World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:17, No:8, 2023

Comparison of Bayesian and Regression Schemes to Model Public Health Services

Sotirios Raptis

Abstract—Bayesian reasoning (BR) or Linear (Auto) Regression (AR/LR) can predict different sources of data using priors or other data, and can link social service demands in cohorts, while their consideration in isolation (self-prediction) may lead to service misuse ignoring the context. The paper advocates that BR with Binomial (BD), or Normal (ND) models or raw data (.D) as probabilistic updates can be compared to AR/LR to link services in Scotland and reduce cost by sharing healthcare (HC) resources. Clustering, cross-correlation, along with BR, LR, AR can better predict demand. Insurance companies and policymakers can link such services, and examples include those offered to the elderly, and low-income people, smoking-related services linked to mental health services, or epidemiological weight in children. 22 service packs are used that are published by Public Health Services (PHS) Scotland and Scottish Government (SG) from 1981 to 2019, broken into 110 year series (factors), joined using LR, AR, BR. The Primary component analysis found 11 significant factors, while C-Means (CM) clustering gave five major clusters.

Keywords—Bayesian probability, cohorts, data frames, regression, services, prediction.

I. Introduction

DIGITAL health and social care allow the linkage of public services and healthcare systems, enabling the evaluation of individual treatment and potential risks as in [1] and [2]. Saving on resources can be based on connecting services using classification and prediction, linking them using BR networks with posterior, prior, and likelihood represented by linked services, or using known fitted distributions used as predictors or targets. Healthcare costa are forecast in the UK to reach, by 2028, a range of several billion pounds a year if they are not better managed. As an indication of that, Bottery discusses that the cost can reach as high as £12 billion by 2030/31 at an average rate of 3.7% a year. The present paper attempts to address this problem using public H&Sc data available on PHS' website at [3] and from the SG [4] posted by June 2019. The data used here were counts of patients (called the "value" attribute in the data) and contained the parameters for each service. PCA was applied to see the most important ones after normalization, as discussed by Lippi [5]. Linkage is similar to mining service patterns from the same category using similarity metrics to those stored in databases, according to Litchfield [6], and further on to prediction or being in the same cohort. Zero padding replaced data imputation of missing data, and Bertsimas [7] discusses works on imputation using Markov models while [8] and [9] use statistical models to approach the missing data. The paper is organized as follows: In the first section, the nature of the data is explained, and the main analysis is given by introducing PCA, and the BR methods (data-driven, ND prior, and BD prior). Indicative comparisons and results are presented and accompanied by comparative plots or tabular forms for numerical comparisons.

Mr. Raptis is with the School of Design and Informatics, Abertay University, Dundee, Scotland, Bell Street, Dundee, DD1 1HG, (e-mail: sotnraptis@yahoo.com).

II. Materials And Methods

The data analyzed had a hierarchy of three levels: (A) services, (B) attributes of the services, and (C) levels of the attributes, as seen in Fig. 1. In Fig. 1, the services are represented by black boxes that are connected to black circles that are the attributes with no further breakdown and one level ("value") that has no further connections, and to red circles that represent attributes with more levels (more unique values). The services share common attributes, and attributes share common values that are represented as links between the boxes and the circles and as links between the red and black circles. Services are also called 'factors' and are assigned acronyms such as 'S-A-Z' to denote a service's name (S), that is its ID, the attribute (A), and the ID of it as (Z). For example, for service (S1), the attribute for age (A) has levels (Z): '13', '15', 'All' and each was tracked as an individual factor (service) or setting indicating the number of patients aged 13 or 15, or any age ('All') tracked over the 39 years. The services with the same (A) are in the same H&Sc 'pack'. The year series of all levels per service (summed over the attributes, and levels) can be visualized for representative ones in Figs. 2 (a)-(f) using the (A)'s defined in Table I. Table I illustrates the separate factors TS per service, that cannot exactly be mapped to the services in Fig. 1 due to the summation. The services shown in Fig. 1 are (1) "Primary1ChildrenBMI-Epidemiological", (2) "SmokingPrevalenceInYoungPeopleSALSUS", (3) "SmokingPrevalenceAndDeprivationSALSUS", (4) "NumberGeneralPracticesRegisteredPatients", (5) "IntensiveHomeCare", (6) "SmokingBehaviorAndSelfRatedHhealthSALSUS", (7) "Primary1BMIDistribution - MainClientGroup", (8) "LowBirthweight". In Fig. 2, the Y-axis shows the attendance, and the X-axis shows the span of 39 years. The breakdown of the service packs into attributes and levels is shown in Fig. 1. The data were heterogeneous (of various formats, dates, and other counts), with missing years, numerical ones (ages), dichotomous (presence, absence of a demographic class or age-band), categorical (classes or text descriptions). For example, ages were kept as ranges, as '*...ages 65+* ' or, as numbers. The gender was a numerical tag: '1' for 'male' and '2' for 'female'. Other records were counts of patients or percentages. The data contained up to six attributes (settings) per service, and each attribute had possible values ('levels'). The attribute ('Value') indicates summed counts across all attributes or no attributes. Some data had up to 20 levels, as in the service 'S20'. This setting allowed levels and attributes to be predicted and tracked as in [10], where the workload in an emergency department is forecasted using an ARMA model. The services can also be studied by associating pairs of administrative and clinical data as co-occurrences or contingency 'dashboards' using the 'matrix' method from NHSS as per Langton [11]. The works of Mishra [12], Guersel [13], and Marshall [14] use simulation algorithms to test the sensitivity of similar patient counts to clinical events ('early diagnosis', 'critical clinical outcomes', etc.) to compute the healthcare (HC) system's sensitivity. Arrival models can predict demand or discharge rates, as per Tovim [15]. Then, the prediction error is the difference from the actual rate. PCA analysis was used to determine important services such as service pack (S20), with the 10 H&Sc factor explaining 56.8% of the variance as in Table II. The second factor explains 28.3%. The attributes and their levels are part of Bayesian relationships, and some are presented in Table II. The Bayesian analytics are compared across the entire time span (39 years). POST.D shows the data-driven posterior for the same target (predicted factor) and the same predictors, with evidence denoted as LKL.D (data-driven under the LKL column) and with the likelihood (LKL) as the joint probability of the rest of the factors from the second factor and onwards. Column AVEPOST is the average posterior probability for the same target (across all levels of it) and years. The columns POST.BD, PRIOR.BD, LKL.BD refer to the use of the binomial models, and LKL.BD, LKL.ND, LKL.D refer to the likelihood of using the three models. Some attributes, like age and gender, are shared among more H&Sc groups. Fig. 1 shows indicative (not all) breakdowns of H&Sc data frames (the (S)s) into their attributes (the (A)s) and of their attributes into their levels (items in (A) lists). Roughly, the same attributes (per H&Sc pack) were found in the same K-means cluster (not shown due to the limited space) after computing the clusters. In Fig. 3, linear relationships can be seen between selected services using the data-driven models (i.e., not the ".ND" or the ".BD" models) as three pairs of co-plots of two (as sub-plots) that contain the original service's year data for the service "S8-SIMDquintiles-1" and 2 (can be more) other service data that are best related to it in the BR context and any (as a third) of the best cross-correlated services that are more than one. In Fig. 3, the plots show the regression between services (cohorts) demands (1) "Primar1BMIDistribution-MainClientGroup-InCareHome-AdultsWithPhysicalDisabilities" as a target (predicted demand) with predictors (2) "MentalWellbeing SSCQ-BirthWeight-LowWeightBirths", (2) "Smoking PrevalenceInCYoungPeopleSALSUS-SmokingBehaviour-RegularSmoker" (3) "SmokinBehaviourAndSelfRatedHealth (SALSUS)-Gender-Female" (4) "NumberGeneralPractices RegisteredPatients-Gender-Male" (5) "SmokingPrevalenceIn YoungPeople(SALSUS)-SelfAssessedGeneralHealth-VeryGood". The plots in Figs. 3 (a) and (b) show the separate linear connections between services (1) and (2) and (1) and (3), while Fig. 3 (c) shows pairs of four comparative plots (sub-plots) for the pairs (1) predicted demand for service "1" paired with the actual one (almost the same), (2) predicted demand for service "1" paired with service "2", (3) predicted demand for service "1" paired with service "3", (4) predicted demand for service "1" paired with service "4". These are shown as pairs of 1 (original target service, which is the POST variable) and three other service plots. Overall, plotted are the pairs (1) (original (POST), one well-correlated (using CC) service), (2) (original, service found as PRIOR), and (3) (original, other service found as LKL). Again, the POST, PRIOR, and LKL are the three quantities involved in the BR setting. We can contrast regression methods (many predictors) to BR models where we have two knowledge sources (as predictors) to combine, that is, the likelihood (LKL), and the prior (PRIOR) to get the target, that is, the posterior (POST). This is not a limitation of the BR approach, though, since the quantities (LKL) and (PRIOR) are not necessarily single variables and can be statistical hypotheses concerning more variables. This is a case that is covered in the present work.

The BR can also be used for prediction. As explained by [16], Bayesian methods interpret data from a study in the light of external evidence and judgment. The external evidence can be taken as a predictor. In the Bayesian context, the predictors can be the prior probability of the evidence $P(hsc_{target})$, where $P(x)$ stands for the unconditional probability of having $P_X = x$. Also, the likelihood $L(x/y)$ that relates the occurrence of data x to a likely model for it, y, that is a knowledge of the model that can be used as an added predictor for x.

III. Bayesian Prediction Using Data-Driven Priors

For a specific factor, hsc_i, we can define a Bayesian predictor on the grounds of other factors from the same data, $hsc_j, i = j, \ni i, j, \in [1, 110]$. The Bayesian prediction using two predictors as conditions can be formulated using the basic formula as in (1):

$$P(hsc_{i,t}/hsc_{\{j,k\},t}) = \\ = P(hsc_{i,t}) \times \\ (\frac{P(hsc_{\{j,k\},t}/hsc_{i,t})}{P(hsc_{\{j,k\}})}), \\ i, j, \in [1, 110], i \notin \{j, k\}, t \in [1981, 2019] \quad (1)$$

The above minimal generic model predicts that the demand for H&Sc factor, i in year, t, will depend on two other factors, j, and k, in the same year. This can be extended to more factors so that the H&Sc factor, i, at year, t, can be predicted from a combination of the demands of the rest of the factors that are the predicting factors, $j, k, l.,,, \ni \{j, k, l, ...\} \in [1, 110] \setminus \{i\}$. In this case, (1) generalizes to (2):

$$
\begin{gathered}
P(hsc_{i,t}/hsc_{j,\ldots,t}, j \in S_{i,t}) = \\
\frac{P(hsc_{i,t}/hsc_{j,t}, j \in S_{i,t})}{\sum_{m \in [1,110]} P(hsc_{m,t}, m \in [1,110])}) \\
t \in [1981, 2019] \\
S_{i,t}\{Set\ of\ predictors\ for\ i\ at\ year\ t\} \subset [1,110]
\end{gathered} \tag{2}
$$

A. Bayesian Prediction Using Normal Priors

We may not be able to have reliable data for one of the predictors in year, t. In the context of H&Sc data (that is, likely zero-padded data), we can define suitable priors from a ready prior (that is not data-driven from other HSc factors). This prior can be a suitable normal distribution (ND), that is adapted to the target model for each factor. If we need to model any data stream using ND, then a good approach to do it is to take a generic model and then take its long-term mean and standard deviation, even if we model the target at a year, $t <= 39$ or by 2019. In this version of the BR, we are actually modeling the target independently of other factors (data) and adapting an ND model for the specific year, t. If we model the prior as an ND and try to predict a year series of demands (that is, we do not seek one year's demand but model demands over more years for some services), then the term above becomes as in (3):

$$
\begin{gathered}
ND_{\vec{\mu},\Sigma}(\vec{x}) = P(\vec{x}) = \frac{1}{\sqrt{2\pi}} * exp^{\frac{(\vec{x}-\vec{\mu}_X)^2}{COV(X)}} \\
\Sigma(\vec{x}) = COV(X) \\
\vec{x} \in X = \{\vec{hsc}_{i,j} j \in S_{i,t}\}, \mu_X = \frac{1}{|S_{i,t}| * \sum_{j=1}^{j=39}(\vec{x}_j)}
\end{gathered} \tag{3}
$$

where X is the local set of the predicting factors for the target factor, i, in the year, t. As we can see, although we do not model the other data (year series of other factors), we need to use the closest ND model we can have, that is, borrow the factor set's mean, μ_X, and set's variance, $\Sigma(\vec{x})$ to develop the generic ND model. The quantity $COV(X)$ is the covariance for the set, $S_{i,t}$ of factors year series by the year, t. This covariance was computed from the covariance of the equivalent ND for the same size of observations (size of $S_{i,t}$) and for a magnitude (statistical average) equal to the average of the observations (that is, factors attendance over the examined year span). The choice of a ready ND prior overcame problems with computing the covariance of extensively zero-padded data (zero columns and zero standard deviation). Thus, for some specific year and almost any combination of factors, the set $S_{i,t}$, will be a subset of all the observations (evidence) we have for that year and for all of the 110 factors. It will also depend on the number of them, that is, the size of that set, denoted as, $|S_{i,t}|$, in (3). This number is the one to take as the predicting set in the BR schema. To make the BR methods comparable, Table III was developed, which has up to four such observations (predictors) per year that can be taken into account. For more, orders or lags, as discussed, the LR or AR methods would not work well. Also, as we raise the data dimensionality, the prior becomes very low, $P \approx 1E-20$.

B. Bayesian Prediction Using the Binomial Model (BD) as a Prior

A second approach was investigated to benefit from the extensive zero-padding, practically producing two data classes ('events'). These are the years with missing data (class #1) and the years with records (class #2). The smooth variation (low fluctuation of the data), seen in most curves in Fig. 2, justified this. This two-class problem was also modeled using the binomial distribution as a prior (BD) for two kinds of events as in (4):

$$
\begin{gathered}
P(eventB = \text{"}missing\ value\text{"}) = \\
P(hsc_{i,t} = 0), i \in [1,110], t \in [1,39] \\
q = \frac{\{No.\ of\ zeros\ by\ year\ t\}}{t} \\
p = P(eventA = \text{"}existing\ record\text{"}) = P(hsc_{i,t} > 0) \\
p = 1 - q \\
P(t,n) = C \times p^n * q^{t-n} = BD(t,n)\ is \\
probability\ of\ (n)\ non-zero\ records\ by\ year\ (t) \\
C = \frac{t!}{(t-n)!}
\end{gathered} \tag{4}
$$

The BD model, $BD(t,n)$, is a function of the year of reference, t, and of the number of events ('successes'), n. The choice of the prior models from (3) and (4) depends on zero-padding. Services with up to 3 years (as in Table I) of nonempty records would be better represented as an occurrence ('existing record') or not ('missing record'). The services with more than 3 years of non-missing records have attendances (people who benefited) in numbers that vary slowly (examples are Figs. 1 (b), (c), and (f)). Both are widely used in the social sciences. As discussed, the ND and the BD are very similar when the number of samples (in our case, the number of years) is very large (that is not the truth, here) or when the odds of having any attendance is $P \approx 1/2$. It can be seen from Fig. 1 that the patterns (a)–(d) may fit this case, while the rest of the patterns are two-event cases (missing/no missing). For most H&Scs, 67%, have more than three non-empty years, and the "non-missing" event is 62% (taken across all 110 factors). More than half of the HSc's data are not suitable for both ND and BD modeling ($P > 1/2$) while for the rest of the data, there is no standard prior (best one) unless they are modeled using other service data (that is, not known distribution models) as in (2). Then, each factor can be modeled as in (4) with individual p's and q's. Each year sequence has its own BD model, and the quantities $p(i,t), q(i,t), P(i,t), i \in [1,110], t \in [2,39]$ are functions of the data IDs ([1,110]) and of the years. The BD model does not check all attendance as the ND model does and only checks for 'successes' (above zero) and 'failures' (zero). This makes different service patterns (year series of attendances) with the same frequency of missing or existing records have the same probability with the BD model. Hence, the BD model is more suitable than the ND when examining hypotheses where the prior or likelihood are used as predicting services and are zero, as well as when the target is not.

The BD model has only two events to examine, as in (4), while the ND model accepts many levels of attendance found in a service. On the other hand, the BD model does not assign a value $p = q = 1/2$, to each event and actually counts strings of 'successes' or 'failures' by a certain year. The two probabilities, above, and in this work are computed from the actual frequencies of these events.

To use the BD model for a specific year, t, we use (4). This probability does not depend on the probability of 'success' or 'failure' by the year, $(t-1)$. The two events do not have equal chances any year and depend on the previous years' records. Therefore, we cannot take $p = q = 1/2$. Using the BD model for a prior, (2) becomes as in (5):

$$
\begin{gathered}
P(hsc_{i,t}/hsc_{j,...,t}) = P(hsc_{i,t}) \times (\frac{\sum_b BD(j,t,b)}{\sum_m \sum_b BD(m,t,b)}) \\
b \in \{0,1\}, j \in S_{i,t}, m \in [1,110] \setminus i, t \in [1981, 2019] \\
S_{i,t} = Set\ of\ predictors\ for\ i\ at\ year\ t \subset ([1,110])
\end{gathered} \quad (5)
$$

In Fig. 4 (c), the BR groups can be seen (set of services linked using BR models) ('S2-Value', 'S10-HouseholdType-Adults', 'S12-TypeOfTenure-Rented') where the posterior is $(POST = 0.3594)$. Both are among the highest. The same holds for row #5 and group ('S5-Gender-Male', 'S20-Gender-Female', 'S12-TypeOfTenure-All') where $(POST = 0.36)$ and for row #7. CC (also likely LR) is not in line with BR in row #6 where the group is ('S20-WeightCategory-Epidemiological (Obese)', 'S10-HouseholdType-Adults', 'S7-AgeBands-75YearsAndOver') have $(POST = 0.488)$. Most of the groups (predicted, predictors) in Table III are good ones (high POST). As it can be seen, the BR in rows #9 (target is 'S10-LowWeightBirths'), and #10 (target 'S20-AlcoholCondition-AllAlcoholic') have a low POST. Low birth weights are well correlated with alcohol-related conditions but do not occur often together (to use BR). In the work of Langton [11], BR can be used to define the LR predictors. The work by Porwal [17] also puts together LR and BR and advocates that BR can select the factors in an LR model to form prediction groups. In the same Fig. 4, labels attached to the edges show the likelihood of the connected services being linearly regressed (related). The arrow points to the regressed (the one that is predicted) service and originates from each of the predictors. The one-to-one connections shown do not mean we only have a single predictor per target. Each arrow is set to link only two services (for clarity in the diagram), and as we can see, some services ("Smoking prevalence in young – SALSUS - Age - All") are the end-points of many arrows. Also, it is possible that a predictor is linked to a target in more ways depending on the rest of the predictors in a linear set. Hence, the service "Smoking prevalence ... Regular Smoker" is connected to the above target in two ways (i.e., it is part of two linear sets that have the same target but not all the same predictors).

The three BR probabilities are computed for the same year. BR uses missing records as a single number (probability of zero), while regression takes all records (weighted sum). As a result, BR and regression schemas may not agree on the degree to which they support a target. An example is the two factors 'S10-SIMDquintiles-1' and 'S20-AlcoholCondition-LiverDisease(ALD)' that are part of two BR groups #6 and #9 with different probabilities, $(0.488 = P_{POST}(15, 16, 17))$, while $1 = CC(15, 17), CC(16, 17)$. We assume that a very high CC is indicative of good LR quality. Indeed, LR is based on the similarity of two (or more) factors, $\vec{x_i}$, $\vec{x_j}$, and, $\vec{x_k}$,.... Their CC, $CC_{j,k,...} = \vec{x_j}^T * \vec{x_k}$, is higher if they are linearly related, $\vec{x_i} = LR(\vec{x_j}, \vec{x_k}, ...)$ or when they are alike. Thus, CC depends on LR since a product is maximized if two or more factors are as similar as possible. The factor (S8) ('drug related discharges' is short for it) has a low $(POST.D = 0.06)$ when it is predicted from the factors 'regular smokers' (S9) and (S8) ('occasional smoker') but all are well correlated (CC = 0.993).

Each year, a new BD model is computed using (5), as shown in Fig. 5, and then an updated binomial probability is assigned to the examined level. The x-axis is the order of different data (110 HSc factors) used to derive the BD models. In Fig. 5, the Y-axis shows the binomial probability for one 'success' (existing record) (given the model). It is thus a Bayesian likelihood for one non-zero record. Fig. 5 (a) shows binomial non-zero probabilities using data from years, $t \in [1981, 2001]$. Fig. 5 (b) shows binomial non-zero probabilities using data from years, $t \in [1981, 2001]$, Fig. 5 (c) shows a normal model for above zero attendance and years, $t \in [1981, 2001]$, and Fig. 5 (d) shows a normal model for above zero attendance and years, $t \in [1981, 2001]$. In Fig. 6 the left plot, (a) shows the three Bayesian distributions ("LKL", "PRIOR", "POST") and the normalized posterior (the fourth one) for the examined attendance, (0). The right plot, (b), in Fig. 6, shows the plots for the same Bayesian distributions for the examined attendance (151564), that is also observed. Both levels (0, 151564), are observed. The distributions are computed using (9). This plot specifically shows the Bayesian distributions for service (ID = 70, "Primary1BMIDistribution-MainClientGroupInCareHome-AdultsWithLearningDisabilities") taking specific values that assume 12 distinct levels, including (0), that is the prevalent (count = 28) one, and eleven others (non-zeros) that only occur once in the 39-year span and belong to the region, > 0. The left panel for the Bayesian distributions of (0), is therefore quite different from the analogous distributions for the rest (> 0), which are represented by a single distribution (almost the same for the 10 other observed attendances) by the panel on the right. The distributions are computed using (9).

Since BD only provides probabilities for the number of 'non-zeros' in the span of 39 years, (5), was applied to the event $b = 1$ ("success"). The plot of the relevant probabilities for one 'success' $(n = 1)$ ("one success") and not ("any success") over the 39-year span is shown in Fig. 5, where the probability of having a non-zero attendance is given, provided that the current BD model used is the one with a specific ID $(ID > 80)$ as seen in the x-axis. It is interesting to observe that the HSc factors with IDs below 50 (overall, the IDs are 110) clearly favor non-zeros, although, most of them did not have records before 1997. It was observed that factors with $(ID > 80)$ that mostly started after 2010, had more than three records (as in Table I). The distribution in Fig. 5 (a) was computed with known data (either known as present or known as missing) until 2001. The first one shows that until 2001, having zero records was more likely than having a record, except for a few factors (local peaks), when there were some early recordings of the attendances for a few services before 1997. The data represented in Fig. 5 show the likelihoods (i.e., what are the odds of attendance given that the underlying distribution is an ND or a BD). The ND as a prior favors 'recent' services (H&Scs with fewer missing records in the past or roughly before 1997), while the BD favors 'early' services (HSc's with fewer missing records in the past). In the second group, (Fig. 5), services for which records started being kept after 1997, in the Bayesian context, likely have higher posteriors as target services since we have non-zero probabilities. Doing the same calculations with the ND model and taking the cumulative probability for having any above-zero attendances in both, above, year spans gives us roughly similar results as shown in Figs. 5 (b) and (d).

Fig. 5 (b) shows the distributions with more updated data (up to years from 2001 to 2019) and shows the three (four is the normalized) BR probabilities (that is, the POST probability of the event (no model, no data) probability in [0,1]). The graph assumes 12 distinct levels, including (0), which is the prevalent (counts = 28) one, and 11 others (non-zeros) that only occur once in the 39-year span and belong to the region $[150282,\ 157998]$. The left panel for the Bayesian distributions of (0) is therefore skewed to high values, while for the rest of the values observed, (>0), the Bayesian distributions are skewed towards low values on the right panel.

Overall, the three paradigms examined are: (a) the use of two or more other H&Sc services as priors and likelihood (2) and co-occurrences (that is, adopt no specific distributions for them), (b) the use of specific distributions as priors and likelihood as predictors of services (ND model as in (3)), (c) the use of the BD model as in (4) and (5). The formulas used to compute the columns PRIOR.ND/BD are:

$$\begin{aligned} POST.ND.PRIOR(j, CandidateAttendance) &= \\ LKL.ND(j, CandidateAttendance) &\times \\ ND.PRIOR(j, CandidateAttendance) &= \\ ND_{hsc_j}(i,t) \times P(hsc(i,t)/ND_k) & \\ CandidateAttendance = HSc(i, year = t) & \end{aligned} \tag{6}$$

$$\begin{aligned} LKL.ND(j, CandidateAttendance) &= \\ = P(CandidateAttendance/ND_{hsc(j,t)}) & \\ ND.PRIOR(j) = ND(\mu_{hsc(i)}, \Sigma_{hsc(i)}) & \end{aligned}$$

where, $(POST.ND.PRIOR(j, CandidateAttendance))$ is the posterior for service, j, having to compute the attendance level *'CandidateAttendance'* using the current ND model $P(x/ND_{hsc(j,t)})$, LKL.ND is the likelihood for the same conditions, and PRIOR.ND for service j, that is, a known distribution for j, that is not shaped (not a function of) by the searched attendance. The Bayesian probabilities, under a known distribution (methods 'b' and 'c' above) for PRIOR or LKL, are computed from the entire year span of the predicting services, denoted as $j, k, ...$, while the target attendance *'CandidateAttendance'* as, i, is assessed on the basis of the predicting ones setting aside the (i), that is from $\{j, k, ...\}\backslash\{i\}$. The predicting services are, thus, computed using (3) for the ND or (4) for the BD model, while their Bayesian combination under ND model is given in (5) by replacing 'BD' with 'ND' with parameters $ND(\mu, \Sigma_{t,\vec{x_t}},)$. The expression $\vec{x}(t)$ is the current data to evaluate the ND in the year, t.

IV. The Pegression Methods

BR is a prediction method for POST using LKL, and PRIOR as predictors. Mainly reported prediction methods by [18], such as Random Forests as per [19], are used to predict the workload in hand surgery operations as in the work of Uematsu [20], whereas BR develops hypotheses including the target and the predictors as clinical factors as per [21], [22] and [10] while [23] uses LR to infer daily patient discharges using twenty patient features, as well as 88 hospital ward-level features. BR models need a reliable PRIOR, or LKL while regression models, as per [24], need the number of predictors or time delays or can define linear predictors on-the-fly, as discussed by [25]. Good candidates for PRIOR or for LKL were partially based on trials using CC and CM since CC checks for numerical similarity and CM for the closest distance. Dunsmuir [26] advocates that CC coupled with ML can reveal specific relationships across data. The regression models define relationships also determined by probabilities as per [27], while BR checks probabilities of co-occurrences (LKL) and ground knowledge (PRIOR) to infer POST. BR computations involved year series of H&Sc observations, $X_i = [x_{i,1}, x_{i,2}, ..., x_{i,N}], i \in [1, 110]\ and\ N = 39$ and related service attributes, or patient parameters (demographics) or other services to create mixed cohorts. Hence, a level of demand, 'a', for POST can relate to another level of demand, 'b', for LKL (first predictor), and to a levels of demand, 'c' for PRIOR (second predictor), using compound hypotheses that involve them. An LR model's sensitivity to year lags or to the number of predictors is equivalent to POST's changes, as seen in Table III, when using different predictors for the same target. CC/CM was not always efficient for LKL and PRIOR due to extensive zero-padding that biased CC/CM towards false alarms (high CC for low POST.D, LKL.D, PRIOR.D). This affected less the ND/BD models that used equivalent data (no zero padding). For example, as seen in Table III, we have very low POST and high CC in the group ('S22-HomeCareClientGroup-LearningDisability', 'S23-SmokingBehaviour-NonSmoker', 'S22-Epidemiological-ClientGroupInCareHome-AdultsWithDisabilities-1"), that is, $POST(\{9, 10, 11\}) = 2.8E-34$, while $CC(9, 10) = 1,\ , CC(9, 11) = 1$.

V. Results

For no models, the priors, as discussed, for both LKL, and PRIOR is computed from other data (factors). That is, the rule $POST = P(hsc_{i,t}/\hat{hsc}_{\{j,k\},t})_{t\in[1,39]}$ applies. The column, AVEPOST is computed keeping the same group of factors, $\{i, j, k\}$, and the POST is averaged with respect to years (from year 1 to 39). The POST is a better-informed probability and represents what is expected when we try to predict, using BR, any attendance for a factor, i, considering that we have data for the factors $j, k, ...$ As it can be seen in Table III, the second factor, j^{th}, represents the evidence (observed data).

The likelihood is the third factor, k, along with any other factor beyond the k-th, if it exists at all. The factors $j, k, ...$, can be a joint probability, $P(hsc_{j,k,year=t})$, that we have for the factors $j, k, ...$, across all possible values (that is, over years till the year, t). This normalized probability(i.e., over all years, t) is, $(\frac{P(hsc_{\{j,k\},t}/hsc_{i,t})}{P(hsc_{\{j,k\}})})$. The joint probability is seen in (2).

Table III shows that the probability of having non-zero attendance for most services is not even (as per prior), as is the theoretical case, and depends on how often two services are co-attended. The POST and LKL probabilities can be computed at any year from 1 to 39 as per (1), where the year, t, is also an input argument. Table III shows that when priors are computed using ND or BD models, then the posterior and priors drop by several scales as they are not observed frequencies (are theoretical) using equivalent (not zero padded) data tailored to the target as in (3) and (5). This was necessary to avoid missing year problems that made the covariance a sparse matrix in the early years. In the data-driven probabilities, PRIOR.D, POST.D, and LKL.D assume roughly values in the scale $1E-2 <= prob <= 1E-1$, using observed evidence that are more confident estimates (higher) than model-driven priors are. In some cases, though, as with the relationship where the service (1) 'S22-HomeCareClientGroup-LearningDisability' is predicted from the service (2) 'S11-TypeOfTenureOwnedOutright', and from the service (3) 'S22-LivingArrangements-Age-All', the posterior is a little higher than CC, and it is the probability of linear similarity. This indicates that no new knowledge is offered by the added factors '3', or '2'. The actual range of the attendances in the data (110-year series) is in the interval $[0, 176944]$ but not all attendances are observed, and the frequentist probability is much higher than using the ND or BD. The frequentist probabilities are less variable (only observed) than the model-based ones are (infinite). The data-driven probabilities have a limited range, based on observed, unique, same year, often zero attendances, yield a POST.D (= 1) for most (but not all), and are almost uniquely linked to predicting factors unless a specific combination of factorsoccurs more than once. This finding (variable combinations can predict the same target) occurred only in a few cases. Some of these are: (1) 'S11-BirthWeight-LiveSingletonBirths', (2) 'S23-WeightCategory-EpidemiologicalOverweightObese', (3) 'S6-EverDrank-EverHadAnAlcoholicDrink" where the probability $P(i/j, k, l...) = 0.045$ when applying (2) and (1) 'S19-Epidemiological-WeightCategory-ClinicalSeverelyObese', (2) 'S23-SelfAssessedGeneralHealth-Bad", (3) 'S22-HouseholdType-Adults' with $P(i/j, k, l...) = 0.628$. Fig. 4 shows the data-driven posterior relationships (in 3-factor sets $i, j.k$), that is, which factors (as posteriors) can be predicted probabilistically in the Bayesian context. The arrows point to the target service, i, and originate from the two predicting factors, j, k. The labels attached to the edges show the likelihood of the connected services being BR related. Fig. 4 is set to show one-to-one connections for clarity, but some services ('S23-Age-All") are the end points of more arrows, and it is possible that a predictor is linked to a target in more ways (along with different predictors). Here, the service 'S23-RegularSmoker' points to the same target as part of different sets/predictors with probabilities $(0.440, 0.814)$. Fig. 4 shows that public services offered to young people who smoke can be predicted from services concerning young smokers of all ages with variable probabilities. Those who are self-assessed as being in good health, as an attribute, is a good predictor ($POST.D = 0.814$) for young smokers, indicating that given that we observe young people who are self-assessed as being in good health, then in the same cohort, the chance that smokers of all ages are in good health is high as well. Young smokers can also be BR predicted by those who are most deprived ($POST.D = 0.729$) which is more expected because social deprivation is a likely cause of smoking and is also confirmed by another likely good predictor, which is young people who are self-assessed as being in bad health ($POST.D = 0.729$). Gender is not likely to affect the prevalence of smoking in young people since the relative probability is rather low for cohorts that are based on age banding ($POST.D \approx 0.4$), and it is related to the prevalence of smoking in different contexts (along with other predictors). More arrows originate from gender-related cohorts that end up in the cohort with a high smoking prevalence. Hence, social deprivation, smoking frequency, age, and gender define young people's cohorts that can variably predict the cohort where smoking is prevalent.

The names of the services in Fig. 4 are defined in Table I.

VI. Discussion

The paper discussed how we can relate H&Sc service attendances using BR prediction to form service cohorts and evaluated three main methods as well as side tools. The dependence of these relationships on classification as well as on service settings was studied using Bayesian statistics, and CC and PCA provided more context, although not fully analyzed due to space limitations. Among the major findings of this work, it was found that, on average, a number of factors in the region ([2, 6]) were well connected using BR. The work mainly reports on data-driven posteriors, POST.D. Model-driven POST.ND, POST.BD posteriors were theoretical approaches to the problem and made it more clear how the zero padding (missing data) can be better modeled using equivalent distributions (if no data are available) using BD or how difficult it is to obtain parameters like the covariance in ND models that need to be approached by equivalent models. The data-driven models made it more clear how the posterior is a better version of the prior (higher confidence), which was not always clear with data models. Using all BR approaches, 3-5 independent factors (as predictors) were practically found and 1 dependent, that is, in groups of up to 4 or 5. Larger groups either did not have co-occurring demands and gave zero data-driven posteriors or had extremely low model-based posteriors (practically zero as well). Similar sizes in the region ([2, 5]) were also discussed by [27]. There were no H&Sc factors that could not be expressed through BR combinations except for those with a single or a few ($<= 2$) years of non-missing records like

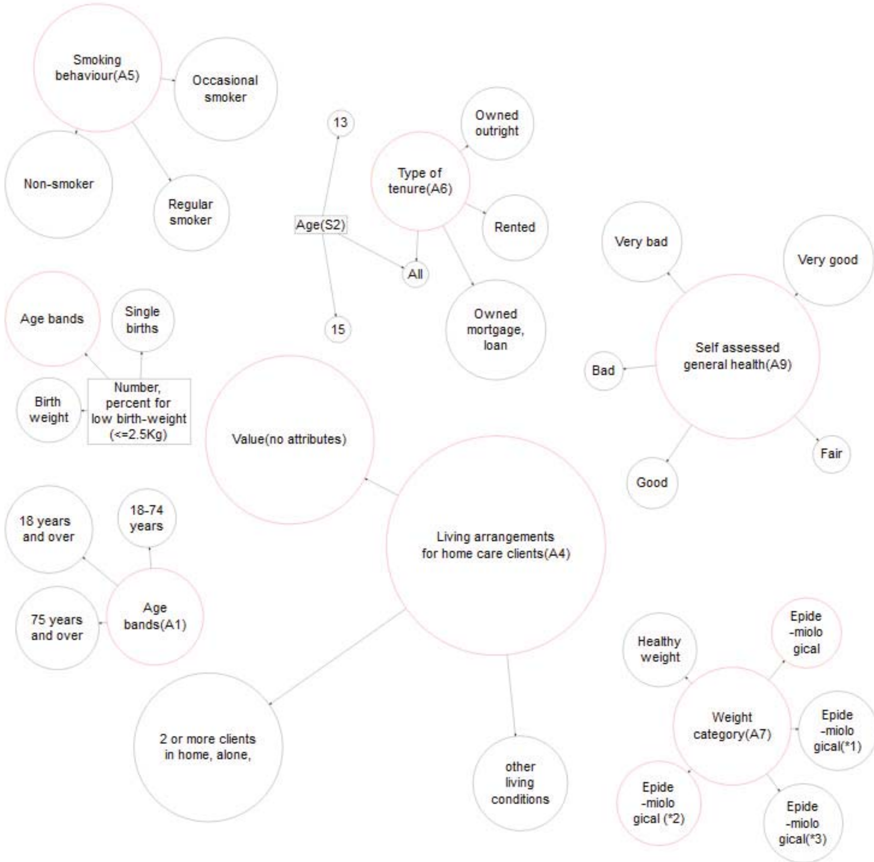


Fig. 1 Block diagram of services, attributes, and levels (values) associations in the data

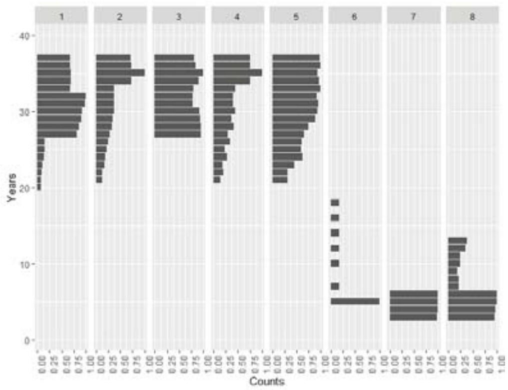
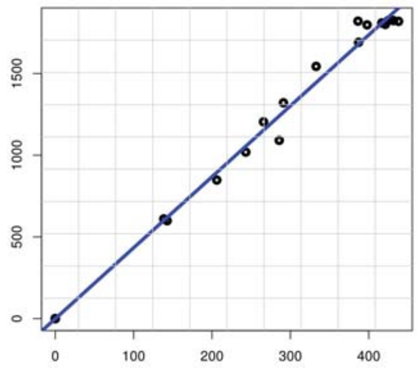
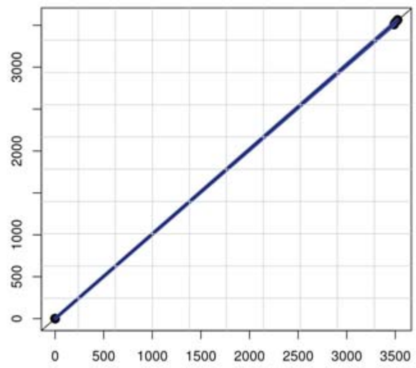


Fig. 2 Representative service plots showing demands summed over all attributes per service offered (22 overall, 8 indicative are shown)

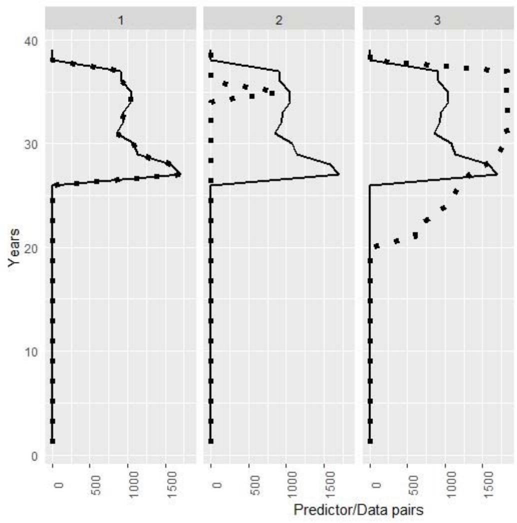
'S17-Health-Fair' (year: 2017) or others with no records after 1997, or, those with a single low attendance before 1997. Most H&Sc factors did not have records then. Also, those H&Sc factors with only very recent records, i.e., after 2017 and not before, like 'S18-OtherLivingConditions-AllLevels', did not relate well (few cases or low CC). The most often observed factors in various BR sets (as BR data-driven predictors) are (S22) (overall, i.e., summed over all attributes and level counts) and smoking-related ones. The factor (S20) (1981-2019) is the target in a combination that had three strong predictors with relatively high probabilities $\{POST.D = 0.5, PIOR.D = 0.36, LKL.D = 0.72\}$ as seen in Table III (row #5). The pack (S2) is also the target in several combinations (rows #3 and #5 in Table III are only indicative). It is also interesting to observe how well the 'S8-Gender-Male' and 'S23-Smoking Behaviour-NonSmoker' correlate (high POST.D) which shows the relevance of smoking to alcohol in males. This can be very helpful in general for the planning of resources (for example, the GPs). Another combination for this target had predictors 'S12-Age-16-64', 'S12-Age-All', with probabilities $POST.D = 0.025, POST.D = 0.046$ (not shown in tables or figures). The less supported group is 'S8-SIMDquintiles-1', 'S23-SmokingBehaviour-RegularSmoker', 'S8-Smoking Behaviour-OccasionalSmoker' ($POST.D = 0.06$) group that shows that alcohol-related admissions are not well jointly related to both smoker categories (regular and occasional smokers). Row #3 has as a target 'S8-Smoking Behaviour-OccasionalSmoker' with three strong data-driven probabilities $\{0.594, 0.35, 0.719\}$ while the target is the same with row #5 ($\{0.5, 0.36, 0.72\}$) which has four predictors with similar three strong probabilities. The factors in the pack (S2) were common as dependent variables and were connected to several other factors, as also discussed by [28], where the 30-day and 48-hour re-admission risks are computed using seven reasons/factors that were not in the PHS data processed



a



b



c

Fig. 3 Plots for indicative linear connections found $(HSc_{factor_i} = LR(HSc_{factor_j}), i <> j)$

and use the ARMA method, which considers re-admission drivers such as the number of re-admissions in the past 12 months. According to [16] and [29], the number of factors is actually, a parameter to adjust, which in our case is fixed, i.e., 110. It was found that among very good independent factors, the strongest belonged to the packs: (S20) and (S23) with many likely dependencies (that is, relatively high posteriors). The Bayesian relationships in rows #3 and #5 connect two or three cohorts: (1) 'S9-WeightCategory-HealthyWeight', (2) (S28), and (3) (S16) Indeed, this can be expected, as such causes are dominant in hospital admissions and are at the root of social problems. Also, a well-matched pack whose factors are often used as independent predictors is (S11), as in rows (#2, #3, #4). This can be so because the mental problems (pack (S15)) cannot be isolated from smoking (packs (S2) and (S10), etc.), and they might be related to a range of alcohol-relevant public services or patient cohorts. One of the factors was 'Percent of people aged 65+' who are admitted as an emergency to hospitals at least twice within 12 months (part of the pack (S2)) alone has connection POST.D probabilities $0.026, 0.035$, respectively, to its Bayesian predictors (S12) and (S16) that is not listed in Table I. In row #9, it can be seen that the pack (S11) and (ID=25 from the 110) are linked to the distance-health pack ('S22-HomeCare-ClientGroup4') which reveals the relationship between remote healthcare and mental problems. The packs (S1) (1998-2010) and (S3) (2008-2019) had very low overlap, and although brought into the same span after zero-padding, they were not found to be well correlated as a pair, but they were with other services. An example is (S13) (2007-2017), while (S1) is well connected, in the BR context, with many but not with (S3). The service pack, (S9), and especially its factor 'Any.TypeOfTenure' is a well-modeled (predicted) factor and creates (where it is common) patient categories (example is 'S9-TypeOfTenure-OwnedOutright') as it can be seen in Fig. 5 with a POST.D probability, 0.521. This knowledge update (an increase in belief) from PRIOR to POST is also roughly seen in the ND and BD models.

TABLE I
INDICATIVE SERVICES OF FIG. 4 WITH DATES OF NO MISSING RECORDS AND THEIR ACRONYMS

H&Sc full names	A^b	B^b	C^c
Alcohol use among young people	S1	1998-2010	3
Headcount of General Practice Workforce	S3	2008-2019	1
Living Arrangements for Home Care Clients	S5	2005-2009	2
Number of single Rooms in care homes	S7	2007-2017	2
Home care services	S9	2005-2009	2
Mental wellbeing by tenure, household type, age, sex, disability	S11	2014-2017	4
Number of care homes by type of provision	S13	2007-2017	6
Places in care homes with en-suite facilities	S15	2007-2017	2
Smoking prevalence among 13 and 15 year olds in Scotland	S17	2001-2019	2
% of children classed healthy weight, overweight, obese,severely obese at Primary 1 review	S19	2002-2015	2
Drug use among 13 and 15 year olds in Scotland	S21	2002-2015	2
Repeated emergency admissions	S2	1998-2010	1
Number of home care clients by care type or disability	S4	2005-2009	3
Intensive Home Care	S6	2002-2011	1
Drug related hospital discharges	S8	1996-2018	2
Number, percent, for low birth weight (¡=2.5Kg) for single births	S10	2000-2019	1
Number of general practices (GPs) with registered patients	S12	2007-2019	1
Occupancy rate in care homes by type of provision	S12	2007-2017	2
Body mass index (BMI) distribution of primary 1 education children	S16	2001-2019	1
Delayed discharges: monthly census	S18	2016-2020	2
Alcohol-related admissions (stays) or discharges	S20	1981-2019	20
Health care clients	S22	2016-2019	2

^{a} Acronyms for services names, ^{b} years of existing records, ^{c} number of attributes tracked per service

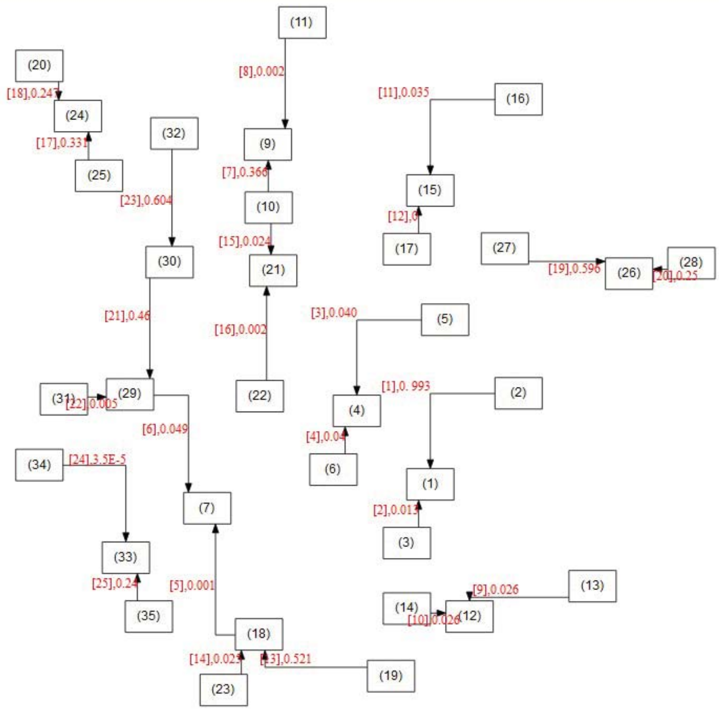


Fig. 4 Data-driven factors (services) relationships diagram for sampled (Smoking-related) factors and names of them as defined in Table I

An innovation of this work is the combination of the three BR approaches and the notion of cohorts that can include services and patients (not only services or only patients) using prediction. As observed in Fig. 5, these relationships change over the years. Table III presents the results for a year, $t = 38$, or the year, 2018. This suggests a learning ratio of $37/39 >= 90\%$ (i.e., 0.9), which is more than the necessary one ([0.6, 0.9]). This is advocated by [30], where the learning ratios are discussed with respect to the quality of the prediction. Lower learning ratios (or fewer years from the past) were possible in the region ([0.2, 0.8]), due to the smoothness of the data. Independently of BR, the PCA analysis was also applied and defined a feature space in the wider H&Scdata space. The most important services that

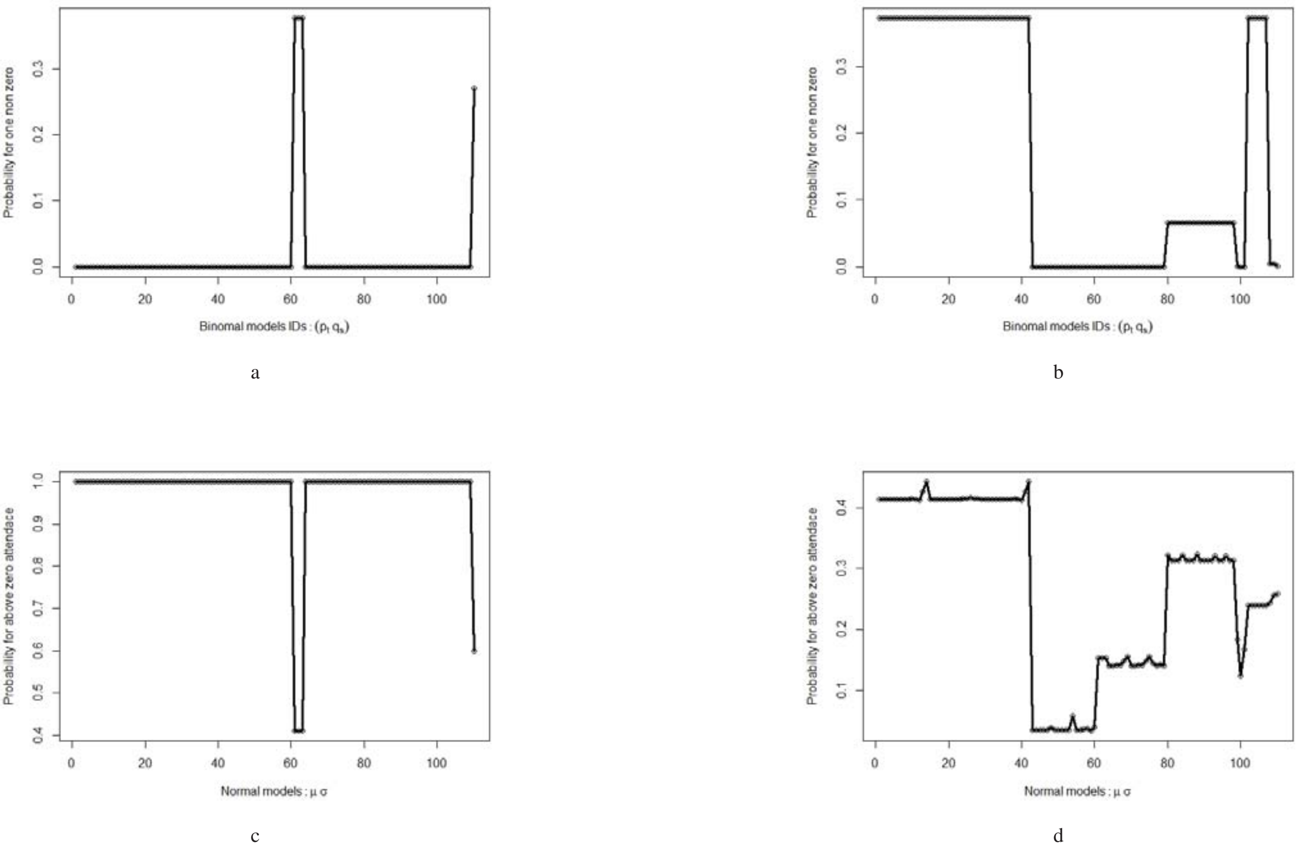
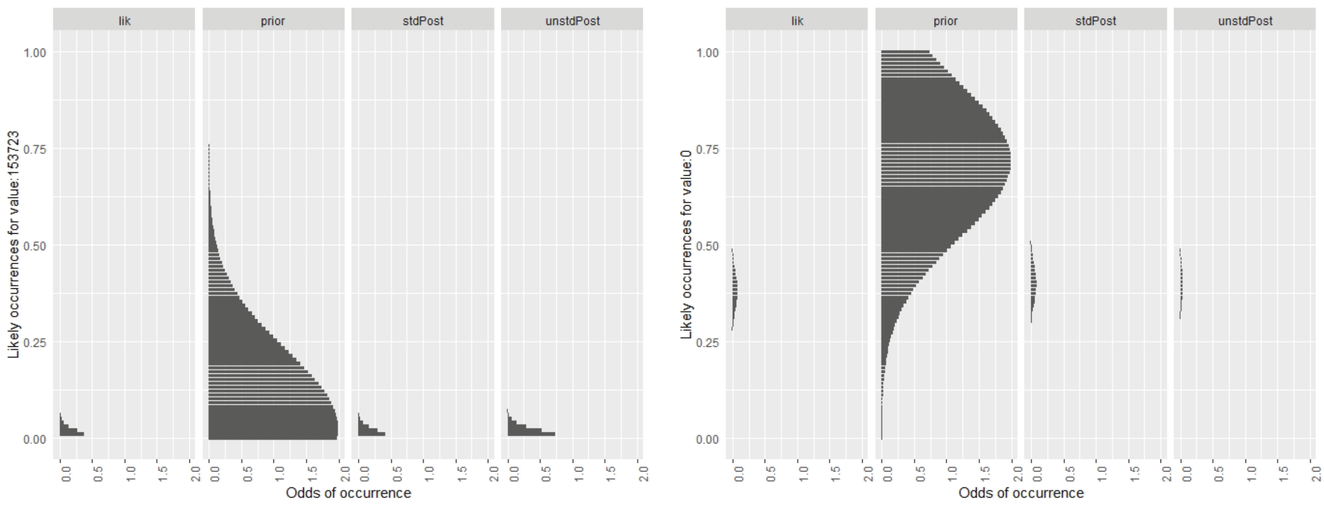


Fig. 5 How specific service attendances that differ in scale (0) and (151564) are modeled using a ND model

were found in the pack (H&Sc data frame) (S23) were five H&Sc factors explaining 56.8% of the data variance, while in the pack (S20), 20 H&Sc factors were found, which explain 28.3% of the variance. The rest of the packs contributed a low percentage to the PCs and were less than 2% or 3%. That is, they contribute to PCs much less. PCA analysis could reduce the data dimensions of the feature space for the classification of the factors so that BR can be applied in a lower dimensional space. PCA was applied to the 110 factors and gave 11 major eigendirections. Alcohol-related factors dominated the major PCs as they are more populated (20 attributes). The reasons for hospital admission due to alcohol are more frequent, thus dominant due to their more likely variance. BR was facilitated by PCA when the services were represented in a PC's sub-space. For example, services ('S11-Gender-All') and ('S9-TypeOfTenure-All') predict (S2)'s services in rows #3 and #4 in Table III using the same data packs (as in row #3). Other factors span more clusters, like (S12), that concern patients and services that are linked to GPs, which can be more diversified since GP visits can be for different reasons. Not all interesting cases are illustrated in figures or tables due to space limitations. PCA looks mainly for independent data (PCs that are orthogonal), while BR looks for data with shared components or common components and uses them as predictors, according to [31]. PCA suggests, as in Table II, that the services pack (S20) has all the PCs, which is also confirmed in Table II, where it can be seen that in many combinations, the pack (S20) is a popular service. It can be seen that BR in Fig. 5 and 6 rather support smoking-related services as being more likely taken in the same years, and PCA rather focuses on alcohol-related services that are quite variant in their usage year patterns. This makes them good PCs. It can be seen in Fig. 6 that given the high odds of having no records for any service (HSc = 0), the odds of having a zero value (that is, the upper part of the curves in Fig. 5 (a)) are higher than the odds of having higher (non-zero) values (illustrated by the right panel in Fig. 5 (b)). As discussed, every time a new year is considered, new models are developed for all three Bayesian distributions. These odds are defined by the time-evolution of the above-mentioned distributions. The left panel in Fig. 5 shows that the likelihood (plot) for zero attendance is below (P = 1/2), while the prior (in the second plot) is normalized, and, as discussed, does not depend on the observed data. The posterior that is shown in the plot is slightly shifted to the upper probabilities considering that the posterior is an update of the prior due to the presence of data as in (6). The results revealed that BR and CC can link up to about four services, while works on LR and AMRA methods were referenced for qualitative (context) comparisons. In



a b

Fig. 6 The 3 Bayesian distributions (“LKL”, ”PRIOR”, ”POST”)

regards to the accuracy and sensitivity of the prediction it was found that data-driven BR is more biased (higher chances) towards zero or non-zero attendances and that model-based (ND, BD) BR versions were not as good at predicting zero but could cover (predict) more cases (not in the data at hand). Model-based approaches under BR better tracked no zero-zero changes with respect to data-driven ones. This can be useful in a progressively accurate classification/prediction schema that narrows down the event space into progressively narrower bands. Also, BR/BD cannot work well on common years (only non-zero record years and no changes) unless the examined period contains some transitions (years of missing records followed by the first year of recording) that are more suited for data-driven approaches (data observed, then data not observed/missing). ND/BR-BD/BR methods worked better on low dimensions (few years) either under PCA or without since data driven needs (no model) need longer strings of years with transitions more likely to happen. PCA yielded the 11 best H&Sc factors, and CM defined five main classes across the 39 years. The BR methods proved that services are uncertain and may depend on factors such as the year the data were recorded, according to [13]. Some H&Sc factors were found to be widely attended, such as the services related to the emergency department that are highly cross-correlated with fewer attended H&Sc factors. The work revealed that the services that are more common as predictors are related to ‘Alcohol Admissions‘ as for example (S20) and home-based services ((S11), (S12), (S14), etc.), confirming that these are common reasons for getting admitted to a hospital and that services may expand and differentiate once a patient is originally admitted for one of these reasons. Moreover, the HC system has grown around services offered to the elderly or to home-based users, since many services in those cohorts are offered from a distance and are BR-related. Depending on the year at hand, though, the *‘....low birth weight (weight* $< 2500g$*‘* is BR related to mental health patients as per [32] and GPs workforce, who help patients who were self-assessed as being well (SALSUS). Among other findings, low birth weights are related to people who are offered housing on a voluntary basis in care homes, and both are related to patients that are registered with GPs and live in adult-type care homes. These offered links across the data that were not expected or clinically justified. The merits of using BR is that it can offer out-of-the-box solutions that may offer insight into hidden data relationships.

It can be seen in Fig. 6 that given the high odds of having no records for any service (HSc = 0) the odds of having a zero value (that is, the upper part of the curves in Fig. 6 (a)) are higher than the odds of having higher (non-zero) values (illustrated by the right panel in Fig. 6 (b)). As discussed, every time a new year is considered, new models are developed for all three Bayesian distributions. These odds are defined by the time-evolution of the above-mentioned distributions. The left panel in Fig. 6 shows that the likelihood (plot) for zero attendance is below ($P = 1/2$) while the prior (in the second plot) is normalized, and, as discussed, does not depend on the observed data. It is based on the ND model, which counts the increased occurrences of missing data for that service. The posterior that is shown in the plot is slightly shifted to the upper probabilities considering that the posterior is an update of the prior due to the presence of data as in (6). Considering the BR method, the ”POST” is a better-informed probability and represents what is expected when we try to predict. With BR, we had three factors involved, the Bayesian likelihood, the prior, and the posterior, while with LR and AR, we could have as many as four (for confident predictions). The BR did not offer the numerical accuracy of the LR or AR as it only checked zero padding, But it was more robust to detect zeros even if one factor’s past samples did not have many zeros. The BR can be tailored (adjusted), though, to model any ”event” as a ”success” or a failure, and it is not limited to yielding exact values as the LR or AR are. This partially offered the previous

TABLE II
H&SCACRONYMS AND DATES GROUPS

Full Description	Acronyms	Years	number of attributes
Alcohol use among young people	S1	1998-2010	3
Headcount of General Practice Workforce	S3	2008-2019	1
Living Arrangements for Home Care Clients	S5	2005-2009	2
Number of single Rooms in care homes	S7	2007-2017	2
Home care services	S9	2005-2009	2
Mental wellbeing by tenure, household type, age, sex, disability	S11	2014-2017	4
Number of care homes by type of provision	S13	2007-2017	6
Places in care homes with en-suite facilities	S15	2007-2017	2
% of children classed healthy weight, overweight, obese,severely obese at Primary 1 review	S17	2001-2019	2
Drug use among 13 and 15 year olds in Scotland	S19	2002-2015	1
HCArrangements	S21	2002-2015	3
Repeated emergency admissions	S2	1998-2010	1
Number of home care clients by care type or disability	S4	2005-2009	2
Intensive Home Care	S6	2002-2011	1
Drug Related Hospital Discharge	S8	1996-2018	1
Number[percent], low birthweight ($<$ 2500g) babies (single births)	S10	2000-2019	2
Number of general practices (GPs) with registered patients	S12	2007-2019	1
Occupancy rate in care homes by type of provision	S14	2007-2017	2
Body mass index (BMI) distribution of primary 1 education children	S16	2001-2019	1
Delayed discharges: monthly census	S18	2016-2020	2
Alcohol-related admissions (stays) or discharges	S20	1981-2019	20
Health care clients	S22	2016-2019	2

H&Sc services groups(as parts of linear prediction equations) comprising a target(first service) and its predictors (left services) in their full names, their acronyms(inside parentheses as ('S.A.Z') triplets, their dates of recording and the numbers of characteristics(attributes) that accompany them]

TABLE III
BEST HSC FACTORS USING PCA

Factor name	PC(%)	Factor name	PC(%)	Factor name	PC(%)
Smoking prevalence in young people(SALSUS) . Age. 13	69	Smoking behaviour and self rated health(SALSUS) .Gender. Female	14	Smoking behaviour and self rated health(SALSUS) . Self assessed general health . Fair	6.46
Smoking prevalence in young people(SALSUS) . Age. 13	5.07	Smoking behaviour and self rated health(SALSUS) .Gender. Male	3.01	Smoking behaviour and self rated health(SALSUS) . Smoking behaviour . Non Smoker	0.7
Smoking prevalence in young people(SALSUS). Age.All	0.59	Smoking behaviour and self rated health(SALSUS) . Self assessed general health . Very good	0.32	Smoking behaviour and self rated health(SALSUS) . Smoking behaviour	0.29
Smoking behaviour and self rated health(SALSUS) .Gender. All	0.11	Smoking behaviour and self rated health(SALSUS) . Self assessed general health . Bad	0.45		

The factors are represented by triplets $\{X.Y.Z\}$ The dominant services are *'Smoking prevalence and deprivation(SALSUS)'* and *'Alcohol-related admissions (stays) or discharges"'*) ***T2*** .

robustness in capturing zeros (no records) in one factor series when there were no past zeros, provided that the other two predicting factors would likely have zeros.

VII. CONCLUSION

The paper discussed how we can relate H&Sc service attendance using prediction to form service cohorts and evaluated several methods. The dependence of these relationships on classification as well as on service settings was studied using LR, AR, and, Bayesian statistics. All three approaches defined relationships that linked the demands of the services and formed groups. CM provided basic knowledge as to how we can limit the closest domain space for prediction for LR and ARMA. The results revealed that LR or ARMA linearity holds for up to about four services and that LR works better than ARMA in regard to the accuracy of the prediction. Also, BR's zero or non-zero attendance odds are better supported (have higher posteriors to be real zeros or non-zero events) with respect to LR/ARMA where there is an error. BR better tracks no zero-zero changes with respect to LR. This robustness comes at the price of BR (BD method) being limited to deciding on ranges and not exactly predicted attendances for the factors. This can be useful in a progressively accurate classification/prediction schema that narrows down the event space into progressively narrower bands. Also, BR cannot work well on common years (only non-zero record years) unless the examined period contains some transitions (years of missing records followed by the first year of recording) that are more suited for linear relationships. LR methods worked better on low dimensions (few or selected years). AR models proved less successful with respect to LR, as seen in the high RMSE, MAE, and MRE errors obtained. The first groupings were found based on CC or CM and were further explored using LR and ARMA, which changed over the years. PCA yielded the 11 best H&Sc factors, and CM defined five main classes across the 39 years. The LR methods proved that services are uncertain and may depend on factors such as the year the data were recorded in as per [13]. Some H&Sc factors were found to be widely attended, such as the emergency department related ones, and highly cross-correlated with fewer H&Sc factors. The work revealed that the services that are more common as predictors for other services are related to 'Alcohol Admissions' as for example (S20) and home-based services (various services: (S11), (S12), (S14), etc.), and confirmed that these are common reasons for getting admitted to a hospital and that services may expand and differentiate once a patient is originally admitted for one of these reasons. Moreover, the HC system has grown around services offered to the elderly or to home-based users, as seen by the plethora of services offered from a distance and their participation in more service groupings. The high specialization of services offered to alcohol-related patients was confirmed by the high linear confidence attached to such H&Sc factors as low birth weights and services related to alcohol. Depending on the year at hand, though, the '....low birth weight (weight $<$ $2500g$)' class can also be regressed (linearly related) with mental health patients as discussed by [32]. It was also found that GPs workforce could be related to patients who were self-assessed as being well (SALSUS). Among other findings, low birth weights are related to the

TABLE IV
LINEAR SETS AND BAYESIAN SETS AS ILLUSTRATED IN THE RELATIONSHIPS DIAGRAM OF FIG. 4

Linear/Bayesian groups of H&Scs	Er1	LR0 LR1 LR12	PR0 PR1 PR2	Er2 Er3	Er2	AR0 AR1 AR2	Er2 Er3	POST.D POST.ND POST.BD	PRIOR.D PRIOR.ND PRIOR..BD	LKL.D LKL.ND LKL.BD	AVEPOST
(1) 1,2,3	0.6 4 1	0.02 0.006 -2E-4	0.993 0.013 0.040	0 1	8. 8 5 2	-2.7E-3 -2.7E-3 8E16	0 1	0.06 1E-5 0.447	0.005 3E-16 0.302	0.103 5.4E-37 1.11E-7	0. 0 8 7
(2) 4,5,6	0.9 4 5	25264 40.6 184	0.04 0.001 0.049	0 1	67609	0. 9 3 -7.3E3 37797	230 4	0.594 417E-27 0.253	0.35 735E-39 0.105	0.077 0.7187 3.5E-178	0. 087
(3) 12,13,14	0.334	-13.6 0.042 0. 263	0.366 0.002 0.026	0 1	7.61	-0.3 0. 9 9 0.195	230 4	0.421 417E-27 0. 2 5 3	0 .35 735E-39 0. 1 0 5	0.7187 3.5E-178 3E-6	0. 2 1 2
(5) 15,16,17	0. 0 6 6	26445 -7.62 0. 973	0.002 0.331 0. 247	2 1	4.37	0.89 -1.5E-1 59E-2	22380 0	0.488 4.7E-4 0. 0 2 6	0.5 6.3E-12 0. 0 7 5	7.40E-6 0 . 9 7 4 3E-69	0. 0 3 5
(7) 18,19,23	0.049	1.654 -6.53 -0.0 0 2	0.5964 0.25 0.46	0.608 1	2.1E-9	0.962 -0.061 -0.041	10309 0	0.5 3E-38 0.026	0.32 8E-42 0.075	0.974 3E-69 0.043	0. 212
(8)21,10,22	0. 389	0.923 -1E-4 5E-4	0.005 0.604 3.5E-5	0.6 1	2.1E-8	0.962 -0.061 -0.041	1.193 1	0.5 1.7E-32 1.7E-32	0.321 1.2E-35 0.075	0.64 2E-09 0.758	0.196
(8)21,10,22	0. 389	0.923 -1E-4 5E-4	0.005 0.604 3.5E-5	0.6 1	2.1E-8	0.962 -0.061 -0.0 4 1	1.193 1	0.5 1.7E-32 0.026	0.321 1.2E-35 0.075	0.64 2E-09 0.758	0.196
(9) 24,25,20	0.013	34.5 9 -7.37 -3E-4	0.24 0.604 0. 977	1 7.6	0. 989	0.195 1E-23 -0. 3 2	3.9 0	0.053 3.4E-4 0.0263	0.083 1.4E-9 0.118	0.641 1E-12 1E-12	0.321
(12) 33,34,35	0.998	24. 2 0.406 -0.139	0.615 0.675 1E-23	0.45 0.83	2E-14	1 -0.03 -0.03	-0.03 0	0.5 1.3E-12 0.221	0.346 2.5E-20 0.121	0.692 1.9E-137 0.171	0.488

The relationships use the IDs defined in figure 4

people who are offered housing on a voluntary basis in care homes, and both are linearly related to the patients that are registered with GPs and live in adult-type care homes. These may offer links across the data that were not expected or even justified. The merits of using ML are that it can offer out of the box solutions that may offer insight into hidden data relationships.

ETHICAL APPROVAL

There are is no animal experimentation in the manuscript and no needed to be obtained. No studies on humans were carried out.

DATA AVAILABILITY

The data used in this work were made freely available on-line by PHS as an open database. The link is provided in the references.

FUNDING

The author received no specific financial support for the authorship and/or publication of this article. The author during the works of this paper was funded by an Abertay University, Dundee stipend.

AUTHORS CONTRIBUTIONS

The author is the only writer of the manuscript and carried out all the research. Other help or contributions are acknowledged as well as the data used.

REFERENCES

[1] ByXu., HRISTINA PASHOVA† AND PATRICK J. HEAGERTY (2017), *Comparing Healthcare Utilization Patterns Via Global Differences in the Endorsement of Current Procedural Terminology Codes* , The annals of applied statistics, Vol. 11, no. 3, 1349–1374, doi: 10.1214/17 − aoas1028
[2] Simon Bottery . https : //www.kingsfund.org.uk/about − us/whos − who/simon − bottery?page = 2 . The King's Fund
[3] *Public Health Scotland (2020)* . Data and intelligence. A – Z Subject Index. https : //www.isdscotland.org/A − to − Z − index/index.asp
[4] Scottish Government (2019). *Statistics Service Health and Social Care Data* . https : //statistics.gov.scot/datahome
[5] Vittorio Lippi (2019). *Incremental Principal Component Analysis: Exact implementation and continuity corrections.*arXiv: 1901:07922v2; stat:ML; 13May2019. https : //arxiv.org/pdf/1901.07922.pdf
[6] Ian Litchfield (2019). *Can pathways of patients with long-term conditions in UK primary care? A study protocol. BMJ* Open, 2018. https : //bmjopen.bmj.com/content/8/12/e019947
[7] Dimitris Bertsimas, Colin Pawlowski, Ying Daisy Zhuo (2018) *From Predictive Methods to Missing Data Imputation: An Optimisation Approach,* Journal of Machine Learning Research 18 (2018) 1-39

[8] E.M. Mirkes (2018), T.J. Coats, J. Levesley, A.N. Gorban (2018). *From Predictive Methods to Missing Data Imputation: An Optimisation Approach.* Journal of Machine Learning Research, 18 (2018),1-39. $http://dx.doi.org/10.1016/j.compbiomed.2016.06.004$

[9] deRooij M. (2018). *Transitional modelling of experimental longitudinal data with missing values.* Adv Data AnalClassif, 12,107–130. $https://link.springer.com/article/10.1007/s11634-015-0226-6$

[10] Muge Capan (2020), Stephen Hoover, et al. (2019), *Time Series Analysis for Forecasting Hospital Census: Application to the Neonatal Intensive Care Unit Multitask learning and benchmarking with clinical time series data,* Appl. Clin. Inform.. 2019,7(2):275–289. $https://dx.doi.org/10.4338\%2FACI-2015-09-RA-0127$

[11] Langton, J.M. (2018), Wong, S.T., Burge, F. et al. (2015), *Population segments as a tool for health care performance reporting: an exploratory study in the Canadian province of British Columbia,* BMC Fam Pract,21-98(2020). $https://doi.org/10.1186/s12875-020-01141-w$

[12] Vimal Mishra (2019), MD, MMCI, Shin-Ping Tu, MD, MPH, Joseph Heim, PhD, Heather Masters, MD, Lindsey Hall, MPH, Ralph R. Clark, MD, Alan W. Dow, MD (2019), *Predicting the Future: Using Simulation Modeling to Forecast Patient Flow on General Medicine Units,* J. Hosp. Med.,2019,1,9-15. doi:10.12788/jhm.3081

[13] Guersel, Gueney (2019), *Healthcare, uncertainty, and fuzzy logic. Digital Medicine,* 2016,2,101-12. $https://www:researchgate:net=publication=310817255Healthcareuncertaintyandfuzzylogic$

[14] Deborah A.Marshall, LinaBurgos-Liz et al. (2015), *Applying Dynamic Simulation Modeling Methods in Health Care Delivery Research—The SIMULATE Checklist:Report of the ISPOR Simulation Modeling Emerging Good Practices Task Force,* Value in Health, volume 18,Issue 2,March 2015,143-144. $https://doi.org/10.1016/j.jval.2014.12.001$

[15] D. Ben-Tovim, J. Filar, et al. (2019), *Hospital Event Simulation Model: Arrivals to Discharge,* 21st International Congress on Modelling and Simulation. Gold Coast,Australia. $https://www.mssanz.org.au/modsim2015/H2/bentovim.pdf$

[16] David J Spiegelhalter, Jonathan P Myles, David R Jones(1999) *An introduction to bayesian methods in health technology assessment* , BMJ 1999; 319, doi: https://doi.org/10.1136/bmj.319.7208.508

[17] Anupreet Porwal, d Adrian E. Rafterya *Comparing methods for statistical inference with model uncertainty* , PNAS 2022 Vol. 119 No. 16 e2120737119, $ttps://www.pnas.org/doi/pdf/10.1073/pnas.2120737119$

[18] Gredell Devin (2019), *Comparison of Machine Learning Algorithms for Predictive Modeling of Beef Attributes Using Rapid Evaporative Ionization Mass Spectrometry (REIMS),* Data. Sci Rep.,9 5721 (2019). $https://pubmed.ncbi.nlm.nih.gov/30952873/$

[19] Bebbington E, Furniss, D . (2015), *Linear regression analysis of Hospital Episode Statistics predicts a large increase in demand for elective hand surgery in England,* J. Plast. Reconstr. Aesthet. Surg, 2015, Feb,68(2),243-51. doi:10.1016/$j.bjps$.2014.10.011

[20] Uematsu, H., Yamashita, K., Kunisawa, S., Otsubo, T., & Imanaka, Y. (2017), *Prediction of pneumonia hospitalization in adults using health checkup data,* PloS one,12(6),e0180159. $https://doi.org/10.1371/journal.pone.0180159$

[21] Juang WC, Huang SJ, Huang FD, Cheng PW, Wann SR. (2017), *Application of time series analysis in modelling and forecasting emergency department visits in a medical centre in Southern Taiwan,* BMJ Open, 2017, Dec 1,7(11),e018628. DOI: $10.1136/bmjopen-2017-018628$

[22] Harutyunyan, H., Khachatrian, H., Kale, D.C. et al. (2019), *Multitask learning and benchmarking with clinical time series data,* Sci .Data,6,96(2019).$https://doi.org/10.1038/s41597-019-0103-9$

[23] Shivapratap Gopakumar (2016), Truyen Tran, Wei Luo, Dinh Phung, *JMIR Medical Informatics* 4(3):e25 . DOI: $10.2196/medinform.5650$.

[24] Bui C., Pham N., Vo A., Tran A., Nguyen A., Le T. (2017), *Time Series Forecasting for Healthcare Diagnosis and Prognostics with the Focus on Cardiovascular Diseases,* Vo Van T.; Nguyen Le T.;

[25] Liew, B.X.W., Peolsson, A., Rugamer, D. et al. (2020), *Clinical predictive modelling of post-surgical recovery in individuals with cervical radiculopathy: a machine learning approach,* Sci.Rep,10,16782(2020). $https://doi.org/10.1038/s41598-020-73740-7$

[26] Dunsmuir WT (2019), *Dangers and uses of cross-correlation in analyzing time series in perception, performance, movement, and neuroscience: The importance of constructing transfer function autoregressive models,* Behav Res Methods,2016,Jun,48(2),783-802. DOI:10.3758/$s13428-015-0611-2$

[27] Yang, C., Delcher, C., Shenkman, E. et al. (2019), *Expenditure variations analysis using residuals for identifying high health care utilizers in a state Medicaid program,* BMC Med Inform Decis Mak, 19,131(2019). $https://doi.org/10.1186/s12911/019/0870/4$

[28] Daniel J. Morgan, Bill Bame, Paul Zimand, et al. (2019), *Assessment of Machine Learning vs Standard Prediction Rules for Predicting Hospital Readmissions,* JAMA Netw Open,2019,Mar,2

[29] Marno Verbeek, *A Guide to Modern Econometrics,* John Wiley & Sons. DOI10.3917/rfs.593.0475. $https://www.researchgate.net/publication/227488993\_A\_Guide\_to\_Modern\_Econometrics$

[30] Aitor Lewkowycz and Ethan S Dyer and Guy Gur-Ari and Jascha Sohl-dickstein and Yasaman Bahri (2020) . *The large learning rate phase of deep learning,* ICLR 2021 Conference . $https://arxiv.org/abs/2003.02218$

[31] Liu C, Zhang X, Nguyen TT, et al. (2021), *Partial least squares regression and principal component analysis: similarity and differences between two popular variable reduction approaches,* General Psychiatry,2022; $35:e100662(2021)$. doi: $10.1136/gpsych-2021-100662$

[32] Lyall DM, Inskip HM, Mackay D, Deary IJ, McIntosh AM, Hotopf M, Kendrick T, Pell JP, Smith DJ. Low birth weight and features of neuroticism and mood disorder in 83545 participants of the UK Biobank cohort . BJPsych Open. 2016 Jan 28,2(1):38-44. DOI:10.1192/ $bjpo.bp$.115.002154. $PMID:27703752$, $PMCID:PMC4995581$