Estimating Cross-Provider Substitution Effects in Primary Care Delivery*

Rosa Gentile
University of Palermo

Paolo Li Donni
University of Palermo

=

Catia Nicodemo
University of Oxford & University of Brunel †

January 3, 2025

Abstract

This study investigates task allocation and substitution effects between general practitioners (GPs) and nurses in primary care settings, leveraging comprehensive electronic healthcare records from the UK. Using a novel copula-based Hidden Markov Bivariate Poisson model, we analyze patient-level data to assess how the allocation of responsibilities impacts service delivery. Our findings reveal significant substitution effects: increased nurse consultations reduce subsequent GP visits and vice versa, highlighting potential efficiencies in workforce planning. The analysis also underscores the importance of continuity of care, with patients more likely to engage consistently with the same type of healthcare provider. By examining utilization patterns across diverse patient cohorts, the study provides actionable insights for optimizing resource allocation in primary care, particularly in the context of GP shortages and evolving healthcare demands. These results have implications for policy development, emphasizing the need for targeted strategies that balance workload distribution while maintaining care quality and accessibility.

JEL classification: *I12, I18, C31, D40*

Keywords: Primary Care, GPs, Nurses, Morbidity, Hypertension, NHS.

*

†Corresponding author: Catia Nicodemo, catia.nicodemo@economics.ox.ac.uk Funding Acknowledgement: Nicodemo receives XXXXXXXXXXXXXXX The views expressed in this publication are those of the author(s) and not necessarily those of the funders.

1 Introduction

Primary care serves as the cornerstone of the National Health Service (NHS), providing the first point of contact for patients. In 2023/24 alone, primary care accounted for approximately 353 million appointments in England, underscoring its critical role in delivering accessible and comprehensive healthcare (The King's Fund, 2023). The system faces mounting pressures driven by demographic shifts, increasing multimorbidity, and workforce shortages, necessitating a re-evaluation of resource allocation and care delivery models (see Fikree et al., 2023; Karthika et al., 2023; Porter et al., 2023, among others). These issues have necessitated the integration of allied health professionals, into primary care teams. This shift raises critical questions about the dynamics of workforce roles, particularly whether GPs and other healthcare workforce operate as substitutes or complements in meeting patient needs. Substitution refers to the extent to which one type of provider, such as nurses, can effectively take over tasks traditionally performed by GPs. Studies have shown that nurses can manage routine care, chronic condition monitoring, and patient education with outcomes comparable to or better than those achieved by GPs (Laurant et al., 2018). This potential substitution effect is particularly relevant for addressing GP shortages and optimizing workforce utilization. Conversely, complementarity arises when the roles of GPs and nurses enhance each other, with their combined efforts meeting the diverse and often complex needs of patients, particularly those with multimorbidities or chronic conditions. There is an extensive debate surrounding the integration of physician associates into primary care and the implementation of the NHS-funded Additional Roles Reimbursement Scheme (ARRS) to support the inclusion of allied health professionals (Halter et al., 2017; Drennan et al., 2017; Malone, 2022).

This study aims to explore these substitution and complementarity effects within the NHS primary care setting. Using a hidden Markov model, we analyze healthcare utilization patterns and transitions across latent states, focusing on the interplay between GP and nurse consultations. The analysis considers how patient characteristics, including age and multimorbidity, influence these dynamics and assesses the implications for workforce planning and resource allocation.

Using comprehensive administrative healthcare records, the study provides a detailed examination of utilization patterns and transitions across four latent states of healthcare

usage. The results of this study highlight distinct patterns of healthcare utilization and the roles of GPs and nurses within primary care. We identify clear substitution and complementarity effects between the two provider types. Substitution is evident as previous nurse consultations reduce the likelihood of subsequent GP visits, suggesting that nurses can take on routine tasks typically managed by GPs. Conversely, complementarity is observed in cases of complex patient needs, where individuals with higher multimorbidity demonstrate increased joint utilization of both GPs and nurses. Notably, the likelihood of transitioning into medium-high or high usage states (States 3 and 4) rises significantly with the number of morbidities and age, underscoring the growing demand for collaborative, multidisciplinary care in managing chronic and complex conditions.

This study advances the literature on workforce optimization and task allocation in primary care by addressing several critical gaps. While previous research, such as Laurant et al. (2018), has established that nurses can effectively substitute for GPs in delivering primary care without compromising patient outcomes, this study goes further by quantifying the substitution effects and examining the dynamics of task allocation between these professionals using comprehensive administrative data. Unlike earlier works that primarily focus on static comparisons, our analysis employs a novel copula-based Hidden Markov Bivariate Poisson (HMBPo) model, extending existing methodologies (e.g. Laudicella and Li Donni, 2022) to capture the interplay between state dependence, unobserved heterogeneity, and time-specific shocks in care utilization. Additionally, by focusing on the UK's NHS—a healthcare system undergoing significant workforce expansion and role diversification—this paper provides timely, policy-relevant evidence on optimizing resource allocation. These contributions not only enhance our understanding of professional substitution in healthcare but also offer a methodological framework applicable to other high-demand settings. By analyzing substitution effects, continuity of care, and the impact of multimorbidity, this research offers actionable insights for policymakers and healthcare planners.

The paper is structured as follows: Section 1 describes the data, outlining its sources, scope, and key variables. Section 2 details the empirical model, explaining the methodological framework employed in the analysis. Section 3 presents the results, highlighting the key findings and their implications. Finally, Section 4 concludes the paper with a discussion of the results and their relevance to policy.

2 Data

We analyzed data from the Aurum Clinical Practice Research Datalink (CPRD)[1] for all patients aged 50 and over from 2008 until 2017. The Aurum CPRD is an extensive database of anonymized primary care electronic health records, encompassing 800 practices and approximately 7% of the UK population (over 3 million patients in 2017). Previous studies have validated that this database is broadly representative of the general UK population in terms of age, sex, and ethnic origin distribution.

Through CPRD's data linkage services, we enhanced our primary care data by integrating several complementary sources from the same analysis period. The Index of Multiple Deprivation (IMD) provides valuable socioeconomic context at the area level, ranking geographical regions based on multiple deprivation indicators, including income, employment, education, health, crime, housing, and living environment. This allows for an analysis of healthcare utilization patterns across various socioeconomic groups. Additionally, we included Hospital Episode Statistics (HES), which offer comprehensive secondary care data, covering emergency department visits, outpatient clinic attendances, day case procedures, and inpatient hospital admissions. We also incorporated mortality data from the Office for National Statistics (ONS).

Our study population focuses on patients diagnosed with hypertension, one of the most prevalent chronic conditions requiring regular primary care monitoring and management. Hypertension affects approximately 30% of adults worldwide (World Health Organization, 2021) and represents a significant burden on healthcare systems. The condition requires consistent primary care oversight due to its progressive nature and role as a major risk factor for cardiovascular disease (Whelton et al., 2018).

We selected a cohort of patients who were diagnosed with hypertension in 2011, ensuring that they had no prior diagnosis of hypotension in the two years leading up to this point. These patients were then followed through their electronic health records until 2017 to track the progression of their condition, any subsequent diagnoses, treatments, and healthcare utilization. This approach allowed for a detailed analysis of the clinical trajectory of hypotension in a population with no previous history of the condition.

We examine two main outcome variables at the patient-year level: the total number of primary care visits and the number of consultations specifically with nurses. These vari-

[1]For more information on CPRD data see the webpage `https://www.cprd.com/`

ables capture the intensity of primary care utilization and allow us to examine potential substitution patterns between different types of care providers.

Our empirical specification includes a comprehensive set of control variables. At the patient level, we control for demographic characteristics including age and gender, as well as health status measured by the number of concurrent morbidities. To account for persistence in healthcare utilization patterns, we include lagged variables for both general practitioner and nurse visits from the previous year $(t-1)$. Socioeconomic status is captured through the IMD. We also control for the specific types of diagnoses to account for variation in care needs across different conditions. Table 1 shows the descriptive statistics of our sample, we have a total number of 15,100 patients observed over seven years (2011-2017). On average, patients had 5.17 GP visits and 1.69 nurse visits. The cohort is nearly evenly split by gender, with an average age of 62.16 years. In terms of socioeconomic status, most patients are from more deprived areas, with 27% in the most deprived category (IMD1) and 23% in the second most deprived (IMD2). On average, patients had 1.93 diagnoses, with common comorbidities including anxiety (16%), cardiovascular disease (18%), cancer (4%), and chronic obstructive pulmonary disease (COPD) (6%).

Variable Name	Mean	SD
Num GP visits	5.17	4.96
Num Nurse visits	1.69	2.31
Gender	0.48	-
Age	62.16	7.81
IMD1 (Most deprived)	0.27	-
IMD2	0.23	-
IMD3	0.19	-
IMD4	0.17	-
IMD4 (Least Deprived)	0.14	-
Total Num of Morbidities	1.93	1.73
Anxiety	0.16	-
CVD	0,18	0.46
Cancer	0.04	-
COPD	0.06	-
Tot Num Obs	105,700	

Table 1: Summary statistics

Figure 1 illustrates the distribution of GP consultations across varying levels of nurse

utilization (Nurse=0, 1, 2, >2). The visualization reveals a positive association between nurse and GP visits, suggesting that patients with higher nurse consultation frequencies also tend to have more GP encounters. This pattern may reflect the complexity of patient needs, with those requiring more frequent nurse care also necessitating increased GP oversight (Laurant et al., 2018). The observed positive correlation could be attributed to several factors: the management of complex chronic conditions requiring coordinated care from both practitioners (**?**), systematic care protocols involving both nurses and GPs (**?**), or the identification of additional health concerns during nurse consultations leading to GP referrals.

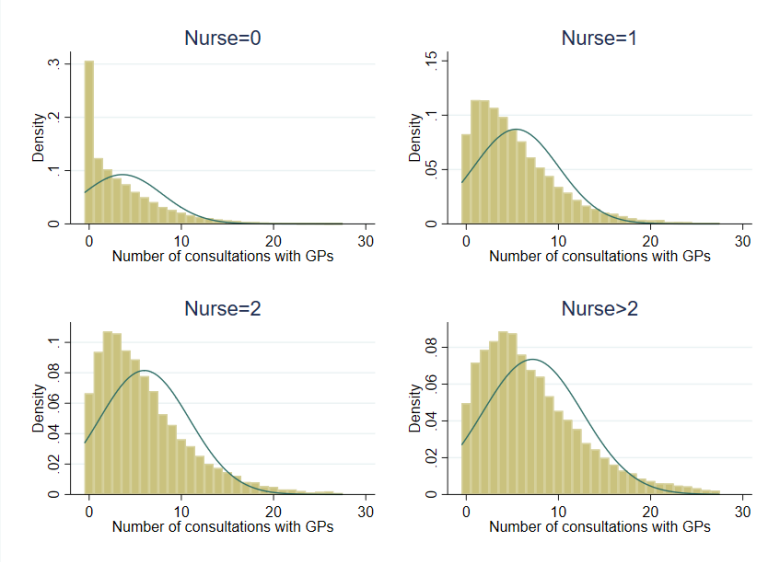


Figure 1: Relationship Between Nurse Utilization and the Number of GP Consultations

3 Empirical Model

To study whether the relationship between nurses and GPs is more complementary or substitutive, we follow the model developed by Laudicella and Li Donni (2022). Suppose we observe two count variables y_{1it} and y_{2it}, measuring the number of GPs and nurses visits an individual i, with $i = 1, \ldots, N$, made at time occasion t, with $t = 1, \ldots, T$. In modelling y_{1it} and y_{2it}, there are three potential sources of correlation to consider. The first source of correlation is related to state dependence: the present number of GPs and nurses' visits could be determined by past utilization. The second source of correlation

is related to unobservable time-varying factors which can affect primary care utilization. For example, an individual with an unobserved health shock may systematically require additional GPs and nurses' care in subsequent periods. In the literature on the determinants of health care utilization, this source of heterogeneity has been modelled using *univariate* latent class models, which assume individuals are drawn from a discrete number of latent groups sharing similar unmeasured characteristics (e.g. Bago d'Uva, 2005; Hyppolite and Trivedi, 2012; Wouterse et al., 2013). Finally, the last source of correlation reflects unmeasured *time-specific* shocks that may affect the utilization of both types of primary care visits in a specific time period. For example, individuals may experience an occasional health deterioration with short-lasting consequences for their utilization of care.

To account for these three sources of correlation, we follow and extend the Hidden Markov Bivariate Poisson (HMBPo) model proposed by Laudicella and Li Donni (2022). The HMBPo model relies on two main assumptions: (i) unmeasured heterogeneity can be modelled using a discrete distribution of latent states ξ_c collected in the vector $\boldsymbol{U}$, with $c = 1, \ldots, k$ following a first-order Markov chain; and (ii) the latent states make GPs and nurses visits conditionally independent given the full set of observable explanatory variables, which also include primary care utilization in previous periods.

In modelling the association between y_{1it} and y_{2it}, the HMBPo model proposed by Laudicella and Li Donni (2022) relies on a bivariate Poisson distribution as a product of Poisson marginals with a multiplicative factor (Lakshminarayana et al., 1999). In this case, the correlation between y_{1it} and y_{2it} is assumed to depend on the multiplicative factor parameter, which is constant across observations, whereas the estimated correlation varies by the pattern of covariates. To relax this assumption, we extend the original HMBPo model using a copula approach to model the bivariate association between y_{1it} and y_{2it} (Frees and Valdez, 1998, see for a review). Using this approach, we can define the normal copula function as

$$C(\nu_{1it}, \nu_{2it}; \rho) = \Phi_2(\Phi^{-1}(\nu_{1it}), \Phi^{-1}(\nu_{2it}), \rho) \tag{1}$$

where $\nu_1 = F(y_{1it}, \mu_{1it})$ and $\nu_2 = F(y_{2it}, \mu_{2it})$ are the cumulative distribution functions of the Poisson probability distribution, Φ^{-1} is the quantile function of the standard normal distribution, and $\Phi_2(.,.,.)$ is the cumulative distribution function of the bivariate normal distribution, where the dependence parameter is parameterized using the inverse

hyperbolic tangent function $\rho = \frac{\exp(2\rho_u)-1}{\exp(2\rho_u)+1} \in (-1, 1)$. The intensity parameter of the Poisson distribution is defined as $\mu_{jit} = \exp\big(\sum_{c=1}^{k} \alpha_{jit}(\xi_c) d_{jit}(\xi_c) + \boldsymbol{x}_{jit}\boldsymbol{\beta}_j + \gamma_{j1} y_{1i,t-1} + \gamma_{j2} y_{2i,t-1}\big)$, with $j = 1, 2$—for GP and nurse visits respectively—while $d_{jit}(\xi_c)$ denotes a dummy variable defining whether the ith unit belongs to the latent state ξ_c of the Markov chain at time t.

The bivariate normal copula function satisfies the Fréchet-Hoeffding lower and upper bounds so that the probability mass function p_{12} can be defined by taking finite differences as follows:

$$\begin{aligned} p_{12}(F(y_{1it}, \mu_{1it}), F(y_{2it}, \mu_{2it}); \rho) \;=\; & C(F(y_{1it}, \mu_{1it}), F(y_{2it}, \mu_{2it}); \rho) - \\ & C(F(y_{1it} - 1, \mu_{1it}), F(y_{2it}, \mu_{2it}); \rho) - \\ & C(F(y_{1it}, \mu_{1it}), F(y_{2it} - 1, \mu_{2it}); \rho) + \\ & C(F(y_{1it} - 1, \mu_{1it}), F(y_{2it} - 1, \mu_{2it}); \rho) \end{aligned} \tag{2}$$

In this framework, the lags of the outcome variables, i.e., $y_{1i,t-1}$ and $y_{2i,t-1}$, model state dependence within and between outcomes, $y_{\cdot it}$, hence the parameters γ capture the first source of correlation described above. The individual random effects α_{1it} and α_{2it} capture the second source of correlation due to unmeasured time-varying heterogeneity; αs are the support points of the discrete distribution of U and vary with the latent state ξ_c. Hence, they capture the persistence in using both types of primary care that is due to individual unmeasured characteristics. Finally, ρ captures the third source of residual correlation due to time-specific shocks, which remain after conditioning on y_{1it-1}, y_{2it-1}, the observable characteristics $\boldsymbol{x}$, and the unobserved heterogeneity captured by the latent state ξ_c.

The HMBPo model with the copula function is completed by two additional equations describing how the Markov process characterizes the distribution of $\boldsymbol{\alpha}$. Individuals are assumed to be drawn from a discrete latent distribution with k classes in each time period. These classes represent different unobserved states in which individuals can be observed with probability $p(\boldsymbol{\alpha}_{it} = \boldsymbol{\xi}_c)$ in state c. The latent states are assumed to follow a first-order Markov chain, such that:

$$p(\boldsymbol{\alpha}_i = \boldsymbol{\xi}) = p^{(1)}(\boldsymbol{\alpha}_{i1} = \boldsymbol{\xi}_c) \prod_{t=2}^{T_i} p^{(t)}(\boldsymbol{\alpha}_{it} = \boldsymbol{\xi}_{\bar{c}} | \boldsymbol{\alpha}_{i,t-1} = \boldsymbol{\xi}_c) \tag{3}$$

where the initial probability refers to the probability of observing an individual in a specific state c in the first period:

$$p^{(1)}\big(\boldsymbol{\alpha}_{i1} = \boldsymbol{\xi}_c | \boldsymbol{z}_{i1}, \boldsymbol{y}_{i0}\big) = \frac{\exp(\alpha_c + \boldsymbol{\beta}_c \boldsymbol{z}_{i1} + \boldsymbol{\gamma}_c \boldsymbol{y}_{i0})}{1 + \sum_{c=2}^{k} \exp(\alpha_c + \boldsymbol{\beta}_c \boldsymbol{z}_{i1} + \boldsymbol{\gamma}_c \boldsymbol{y}_{i0})}, \tag{4}$$

where $\boldsymbol{z}_{i1}$ is a vector of observable covariates. Estimating the coefficients from the initial condition and including the effect of $\boldsymbol{y}_{i0}$ in the unobserved process in the first period allows for the well-known initial condition problem (Heckman, 1981).

Finally, the transition probabilities model the unobserved heterogeneity over time. The original HMBPo model assumes that the probability to switch from state c to state $\bar{c}$ is homogenous, e.g. the same transition probability applies to each individual and time period (Laudicella and Li Donni, 2022). However, individuals with different sociodemographic characteristics may experience different time trajectories in their latent state, for instance some age and socioeconomic groups may be more likely to change state than others. Examining heterogeneity in the trajectories of the latent states may be useful to identify sociodemographic groups with the highest probability of changing their latent utilisation profiles, and thus to inform policies aiming at reducing inequalities in access to care. To this end, we extend the original HMBPo model and allow transition probabilities to differ according to observable individual characteristics, such as age, income, and education. Hence, we move to a more general case with individual-specific transition matrix Π_{it}, which collects the following $k \times (k-1)$ probabilities:

$$p^{(t)}\big(\boldsymbol{\alpha}_{it} = \boldsymbol{\xi}_{\bar{c}} | \boldsymbol{\alpha}_{i,t-1} = \boldsymbol{\xi}_c, \boldsymbol{z}_{it}\big) = \frac{\exp(\alpha_{c\bar{c}} + \boldsymbol{\beta}_{c\bar{c}} \boldsymbol{z}_{it})}{1 + \sum_{\bar{c}}^{k} \exp(\alpha_{c\bar{c}} + \boldsymbol{\beta}_{c\bar{c}} \boldsymbol{z}_{it})}, \text{with } c, \bar{c} = 1, \dots k, \text{ and } c \neq \bar{c} \tag{5}$$

with $c, \bar{c} = 1, \dots k$, $c \neq \bar{c}$ and $t = 2, \dots, T_i$. Transition probabilities can be collected on the matrix $\boldsymbol{\Pi}_i$. Notice that $\boldsymbol{z}$ is a vector of observable covariates influencing the transition from c to state $\bar{c}$; the homogeneity assumption is equivalent to setting $\boldsymbol{\beta}_{c\bar{c}} = 0$ in Eq. (5).

The marginal log-likelihood for the proposed model can be obtained as follows:

$$\ell(\boldsymbol{\theta}) = \sum_{i}^{N} \log \Big[p(\boldsymbol{y}_i | \boldsymbol{x}_i, \boldsymbol{y}_{i0}, \boldsymbol{z}_i) \Big] \tag{6}$$

where $\boldsymbol{\theta}$ be the vector collecting all the non redundant model parameters corresponding to the vectors $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and ρ. Given the structure of the copula based HMBPo model described by equations (2), (4) and (5), the probability of observing a given pattern of primary care utilisation described by $\boldsymbol{y}_{1i}$ and y_{2i} is equal to:

$$p(\boldsymbol{y}_i|\boldsymbol{x}_i, \boldsymbol{y}_{i0}, \boldsymbol{z}_i) = \sum_{\boldsymbol{\alpha}_{i1}} \dots \sum_{\boldsymbol{\alpha}_{iT}} \Bigg[p^{(1)}(\boldsymbol{\alpha}_{i1} = \boldsymbol{\xi}_c|\boldsymbol{z}_{i1}, \boldsymbol{y}_{i0}) \prod_{t=2}^{T} p^{(t)}(\boldsymbol{\alpha}_{it} = \boldsymbol{\xi}_{\overline{c}}|\boldsymbol{\alpha}_{i,t-1} = \boldsymbol{\xi}_c, \boldsymbol{z}_{it}) \times \\ \times \prod_{t=1}^{T} p_{12}(\boldsymbol{y}_{it}|\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1}) \Bigg] \tag{7}$$

which implies the sum over all the possible configurations of $\boldsymbol{\alpha}_{i1}$ given the latent Markov process (MacDonald and Zucchini, 1997). Probability $p(\boldsymbol{y}_i|\boldsymbol{x}_i, \boldsymbol{y}_{i0}, \boldsymbol{z}_i)$ can be efficiently computed by using a forward recursion based on the Baum-Welch algorithm, which is a special case of the EM (Dempster et al., 1977) algorithm used to find the unknown parameters of a latent Markov model. The EM algorithm alternates two steps until convergence. In the first step, named E-step, the conditional expected value of the complete data log-likelihood is computed given the observed data and the corresponding current estimate $\widetilde{\boldsymbol{\theta}}$. Finally the M-step maximizes the preceding expected value with respect to $\boldsymbol{\theta}$. In practice, to estimate this model a few steps of EM algorithm are usually implemented to refine starting values, which are used in the direct numerical maximization of the likelihood in equation (6).

From the copula based HMBPo estimates, the share ω_c of individuals in the population belonging to the unobserved state $c$can be obtained by using the posterior distribution of latent states ξ_c. The posterior probability $p\big(\boldsymbol{\alpha}_{it} = \boldsymbol{\xi}_c|\boldsymbol{y}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{it-1}, \boldsymbol{z}_{it}\big)$ is obtained via Bayes's formula and backward recursion involved by equation (3). This is computed for each individual i at time occasion t. Using this information the posterior weight ω_c for each latent state is obtained as follows:

$$\omega_c = \frac{1}{N} \sum_{i=1}^{N} \Bigg[\frac{1}{T} \sum_{t=1}^{T} p\big(\boldsymbol{\alpha}_{it} = \boldsymbol{\xi}_c|y_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{it-1}, \boldsymbol{z}_{it}\big) \Bigg], \text{ with } c = 1, \dots, k \tag{8}$$

The proposed model allows us to disentangle the sources of correlation between GPs and nurses' visits and to quantify the extent to which their relationship is complementary or substitutive. This has important implications for health policy, as it provides insights

into the optimal allocation of resources in primary care and the design of interventions to improve health outcomes.

4 Results

We start estiamting the copula normal HMBPo model under different number of latent states in order to explore how many groups are adequate to describe the unobeserved heteroegenity in the population. A standard approch to identify the proper number of latent states is to compare the estimates from HMBPo model using the Bayesian Information Criterion (BIC). Usually model with lower BIC is preferred to others. In our case the XXXX

Table 2 presents the estimated coefficients for GP and nurse visits, with a focus on the dynamics introduced by lagged consultation variables. A notable finding is the significant substitution effect observed between GP and nurse consultations. Specifically, lagged nurse visits ($Num\ Visits\ Nurse_{t-1}$) have a negative and statistically significant impact on subsequent GP visits ($\gamma_{nurse,t-1} = -0.0378, p < 0.01$), indicating that prior consultations with a nurse reduce the likelihood of a GP visit. Similarly, lagged GP visits ($Num\ Visits\ GP_{t-1}$) show a negative and significant effect on nurse visits ($\gamma_{GP,t-1} = -0.0306, p < 0.01$).

Additionally, we find positive state dependence in consultation patterns, evidenced by positive coefficients on the first lag of the dependent variable ($\gamma_{own,t-1} > 0$). For GP visits, the lagged GP visit coefficient is 0.0310 ($p < 0.01$), while for nurse visits, the lagged nurse visit coefficient is 0.1156 ($p < 0.01$). This finding underscores the importance of continuity of care, indicating that patients who have consulted either practitioner type are more likely to have follow-up consultations with the same type of healthcare provider.

These findings underline the interplay between the two types of healthcare providers, where consultations with one type decrease the probability of seeking care from the other in subsequent periods, while continuity in care with the same provider type is reinforced. This dynamic reflects potential substitution effects, which may arise from patients addressing overlapping healthcare needs during prior consultations or from system-level strategies aimed at optimizing provider workload.

Finally, the positive coefficients for the total number of diagnoses ($Tot\ Num\ of\ Diagn$) across both models (0.0729 for GP visits and 0.0573 for nurse visits, both $p < 0.01$) high-

light that individuals with more diagnosed conditions are more likely to utilize both types of care. These results, in combination with the lagged effects, provide a comprehensive picture of healthcare utilization patterns and suggest areas for optimizing resource allocation.

These results have important implications for workforce planning and resource allocation in primary care settings, particularly in the context of recent NHS initiatives aimed at expanding the primary care workforce. The observed substitution effects between GP and nurse consultations suggest potential opportunities for workload optimization. Specifically, the negative lagged effects ($\gamma_{nurse,t-1}$ and $\gamma_{GP,t-1}$) imply that enhancing the capacity and accessibility of one type of practitioner could reduce the demand for the other. For example, increasing the availability of nurse consultations could alleviate pressure on GP services, offering a more balanced distribution of workload and potentially improving overall system efficiency.

At the same time, the positive association in consultation patterns ($\gamma_{own,t-1} > 0$) underscores the critical importance of continuity of care. Patients are more likely to have follow-up consultations with the same type of healthcare provider, highlighting the value of sustained patient-provider relationships. This finding supports recent policy shifts in the NHS to strengthen team-based care models and encourage long-term interactions between patients and their primary care providers. Ensuring that patients can consistently access the same practitioner type—be it GP or nurse—can foster trust, improve health outcomes, and reduce the need for duplicative care.

Moreover, the positive association between the total number of diagnoses ($Tot\ Num\ of\ Diagn$) and consultation frequency suggests that individuals with more complex health needs disproportionately utilize primary care services. This reinforces the importance of targeted workforce expansion strategies that focus on high-need patient groups. Deploying more specialized nurse practitioners or condition-specific care teams, for example, could better address the needs of patients with multiple diagnoses, such as those with chronic conditions like COPD or cancer.

In light of these findings, the NHS's recent efforts to recruit additional nurses and allied health professionals are particularly relevant. By leveraging the substitution effects identified in this analysis, primary care teams can allocate resources more effectively, ensuring that both GPs and nurses are utilized to their full potential. Furthermore, policies aimed at integrating new staff into existing care pathways should prioritize mechanisms to

preserve continuity of care, such as enabling flexible scheduling or incentivizing follow-up appointments with the same provider.

Overall, these insights provide a robust evidence base for aligning workforce planning initiatives with patient utilization patterns, thereby improving the efficiency and quality of care delivery in primary care settings.

Table 2: Estimated Coefficients for GP and Nurse Visit Models: Substitution and State Dependence Effects

	Num GP visits	Num Nurse visits
Gender	0.0700 ***	0.0140
	(0.01)	(0.01)
Age	0.0006	0.0029 ***
	(0.001)	(0.001)
Imd2	-0.0035	0.0058
	(0.01)	(0.01)
Imd3	-0.0107	0.0396 ***
	(0.01)	(0.01)
Imd4	-0.0386 ***	-0.0043
	(0.01)	(0.01)
Imd5	-0.0126	-0.0231
	(0.01)	(0.01)
Tot Num of Morbidities	0.0729 ***	0.0573 ***
	(0.002)	(0.003)
Anxiety	0.0539 ***	-0.1042 ***
	(0.01)	(0.01)
CVD	0.0182 ***	0.0160 **
	(0.01)	(0.01)
Cancer	0.0931 ***	0.0418 ***
	(0.01)	(0.02)
COPD	0.0073	0.1911 ***
	(0.01)	(0.01)
Num Visits Nurse$_{t-1}$	-0.0378 ***	0.1156 ***
	(0.001)	(0.002)
Num Visits GP$_{t-1}$	0.0310 ***	-0.0306 ***
	(0.001)	(0.001)
Constant†	0.7971***	-0.6391***

† Weigheted at support points
Standard errors are in parentheses. * significant at 10%, ** significant at 5%,*** significant at 1%.

Table 3 presents the estimated support points (ξ_c) for the latent states in the context

Latent	GP visit		Nurse visit	
State	$\alpha_1(\xi_c)$	SD	$\alpha_2(\xi_c)$	SD
1	-1.6834***	(0.03)	-3.9176***	(0.08)
2	0.6785***	(0.02)	-0.6842***	(0.03)
3	1.5241***	(0.02)	0.2494***	(0.03)
4	2.1223***	(0.02)	1.003***	(0.03)

Table 3: Estimated support points ξ_c for the subject-specific parameters [2]

of GP and nurse visits, along with their associated standard deviations (SD). The results provide valuable insights into the heterogeneity of healthcare utilization patterns across the latent states, revealing how these visits relate to an underlying latent variable across four states. In State 1, both types of visits show strong negative associations, suggesting potential substitution. In contrast, States 3 and 4 display positive associations for both GP and Nurse visits, implying a complementary relationship. These results indicate that the relationship between GP and Nurse visits is conditional on the latent state, with substitution occurring in some cases and complementarity in others, highlighting the complex dynamics at play.

Figure 2 shows the joint utilization of GP and nurse consultations for a given latent state. Based on the estimated model, we categorize individuals into four distinct latent states, reflecting different levels of healthcare utilization. Individuals in State 1 are classified as low users, as they typically make fewer consultations with both the GP and the Nurse. In State 2, individuals are classified as medium-low users, with a higher likelihood of GP consultations, although the frequency of visits remains low for both types. State 3 represents medium-high users, where individuals exhibit more frequent consultations with both the GP and Nurse, though the number of visits is still moderate compared to higher states. Finally, State 4 consists of high users, characterized by a greater frequency of both GP and Nurse visits, despite the smaller size of this state. These classifications provide insights into varying levels of healthcare utilization and highlight the heterogeneity in how individuals access healthcare services. The size of these four latent states in the population is equal to 0.1541, 0.3572, 0.3728 and 0.1159, respectively. Interestingly, the size of those who are low users and heavy users is relatively larger than the other two groups. This is not surprising since the population is composed by patients

[2]Standard errors are in parentheses. * significant at 10%, ** significant at 5%,*** significant at 1%.

aged 50 or older. In this population, the number patients with no health condition (good health) or very sick are likely to be relatively smaller as compared to intermediate health status.

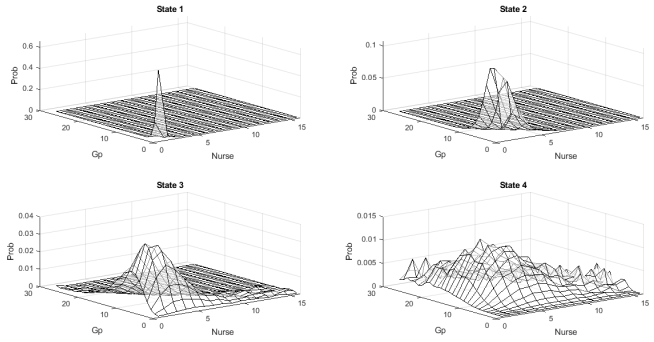


Figure 2: Predicted joint distribution of GP and Nurse consultations by latent types

4.1 Time-varying heterogeneity across states

Figure 3 shows the class membership probabilities at each time point. In practice, this figure provide information on the size of the four latent groups of the population in each period of time. The plot reveals three key shifts in the composition of probabilities across the states. At the first time point, the highest probability is associated with State 3, with an individual having approximately a 37% chance of belonging to this state, while the lowest probability is linked to State 4. From the third period onward, a reversal occurs between State 1 and State 4, with State 4 now having the lowest probability of membership. In the sixth and final period, another shift occurs between State 3 and State 2, with State 2 now having the highest probability of belonging.

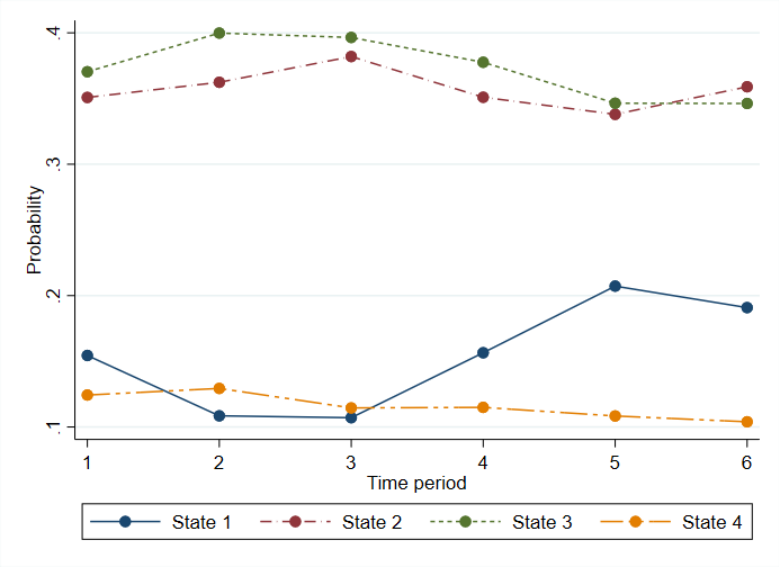


Figure 3: Estimated average probabilities of latent states in each period

These time-varying heterogeneity can be unpacked in order to evaluate how the use of healthcare varies with the number of morbidities an individual has. Typically, individuals with one morbidity, such as hypertension, are more likely to develop additional comorbidities, including conditions such as diabetes, cardiovascular disease, and others (Roberts et al., 2014; Cushman, 2003; Martín-Fernández et al., 2023). As the number of morbidities increases, people often require more frequent healthcare services, leading to higher utilization of GP and nurse consultations. This pattern highlights the interconnected nature of chronic diseases and underscores the importance of managing multiple health conditions to prevent further complications and reduce healthcare burdens.

Using the estimated from the HMBPo model we can recover the posterior probabilities described in (8), to assign each individual to a specific latent state and/or to evaluate how these assignment is affected by observed health status. Figure 4 illustrates the composition of states and how they vary with the number of secondary diagnoses, reflecting the role of multimorbidity in healthcare utilization. The probability of belonging to State 1 (low users) decreases as the number of secondary diagnoses, or morbidities, increases. This is because people in the first state typically have fewer health issues and are considered to be in better health. As the number of comorbidities rises, the likelihood of remaining in this state diminishes, reflecting an increased need for healthcare services. In State 2, the probability decreases with the number of diagnoses until an inversion point is

reached at higher levels, indicating that individuals with more complex health conditions may initially experience a lower rate of healthcare use, but eventually require more care as their conditions worsen. In State 3, the probability of being in this state increases with the number of diagnoses, suggesting that individuals with multiple morbidities are more likely to belong to the intermediate groups of medium-low or medium-high users. In State 4, the probability increases with more diagnoses up to a threshold (around five), after which it declines. This reversal likely reflects the smaller size and greater heterogeneity of this high-usage group, where individuals with severe or multiple conditions may experience varied healthcare needs.

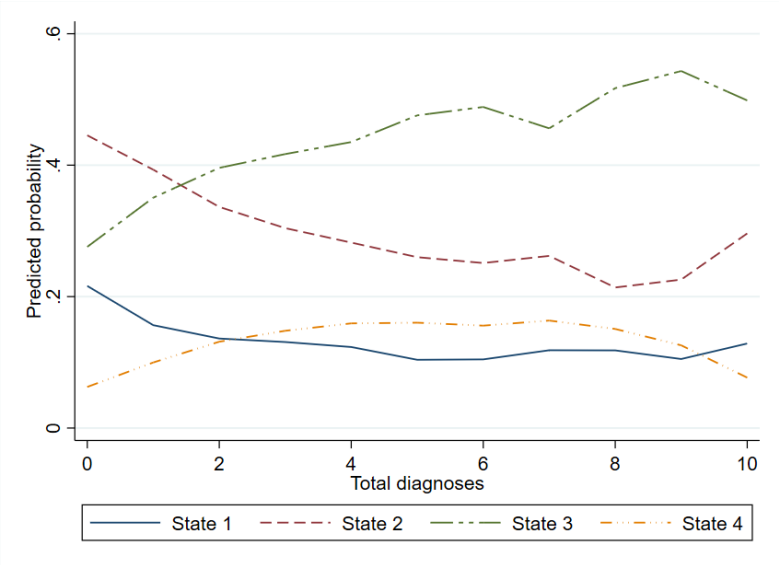


Figure 4: Association between number of secondary diagnoses and probability of belonging to a different state

Figure 5 describes how the ex-post probability of belonging to a specific state is associated by the number of morbidities. The findings show that as the number of morbidities increases, the probability of being in State 3 (medium-high users) and State 4 (high users) also rises. This suggests that individuals with more morbidities are more likely to require frequent GP and nurse consultations, reflecting the greater healthcare needs associated with multiple or more severe health conditions.

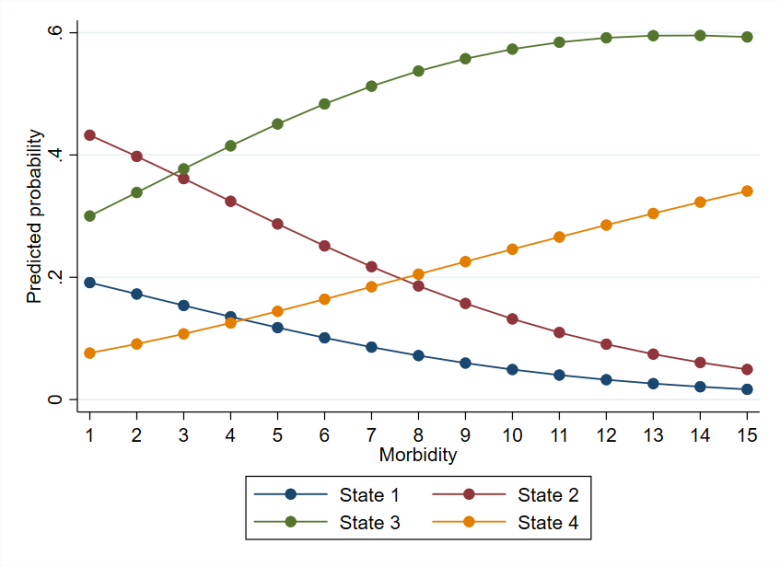


Figure 5: Association between number of morbidity and probability of belonging to a different state

We also unpack this association by analysing the composition of the states based on the type of morbidity, as classified by the Cambridge Morbidity Indicator of 36 types of morbidities (Head et al., 2021). Figure 6 shows that for all types of morbidity considered in the analysis, individuals are more likely to be in the medium-high user state (State 3). This suggests that individuals with various morbidities, regardless of the specific type, tend to experience higher healthcare utilization, reflecting the increased need for medical care associated with more complex health conditions. In particular, learning disabilities emerge as a significant morbidity that is frequently comorbid with psychosis and schizophrenia, and individuals in this group show particularly high levels of joint healthcare utilization. This pattern highlights the need for tailored healthcare interventions for individuals with these overlapping conditions, as they tend to require more frequent and intensive care. Understanding the relationship between these morbidities and healthcare usage is crucial for informing healthcare resource allocation and intervention strategies, especially as individuals with multiple or severe conditions are more likely to fall into higher usage categories.

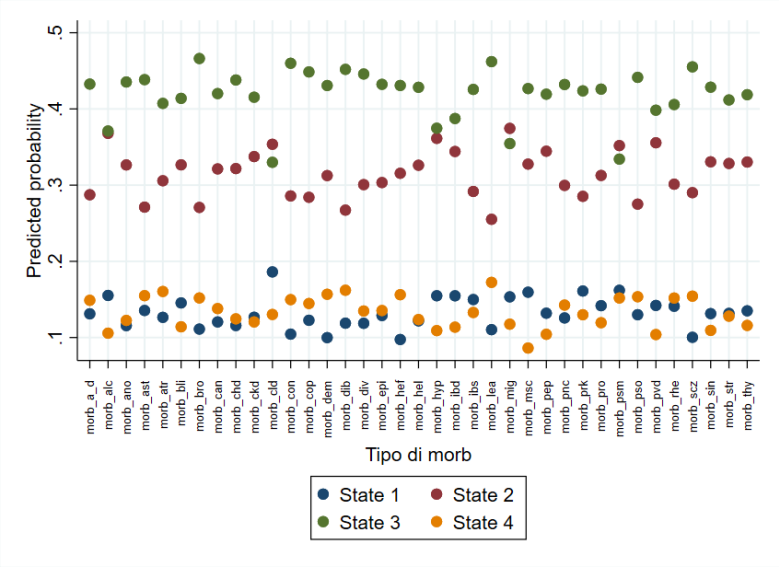


Figure 6: Type of morbidity and probability of belonging to a different state

4.2 Transition across states

An important piece of information the HMBPo model provides refers to the probability an individual has to switch from one state to another. In this section, we examine the transition probabilities across the different states based on age. Understanding these transitions is crucial for predicting healthcare utilization patterns and identifying how age influences movement between different levels of healthcare usage.

Figure 7 illustrates that, for each initial state, the probability of transitioning to other states increases sharply between the ages of 47 and 50. Furthermore, for all initial states, the likelihood of transitioning to State 3 (medium-high users) significantly rises after the age of 60. This indicates that individuals over 60 are more likely to transition into State 3, regardless of their initial state, as they become more frequent users of healthcare services. This pattern underscores the effect of aging on healthcare utilization, with older individuals requiring increased GP and nurse consultations.

For each initial state, the probability of transitioning to other states rises sharply between the ages of 47 and 50. Additionally, for all initial states, the probability of transitioning to State 3 (medium-high users) begins to increase significantly from the age of 60 onward. This suggests that individuals over the age of 60 are likely to transition into State 3, regardless of their initial state, as they become medium-high users of healthcare

services.

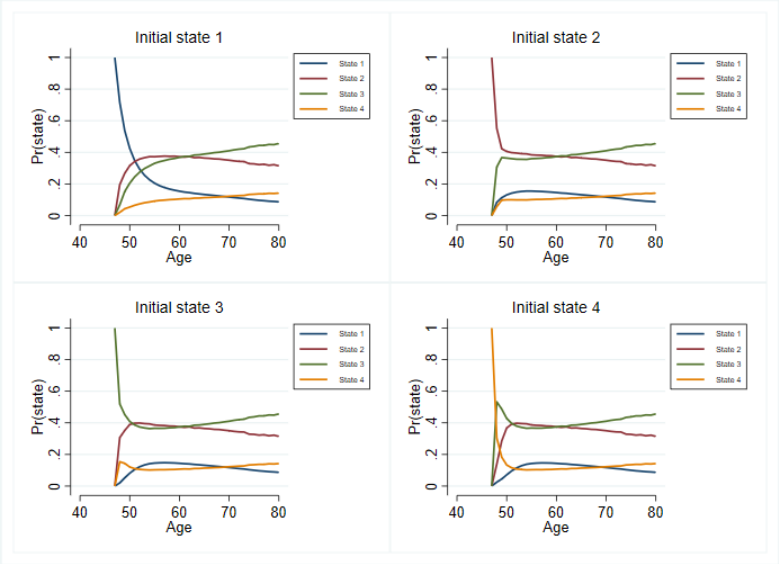


Figure 7: Transition probabilities for patients in a specific initial state at age 47

Instead, Figure 8 presents the transition probabilities based on the number of secondary diagnoses. For each initial state, the probability of transition increases rapidly as the number of secondary diagnoses rises from 0 to 5. Additionally, individuals with more than 5 diagnoses are more likely to transition into State 3 (medium-high users). The only exception observed in the graph is for individuals initially in State 3 (medium-high users), who, despite a slight increase in transition probability for higher numbers of diagnoses, remain predominantly in State 3. This indicates that individuals in this state are less likely to transition to a higher usage category, even with an increasing number of secondary diagnoses.

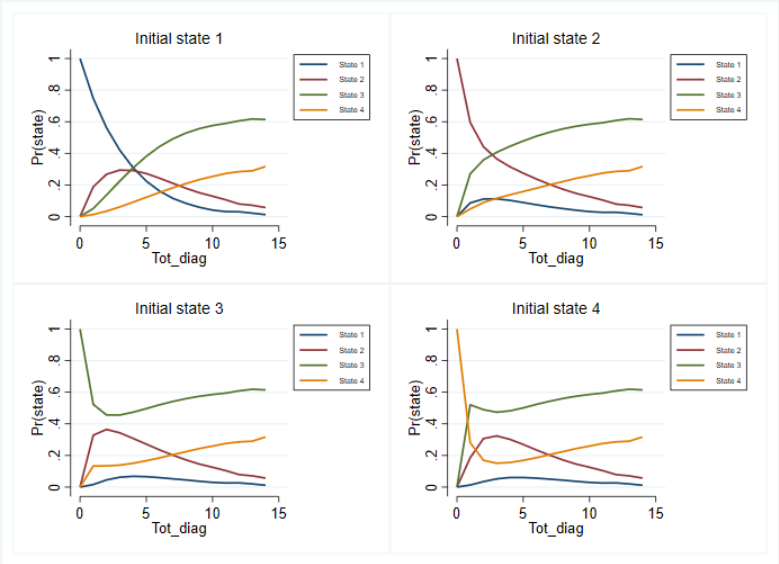


Figure 8: Transition probabilities for patients in a particular initial state without secondary diagnoses

Figure 9 illustrates how the transition probabilities change when individuals are initially 47 years old and have no secondary diagnoses. In this scenario, the pattern is similar to the previous case, but with the probability of transition increasing up to two secondary diagnoses. Additionally, the probability of switching into State 3 (medium-high users) rises steadily as the number of secondary diagnoses increases, starting from a value of 2 and continuing through to 10 secondary diagnoses. Noticeably, the probability of switch from any state to State 4 is always increasing with the number of diagnoses. This suggests that even at a younger age with fewer diagnoses, individuals are more likely to move into higher healthcare usage categories as their number of secondary diagnoses increases.

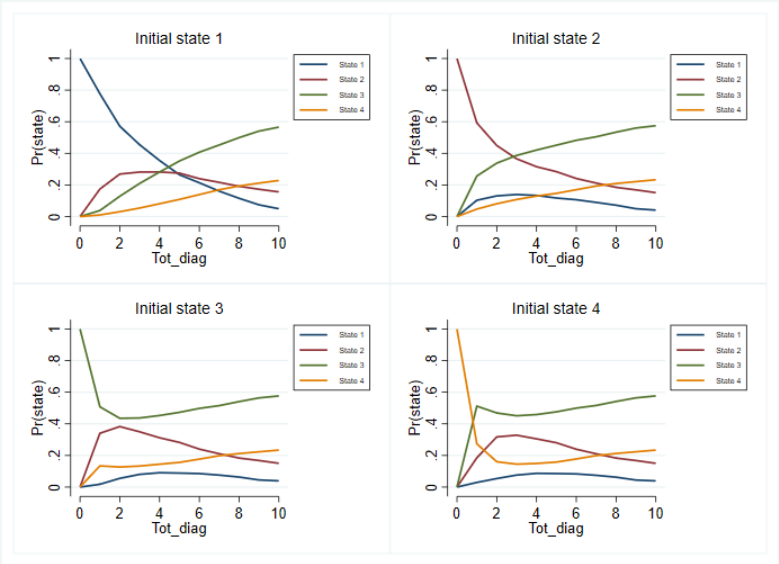


Figure 9: Transition probabilities for patients in a particular initial state without secondary diagnoses and at age 47

5 Discussion and Conclusions

This study provides important insights into healthcare utilization in primary care, focusing on the interactions between GPs and nurses and how these dynamics are influenced by age, multimorbidity, and specific health conditions. Using a hidden Markov chain framework, we identified distinct patterns of healthcare utilization and examined transitions between states of low, medium, and high usage. The results reveal critical substitution and complementarity effects between GPs and nurses, emphasizing the growing healthcare demands of individuals with multiple morbidities and aging populations. Individuals with more morbidities are significantly more likely to transition into medium-high and high usage states, underscoring the interconnected nature of chronic diseases and the escalating demand for primary care services. Furthermore, specific morbidities, such as learning disabilities combined with psychosis and schizophrenia, were shown to drive particularly high utilization of both GP and nurse services. These results highlight the importance of tailored care for vulnerable groups, whose complex health conditions necessitate more frequent and integrated care pathways.

From a policy perspective, these findings have important implications for workforce planning within the NHS. The substitution effects observed between GPs and nurses

suggest that expanding the roles of allied health professionals, such as advanced nurse practitioners and physician associates, could alleviate GP workloads. These findings support ongoing NHS efforts to integrate additional clinical roles into primary care and strengthen team-based models of care delivery. This research emphasizes the critical role of workforce optimization in addressing the increasing demands on primary care systems. The NHS's efforts to recruit and integrate allied health professionals under the Additional Roles Reimbursement Scheme (ARRS) are well-aligned with our findings, which demonstrate the potential of nurses and other roles to complement and substitute for GPs in managing routine care. This approach can enhance efficiency, address workforce shortages, and better meet the needs of an aging, multimorbid population.

The study also highlights the importance of tailoring interventions for high-need patient groups, particularly those with complex comorbidities. As healthcare demands continue to grow, policymakers must focus on strategies that promote continuity of care, strengthen multidisciplinary teams, and ensure equitable access to primary care services. Future research should evaluate the long-term impacts of workforce initiatives on patient outcomes, healthcare efficiency, and system sustainability, providing a foundation for informed policy development. By aligning resource allocation with utilization patterns, the NHS can create a more resilient and responsive primary care system.

References

Bago d'Uva, T. (2005), 'Latent class models for use of primary care: Evidence from a british panel', *Health Economics* **14**(9), 873–892.

Cushman, W. C. (2003), 'The burden of uncontrolled hypertension: morbidity and mortality associated with disease progression', *The Journal of Clinical Hypertension* **5**(3), 14–22.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977), 'Maximum likelihood from incomplete data via the EM algorithm', *Journal of the Royal Statistical Society. Series B (Methodological)* **39**(1), 1–38.
URL: *http://www.jstor.org/stable/2984875*

Drennan, V. M., Gabe, J., Halter, M., de Lusignan, S. and Levenson, R. (2017), 'Physician associates in primary health care in england: A challenge to professional boundaries?', *Social Science & Medicine* **181**, 9–16.

Fikree, S., Hafid, S., Lawson, J., Agarwal, G., Griffith, L. E., Jaakkimainen, L., Mangin, D. and Howard, M. (2023), 'The association between patients' frailty status, multimorbidity, and demographic characteristics and changes in primary care for chronic conditions during the covid-19 pandemic: a pre-post study', *Family practice* **40**(4), 523–530.

Frees, E. W. and Valdez, E. A. (1998), 'Understanding relationships using copulas', *North American Actuarial Journal* **2**(1), 1–25.

Halter, M., Drennan, V. M., Joly, L. M., Gabe, J., Gage, H. and de Lusignan, S. (2017), 'Patients' experiences of consultations with physician associates in primary care in england: A qualitative study', *Health Expectations* **20**(5), 1011–1019.

Head, A., Fleming, K., Kypridemos, C., Schofield, P., Pearson-Stuttard, J. and O'Flaherty, M. (2021), 'Inequalities in incident and prevalent multimorbidity in england, 2004–19: a population-based, descriptive study', *The Lancet Healthy Longevity* **2**(8), e489–e497.

Heckman, J. J. (1981), Heterogeneity and state dependence, *in* S. Rosen, ed., 'Studies in labor markets', University of Chicago Press, Chicago, IL, chapter 3, pp. 91–140.

Hyppolite, J. and Trivedi, P. (2012), 'Alternative approaches for econometric analysis of panel count data using dynamic latent class models (with application to doctor visits data)', *Health Economics* **21**(Suppl. 1), 101–128.

Karthika, M., Abraham, J., Kodali, P. and Mathews, E. (2023), 'Emerging trends of chronic diseases and their care among older persons globally', *Handbook of Aging, Health and Public Policy: Perspectives from Asia* pp. 1–24.

Lakshminarayana, J., Pandit, S. and Rao, K. S. (1999), 'On a bivariate poisson distribution', *Communications in Statistics - Theory and Methods* **28**(2), 267–276.

Laudicella, M. and Li Donni, P. (2022), 'The dynamic interdependence in the demand of primary and emergency secondary care: A hidden markov approach', *Journal of Applied Econometrics* **37**(3), 521–536.

Laurant, M., van der Biezen, M., Wijers, N., Watananirun, K., Kontopantelis, E. and van Vught, A. J. (2018), 'Nurses as substitutes for doctors in primary care', *Cochrane Database of Systematic Reviews* (7).

MacDonald, I. L. and Zucchini, W. (1997), *Hidden Markov and other models for discrete-valued time series*, Vol. 110, CRC Press, Boca Raton, FL.

Malone, R. (2022), 'The role of the physician associate: an overview', *Irish Journal of Medical Science (1971-)* **191**(3), 1277–1283.

Martín-Fernández, J., Alonso-Safont, T., Polentinos-Castro, E., Esteban-Vasallo, M. D., Ariza-Cardiel, G., González-Anglada, M. I., Sánchez-Perruca, L., Rodríguez-Martínez, G., Rotaeche-del Campo, R. and Bilbao-González, A. (2023), 'Impact of hypertension diagnosis on morbidity and mortality: a retrospective cohort study in primary care', *BMC Primary Care* **24**(1), 79.

Porter, J., Boyd, C., Skandari, M. R. and Laiteerapong, N. (2023), 'Revisiting the time needed to provide adult primary care', *Journal of general internal medicine* **38**(1), 147–155.

Roberts, E. R., Green, D. and Kadam, U. T. (2014), 'Chronic condition comorbidity and multidrug therapy in general practice populations: a cross-sectional linkage study', *BMJ Open* **4**(7), e005429.

The King's Fund (2023), 'NHS activity in a nutshell: How is the NHS performing?', https://www.kingsfund.org.uk/insight-and-analysis/data-and-charts/NHS-activity-nutshell. Accessed: 2024-10-24.

Whelton, P. K., Carey, R. M., Aronow, W. S., Casey, D. E., Collins, K. J., Himmelfarb, C. D., DePalma, S. M., Gidding, S., Jamerson, K. A., Jones, D. W., MacLaughlin, E. J., Muntner, P., Ovbiagele, B., Smith, S. C., Spencer, C. C., Stafford, R. S., Taler, S. J., Thomas, R. J., Williams, K. A., Williamson, J. D. and Wright, J. T. (2018), '2017 acc/aha/aapa/abc/acpm/ags/apha/ash/aspc/nma/pcna guideline for the prevention, detection, evaluation, and management of high blood pressure in adults', *Journal of the American College of Cardiology* **71**(19), e127–e248.

World Health Organization (2021), 'Hypertension fact sheet'.
URL: *https://www.who.int/news-room/fact-sheets/detail/hypertension*

Wouterse, B., Huisman, M., Meijboom, B. R., Deeg, D. J. and Polder, J. J. (2013), 'Modeling the relationship between health and health care expenditures using a latent Markov model', *Journal of Health Economics* **32**(2), 423–439.