SORT 49 (2) July-December 2025, 1-20 DOI:

Bayesian estimation for conditional probabilities associated to directed acyclic graphs: study of hospitalization of severe influenza cases

Lesly Acosta[1,*] and Carmen Armero[2]

Abstract

This paper presents a Bayesian framework to estimate joint, conditional, and marginal probabilities in directed acyclic graphs to study the progression of hospitalized patients with confirmed severe influenza. Using data from the PIDIRAC retrospective cohort in Catalonia, we model patient pathways from admission to discharge, death, or transfer. Transition probabilities are estimated using a Bayesian Dirichlet-multinomial approach, while posterior distributions for absorbing states or inverse probabilities are assessed via simulation. Bayesian methodology quantifies uncertainty through posterior distributions, offering insights into disease progression and in improving hospital planning. These findings support more effective patient management and informed decision making during seasonal influenza outbreaks.

MSC: *62F15, 62P10.*

Keywords: *Confirmed influenza hospitalization, Directed acyclic graphs (DAGs), Dirichlet-multinomial Bayesian inferential process, Healthcare decision-making, Transition probabilities.*

1. Introduction

According to the World Health Organization (WHO): "Seasonal influenza (the flu) is an acute respiratory infection caused by influenza viruses common in all parts of the world". Following official estimates, about 1 billion cases of seasonal influenza occur worldwide each year. This includes between 3 and 5 million cases of severe illness and between 290 and 650 thousand respiratory deaths caused by the disease (Pietrasik, 2023). Influenza rates are very high in children but its mortality records are shocking in elderly

* Corresponding author: Lesly Acosta; lesly.acosta@upc.edu

[1] Department of Statistics and Operations Research, Universitat Politècnica de Catalunya, Barcelona-TECH.

[2] Department of Statistics and Operations Research, Universitat de València, Burjassot, Spain.

Received: April 2025

Accepted: October 2025

populations as well as in people affected by chronic diseases (Paget et al., 2022). It is a worldwide cause of hospital admissions and mortality in these latter groups (Macias et al., 2021).

Influenza prevention and control remains a serious public health challenge, despite the availability of vaccines and antiviral treatments (Carrillo-Santisteve et al., 2012). The European Centre for Disease Prevention and Control (ECDC) is an agency of the European Union (EU) that collects epidemiological and virological data from member countries of the European Economic Area (EEA). Surveillance data come from the sentinel influenza surveillance systems of each associated country, which may cover substantial parts of the population or even have a universal surveillance system (Snacken and Brown, 2015).

Although most people affected by the seasonal influenza recover within one or two weeks without medical attention, it can cause serious illnesses and mortality, especially among population at higher risk. Severe influenza complications can result in hospitalization, possibly with admission to the ICU or even death (Acosta et al., 2021; Soldevila et al., 2021). According to ECDC, around 10 to 30% of Europe's population is infected annually with influenza, causing hundreds of thousand hospitalizations. A systematic review of the clinical burden of influenza disease in older people was done by Langer et al. (2023) with data from January 2012 to February 2022.

There are almost no studies that quantitatively analyse the different health conditions that hospitalised patients with severe influenza can experience from admission to discharge. Knowledge of these pathways would be a valuable tool for improving hospital resource planning and organization of the seasonal period of influenza, the winter. In Europe, influenza generally causes annual epidemics that affect up to 20% of the population.

Graph theory is a very theoretical mathematical subject with an enormous power to visualize the basic functioning of scenarios that operate in environments with many sources of uncertainty. Our approach to graphs is essentially graphical and structural. In particular, the evolution of a hospitalized patient from admission to discharge can be represented graphically by means of a probabilistic directed acyclic graph (DAG) (Cowell et al., 1999; Barber, 2012) with nodes defined by random events associated to the different health conditions of the inpatients and arrows connecting two consecutive nodes without any possibility of return. Transition probabilities between nodes are conditional probabilities that provide valuable clinical information on the state of health in which individuals move from their current status. They are the basis for assessing the uncertainty associated with the different trajectories of the study, the final (absorbing) states, and inverse probabilities that inform previous events.

This paper presents a general inferential procedure for estimating joint, conditional, and marginal probabilities in probabilistic DAGs associated to random events, which we apply to assess the different pathways that a patient with severe influenza may follow from their admission to the hospital to their discharge, as fully cured, dead or sent to a long-stay facility. These latter type of institutions usually welcome patients with chronic

diseases and comorbidities who have little hope of cure, but require continued clinical care.

All the inferential processes in this work are framed within the Bayesian inferential methodology. This will allow to directly quantify the uncertainty associated with the relevant outcomes through probability distributions. In particular, posterior distribution associated to transition probabilities or absorbing states will allow us to better understand the hospital evolution of patients with severe influenza. Overall, they may be a useful tool in the effective management of patients hospitalised with influenza during peaks of influenza epidemic activity.

This paper is organized as follows. Section 2 presents a description of the motivated problem, named the PIDIRAC cohort study, and data as well as the DAG that represents the evolution of a severe influenza patient in hospital. Section 3 introduces the general Bayesian modeling for assessing conditional probabilities for adjacent and non adjacent states of a DAG. Section 4 applies the general approach in Section 3 to the data of the PIDIRAC study. Section 5 provides a more general framework for modelling conditional probabilities through covariates and illustrates its application in a particular probability distribution from the PIDIRAC project. Finally, Section 6 contains some general conclusions and comments.

2. PIDIRAC retrospective cohort study

This study focuses on hospitalized severe influenza patients. Data for the analysis were collected from a retrospective cohort study of hospitalized, laboratory confirmed, influenza (SHLCI) patients registered from 1 October 2017 to 22 May 2018 by the 14 hospitals included in the Primary Care Influenza Surveillance System of Catalonia (PIDIRAC). Catalonia is an autonomous community of Spain located in the northeast of the Iberian Peninsula with a population of around 8 million inhabitants. The median and interquartile range age of the hospitalized patients was 72 and $59-83$ years respectively, with 563 (43%) of all being female.

All severe influenza patients who came to the hospital were initially attended by a physician who, depending on the patient's state of health, recommended admission to an intensive care unit (ICU, denoted by I when treated as a variable) or to a specific hospital ward ($W1$). Some of the patients who were initially directed to $W1$ were later transferred to the ICU, derived to a long-term care facility (L), died (D) or were cured and discharged from the hospital and sent home (H). Patients in ICU can die or, if they improved, be sent to a second type of ward ($W2$) of the hospital, from where they can move to H, L or D.

Consider a graph $G = (S, \mathcal{C})$ with nodes $S = \{A, I, W1, W2, D, H, L\}$ the different states or services in the hospital and ordered arcs

$$\mathcal{C} = \{AI, AW1, W1I, IW2, W1D, W1H, W1L, W2D, W2H, W2L\}$$

connecting the different adjacent nodes. See Figure 1 for a DAG that represents the evolution of hospitalized patients with severe influenza through nodes describing their

different health states and/or hospital services used and the different directed arcs connecting neighbouring nodes. The nature of the different states is different: The admission of a patient to the hospital is the initial state A, states I, $W1$, and $W2$ are transient, and states D, H, and L absorbing.

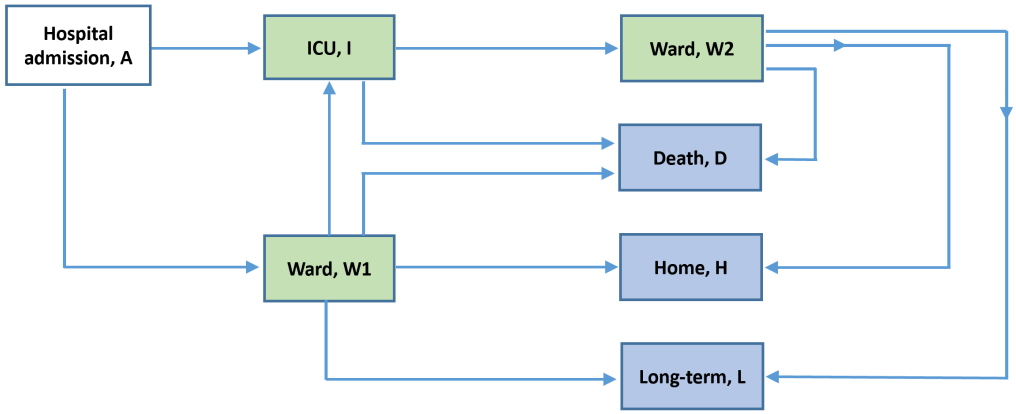

Figure 1. *Directed acyclic graph for a description of the progress of patients with severe influenza in hospital. Transient states in green and absorbing states in blue.*

Over the study period, the number of patients with a diagnosis of severe influenza admitted to the hospital was 1306. Of these patients, 1208 were initially referred to the ward and 98 were sent to the ICU. A total of 82 patients in $W1$ were subsequently transferred to the ICU. Of the total of 1126 patients on the ward not sent to the ICU, 946 were discharged and sent home, 55 were sent to L, and 125 died. A total of 35 patients died during their stay in ICU; and the rest, 145, were sent to $W2$ from where 118 were discharged and sent home, 12 were sent to a long-stay facility and 15 died.

3. Bayesian modeling of conditional probabilities

The probabilistic approach to our model is concentrated on conditional probabilities that assess the uncertainty associated to the different paths between the states of the process: probabilities of visiting a certain state from another given state, directly without intermediate states or not; probabilities of ending in each of the absorbing states; and even, inverse probabilities that focus on assessing the uncertainty associated with a previous state knowing that the process has ended up visiting a certain state subsequently.

3.1. Dirichlet-multinomial learning process for assessing direct conditional probabilities

Assume a graph with a finite set S of states and consider (without loss of generality) that from the state $i \in S$ it is possible to directly visit the states $\{1,\ldots,J\} \in S$ with probabi-

lities $\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{iJ})^\intercal$, where θ_{ij}, $i \neq j$, represents the probability that the individual's departure from state i will be to visit state j directly, with $\theta_{iJ} = 1 - \sum_{j=1}^{J-1} \theta_{ij}$.

Let $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{iJ})^\intercal$ be the random multinomial vector whose component j, Y_{ij}, describes the number of individuals that move directly from state i to state j for the n_i individuals that are in the state i and eventually leave it, where $Y_{iJ} = n_i - \sum_{j=1}^{J-1} Y_{ij}$. The parametric vector associated to this multinomial distribution is $\boldsymbol{\theta}_i$. As an illustrative example, in the PIDIRAC study, a total of 145 patients were recorded in state $W2$, who subsequently transitioned to states H, L, or D. In this case, the multinomial outcome vector of interest, $\boldsymbol{Y}_{W2}$, is three-dimensional, with components Y_{W2H}, Y_{W2L}, and Y_{W2D} representing the numbers of individuals in $W2$ who moved to H, L, and D, respectively. The corresponding parameter vector for inference, $\boldsymbol{\theta}_{W2}$, consists of the transition probabilities from $W2$ to H, L, and D, denoted by θ_{W2H}, θ_{W2L}, and θ_{W2D}, respectively.

The basic Bayesian inferential process for multinomial probabilities $\boldsymbol{\theta}_i$ account for the Dirichlet family of distributions as the conjugate family for the sampling multinomial distribution, $\boldsymbol{Y}_i \mid \boldsymbol{\theta}_i \sim \text{Mn}(n_i, \boldsymbol{\theta}_i)$. This implies that if a Dirichlet distribution $\text{Di}(\boldsymbol{\alpha}_i)$ with parameters $\boldsymbol{\alpha}_i = (\alpha_{i1}, \ldots, \alpha_{iJ})^\intercal$ is chosen as a prior distribution $\pi(\boldsymbol{\theta}_i)$ for $\boldsymbol{\theta}_i$, the corresponding posterior distribution $\pi(\boldsymbol{\theta}_i \mid \mathcal{D})$, where $\mathcal{D}$ stands for the data, will also be a Dirichlet distribution, $\text{Di}(\boldsymbol{y}_i + \boldsymbol{\alpha}_i)$, where $\boldsymbol{y}_i$ is the vector of the observations from $\boldsymbol{Y}_i$. Just as the binomial distribution is the univariate version of the multinomial distribution, the Beta distribution is the univariate version of the Dirichlet distribution. In this regard, the posterior marginal distribution associated to probability θ_{ij}, $\pi(\theta_{ij} \mid \mathcal{D})$, is a Beta distribution with parameters $y_{ij} + \alpha_{ij}$ and $\alpha_i^+ + n_i - (y_{ij} + \alpha_{ij})$, where $\alpha_i^+ = \sum_{j=1}^{J} \alpha_{ij}$ (Congdon, 2005; Armero et al., 2021).

There are many studies where prior information cannot be provided or is available but it is not intended to be used in the inferential process in order to make the analysis more "objective" (Alvares et al., 2018). In these scenarios, the Bayesian inferential protocol establishes as necessary the specification of a prior distribution that serves as a starting point for the inference and that disturbs and distorts as little as possible the information provided by the data. The choice of such scarce informative distribution is not unanimous in the scientific literature and has generated a lot of controversy. We will not go into this issue and will use the Perk's prior Dirichlet distribution (Perks, 1947; Berger, Bernardo and Sun, 2015), whose parameters share the unit of probability equally and is the most widely used Dirichlet distributions for this type of problem. The corresponding posterior distribution is $\text{Di}(\boldsymbol{y}_i + \mathbf{1}/J)$, where $\mathbf{1}$ is now a vector of ones of dimension J.

3.2. Bayes inference for non-direct conditional probabilities

A little more notation should be added to the study to analyse conditional probabilities associated with non-contiguous states. We define $\theta_{ik_1 \ldots k_K j}$ as the probability associated with the path that begins in state i, then visits state k_1, immediately after state k_2, and so on up to k_K, and finally ends in j, where each of the visited states is temporally following its immediate preceding. In the event the state i is temporally earlier than state j, $\theta_{i\bullet j}$

shall stand for the probability of visiting j from i via any path $\mathcal{P}(i,j)$ that connects both states,

$$\theta_{i\bullet j} = \sum_{\mathcal{P}(i,j)} \theta_{\mathcal{P}(i,j)}. \tag{1}$$

Analogously, if state i is temporally later than state j, $\theta_{i\bullet j}$ will represent the probability of having departed from state j knowing that the process has visited the posterior state i. This probability is calculated from Bayes's theorem as follows

$$\theta_{i\bullet j} = \frac{\theta_{j\bullet i}\,\theta_j}{\theta_i},$$

where now θ_i (θ_j) indicates the probability of visiting state i (j) from the initial state of the process.

The conjugate Dirichlet-multinomial learning process is suitable for the computation of posterior distributions associated to jumping probabilities between a state and its immediate next states, but not for posterior distributions associated to probabilities between non-neighbouring states or inverse probabilities. In this sense, we will assume a Markovian structure for the transition probabilities as follows:

$$\theta_{ik_1\ldots k_K j} = \theta_{ik_1}\,\theta_{k_1k_2}\ldots.\,\theta_{k_K j}. \tag{2}$$

This condition allows the simulation of the posterior distribution of the probability associated to this particular trajectory $\pi(\theta_{ik_1\ldots k_K j} \mid \mathcal{D})$. In fact, a simulated random sample $\{\theta^{(m)}_{ik_1\ldots k_K j}, m = 1,\ldots,M\}$ from this posterior distribution is constructed as follows

$$\theta^{(m)}_{ik_1\ldots k_K j} = \theta^{(m)}_{ik_1}\,\theta^{(m)}_{k_1k_2}\ldots.\,\theta^{(m)}_{k_K j},$$

where each simulated value $\theta^{(m)}_{k_ik_j}$ is generated from the Beta marginal posterior distribution, $\pi(\theta_{k_ik_j} \mid \mathcal{D})$, of the transition probability from state k_i to its immediate next state k_j.

In this way, we can compute a random sample from $\pi(\theta_{i\bullet j} \mid \mathcal{D})$ taking into account that

$$\pi(\theta_{i\bullet j} \mid \mathcal{D}) = \pi(\textstyle\sum_{\mathcal{P}(i,j)} \theta_{\mathcal{P}(i,j)} \mid \mathcal{D}). \tag{3}$$

In the case that i is subsequent to j, we can generate a random sample from $\pi(\theta_{i\bullet j} \mid \mathcal{D})$ bearing in mind that

$$\pi(\theta_{i\bullet j} \mid \mathcal{D}) = \pi\Big(\frac{\theta_{j\bullet i}\,\theta_j}{\theta_i} \mid \mathcal{D}\Big). \tag{4}$$

The software R, version 4.4.2 (R Core Team, 2025), has been used to implement the computation of all the transition probabilities results reported in this work. The code and the suplementary material can be found in a GitHub repository (https://github.com/LAcosta15/CALA_SupplementaryMaterial).

4. Hospitalization of severe influenza cases

4.1. Probabilities associated with direct transitions between states

We begin the inferential process of the PIDIRAC study by estimating the probability distribution associated with visiting each of the states that can be accessed from an immediately preceding state. In our research these would be the random vector $\boldsymbol{\theta}_A = (\theta_{AW1}, \theta_{AI})^\intercal$ that assesses the probability that a patient admitted to the hospital will be referred to ward $W1$ or ICU, $\boldsymbol{\theta}_I = (\theta_{IW2}, \theta_{ID})^\intercal$ that accounts for the probability that a patient in the ICU moves to ward $W2$ or dies, $\boldsymbol{\theta}_{W1} = (\theta_{W1I}, \theta_{W1D}, \theta_{W1H}, \theta_{W1L})^\intercal$ as the vector that indicates the possible visits to I, D, H and L from $W1$, and the vector $\boldsymbol{\theta}_{W2} = (\theta_{W2D}, \theta_{W2H}, \theta_{W2L})^\intercal$ formed by the probability that a patient ends up in D, H or L since his second stay on the ward, $W2$. In all these cases, we will use the Perk prior distribution introduced above as the non-informative prior distribution of all the inferential processes in our study.

From admission to the ICU or ward $W1$

Data for learning about the probability $\boldsymbol{\theta}_A = (\theta_{AW1}, \theta_{AI})^\intercal$ that a patient admitted to the hospital will be referred to ward $W1$ or ICU, respectively, refer to the number of patients admitted to the hospital, 1306, and how many of them were sent to the ward $W1$ or ICU, 1208 and 98, respectively. The posterior distribution of $\boldsymbol{\theta}_A$ is the Dirichlet distribution

$$\pi(\boldsymbol{\theta}_A \mid \mathcal{D}) = \text{Di}(1208.5, 98.5),$$

with posterior expectations 0.925 and 0.075 for θ_{AW1} and θ_{AI}, respectively, thus showing that about the 92.5% of hospital patients are transferred directly to $W1$, and the rest, around 7.5%, are sent directly to the ICU. Posterior 95% credible intervals for these probabilities are $(0.910, 0.938)$ and $(0.062, 0.090)$, respectively. These are very narrow intervals indicating very little uncertainty about both probabilities. It is interesting to note that these credible intervals provide a direct measure of the uncertainty of θ_{AW1} and θ_{AI} that is not possible to consider in the frequentist framework. Figure 2 shows the marginal beta posterior distribution associated for each of the two probabilities, $\pi(\theta_{AI} \mid \mathcal{D}) = \text{Be}(98.5, 1208.5)$ and $\pi(\theta_{AW1} \mid \mathcal{D}) = \text{Be}(1208.5, 98.5)$.

From ward $W1$ to ICU, death, home or a long-stay facility

Of the 1208 patients who were initially sent to ward $W1$, 82 were transferred to ICU, 946 discharged home, 55 derived to a long-term care facility L, and 125 died. This information generates the following posterior distribution for $\boldsymbol{\theta}_{W1} = (\theta_{W1I}, \theta_{W1D}, \theta_{W1H}, \theta_{W1L})^\intercal$, the vector of the probability associated to a possible visit from $W1$ to I, D, H and L, respectively

$$\pi(\boldsymbol{\theta}_{W1} \mid \mathcal{D}) = \text{Di}(82.25, 125.25, 946.25, 55.25),$$

with posterior expectation and 95% credible interval for the posterior distribution associated with each one of the probabilities in $\boldsymbol{\theta}_{W1}$ in Table 1.

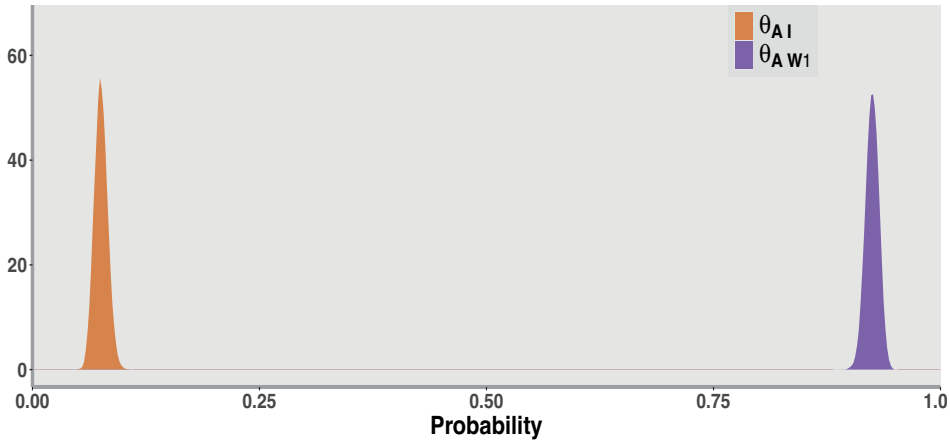

Figure 2. *Posterior marginal distribution for the probability of a patient admitted to the hospital being sent to ICU, θ_{AI}, or ward $W1$, θ_{AW1}.*

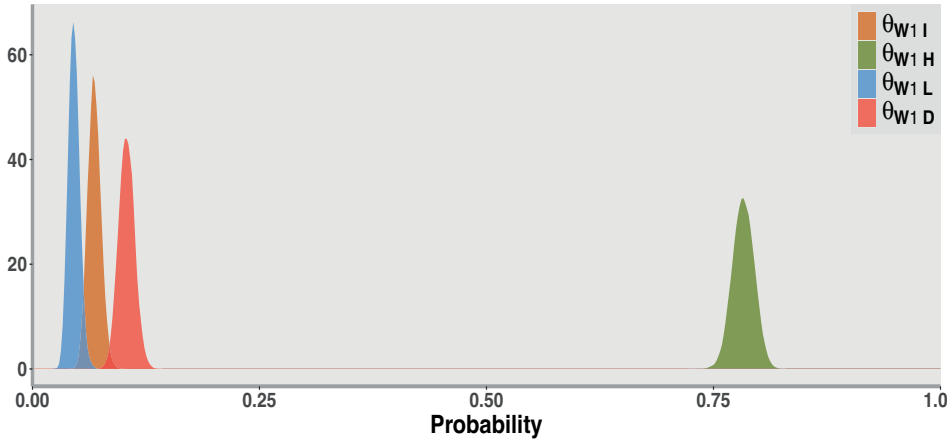

Figure 3. *Posterior conditional distribution for the probability of a patient in $W1$ is sent to a long stay facility, θ_{W1L}, to ICU, θ_{W1I}, to death, θ_{W1D}, or home, θ_{W1H}.*

Table 1. *Posterior distribution, mean, and 95% credible interval for the probabilities in $\boldsymbol{\theta}_{W1}$.*

Probability	Marginal	Mean	CI 95%
θ_{W1I}	Be(82.25, 1126.75)	0.068	(0.055, 0.083)
θ_{W1D}	Be(125.25, 1083.75)	0.103	(0.087, 0.121)
θ_{W1H}	Be(946.25, 262.75)	0.783	(0.759, 0.806)
θ_{W1L}	Be(55.25, 1153.75)	0.046	(0.035, 0.058)

Figure 3 shows the posterior distribution of each of the probabilities associated with $\boldsymbol{\theta}_{W1}$. It is interesting to note the homogeneity of the distributions associated with moving to states I, D and L and the difference in magnitude and variability of that associated with H. Approximately 10% of the patients in the ward $W1$ die, 5% are sent to the ICU, and around 4% are transferred to a long-term institution, probably with very little chance of cure. On the other hand, about 78% of patients admitted to the ward are discharged, although the uncertainty associated with this probability is greater than the previous three.

From ICU to death or to a second hospital ward, $W2$

Among the 180 patients who went through the ICU 35 died, and the rest, 145, were referred to $W2$. As a result, the posterior distribution for $\boldsymbol{\theta}_I = (\theta_{IW2}, \theta_{ID})^\intercal$ that accounts for the probability that a patient in the ICU moves to ward $W2$ or dies is as follows,

$$\pi(\boldsymbol{\theta}_I \mid \mathcal{D}) = \text{Di}(145.50, 35.50),$$

with posterior expectation 0.804 and 95% credible interval (0.743, 0.858) for θ_{IW2}, and 0.196 and (0.142, 0.257) for θ_{ID}, respectively. Figure shows both posterior distributions, very different in their location but similar in shape and variability.

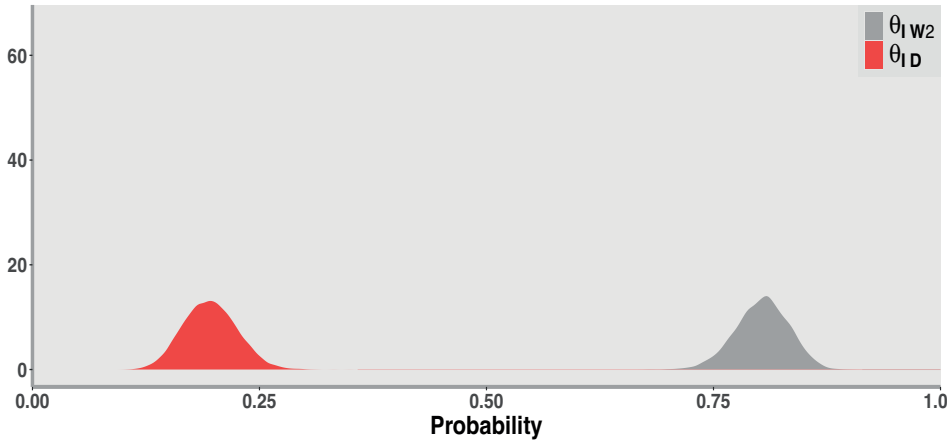

Figure 4. *Posterior distribution of the probability that a patient leaving ICU is transferred to ward, θ_{IW2} or dies, θ_{ID}.*

From ward W2 to death, home or a long-term care facility

A total of 145 patients passed through ward $W2$. Of these, 15 died, 118 were discharged and sent home, and the remaining 15 were sent to a long-term care facility. With this information, the posterior distribution associated to $\boldsymbol{\theta}_{W2} = (\theta_{W2D}, \theta_{W2H}, \theta_{W2L})^\intercal$ is:

$$\pi(\boldsymbol{\theta}_{W2} \mid \mathcal{D}) = \text{Di}(145.33, 118.33, 12.33),$$

with posterior mean and 95% credible interval for the posterior distribution associated to each one of the probabilities in $\boldsymbol{\theta}_{W2}$ given in Table 2.

Table 2. *Posterior marginal distribution, mean and 95% credible interval for the probabilities in* $\boldsymbol{\theta}_{W2}$.

Probability	Marginal	Mean	CI 95%
θ_{W2D}	Be(145.33, 130.67)	0.105	(0.061, 0.159)
θ_{W2H}	Be(118.33, 157.67)	0.810	(0.743, 0.870)
θ_{W2L}	Be(12.33, 263.67)	0.085	(0.045, 0.135)

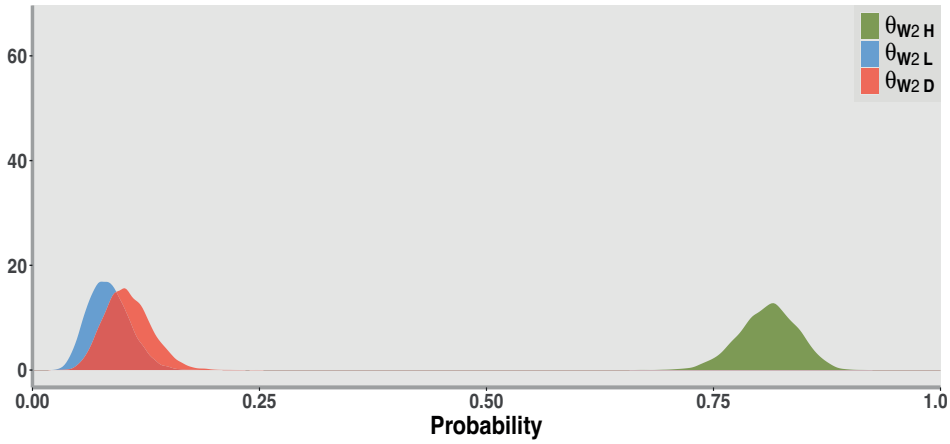

Figure 5. *Posterior distribution of the probability that a patient in ward W2 dies,* θ_{W2D}*, is sent home with a discharge,* θ_{W2H}*, or to a long-stay facility,* θ_{W2L}.

	W1	I	W2	D	H	L
Adm	0.925 [0.910; 0.938]	0.075 [0.062; 0.090]	–	–	–	–
W1	–	0.068 [0.055; 0.083]	–	0.103 [0.087; 0.121]	0.783 [0.759; 0.806]	0.046 [0.035; 0.058]
I	–	–	0.804 [0.743; 0.858]	0.196 [0.142; 0.257]	–	–
W2	–	–	–	0.105 [0.061; 0.159]	0.811 [0.743; 0.870]	0.085 [0.045; 0.135]

Figure 6. *Posterior mean and 95% credible interval associated with transition probabilities between contiguous and consecutive health states. Information relating to final states is shown in blue.*

Approximately, 81% of the patients leaving $W2$ are discharged and sent home, 10% die, and the rest, about 9%, are sent to a long-term care facility. Figure 5 shows the corresponding Beta distributions: very different in location, but with very similar variability and shape.

As a summary, Figure 6 presents the mean and a 95% credible interval of the posterior beta distributions for transition probabilities between contiguous and consecutive health states. We have chosen a matrix format because it most clearly visualises the movement of patients between different health states.

4.2. Probability of leaving hospital on discharge, dying, or being sent to a long-term care facility.

From a clinical point of view, it is important to assess the probability that a patient who enters the hospital with severe influenza will eventually be discharged home, $\theta_{A\bullet H}$, die, $\theta_{A\bullet D}$, or be sent to a long-term care facility, $\theta_{A\bullet L}$. These terminal probabilities are defined in terms of the different trajectories that connect hospital admission A to the absorbing state, D, H or L, that determine the patient's condition on discharge from hospital.

$$\begin{aligned}
\theta_{A\bullet D} &= \theta_{AIW2D} + \theta_{AID} + \theta_{AW1IW2D} + \theta_{AW1ID} + \theta_{AW1D} \\
&= \theta_{AI}\,\theta_{IW2}\,\theta_{W2D} + \theta_{AI}\,\theta_{ID} + \theta_{AW1}\,\theta_{W1I}\,\theta_{IW2}\,\theta_{W2D} \\
&\quad + \theta_{AW1}\,\theta_{W1I}\,\theta_{ID} + \theta_{AW1}\,\theta_{W1D}. \\
\theta_{A\bullet H} &= \theta_{AW1H} + \theta_{AW1IW2H} + \theta_{AIW2H} = \theta_{AW1}\,\theta_{W1H} \\
&\quad + \theta_{AW1}\,\theta_{W1I}\,\theta_{IW2}\,\theta_{W2H} + \theta_{AI}\,\theta_{IW2}\,\theta_{W2H}. \\
\theta_{A\bullet L} &= \theta_{AW1L} + \theta_{AW1IW2L} + \theta_{AIW2L} = \theta_{AW1}\,\theta_{W1L} \\
&\quad + \theta_{AW1}\,\theta_{W1I}\,\theta_{IW2}\,\theta_{W2L} + \theta_{AI}\,\theta_{IW2}\,\theta_{W2L}.
\end{aligned} \tag{5}$$

The posterior distribution for these probabilities, $\pi(\theta_{A\bullet D} \mid \mathcal{D})$, $\pi(\theta_{A\bullet H} \mid \mathcal{D})$, and $\pi(\theta_{A\bullet L} \mid \mathcal{D})$, is not analytical but, as mentioned above, it can be approximated via simulation by generating approximate samples of the posterior distribution of each direct transition probability following (3).

Figure 7 displays the posterior distribution for the probability that a patient hospitalized with severe influenza will die in the hospital, be cured, discharged and sent home, or be transferred to a long-term care unit. Posterior mean and 95% credible intervals are reported in Table 3.

Table 3. *Posterior expectation and 95% credible interval for the probability associated to the enter in the absorbing states L, H, and D.*

Probability	Mean	CI 95%
$\theta_{A\bullet L}$	0.052	(0.040, 0.065)
$\theta_{A\bullet H}$	0.814	(0.784, 0.843)
$\theta_{A\bullet D}$	0.134	(0.116, 0.155)

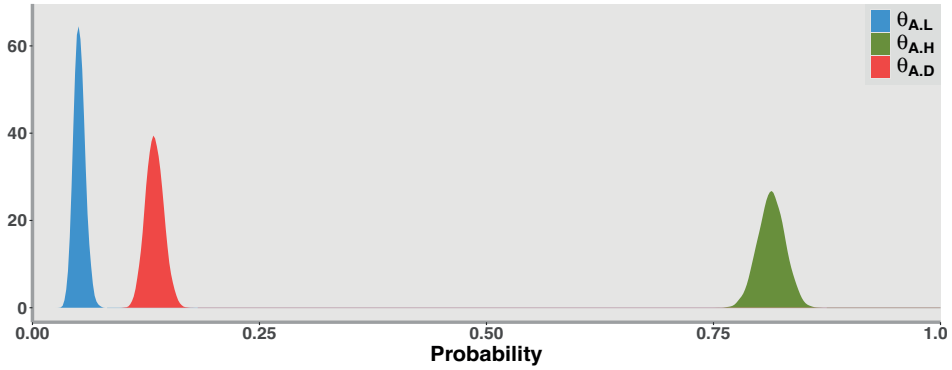

Figure 7. *Posterior distribution for the probability of dying in hospital. being sent to a long-term institution or being discharged cured and sent home.*

About 81% of patients who are hospitalised due to severe influenza will be discharged cured and sent home, about 13% will die and about 5% will be sent to a long-term care facility. As the credible intervals above and the posterior distributions for $\theta_{A\cdot L}$, $\theta_{A\cdot H}$, and $\theta_{A\cdot D}$ in Figure 7 indicate, the uncertainty associated with each of these estimates is quite small, especially that associated with dying in hospital or being sent to a long-term institution.

As a kind of general summary, Figure 8 presents, overlays on the different stages of health, the posterior mean and a 95% credibility interval for the probability that, after being admitted to hospital, a patient will visit each node of the hospital. For example, the ICU state has an approximate probability of 0.138. This probability will be that of passing through the ICU, either directly when she/he is admitted or through $W1$, that is $\theta_{A\cdot I} = \theta_{AI} + \theta_{AW1I}$.

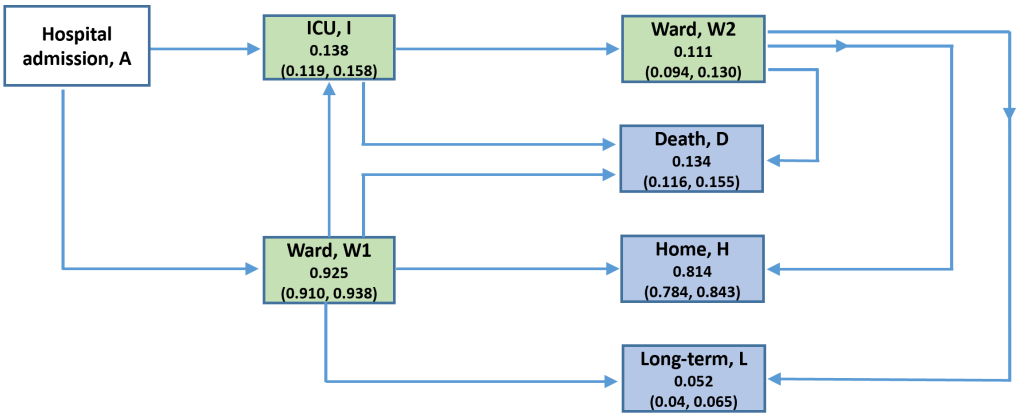

Figure 8. *Posterior mean and 95% credible interval for the probabilities associated with visiting each node from hospitalization.*

As relevant information that we have not mentioned previously, we would like to point out that visiting $W2$ means leaving the ICU alive, which occurs with an approximate probability of 0.11. We would also like to note that approximately 14% of hospitalised patients pass through the ICU.

4.3. Probability that a patient who has died, or has been discharged and sent home or has been sent to a long-term institution has spent time in the ICU

In our study, it may be interesting to simulate the posterior distribution of some inverse probabilities, such as the probability that a patient who died in the hospital, was sent to a long-term care facility or was discharged cured had previously been in the ICU, $\theta_{D\,\cdot\,I}$, $\theta_{L\,\cdot\,I}$, or $\theta_{H\,\cdot\,I}$. We start with the posterior distribution for $\theta_{D\,\cdot\,I}$. Following (4)

$$\theta_{D\,\cdot\,I} = \frac{\theta_{I\,\cdot\,D}\,\theta_I}{\theta_D},$$

where θ_D (θ_I) is the probability that a hospitalized patient dies in hospital (enters the UCI) which we have previously represented as $\theta_{A\,\cdot\,D}$ ($\theta_{A\,\cdot\,I}$), with

$$\theta_{I\,\cdot\,D} = \theta_{ID} + \theta_{IW2D} = \theta_{ID} + \theta_{IW2}\,\theta_{W2D},$$
$$\theta_{A\,\cdot\,I} = \theta_{AI} + \theta_{AW1I} = \theta_{AI} + \theta_{AW1}\,\theta_{W1I},$$

and $\theta_{A\,\cdot\,D}$ expressed in (5) in terms of transition probabilities between contiguous states. This procedure can be also followed to simulate the posterior distribution of the rest of inverse probabilities, $\theta_{L\,\cdot\,I}$ and $\theta_{H\,\cdot\,I}$.

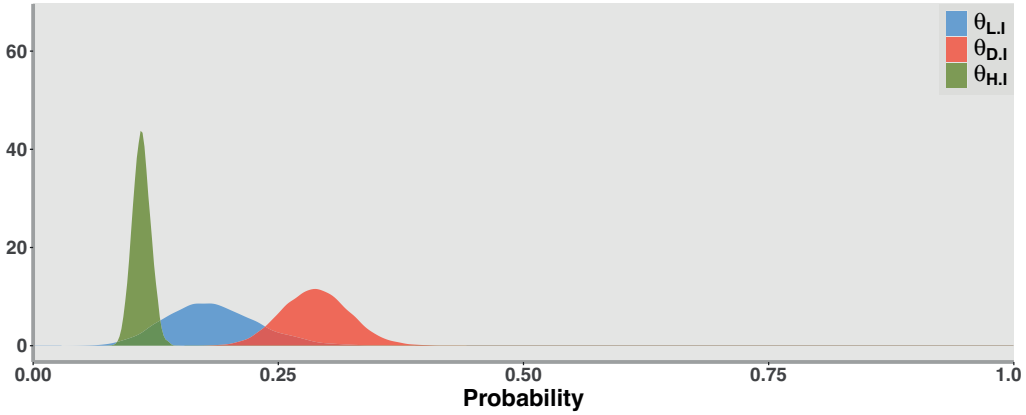

Figure 9. *Posterior distribution for the probability that a patient who has died, or has been discharged and sent home or has been sent to a long-term institution has spent time in ICU.*

Figure 9 shows the posterior distribution of the three previous probabilities. We can see that people who were discharged cured from the hospital are the least likely to have

spent time in the ICU, followed by patients who were sent to a long-stay ward, and finally people who died in hospital. We can also observe that the distribution associated with leaving the hospital cured is the one with the least uncertainty. The corresponding posterior means and 95% credible intervals are reported in Table 4.

Table 4. *Posterior mean and 95% credible interval for the probability that a patient has been in ICU given that she/he has been finally L, H, or D.*

Probability	Mean	CI 95%
$\theta_{L\bullet I}$	0.183	(0.100, 0.280)
$\theta_{H\bullet I}$	0.111	(0.093, 0.128)
$\theta_{D\bullet I}$	0.288	(0.224, 0.357)

5. Introducing covariates

The probabilities associated with transitions between states will surely depend on the different characteristics of the patients in the study population. Therefore, introducing covariates into the statistical modelling will improve the quality and accuracy of the analysis. In our case, including covariates in the model means entering the world of logistic or multinomial regression depending on the number of states, two or more than two respectively, to which a given state can visit in a single transition. In the general case where states $\{1,\ldots,J\}$ are accesible directly from state i with probabilities $\boldsymbol{\theta}_i = (\theta_{i1},\ldots,\theta_{iJ})'$ and $\boldsymbol{x}_i = (x_{i1},\ldots,x_{iQ})'$ is a vector of covariates related to state i, a multinomial regression model states that

$$\log\Big(\frac{\theta_{ij}}{\theta_{i1}}\Big) = \beta_{0ij} + \sum_{q=1}^{Q} \beta_{qij}\, x_{iq} = \eta_{ij}, \;\; j = 2,\ldots,J, \tag{6}$$

where β_{qij} are the regression coefficients of the model and $\sum_{j=1}^{J} \theta_{ij} = 1$. This modelling implies that:

$$\begin{aligned}
\theta_{i1} &= 1/(1+\textstyle\sum_{j=2}^{J} \mathrm{e}^{\eta_{ij}}),\\
\theta_{ij} &= \mathrm{e}^{\eta_{ij}}/(1+\textstyle\sum_{j=2}^{J} \mathrm{e}^{\eta_{ij}}),\; j = 2,\ldots,J.
\end{aligned}$$

To complete the Bayesian model, it is required to construct a prior distribution for the model parameters, in this case the regression coefficients. We assume an inferential scenario of prior independence and very little initial information. Since we are working on the logarithmic scale, with positive and negative real values, we choose normal prior distributions centred on zero and with a wide variance. When the number of destination states from the generic i is 2, we are in the framework of logistic regression, with a logit link function for the probability associated with one of the two states directly accessible from i.

Table 5. *Number of patients who end up in the absorbing states D, H, and L from the ICU based on their age and gender.*

	Women		Men	
	Age<60	Age≥60	Age<60	Age≥60
Death, D	1	14	8	27
Home, H	25	23	28	42
Long-stay, L	1	5	2	4

To illustrate the potential of introducing covariates into the PIDIRAC study, we look at the transition that the process can make from the ICU (I) to Home (H), Death (D) or a long-stay faciliy (L) and consider two covariates relevant to the study: gender and age, the latter of which we only have in binary form, under 60 years of age or 60 years of age or older. It is worth noting that we know that ward $W2$ is an intermediate state between a person leaving the ICU without dying and the three previous states. For the sake of simplicity, we will omit this in the example, considering only moving from the ICU to the three aforementioned absorbing states. Table 5 shows the number of patients who ended up in H, L and D from the ICU in relation to their age and sex. We can see that category L has very few observations, which seems too little to feed inferences in this group.

It is well known that people who are sent to long-term care facilities often have very serious health problems with a high probability of death. That is why we have combined categories D and L into a single category, which we will represent as DL. We analyze these data by means of a logistic regression model for the probability associated to the DL category in the logit scale:

$$\text{logit}(\theta_{I\,\cdot\,DL}) = \beta_0 + \beta_1 I_{(Man)} + \beta_2 I_{(Age>=60)}. \tag{7}$$

This model implies the probability $\theta_{I\,\cdot\,DL} = \mathrm{e}^{\beta_0}$ in the group of women under the age of 60, $\theta_{I\,\cdot\,DL} = \mathrm{e}^{\beta_0+\beta_2}$ for women aged 60 or over, $\theta_{I\,\cdot\,DL} = \mathrm{e}^{\beta_0+\beta_1}$ in men less than 60 and $\theta_{I\,\cdot\,DL} = \mathrm{e}^{\beta_0+\beta_1+\beta_2}$ in men aged 60 or more.

Bayesian inference completes this model with a prior distribution on the model parameters, $\pi(\beta_0,\beta_1,\beta_2)$. We deal with an environment of prior independence and little initial information, in particular

$$\pi(\beta_0,\beta_1,\beta_2) = \pi(\beta_0)\pi(\beta_1)\pi(\beta_2),$$

with $\pi(\beta_i) = \mathcal{N}(0, 10^2)$. The posterior distribution $\pi(\beta_0,\beta_1,\beta_2 \mid \mathcal{D})$ was approximated by using Markov chain Monte Carlo (MCMC) methods through JAGS Software (Plummer, 2003). The MCMC algorithm ran for three Markov chains with 200000 iterations after a burn-in period with 20000. The effective iterations were thinned by storing every 30th iteration in order to decrease autocorrelation in the sample. Convergence of the chains to the posterior distribution is assessed through the potential scale

reduction factor, *Rhat*, and the effective number of independent simulation draws, n_{eff}. In all cases, the *Rhat* values are equal or near 1 and $n_{eff} > 100$, thus indicating that the distribution of the simulated values between and within the three chains is practically identical and also that sufficient MCMC samples have been obtained, respectively.

Figure 10 shows the approximate posterior distribution for the probability that a person in $W2$ will die, be discharged and sent home, or be referred to a long-stay service, by gender and age.

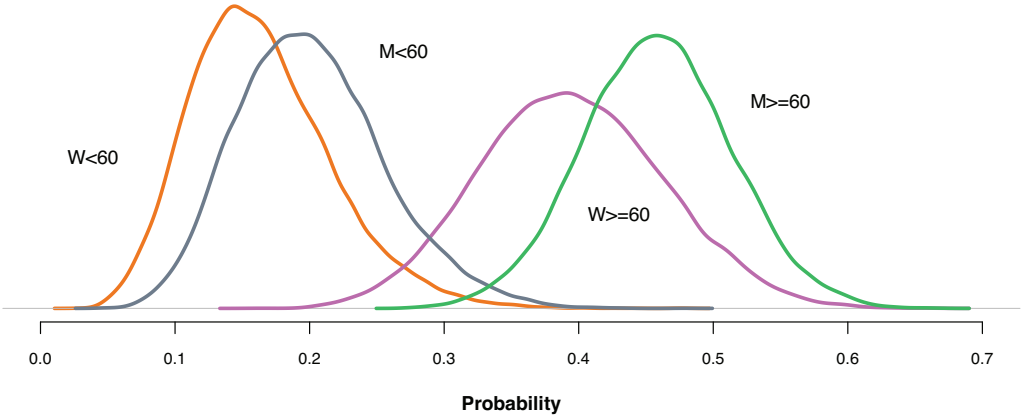

Figure 10. *Posterior distribution for the probability of dying or being sent to a long-term care facility after ICU in women under 60 years (W<60) or aged 60 or over (W>=60), and men in the same age categories (M<60 and M>=60).*

Table 6. *Number of patients who end up in the absorbing states D, H, and L from the ICU based on their age and gender.*

Group	Mean	Sd	CI 95%
$W < 60$	0.163	0.052	(0.078, 0.279)
$W >= 60$	0.395	0.068	(0.266, 0.532)
$M < 60$	0.200	0.054	(0.107, 0.320)
$M >= 60$	0.457	0.055	(0.352, 0.567)

It is interesting to note that age and gender area very relevant factors in the possible recovery of patients hospitalised with severe influenza. In the group of people under 60, the approximate probability of dying or being sent to a care home is approximately 0.163 for women and 0.20 for men. These probabilities increase dramatically in people aged 60 or over: 0.395 for women and 0.457 for men.

6. Conclusions

Probabilistic DAGs are very helpful representations of complex environments with stochastic dependencies. Bayesian inference in DAGS defined by random events is a powerful framework for understanding and assessing the prevalence and uncertainty associated with the different trajectories in the system.

The graphical representation of the evolution of patients admitted to hospital as a consequence of severe influenza through a DAG and the subsequent statistical analysis provides valuable clinical information on the severity of the disease as well as on the utilisation of healthcare resources. This information is key to hospital resource planning because it helps identify the human, material and financial resources needed to ensure quality and efficient hospital care.

Our work provides a flexible framework that allows the inclusion of potentially relevant additional information in terms of demographic (sex, age) or clinical (comorbidities) covariates, as well as other types of information such as length of stay in the different services, or even considering the potential variability between the different hospitals participating in the study. These last two proposals would correspond to topics in multistate models and hierarchical Bayesian models, respectively.

The application of Bayesian methods usually requires the use of complex computational tools because the underlying posterior distribution is not analytical. In our case, this is not so, because the inferential process associated with multinomial probabilities is completely analytical: the Dirichlet distribution is the conjugate family with respect to the multinomial probabilistic model. This scenario facilitates the simplicity and transparency of the implementation of the procedures involved. Furthermore, the use of basic procedures for generating observations from the resulting posterior distributions allows us to obtain very good approximations of non-analytical posterior distributions, as is the case with probabilities associated with non-adjacent paths or inverse transitions. However, when we introduce complexity into the modelling, such as the inclusion of covariates, we need to use intensive computational procedures, MCMC methods in our case.

Acknowledgements

The authors are very grateful to two anonymous reviewers, whose valuable comments and suggestions have contributed significantly to improving the quality of the paper. This paper is partially supported by the project PID2023-148158OB-I00, funded by Ministerio de Ciencia, Innovación y Universidades (Spain) and the project PID2022-136455NB-I00, funded by Ministerio de Ciencia, Innovación y Universidades (Spain) (MCIN/AEI/10.13039 /501100011033/FEDER, UE) and the European Regional Development Fund.

References

Acosta, L., Soldevila, N., Torner, N., Martínez, A., Ayneto, X., Rius, C., Jané, M., Domínguez, A. and the Influenza Surveillance Network of Catalonia, PIDIRAC. (2021). Influenza Vaccine Effectiveness in Preventing Severe Outcomes in Patients Hospitalized with Laboratory-Confirmed Influenza during the 2017-2018 Season. A Retrospective Cohort Study in Catalonia (Spain). *Viruses* 13, 14655.

Alvares, D., Armero, C., and Forte A. (2018). What does objective mean in a Dirichlet-Multinomial process? *International Statistical Review (ISI)* 86(1), 106-18.

Armero, C., García-Donato, G., Jiménez-Puerto, J., Pardo-Gordó, S., and Bernabeu, J. (2021). Bayesian classification for dating archaeological sites via projectile points. *SORT* 45(I), 33–46.

Barber, D. (2012). *Bayesian reasoning and machine learning.* Cambridge University Press.

Berger, J. O., Bernardo, J. M. and Sun, D. (2015). Overall objective priors. *Bayesian Analysis*, 10(1), 189-221.

Carrillo-Santisteve, P., Ciancio, B. C., Nicoll, A., & Luigi Lopalco, P. (2012). The importance of influenza prevention for public health. *Human Vaccines & Immunotherapeutics*, 8(1), 89-95.

Congdon, P. (2005). *Bayesian Models for categorical Data*. WILEY.

Cowell, R. G., Dawid, A., Lauritzen, S., Spiegelhalter, D. (1999), *Probabilistic Networks and Expert Systems*. Springer.

European Centre for Disease Prevention and Control (ECDC). Disease facts about seasonal influenza. Available online: https://www.ecdc.europa.eu/en/seasonal-influenza/facts [last access: March 2025].

Langer, J., Welch, V.L., Moran, M.M., Cane, A., Lopez, S.M., Srivastava, A., Enstone, A.L., Sears, A., Markus, K.J, Heuser, M., Kewley, R.M., and Whittle, I.J. (2023). High Clinical Burden of Influenza Disease in Adults Aged $\geq$ 65 Years: Can We Do Better? A Systematic Literature Review. *Adv Ther* 40(I), 1601–1627. https://doi.org/10.1007/s12325-023-02432-1

Macias, A. E., McElhaney, J. E., Chaves, S. S., Nealon, J., Nunes, M. C., Samson, S. I., Seet, B. T., Weinke, T., and Yu, H. (2021). The disease burden of influenza beyond respiratory illness. *Vaccine*, 39, Supplement 1, A6-A14.

Paget, J., Iuliano, A. D., Taylor, R. J., Simonsen, L., Viboud, C., Spreeuwenberg, P. (2022). Estimates of mortality associated with seasonal influenza for the European Union from the GLaMOR project. *Vaccine*, 40(9), 1361-1369,

Perks, W. (1947). Some observations on inverse probability including a new indifference rule. *J. Inst. Actuaries*, 73(2), 285-334.

Pietrasik, T. (2023). Influenza (Seasonal). *World Health Organization (WHO)*. Available online: https://www.who.int/en/news-room/fact-sheets/detail/influenza-(seasonal) [last access: March 2025].

Plummer, M. (2003). JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing* (DSC 2003), Vienna, 20-22 March 2003, 1-10.

R Core Team (2025). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Snacken, R., and Brown C. (2015). New developments of influenza surveillance in Europe. *Eurosurveillance* 20(4), pii=21020. Available online: https://doi.org/10.2807/ese.20.04.21020-en.

Soldevila, N., Acosta, L., Martínez, A., Godoy, P., Torner, N., Rius, C., Jané, M., Domínguez, A. & the Surveillance of Hospitalized Cases of Severe Influenza in Catalonia Working Group. (2021). Behavior of hospitalized severe influenza cases according to the outcome variable in Catalonia, Spain, during the 2017-2018 season. *Scientific Reports* 11:13587.