 t) - (c_0 + c_1 t)dt$$

$$v = \frac{1}{A}\int e^{pt}(\lambda_0 + \lambda_1 t) - (c_0 + c_1 t)dt + k \qquad \text{Eqn. (A.9)}$$

Using *integration by parts*, $\int xydt = x\int ydt - \int x'\left(\int ydt\right)dt$ with $x = (\lambda_0 + \lambda_1 t) - (c_0 + c_1 t)$ and $y = e^{pt}$, gives

$$v = \frac{1}{A}\left[\left((\lambda_0 + \lambda_1 t) - (c_0 + c_1 t)\right)\frac{1}{p}e^{pt} - \int (\lambda_1 - c_1)\frac{1}{p}e^{pt}dt\right] + k$$

$$v = \frac{1}{Ap}[(\lambda_0 + \lambda_1 t)e^{pt} - (c_0 + c_1 t)e^{pt} - \int (\lambda_1 - c_1)e^{pt}dt] + k$$

$$v = \frac{1}{Ap}\left[(\lambda_0 + \lambda_1 t)e^{pt} - (c_0 + c_1 t)e^{pt} - \frac{\lambda_1}{p}e^{pt} + \frac{c_1}{p}e^{pt}\right] + k$$

$$v = \frac{1}{Ap}e^{pt}\left[(\lambda_0 + \lambda_1 t) - (c_0 + c_1 t) - \frac{\lambda_1}{p} + \frac{c_1}{p}\right] + k \qquad \text{Eqn. (A.10)}$$

Given $W = uv$, incorporating Eqn. (A.8) and Eqn. (A.10)

$$W = Ae^{-pt}\left[\frac{1}{Ap}e^{pt}\left[(\lambda_0 + \lambda_1 t) - (c_0 + c_1 t) - \frac{\lambda_1}{p} + \frac{c_1}{p}\right] + k\right]$$

$$W = \frac{1}{p}\left[(\lambda_0 + \lambda_1 t) - (c_0 + c_1 t) - \frac{\lambda_1}{p} + \frac{c_1}{p}\right] + Be^{-pt} \qquad \text{Eqn. (A.11)}$$

where *B* is an unknown constant.

Model calibration

B can be estimated given information on the latest known waiting list size (i.e. that at the time of the latest data). If this is $W(0) = W_0$, then, with Eqn. (A.11)

$$W_0 = \frac{1}{p}\left[(\lambda_0 + \lambda_1 t) - (c_0 + c_1 t) - \frac{\lambda_1}{p} + \frac{c_1}{p}\right] + B$$

$$B = W_0 - \frac{1}{p}\left[\lambda_0 - c_0 - \frac{\lambda_1}{p} + \frac{c_1}{p}\right] \quad \text{Eqn. (A.12)}$$

Putting Eqn. (A.12) into Eqn. (A.11) gives the formula for waiting list size over time

$$W(t) = \frac{1}{p}\left[(\lambda_0 + \lambda_1 t) - (c_0 + c_1 t) - \frac{\lambda_1}{p} + \frac{c_1}{p}\right] + \left[W_0 - \frac{1}{p}\left[\lambda_0 - c_0 - \frac{\lambda_1}{p} + \frac{c_1}{p}\right]\right] e^{-pt} \quad \text{Eqn. (A.13)}$$

This is the formula, equivalent to Eqn. (1), which is used to derive the results in Figure 2 (top-middle panel). Other plotted measures are calculated using the equations already stated in Section 3 of the main paper.

p can be estimated using information on the latest known reneging rate (i.e. that from a recent period of the data defined by *n*), calculated by the number reneging over that period divided by the sum of numbers reneging and being treated over that period. If this is R_0, then

$$R_0 = \frac{pW_0}{pW_0 + c_0}$$

which yields

$$p = \frac{R_0 c_0}{W_0(1 - R_0)} \quad \text{Eqn. (A.14)}$$

Fitting for W_0 and *p* this way ensures their values are equal to those of the latest data at the start of the forecast period.

There is no ‘free parameter’ to ensure that mean treatment wait equates approximately to the last known value. In order to ensure alignment, the projections for this metric are scaled linearly.