

Modelling the neonatal system: a joint analysis of length of stay and patient pathways

Shola Adeyemi and Eren Demir

shola@statsxperts.com
Statsxperts Consulting Ltd, Haverhill, CB9 8PP, UK.

e.demir@herts.ac.uk
University of Hertfordshire, Hertfordshire Business School,
Hertfordshire, UK

Acknowledgments

None

Funding or sources of support

None

Conflict of interest

None

Human Studies and Subjects

The article does not require any human/animal subjects to acquire such approval.

Abstract

In the United Kingdom (UK), 1 in 7 babies require specialist neonatal care after birth, with a noticeable increase in demand. Coupled with budgeting constraints and lack of investment means that neonatal units are struggling. This will inevitably have an impact on baby's length of stay (LoS) and the performance of the service. Models have previously been developed to capture individual babies' pathways to investigate the longitudinal cycle of care. However, no models have been developed to examine the joint analysis of LoS and babies' pathways. LoS at each stage of care is a critical driver of both the clinical outcomes and economic performance of the neonatal system. Using the generalized linear mixed modelling approach, extended to accommodate multiple outcomes, the association between neonate's pathway to discharge and LoS is examined. Using data about 1002 neonates, we noticed that there is a high positive association between baby's pathway and total LoS, suggesting that

discharge policies needs to be looked at more carefully. A novel statistical approach that examined the association of key outcomes and how it evolved over time is developed. Its applicability can be extended to other types of long-term care or diseases, such as heart failure and stroke.

Keywords

Patient level modelling, Joint analysis, Neonatal care, Length of stay

Introduction

In 2017-18, out of the 750,000 births in the UK, approximately 105,000 babies (1 in 7 of all births) were admitted to neonatal services (1), up from 59,711 in 2005-6 (59% increase) (2). These babies are born premature, under 37 weeks, have a low birth weight (under 2,500g) or are born at term but with medical problems. The increasing number of babies needing neonatal care is mainly due to technological and therapeutic advances. These advances have implied a decreasing neonatal mortality rate for very low birth weight infants (under 1,500g) and a falling incidence of preterm prematurity (3).

Neonatal units face critical shortage of nurses, doctors and the full range of professionals needed to deliver safe care of the quality that these vulnerable babies need and deserve (1). Staff shortages has led to a large number of transfers of babies, putting infants at unnecessary risk and adding to their families' stress and worry. According to the National Neonatal Audit Programme, 10% of babies in neonatal care experience at least one transfer, of which some could be inappropriate. Of the 95,222 babies included in their data in 2015, there were 14,308 transfers in total involving 9,523 babies (4). In another study it was found that 10% of neonatal units exceeded their capacity in intensive care for more than 50 days over a 6-month period (5; 6).

The structure of neonatal services is made up of two layers affecting its ability to cope with the increasing demand. The first is the three levels of care provided for babies (depending on severity), namely, intensive care, high dependency care and special care. Babies within each level of care are subjected to various movements between different intermediate states from admission to discharge (either by death or alive). The second is that most babies are not discharged from neonatal services in the level of care they initially started with, where they are subjected to transitions between states in different levels of

care while in the unit. These two layers of complexity means that neonatal units are a multi-state, multi care level of service in which babies spend some time in some states and care levels and then flow to others before they are discharged from the service.

There has been limited number of research into modelling neonatal services. Understandably the majority has focused on developing survival models predicting time to discharge (7-12). As an extension of these studies, Seaton et al. (13) developed a Cox multistate model for neonatal care pathway, where the three levels of care have been considered (intensive care, high dependency and special care). The probabilities of receiving each of the levels of care or having died or been discharged from neonatal care were estimated (adjusted for gestational age). These models have primarily focused on predicting length of stay as an outcome variable, and very few research has been conducted around modelling individual babies' pathway. Note that a patient pathway is the route that a patient takes from their first contact with the NHS to the completion of their treatment. It includes the patient's referral, consultation, diagnosis, treatment and discharge. For example, a baby could begin their journey in ICU for x number of days, followed by HDU for y days, and finally in SCU for z days before discharge. This is the pathway for this baby.

To better understand the complexity within neonatal services, Adeyemi et al. (14) developed a series of generalized linear mixed models. A multinomial logit with exponential random effects was developed to investigate the flow of babies within a neonatal system. It was assumed that the movement of babies from one level of care to the other until discharge is multinomial. The approach identified interesting pathways, such as those that resulted in high probability of death (or survival), incurring the least (or highest) cost of care or ones with the least (or highest) length of stay. For instance, babies passing from high dependency to special care were more likely to end with a discharge home for babies with standardised random effects between -0.6 and 1.8.

Information about each baby's pathway is utilised, such as stages of care, e.g. starting point of a baby is ICU followed by HDU and finally SCU before discharge. This multiple information about each baby is fed into the statistical model, instead of relying on a single piece of information. The model then allows us to generate unique set of results for each baby, such as probability of discharge. Typically, in regression analysis an overall average effect (or fixed effect) is estimated, whereas in longitudinal data

analysis we are able to estimate individual babies' deviation from the average effect, a "measure of frailties", known as random effects. Random effects are then standardised for ease of understanding.

This model was further extended based on the assumption that movements of babies from one level of care to another (e.g. intensive to high dependency) is an improvement, hence the ordinal assumption was made instead of multinomial (15). A nonproportional cumulative logit random effects modelling framework was developed to capture individual babies experience in the care process, by relating probabilities of transition and discharge to babies history, such as states visited in the care process with their characteristics, including total length of stay, gender, gestation age and weight. The approach estimated the thresholds at which babies moved from one unit of care to another. For instance, a threshold separated babies passing through intensive care to high dependency from babies passing through intensive to special care, which indicated whether their condition is improving or deteriorating.

So far models have been developed to capture individual babies' pathways to investigate the longitudinal cycle of care or length of stay (predicting time to discharge). An elegant way would be to jointly model length of stay (LoS) and babies flow paths. The time spent in hospital at each stage of care 'LoS' is critical to the performance of neonatal services as it regulates the pace at which babies "flow" in the system. This pace determines the number of babies which can be treated in a ward per unit of time and the level of resource utilisation in these services. Furthermore, if we factor in the fact that intensive care babies belong to the "high cost" category of care, then LoS becomes a major factor of economic and clinical performance of neonatal services.

Given the importance of LoS as a proxy measure that determines the clinical and economic performance of health system and patient pathways to capture individual operational and clinical pathways through the health care system (i.e. inner workings), the current study has two objectives. First, to introduce the joint analysis of babies' pathways and LoS within a neonatal setting. This will enable us to measure the association between these outcomes and how it evolves over time with the inclusion of key risk adjusting variables. Suppose that total LoS has a positive association with flow paths, which may indicate that increasing LoS increases the odds of a baby being discharged alive. Using this information neonatal managers (and/or clinicians) could act accordingly, such as reassess their discharge policies, thus an evidence-based approach for decision making.

Second, we aim at advancing our knowledge about the application of modelling techniques to patient pathways and LoS. By doing so, the aim is to widen the remit of applications of modelling techniques in health care management and assess the pertinence of the insights gained from such exercises. Given the importance of modelling patient pathways and LoS, our research can be of great interest to providers and commissioners of care in England and other countries.

The Neonatal System

Neonatal units in hospitals specialise in the care of babies born early, with low weight or who has a medical condition that requires specialised treatment. There are, namely special care (SC), high dependency care (HDC) and intensive care (IC) unit. SC provides the basic level 1 to newborn babies at low risk. They have the capabilities to perform neonatal resuscitation at delivery as well as providing routine postnatal care of healthy newborn babies. HDC (level 2) provides care to babies who are ill with problems that are expected to be resolved rapidly. IC (level 3) is defined as having staff continuously available (neonatologists, neonatal nurses, respiratory therapists) and equipment to provide life support for as long as needed. The care process is depicted in Figure 1. Discharge from each unit could be by death, transfer to another hospital or home. In this system, there is no backward flow of babies (e.g. SC to HDC), although this is clinically possible as a baby in SC may relapse and be returned to HDC (or IC). From the data provided by University College London Hospital (UCLH), we notice that all the movements are forward, hence the models proposed here do not capture backward flow.

To perform this analysis, one-year data were collected from the North Central London Peri-natal Network (NCLPN). The Standardised Electronic Neonatal Database (SEND) was established in 2006 to collect information from the network. The SEND system allows the evaluation of the activity across the network and sharing of clinical information as neonates transfer between units take place.

The NCLPN is made up of five hospitals: UCLH provides the perinatal center for level 3, two level 2s, Barnet and Whittington as well as two level 1 ones, Royal Free Hampstead and Chase Farm. There is also the highly specialized center Great Ormond Street Hospital (GOSH) for treating very sick infants including neonatal surgery and some other specialized services. The network was set up to deliver a single neonatal service across London's north central area. The underlying aim of the network is to improve standards and achieve capacity so that 95% of neonates may be cared for within the network.

Figure 1: Pathways within the neonatal system

Data exploration

There were around 1002 neonates admitted to UCLH, where 993 were complete cases (no-missing information) and 9 missing. From the 1002 neonates, 830 was booked for delivery at the neonatal unit of UCLH and 172 from other hospitals, either outside or inside the NCLPN. Generally, babies with booked places in hospitals other than UCLH are admitted on account of their health conditions, which often requires intensive care or high dependency treatment. This admission could also be caused by lack of cots in the hospital initially booked. Table 1 provides the details about the reasons of admission for all levels of care.

The reasons for admission could be classified into three categories: back transfer, specialist care and neonatal care. Babies are generally transferred out of hospital for medical reasons or due to lack of resources (e.g. cots, trained nurses) and they could be transferred back to the hospital where they came from. This reason of admission is known as back transfer, which enables babies to continue receiving their care in a hospital closer to home. At UCLH, back transfers for continuing neonatal care represent a small proportion of admissions, about 4% (see Table 1). The second category of reasons for admission is the need for specialist care treatment (e.g. cardiac care, surgical care) which is not provided in all the hospitals of a particular neonatal network. The most frequent reason for admission consists in the need for neonatal care, i.e., IC, HDC, SC and transitional care.

Roughly, there are four main types of referral: children ward, labour ward, postnatal ward and theatre. From the 1002 neonates, 249 were referred to the postnatal ward, 223 to labor ward, 149 to theatre, 2 to children ward, 33 other referrals and 346 missing. Most babies are discharged home (approximately 71%); only few die (1.9%); the others are referred to wards within the hospital (7.2%) or to other hospitals (18.6%) for cardiac care, critical care, special care or surgery. Few records have missing discharge destinations.

A particular interest of this paper is the pathway frequencies leading to discharge illustrated in Table 2. Around 265 neonates were admitted to IC, where these births had taken place at either UCLH or back transfers from other hospitals. One hundred and thirteen babies, about 42% of IC admissions, were later transferred directly to SC and 78 sent to other hospitals for continuing care, while five babies

(approximately 1.9% of total IC admission) were discharged home. Around 49 babies experienced movement from IC to HDC then to SC, of which 24 and 25 were discharged to other hospital and home, respectively. Total admission into SC was 889, where 684 were direct admissions. Discharge from the system to other hospital is 258, out of which 78 from IC, 2 babies from HDC and 178 from SC. Death occurred in only 19 cases and are all from IC unit.

We further investigate the LoS distribution for each care unit. Figure 2 shows the distribution of LoS in IC (IC LoS), HDC (HDC LoS), SC (SC LoS) and the total LoS (tLoS), i.e., the sum of LoS at each level of care. The distribution of the observed LoS shows an exponential decline for the various levels of care. According to the level of care considered, LoS varies from few hours to weeks. The mean and standard error of IC LoS, HDC LoS, SC LoS and tLoS are 7.32 (0.41), 4.06 (0.77), 0.49 (0.08) and 11.87 (0.98) days, respectively. We observe substantial amount of variation between levels of care, hence it will be of an interest to examine the association between LoS and flow paths and how it evolves over time, which is the focus of this paper.

Figure 2: Histogram and a kernel density plot of LoS for each type of care

Method

Statistical problems where various outcomes of a mixed nature are observed have been around for about half a century and are rather common at present (16). Perhaps the most common situation in health service research or other fields, is that of the joint occurrence of a continuous, often normally distributed and a binary or ordinal outcome. Emphasis can be placed on the determination of the entire joint distribution of both outcomes or in specific aspects, such as the association in general or correlation between both outcomes. For the problem sketched above, there are three broad approaches. The first one postulates a marginal model for the outcome of babies' pathway (e.g. discharged alive or by death) and then formulates a conditional model for LoS, given the pathway. For the former, one can use logistic regression, while for the latter conditional normal models are a straightforward choice, i.e., a normal model with the baby's pathway (binary outcome) used as a covariate (17). The second family starts from the reverse factorization, combining a marginal model for LoS with a conditional one for babies' pathway. Conditional models have been discussed by (18), (19) and (20).

The third model family directly formulates a joint model for the two outcomes. In this context, one often starts from an underlying bivariate continuous latent variable of which one component is explicitly observed and the other one is observed in dichotomized or generally discretized version only (21). Molenberghs and Geys (22) presented a model based on a Plackett-Dale approach, where a bivariate Plackett distribution is assumed, of which one margin is directly observed and the other only after dichotomization. General multivariate exponential family-based models have been proposed by (23) and (24). Of course, these developments have not been limited to bivariate joint outcomes. One can obviously extend these ideas and families to a multivariate continuous outcome and (or) a multivariate categorical outcome. For the first and second families, one then starts from conditional and marginal multivariate normal and appropriately chosen multinomial models. Such a model within the first family has been formulated by (25). The literature on joint modelling of outcomes of various natures is diverse and growing. A broad ranging review of hierarchical models for joint continuous and discrete models can be found in (26).

Joint Analysis of Neonatal Pathways and Length of Stay

Let m be the dimension of the problem, that is, the number of outcomes that need to be modelled jointly. Further, let Y_{prj} denote the j -th measurement taken on the p -th baby for the r -th outcome, $p = 1, \dots, N, r = 1, \dots, m$, and $j = 1, \dots, n_{ri}$. Note that we do not necessarily need to assume the same number of measurements is available for all babies, nor for all outcomes. Let Y_{pr} be the vector of n_{rp} measurements taken on baby p for outcome r .

Our model assumes that each Y_{pr} satisfies a mixed model. Let $f_{rp}(y_{rp} | \theta_{rp}, \psi)$ be the density of Y_{pr} , conditional on a q_r dimensional vector θ_{rp} of random effects for the r -th outcome on baby p . The vector ψ contains all fixed effects β_r and possibly also a scale parameter (\mathbf{D} and Σ variance-covariance matrices of both the random and fixed effects, respectively) needed in the model for the r -th outcome. We do not necessarily need to assume the same type of model for all outcomes. A combination of linear, non-linear and generalized linear mixed models is possible. It is also not assumed that the same number q_r of random effects is used for all m outcomes. It may be assumed that, conditionally on the random effects $\theta_{1p}, \theta_{2p}, \dots, \theta_{mp}$, the m outcomes $Y_{1i}, Y_{2i}, \dots, Y_{mi}$ are independent. This is called the marginal

formulation. Other assumptions can be made, such as the hierarchical formulation.

We focus on the bivariate mixed effects model for continuous and binary endpoint, i.e., pathways leading to discharge alive (e.g. other hospital, home) or death are modelled as a binary response while total length of stay is modelled as a continuous response. These are jointly modelled as functions of covariates. This can be written in the form of

$$Y_{pr} = \mu_{pr}(\eta_{pr}) + \epsilon_{pr} = h(X_{pr}\beta_r + Z_{pr}\theta_{pr}) + \epsilon_{pr}$$

$$r = 1, 2; p = 1, 2, \dots, N, i = 1, 2, 3, \dots, k \quad (1)$$

where μ_{pr} is specified in terms of the unknown fixed effects β_r and 2-dimensional random effects.

X_{pr} and Z_{pr} are the known covariate values corresponding to patient p and ϵ_{pr} is the residual error. The components of the residual error structure ϵ_{pr} have the appropriate distribution with variance depending on the mean-variance relationship of the outcomes and can contain in addition a correlation matrix R_{pr} and an overdispersion parameter.

The components of the inverse link function $h(\cdot)$ depend on the nature of the outcomes in Y_{pri} . In this context we propose a logistic model for pathway endpoint Y_{1pi} (binary) and exponential for total length of stay Y_{2pi} (tLoS), since it came out to be the most plausible assumption (27; 6) agreeing with the general notion that LoS is often always exponentially distributed (see Figure 2). However, other distributions for tLoS have also been specified to investigate the sensitivity of our modelling results to different LoS distributions.

Therefore, our modeling framework can be expressed as in the form

$$\begin{pmatrix} Y_{1pi} \\ Y_{2pi} \end{pmatrix} = \begin{pmatrix} X_{1p} & 0 \\ 0 & X_{2p} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} Z_{1p} & 0 \\ 0 & Z_{2p} \end{pmatrix} \begin{pmatrix} \theta_{1p} \\ \theta_{2p} \end{pmatrix} + \begin{pmatrix} \epsilon_{1p} \\ \epsilon_{2p} \end{pmatrix} \quad (2)$$

with

$$\begin{pmatrix} \theta_{1p} \\ \theta_{2p} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix}\right] \quad (3)$$

and

$$\begin{pmatrix} \epsilon_{1p} \\ \epsilon_{2p} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix} \right] \quad (4)$$

Here, θ_{1p} and θ_{2p} are the random intercepts for the binary outcome and tLoS, respectively. We are particularly interested in the correlation structure of the data (i.e. the association between tLoS and flow paths) (see Appendix for further mathematical formulation).

Results: Joint modelling of pathways and tLoS in a neonatal unit

In this section, a joint modelling strategy is presented for the neonatal data illustrated, to answer how pathways to discharge (death, home, other hospital) and tLoS are associated. The evolution of tLoS is associated with the evolution of pathways is derived using equation (12). When the numerator of equation (12) is zero or close to zero both pathways and total length of stay are assumed to be completely independent at any point of time. The opposite situation applies when some of the enumerated covariance parameters lead to unit correlations. For example, the random slopes for pathways can be perfectly correlated with the random slopes for total length of stay. We then have a so-called shared-parameter model (29).

Parameters of the model are obtained using pseudo-likelihood (28) based inference because this concerns models with both continuous and binary outcomes. The parameter estimates were obtained using the SAS system. The fixed effects parameter estimates based on the exponential and log-normal assumptions for length of stay can be found in Table 3. Judging from the loglikelihood value the exponential assumption for LoS gives the best fit.

The variance-covariance (**D**) parameters for the model are transformed to a correlation matrix and ρ is calculated to be 0.8322 and 0.7622 for the exponential and Log-normal LoS assumption, respectively. This suggests that the correlation between the random intercept for the pathways and the random intercept for total length of stay is a strong positive association between the evolution of both pathways and total length of stay. In other words, increasing LoS increases the odds of babies passing through higher states towards being discharged alive. Higher states refer to babies who start treatment at the highest possible state (i.e. ICU) until discharged alive in a lower state (i.e. SCU).

Similarly, the population average effects of pathways are statistically significant suggesting that babies with increased length of stay and who pass through higher states (i.e. IC→HD→SC) are likely to be discharged alive from the system with log odds = 9.3920 ($p < 0.0044$) and log odds = 7.1202 ($p < 0.0001$), respectively. Not surprisingly, passing through the SC increases the chance of being discharged alive from the system (log odds = 0.8108, $p < 0.0269$). As expected, higher gestation age and birth weight increase the chances of progression to higher states (e.g. pathway IC to SC leading to discharge home) in the system, thereby increasing the chance of being discharged alive from neonatal care. Furthermore, small gestation age and birth weight reduces chances of progression to higher states, confirmed by the negative parameter estimates for gestation age and tLOS and birth weight and tLOS. This is not surprising as some of the babies with low gestational and birth weight died shortly after admission.

This explains the negative parameter estimates for gestation age and tLOS (-0.1186, -0.0946) and birth weight and tLOS (-0.2622, -0.2664) in Table 3, under both the exponential and log-normal models, respectively. This is also confirmed in Figure 3, where babies who had similar tLoS range but discharged by death had lower average gestation age and birth weight compared to those discharged home. The Figure also confirms the positive association of tLOS and pathways explained by the correlation estimates, 0.8322 and 0.7622 for the exponential and Log-normal LoS assumption, respectively.

These are valuable pieces of information for senior decision makers, which may enable them to act accordingly, such as reassessing their discharge policies and/or put in place additional support before discharge. It will help commissioners, clinicians and managers in making informed optimal decisions about transfer and discharge of patients (babies). Also, because babies' frailties are modelled as random effects, it is easier for managers to evaluate the level of frailties of discharging babies viz-a-viz some model established thresholds. For instance, Demir et. al (30) observed for this dataset that a LoS of 31 days serves as the threshold below which the probability of discharge by death is most probable, while babies who survive these first 31 days have a high chance of being discharged home.

These results also point to the fact that adequate observation time for babies could help to limit early discharge and help babies conditions positively. Though, there are always cost implication for additional days spent in the system, however future research should be geared towards assessing

quality of life gained by babies for any additional day(s) spent in the hospital. An assessment of this in view of long-term health outcomes need to be investigated.

Note that for the first time, the association between levels of care and LoS is investigated simultaneously within a single modelling framework. This can be applied to any disease area or department/specialty. The key highlight from this modelling exercise is the importance of LoS within each level of care. Reduction in LoS due to lack of resources (neonatal cots and staff) is a phenomenon in the NHS, not just in neonatal services but across the board. We are emphasizing the importance of this indicator and that neonatal services must be extra cautious in terms of how long a baby stays in each level of care before discharge. Furthermore, the current modelling exercise is to build an explanatory model, to explain the inherent relationships amongst the levels of care and the outcomes rather than prediction. A further research would be to extend this into a risk-based predictive model which can be used by clinicians in practice.

Figure 3: LoS versus average gestation age and average birth weight

Discussion

Studying individual babies' pathway in care coupled with length of stay can be highly beneficial for two reasons, (1) LoS is regarded as a major driving factor in health care systems. It regulates the possible number of patients that could potentially go through neonatal services, and (2) LoS is an important indicator of health care resource consumption, including human resources (neonatologist and nurse), and non-human resources (e.g. a cot, equipment). These have a direct impact on most of the cost associated with care of babies and day to day running of neonatal services.

This article addresses a highly topical issue in health care, focusing on policies around patient pathways and LoS, and how these two indicators influence patient outcomes and performance of neonatal services. Given that demand for neonatal care is increasing with limited finances (due to budget cuts in the NHS), the importance of this research will be highly appreciated by key decision makers (e.g. neonatal service manager) to improve the effectiveness of services and achieve better outcomes for babies, mothers, NHS staff and beyond.

The entire modelling landscape for neonatal care is revolved around a single variable of interest, such as predicting LoS of neonates; a queueing model to establish the required number of cots at all care

units, or individual babies movements within neonatal care to estimate probability of discharge (by death or alive). However, no study has ever considered modelling both LoS and neonates' pathway within a single modelling framework. Jointly analysing neonate's pathway and LoS provided us with some valuable information and insight into the inner workings of neonatal services.

For instance, the high positive association between total LoS and neonate's pathway suggests that increasing LoS also increases the odds of a baby being discharged alive through higher states (from intensive care to special care). Key decision makers, particularly neonatologists, may need to re-evaluate their discharge policies more carefully. Furthermore, small gestation age and birth weight reduces length of stay and the chances of progression to higher states (e.g. pathways leading to discharge home), confirmed by the negative parameter estimates for gestation age and tLOS and birth weight and tLOS.

No research is ever perfect and this one is no exception. The study has a couple of limitations, (i) data around individual pathways of neonates (full cycle of longitudinal care) with patient outcomes may not always be available. In some cases, if the data is available on case notes (but not in a health information system) it can be cumbersome and a laborious exercise to extract this vital piece of information, and (ii) computational problems may arise when the number of observations (including neonate pathways and outcomes) is very large. It could either take a long time to estimate the parameters or the maximum likelihood estimation of the generalized linear mixed model may not converge. If this occurs, the bootstrapping approach can be carried out using a sample of data at a time.

We have modelled a single hospital unit in London, which may not be generalizable to all neonatal services in England and other parts of the world. As neonatal units are part of a much larger network of providers in the NHS, critical drivers and factors could be missing, for example resources around transportation and admission policies between neonatal units are not captured. A modelling framework at a network level is currently been developed by the team (north central London perinatal network), along with a discrete event simulation model for comparative purposes.

The joint modelling analysis formulated in this research offer neonatal unit managers and physicians a statistical tool to appreciate the complexity of their system. It enables them to better understand the inner workings and the critical drivers that affects the performance and behaviour of neonatal services. The

methods presented in this study can be considered to be “tools for thinking”. This study will enable key decision makers to evaluate implications of change (e.g. reduction or increase in LoS) before it is implemented in practice, thus avoiding the trap of “doing things and hoping for the best”.

This research primarily focused on the association between baby’s pathway and length of stay, and the evolution of these two outcomes over time, adjusted with the inclusion of variables (e.g. gestation age and weight). Additional variables related to quality of care and its influence on the quality of life of neonates could have been investigated, particularly the impact on readmission of babies and their different care states (i.e. IC, HDC and SC). Patient readmission is a huge concern and better understanding the root causes (and risk factors) might prevent readmissions, thus reduce the burden on hospital services, patient and parents. Furthermore, it will be immensely interesting to determine the effect of integration of services, policies around admission and allocation of resources, and the impact of restructuring of neonatal services on key performance metrics.

Budgeting constraints in the NHS due to tough economic conditions in England are expected to be with us for a while. Against this background, the increasing demand for the NHS along with the introduction of new procedures & technologies, means that a paradigm shift is needed in the decision-making process in order to deal with this reality. In this respect, healthcare modelling is well positioned to play a greater role in influencing and shaping these processes and policies. Modelling will also enable healthcare systems around the world to face up to the challenges and inform better decision making to improve the quality of care, thus better patient outcomes, and the opportunity for providers to achieve all the necessary targets set by their Governments.

Appendix

A first order approximate expression for the variance-covariance matrix of Y_{pr} (*equation 2 above*) can be derived as (28)

$$V_{pr} = \text{var}(Y_{pr}) \simeq \Delta_{pr} Z_{pr} D Z'_{pr} \Delta'_{pr} + \Phi_{pr} \quad (5)$$

with

$$\Delta_{pr} = \left(\frac{\partial \mu_{pr}}{\partial \eta_{pr}} \right) |_{\theta_{pr}}$$

and

$$\Phi_{pr} \simeq \Xi_{pr}^{1/2} A_{pr}^{1/2} R_{pr}(\alpha) A_{pr}^{1/2} \Xi_{pr}^{1/2}$$

where A_{pr} is a diagonal matrix containing the variances following from the generalized linear model specification of Y_{pr} , given the random effects $\theta_{pr} = 0$. Similarly, Ξ_{pr} is a diagonal matrix with the over-dispersion parameters along the diagonal. The first term on the right-hand side of equation (5) corresponds to the random effects structure of equation (1); the second term on the right-hand side of equation (5) captures the variance-covariance in the residual error ϵ_{pr} .

Using equation (2) and the correlation structure in equation (5) we will jointly model tLoS (continuous) and pathways (binary) by means of a generalized linear mixed effects model, that is:

$$\begin{pmatrix} Y_{1pi} \\ Y_{2pi} \end{pmatrix} = \begin{pmatrix} \beta_{10} + \beta_{11}X_{11} + \beta_{12}X_{12} + \cdots + \beta_{1q}X_{1q} + \theta_{p1} \\ \frac{\exp(\beta_{20} + \beta_{21}X_{21} + \beta_{22}X_{22} + \cdots + \beta_{2q}X_{2q} + \theta_{p2})}{1 + \exp(\beta_{20} + \beta_{21}X_{21} + \beta_{22}X_{22} + \cdots + \beta_{2q}X_{2q} + \theta_{p2})} \end{pmatrix} + \begin{pmatrix} \epsilon_{i1p} \\ \epsilon_{i2p} \end{pmatrix} \quad (6)$$

where the random effects in equation (3) are normally distributed as

$$\begin{pmatrix} \theta_{1p} \\ \theta_{2p} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_1^2 & \rho\tau_1\tau_2 \\ \rho\tau_1\tau_2 & \tau_2^2 \end{pmatrix} \right] \quad (7)$$

where ϵ_{i1p} and ϵ_{i2p} are independent. The random effects θ_{1p} and θ_{2p} are used to accommodate the longitudinal structure in the data. The variances of Y_{1pi} and Y_{2pi} can be derived from equation (5), in which

$$Z_{pi} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \Delta_{pi} = A_p = \begin{pmatrix} 1 & 0 \\ 0 & v_{p2i} \end{pmatrix} \quad (8)$$

$$D = \begin{pmatrix} \tau_1^2 & \rho\tau_1\tau_2 \\ \rho\tau_1\tau_2 & \tau_2^2 \end{pmatrix}, \quad \Xi_{pi} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 1 \end{pmatrix} \quad (9)$$

where

$$v_{p2i} = \pi_{p2i}(\theta_{p2} = 0)[1 - \pi_{p2i}(\theta_{p2} = 0)]$$

with

$$\pi_{p2i} = \frac{\exp(\beta_{20} + \beta_{21}X_{21} + \beta_{22}X_{22} + \cdots + \beta_{2q}X_{2q} + \theta_{p2})}{1 + \exp(\beta_{20} + \beta_{21}X_{21} + \beta_{22}X_{22} + \cdots + \beta_{2q}X_{2q} + \theta_{p2})}$$

As a result, the approximate variance-covariance matrix of the two measurements for baby p at pathway i is equal to

$$V_{pi} = \begin{pmatrix} \tau_1^2 + \sigma^2 & \rho\tau_1\tau_2\nu_{p2i} \\ \rho\tau_1\tau_2\nu_{p2i} & \nu_{p2i}\tau_2^2 + \nu_{p2i} \end{pmatrix}. \quad (10)$$

The correlation $\rho_{Y_1 Y_2}$ between the tLoS (continuous) and pathways (binary) is induced by the incorporation of a correlation ρ between the two random effects and is approximately equal to

$$\rho_{Y_1 Y_2} = \frac{\rho\tau_1\tau_2\nu_{p2i}}{\sqrt{\tau_1^2 + \sigma^2} \sqrt{\nu_{p2i}\tau_2^2 + \nu_{p2i}}}. \quad (11)$$

This could be further approximated by

$$\rho_{Y_1 Y_2} = \frac{\rho\tau_1\tau_2}{\sqrt{\tau_1^2 + \sigma^2} \sqrt{\tau_2^2 + \pi^2/3}} \quad (12)$$

where $\pi^2/3$ is the variance of the standard logistic density.

References

1. NNAP (2018). National Neonatal Audit Programme (NNAP) 2018 annual report on 2017 data. <https://www.hqip.org.uk/resource/national-neonatal-audit-programme-2018-annual-report-on-2017-data/>. Accessed 15 May 2019. London.
2. National Audit Office (2008). Caring for Vulnerable Babies: The reorganisation of neonatal services in England. Retrieved from <https://www.nao.org.uk/wp-content/uploads/2007/12/0708101.pdf>. Accessed on 03 July 2019.
3. Tucker J. Patient volume, staffing, and workload in relation to risk-adjusted outcomes in a random stratified sample of UK neonatal intensive care units: a prospective evaluation. Lancet. 2002;359:99-107.
4. NNAP (2018). National Neonatal Audit Programme (NNAP) 2016 annual report on 2015 data. Retrieved from <https://www.hqip.org.uk/resource/national-neonatal-audit-programme-2018-annual-report-on-2017-data/>. Accessed 20 May 2019. London.
5. Asaduzzaman Md, Chaussalet TJ, Robertson NJ. A loss network model with overflow for capacity planning in a neonatal unit. Annals of Operations Research. 2009;178:67-76.

6. Asaduzzaman Md, Chaussalet TJ. An overflow loss network model for capacity planning of a perinatal network. *Journal of the Royal Statistical Society Series A*. 2011;174:403–417.
7. Altman M, Vanpee M, Cnattingius S, Norman M. Moderately preterm infants and determinants of length of hospital stay. *Archives of Disease in Childhood Fetal and Neonatal*. 2009;94:414–418.
8. Lee HC, Bennett MV, Schulman J, Gould JB. Accounting for variation in length of NICU stay for extremely low birth weight infants. *Journal of Perinatology*. 2013;33:872–876.
9. Manktelow B, Draper ES, Field C, Field D. Estimates of length of neonatal stay for very premature babies in the UK. *Archives of Disease in Childhood Fetal and Neonatal*. 2010;95:288–292.
10. Bender GJ, Koestler D, Ombao H, McCourt M, Alskinis B, Rubin LP, et al. Neonatal intensive care unit: predictive models for length of stay. *Journal of Perinatology*. 2013;33:147–153.
11. Manktelow BN, Seaton SE, Field DJ, Draper ES. Population-based estimates of in-unit survival for very preterm infants. *Pediatrics*. 2013;131:425–432.
12. Hinchliffe SR, Seaton SE, Lambert PC, Draper ES, Field DJ, Manktelow BN. Modelling time to death or discharge in neonatal care: an application of competing risks. *Paediatric and Perinatal Epidemiology*. 2013;27:426–433.
13. Seaton SE, Barker L, Draper ES, Abrams KR, Modi N, Manktelow BN. Modelling Neonatal Care Pathways for Babies Born Preterm: An Application of Multistate Modelling. *PLOS One*. 2016;11: e0165202.
14. Adeyemi S, Chaussalet TJ, Xie H, Asaduzaman MD. Random effects models for operational patient pathways. *Journal of Applied Statistics*. 2010;37:691–701.
15. Adeyemi S, Chaussalet TJ, Demir E. Nonproportional random effects modelling of a neonatal unit operational patient pathways. *Statistical Methods and Applications*. 2011;20:507–518.
16. Molenberghs G, Verbeke G. Models for Discrete Longitudinal Data. New York. Springer Series in Statistics. 2005.
17. Tate RF. Correlation between a discrete and a continuous variable. *Annals of Mathematical Statistics*. 1954;25:603-607.
18. Cox DR, Wermuth N. Response models for mixed binary and quantitative variables. *Biometrika*. 1992;79:441–461.
19. Cox DR, Wermuth N. A note on the quadratic exponential binary distribution. *Biometrika*. 1994;81:403–408.
20. Krzanowski WJ. Principles of Multivariate Analysis. Oxford: Clarendon Press. 1988.
21. Tate RF. The theory of correlation between two continuous variables when one is dichotomized. *Biometrika*. 1955;42:205–216.
22. Molenberghs G, Geys H, Buyse M. Evaluation of surrogate endpoints in randomized experiments with mixed discrete and continuous outcomes. *Statistics in Medicine*. 2001;20:3023–3038.
23. Zhao LP, Prentice RL, Self SG. Multivariate mean parameter estimation by using a partly exponential model. *Journal of the Royal Statistical Society - Series B*. 1992;54:805– 811.

24. Sammel MD, Ryan LM, Legler JM. Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society - Series B*. 1997;59:667–678.
25. Olkin I, Tate RF. Multivariate correlation models with mixed discrete and continuous variables. *Annals of Mathematical Statistics*. 1961;32:448–465.
26. Aerts M, Geys H, Molenberghs G, Ryan L. Topics in Modelling of Clustered Data. New York, Chapman and Hall/CRC, Chapter 14, 2002.
27. Xie H, Chaussalet TJ, Millard PH. Continuous-time Markov models for the flow of elderly people in institutional long-term care. *Journal of the Royal Statistical Society - Series A*. 2005;168:51–61.
28. Faes C, Aerts M, Molenherghs G, Geys H, Teuns G, Bijnens L. A high-dimensional joint model for longitudinal outcomes of different nature. *Statistics in Medicine*. 2008;27:4408–4427.
29. Gruttola V, Tu XM. Modelling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics*. 1994;50:1003-1014.
30. Demir E, Lebcir M, Adeyemi S. Modelling length of stay and patient flows: methodological case studies from the UK neonatal care services. *Journal of the Operational Research Society*. 2013;5:532–545.