BMJ Open

Predictors of district nursing care utilisation for community-living people in the Netherlands: an exploratory study using claims data

Jessica Desirée Veldhuizen ![ORCID],[1] Misja Chiljon Mikkers,[2,3] Marieke J Schuurmans,[2,4] Nienke Bleijenberg[1,5]

ABSTRACT

Objective To explore predictors of district nursing care utilisation for community-living (older) people in the Netherlands using claims data. To cope with growing demands in district nursing care, knowledge about the current utilisation of district nursing care is important.

Setting District nursing care as a part of primary care.

Participants In this nationwide study, claims data were used from the Dutch risk adjustment system and national information system of health insurers. Samples were drawn of 5500 pairs of community-living people using district nursing care (cases) and people not using district nursing care (controls) for two groups: all ages and aged 75+ years (total N=22 000).

Outcome measures The outcome was district nursing care utilisation and the 114 potential predictors included predisposing factors (eg, age), enabling factors (eg, socioeconomic status) and need factors (various healthcare costs). The random forest algorithm was used to predict district nursing care utilisation. The performance of the models and importance of predictors were calculated.

Results For the population of people aged 75+ years, most important predictors were older age, and high costs for general practitioner consultations, aid devices, pharmaceutical care, ambulance transportation and occupational therapy. For the total population, older age, and high costs for pharmaceutical care and aid devices were the most important predictors.

Conclusions People in need of district nursing care are older, visit the general practitioner more often, and use more and/or expensive medications and aid devices. Therefore, close collaboration between the district nurse, general practitioner and the community pharmacist is important. Additional analyses including data regarding health status are recommended. Further research is needed to provide an evidence base for district nursing care to optimise the care for those with high care needs, and guide practice and policymakers' decision-making.

Strengths and limitations of this study

► The results of this study confirm current knowledge that people in need of district nursing care are older and have higher median healthcare costs regarding general practitioner consultations, pharmaceutical care and aid devices.

► Because people in need of district nursing care visit the general practitioner more often and use more and/or expensive medications and aid devices, the results of this study underline that close collaboration between the district nurse, general practitioner and the community pharmacist is important.

► The random forest algorithm is robust to outliers, noise, overfitting and is capable of dealing with large amounts of observations as well as potential predictors.

► Because of the nature of claims data, it was not possible to include potentially relevant predictors based on the literature since this was not readily available in the dataset.

INTRODUCTION

The worldwide population of 906 million older people aged 60 years and older in 2015 will increase rapidly to approximately 2.1 billion older people in 2050.[1] With increasing age and associated adverse consequences like frailty, disability and (multiple) chronic diseases, demands for care will grow and healthcare costs will increase.[2–4] To meet the needs of older people and decrease their costs of care, policies of governments aim towards ageing in place, shifting care from institutions to home and prioritising community-based services.[4] This leads to an increased demand for home healthcare. Home healthcare comprises all care delivered at home, including in-home nursing care, as well as housekeepers, mobile meals, physical therapy, occupational therapy, social work services and care provided by the general practitioner (GP) at home. Personal care, psychosocial care and technical nursing care to community-living older people are mostly performed by district nursing care (DNC).[5] Demands on DNC will increase due to the ageing population, the increase of complexity

of care and the shortage of DNC professionals.[2 4 6–8] In order to cope with these growing demands, knowledge about the current users of DNC is important.

The context of DNC can be confusing due to the international differences in DNC practices and the variety of titles and names used for district nurses.[9–11] In general, DNC refers primarily to the formal nursing services and personal care provided by nurses or health assistants behind someone's front door, enabling people to remain living in their home environment.[5 9 11] Two comprehensive studies have been conducted to identify differences in district nursing practices between countries in Europe.[9 12] Both studies conclude that in each country, the healthcare system is embedded in a sociocultural and political context, asking for country-specific actions in order to cope with the growing demands on DNC.[9 12] In the Netherlands, DNC has a technical, supportive, rehabilitative or preventive nature, covering technical nursing care, psychosocial care and personal care.[5] It is directed at both individual patients and community populations.[5] From 2015, DNC in the Netherlands is a shared responsibility of municipalities, insurers and the national government. It is funded on a fee-for-service basis, and it is financed through the Health Insurance Act.[13] DNC is provided by district nurses, vocational nurses, healthcare assistants and health aides. In 2017, a total of 557 005 people received DNC, of which 343 300 people were women (61,6%), 276 115 people lived alone (49.6%) and 353 570 people were 75 years or older (63.5%).[14]

To study the utilisation of home healthcare, including DNC, the behavioural model of healthcare service utilisation by Andersen[15] is a widely applied model.[16 17] The latest version of the model suggests that contextual and individual characteristics influence the use of health services.[18] These characteristics are dependent on (1) predisposing factors (ie, characteristics that exist before a person's illness, such as age, gender and health beliefs), (2) enabling factors (ie, logistical aspects such as income, health insurance and travel time to available facilities), and (3) need factors (ie, the level of disability and illness as perceived by the individual).[18]

Two recent studies conducted in the Dutch context focusing on predictors in home healthcare including DNC identified age and gender as predisposing factors,[17] income[17] or social environmental characteristics[19] as enabling factor, and impairment,[17] physical functioning and daily functioning[19] as need factors. Both studies focus on home healthcare, which includes social work, household care, and/or care delivered by municipals, rather than district nursing specifically. Little is known regarding predictors associated with DNC utilisation compared with no DNC utilisation.

The aim of this study is to explore predictors of DNC utilisation for community-living (older) people in the Netherlands using claims data. The results of this study may contribute to better awareness and understanding of older people in need of DNC. This insight may also guide professionals, researchers, and policymakers in providing

care and further research regarding the use and potential demands of DNC.

MATERIALS AND METHODS
Study design and data sources

To explore predictors of DNC for community-living (older) people, an exploratory study was conducted, using a nationwide patient-level dataset including healthcare claims data of all insured people in the Netherlands. In this study, people who use DNC were compared with people who do not use DNC.

The dataset used in this study has been created by combining three national datasets: data from the Dutch risk adjustment system by the Dutch Healthcare Authority, claims data of the Dutch national information system of health insurers (Vektis) and data with patient characteristics (Vektis characteristics). In the Dutch risk adjustment system, health insurers are compensated for predictable, health-related cost differences among insured people. Data regarding this risk adjustment (eg, socioeconomic status, persons per household) are included in this dataset. The claims dataset includes all expenses that were claimed for all delivered care in 2017. The patient characteristics dataset includes variables as age and gender. The three datasets include data from all people living in the Netherlands. All data were collected during 2017. The datasets were linked and merged on person level using the pseudo-anonymised identification number. For the flow chart of the merging of the datasets and selection of variables, see online supplemental appendix 1. Only observations that were available in all datasets were included. Removed variables were irrelevant for the scope of the study (eg, regarding persons living in a nursing home) or were already available in one of the other datasets. New variables were created when it was needed (ie, to create dummy variables due to the format of the variable). Observations were removed if gender was missing (148 802 observations; 0.7%) or if the outcome was negative (22 observations; 0.0%). A negative outcome was possible if an administrative adjustment regarding DNC costs was made with respect to the previous year (2016), and no DNC costs were made in 2017. Due to people changing their health insurer during 2017, duplicates (581 210 observations; 3.3%) were identified and removed from the dataset. The final dataset comprised data from 16 833 188 persons on 115 variables (online supplemental appendix 1). To guide reporting, the Reporting of studies Conducted using Observational Routinely-collected data statement was followed (online supplemental appendix 2).[20]

Participants and sample

Because of prior knowledge that DNC is mainly provided to older people, DNC utilisation was predicted for two groups: those aged older than 75 years (75+) and the total population including all ages. Because of the long running time of the proposed analyses, calculations were
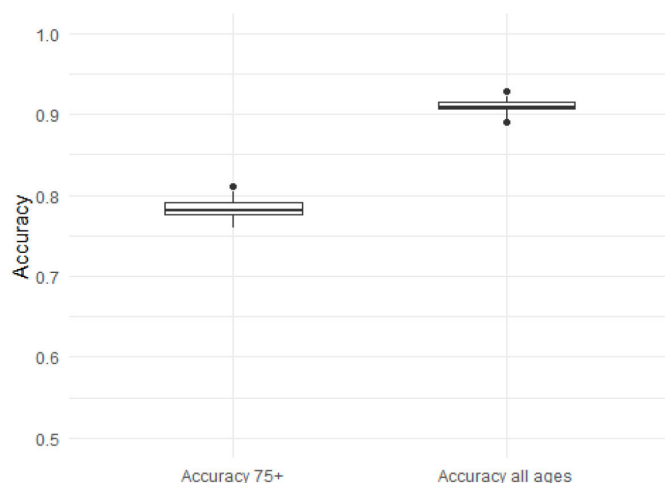
Figure 1 Accuracy of the random forest models for 75+ years and all ages. K-fold cross-validation has been performed, using 10 folds and 5 repeats. The boxplots show the distribution of the accuracies from the 50 models that were calculated in the test sets.

shows the accuracy of all models for all samples. A sample of 5.000 and 10.000 showed the best accuracy with small CIs and minor differences between both models. Due to the long runtime of a sample of 10.000 observations, a sample of 5.000 was preferred. An increase of more observations than 5000 does not substantially improve the accuracy estimates of our models. A sample of 1.0% of the total population of DNC users came closest to the favourable sample. With 544.304 people receiving DNC in our final dataset, the sample was rounded up to 5.500 (1.01%). An equal amount of non-DNC users was drawn from the final dataset as controls, resulting in a total of 11 000 observations. Two separate samples of 11 000 observations were drawn for both groups (75+ years and all ages). The samples were drawn randomly from the total dataset, including 16 833 188 community-living persons in the Netherlands.

Patient and public involvement
Patients or the public were not involved in the design of the study.

Outcome variable and potential predictors
The outcome variable was DNC utilisation in 2017, operationalised as all people who claimed expenses regarding DNC they received in 2017. Next to the outcome variable, 114 potential predictors were included. Because of the nature of included datasets, only data were available on patient characteristics and healthcare utilisation costs. No data were available regarding health status or diagnosis.
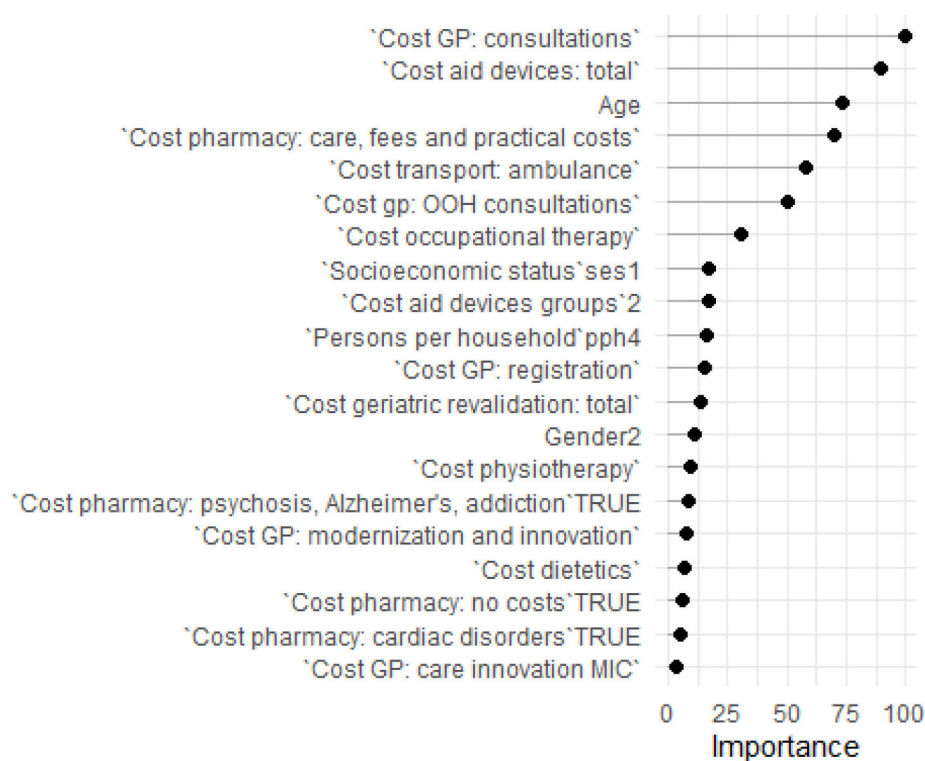
made to determine a sample of sufficient size with a feasible length of the analysis running time. Samples of different sizes (250, 500, 1.000, 2.500, 5.000, 10.000) were drawn. A prediction model was made using the random forest algorithm by Breiman[21] and the implementation by Wright and Ziegler.[22] Subsequently, the accuracies of the models were calculated for all samples. K-fold cross-validation has been performed, using 10 folds and 5 repeats. Figure 1 in online supplemental appendix 3

Figure 2 Predictor importance 75+ years. cost aid devices groups 2, compression stockings; gender2: female; GP, general practitioner; MIC: multidisciplinary integrated care; OOH, out-of-office hours (evening, night, weekend); pph4, persons per household group 4 (living alone); ses1, low socioeconomic status.
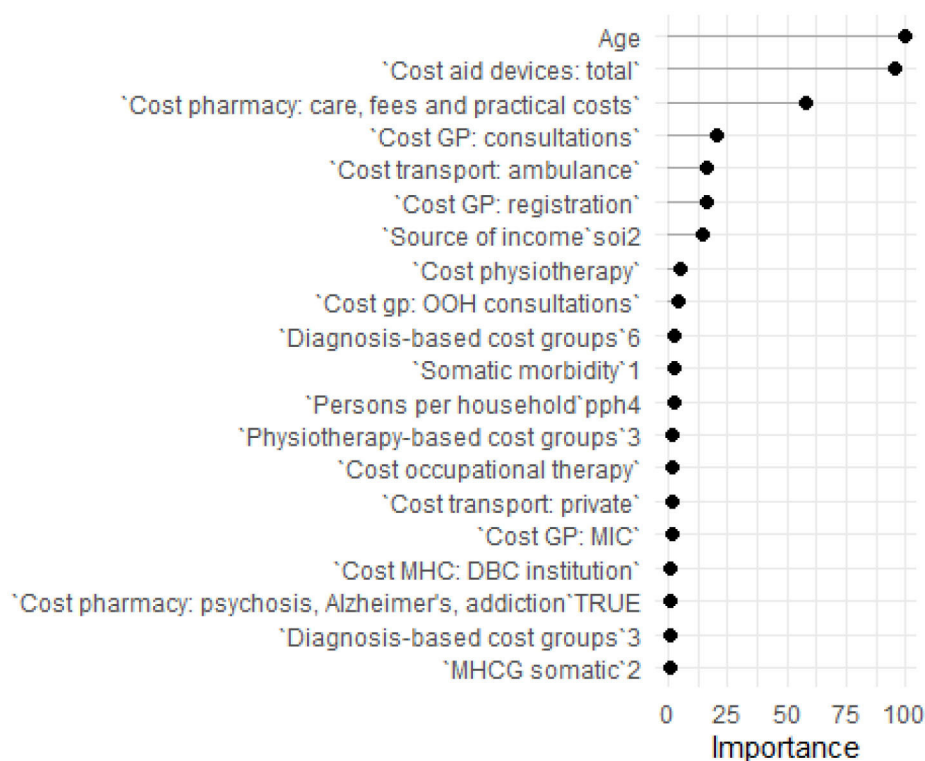
Figure 3 Predictor importance all ages. DBC, diagnosis treatment combination; diagnosis-based cost groups 3, unspecified; diagnosis-based cost groups 6, unspecified; GP, general practitioner; MHCG: multi-year high-cost group somatic 2, at least twice MHC costs in last 5 years: top 10%; MIC, multidisciplinary integrated care; OOH, out-of-office hours (evening, night, weekend); physiotherapy-based cost groups 3, unspecified; pph4, persons per household group 4 (living alone); soi2, source of income 2 (65 years and older); somatic morbidity 1, morbidity.

These predictors were divided into predisposing factors (age, gender), enabling factors (socioeconomic status, persons per household and source of income) and need factors. The need factors included the total costs regarding healthcare utilisation and were divided into nine categories: aid devices, pharmaceutical care, GP care, mental healthcare, paramedic care, oral care, transport and care abroad, other and cost-based groups, which include somatic morbidity, operationalised as having at least two types of costs for somatic pharmacy, diagnosis groups, physiotherapy groups or aid devices. The healthcare utilisation predictors comprised a sum of all costs a person made in 2017. An overview of all variables and their operationalisation can be found in online supplemental appendix 1.

Statistical analysis

To explore predictors of DNC utilisation, a predictive algorithm was performed for both groups (75+ years and all ages). The random forest algorithm by Breiman[21] and the implementation by Wright and Ziegler[22] were used. The random forest algorithm is a powerful, non-parametric statistical method for exploring large amounts of potential predictors.[21] In a random forest algorithm, many decision trees are made, in which each node is split using the best among a subset of randomly chosen predictors.[21] In this study, the random forest algorithm made

500 decision trees. Next, the predictions of all decision trees are aggregated.[21]

Although the random forest algorithm is capable of dealing with large amounts of predictors, predictive performance may be affected by adding irrelevant predictors. Therefore, recursive feature elimination has been performed to select predictors for model estimation. In an initial, analysis was calculated with how many predictors the accuracy of the model would be the highest, comparing models with 1, 20, 40, 60, 80 and 100 predictors. The models with 40 predictors for both 75+ years and all ages had the highest accuracy (0.79 and 0.91, respectively) (see online supplemental appendix 4).

To evaluate the performance of the models, the prediction accuracy has been calculated using a k-fold cross-validation, using 10 folds and 5 repeats. In every fold, 10% of the sample is used as a test dataset and 90% as training dataset. With 10 folds and 5 repeats, the prediction accuracy has been calculated for the 50 models using the test sets.

To conclude what predictors (features) contribute most in the decision-making in the model, the feature importance method was applied.[21 22] To measure the importance of the predictors in the models, the permutation accuracy importance measure has been used. To estimate the importance of a predictor, the random forest algorithm calculates how much the prediction error increases

Table 1 Descriptive statistics of important predictors with an importance of 25 or higher

	75+		All ages	
	No DNC	**DNC**	**No DNC**	**DNC**
Cost GP: consultations				
Median (min–max)	51 (0–2020)	129 (0–2884)		
IQR (Q1–Q3)	83 (18–102)	157 (–69-226)		
Cost aid devices: total				
Median (min–max)	0 (0–32 492)	315 (0–15 841)	0 (0–7604)	305 (0–71 398)
IQR (Q1–Q3)	143 (0–143)	810 (50–859)	0 (0–0)	879 (33–911)
Age				
Median (min–max)	80 (76–107)	84 (76–103)	41 (0–103)	79 (0–104)
IQR (Q1–Q3)	6 (78–84)	8 (80–88)	37 (22–59)	70 (16–86)
Cost pharmacy: care, fees and practical costs				
Median (min–max)	306 (0–35 052)	944 (0–71 390)	31 (0–133 641)	918 (0–63 177)
IQR (Q1–Q3)	634 (116–750)	1218 (421–1639)	140 (0–140)	1339 (396–1736)
Cost transport: ambulance*				
Median (min–max)	0 (0–4846)	0 (0–9608)		
IQR (Q1–Q3)	0 (0–0)	690 (0–690)		
Cost GP: OOH consultations*				
Median (min–max)	0 (0–1088)	0 (0–1998)		
IQR (Q1–Q3)	0 (0–0)	141 (0–141)		
Cost occupational therapy*				
Median (min–max)	0 (0–1232)	0 (0–1197)		
IQR (Q1–Q3)	0 (0–0)	0 (0–0)		

Only results are presented for those variables with a variable importance of 25 or higher. Aid devices include Continuous Positive Airway Pressure (CPAP) equipment, compression stockings, materials for stoma patients, nebuliser with accessories, materials for urine collection, syringes with accessories (excluding diabetes), oxygen delivery devices and accessories, nutritional aids and materials (excluding infants), phlegm suction equipment and portable infusion pumps.
*The differences in the medians and overall costs seem minimal. The histogram in figure 4 shows the differences between DNC and no DNC.
DNC, district nursing care; GP, general practitioner; OOH, out-of-office hours (evening, night, weekend).

when data for that predictor are permuted while all others are left unchanged.[23] The relative importance of predictors was ranked by tallying the number of inclusions of the predictor in all decision trees. The importance of the predictors was normalised on a scale from 0 to 100 and was plotted on the horizontal axis, with the ranked predictors on the vertical axis.

To provide a description of DNC users and non-DNC users, descriptive statistics were calculated for the predictors having a predictor importance of 25 out of 100 or higher. The cut-off of 25 was based on the elbow of the curve in the predictors importance plot. Because of the expected non-normal distribution, the median and IQR were calculated for the predictors. Histograms of the most important predictors were calculated to provide a visual interpretation of the data.

To calculate the magnitude and direction of the predictors on the probability of having district nursing, in addition to the random forest analysis, two binary (Bayesian) logistic regression analyses were conducted. The selected variables from the random forest analysis were included

as independent variables. The use of DNC is taken as the outcome variable. ORs were provided to interpret the association between the independent variables and the outcome. For each variable estimate, a credibility interval was calculated. A credibility interval shows the probability (95%) that the true population value falls within this interval.[24] All calculations were made using R V.3.5.3.[25]

RESULTS

From the total population of 16 833 188 persons included in the dataset, 544 304 people received DNC in 2017 (3,2%). Of those aged older than 75 years, 328 767 (26.2%) received DNC. From the total users of DNC, 60.4% were older than 75 years.

Accuracy of the models with 40 predictors

Using recursive feature elimination with 10-fold cross-validation and 5 repeats, 50 models were built with 40 predictors. The median accuracies for the models of 75+ and all ages were, respectively, 0.79 and 0.91 (figure 1).
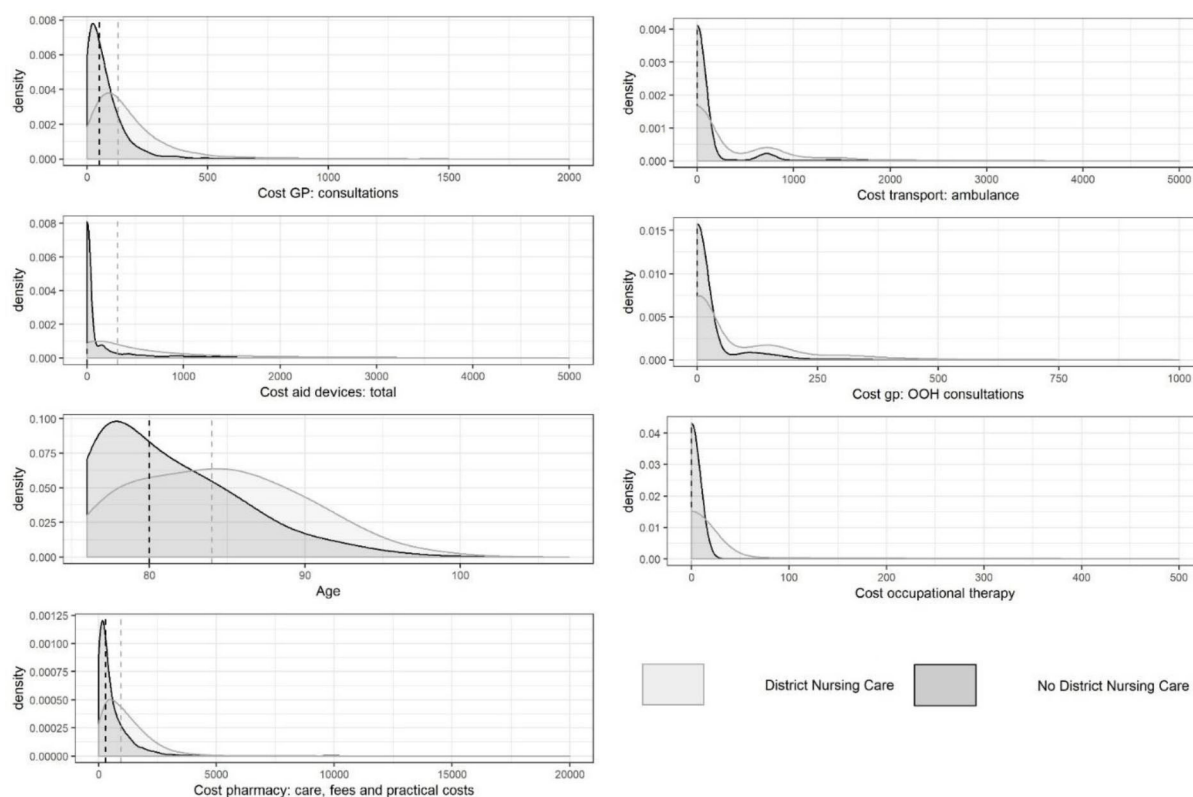
Figure 4 Histogram of important predictors with an importance of 25 or higher for those aged 75+ years. These plots do not contain outliers of extremely high costs (horizontal axis) since the outliers made the plots not readable. The plots with outliers on the horizontal axis can be found in online supplemental appendix 5, figure 1. GP, general practitioner; OOH, out-of-office hours (evening, night, weekend).

Predictor importance

In total, seven predictors had an importance of >25 and higher for the prediction of DNC utilisation in people aged 75 years and older. These were the total costs of GP consultations during office hours, the total costs of aid devices, age, total costs for the use of pharmaceutical care, total costs of ambulance transportation, total costs of GP consultation after office hours and the total costs of occupational therapy (figure 2). Other patient characteristics in the top 20 were low socioeconomic status, living alone and female gender.

For the total population with all ages, three predictors of DNC utilisation had an importance of 25 and higher: age, the total costs of using aid devices and the total costs for the use of pharmaceutical care (figure 3). The top five most important predictors included the costs of GP consultations and the costs of ambulance transport. Other patient characteristics in the top 20 are somatic morbidity and living alone.

For those aged older than 75 years, all people receiving DNC had a higher median age and higher median total costs regarding GP consultation, use of aid devices and pharmaceutical care compared with those not receiving DNC (table 1). Regarding ambulance transportation, GP consultation after office hours and occupational therapy, people using DNC had higher overall costs in a year compared with those not using DNC (figure 4). For the group with all ages, people receiving DNC had a higher

median age and higher median total costs regarding use of aid devices and pharmaceutical care compared with those not receiving DNC (table 1, figure 5).

Logistic regression and final model accuracy

The positive coefficients of the Bayesian logistic regression confirm that higher age and costs are important predictors of DNC utilisation (table 2). All variables in the model are credibly associated with the probability of DNC utilisation, with age having the highest association with the outcome DNC. The logistic regression models have an out-of-sample accuracy of 0.744 for those aged 75 years and 0.873 for all ages, which comes close to the accuracy of the random forest analyses (with accuracies of 0.79 and 0.91, respectively).

DISCUSSION

This paper identified important predictors of DNC utilisation using claims data. For the population of older people (75+ years), seven predictors were the most important, being higher total costs of GP consultations during office hours, higher total costs of aid devices, older age, higher total costs for pharmaceutical care, higher total costs of GP consultations after office hours, higher total costs for transport by ambulance and higher total cost of occupational therapy. For the total population, older age, higher total costs for pharmaceutical care and higher total costs
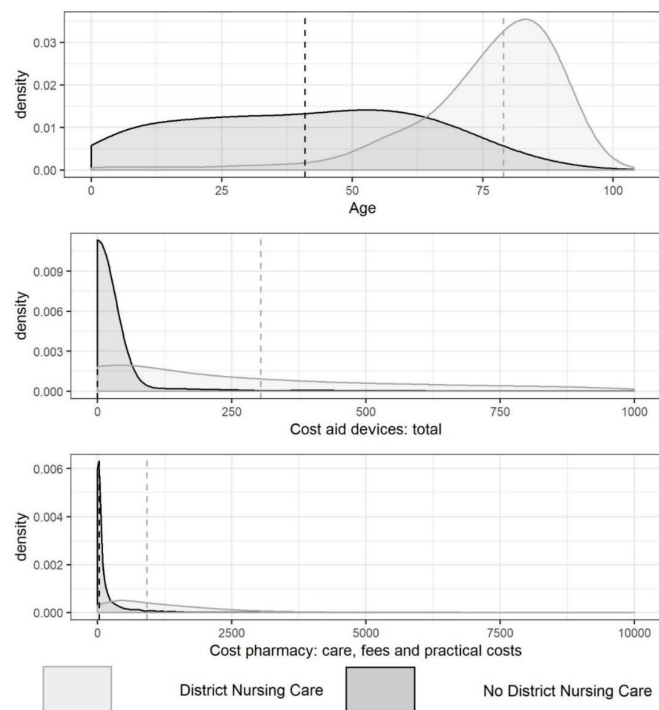
Figure 5 Histogram of important predictors with an importance of 25 or higher for all ages. These plots do not contain outliers of extremely high costs (horizontal axis) since the outliers made the plots not readable. The plots with outliers on the horizontal axis can be found in online supplemental appendix 5, figure 2.

for aid devices were the three most important predictors. The logistic regression models with these variables have an accuracy of 0.87 for those aged 75+ years and 0.74 for all ages. Differences were found in the order of the predictor importance for the total population compared with those aged 75+ years. In the total population, age was

the most important factor. When looking at 75+ years, age became less important, while the cost of GP consultations and the cost of using aid devices became more important.

Allocating the identified predictors in the behavioural model of healthcare service utilisation by Andersen,[18] age was an important predisposing factor. None of the predictors with an importance of 25 or higher were enabling factors. Identified need factors were the costs for GP consultations during and after office hours, aid devices, pharmaceutical care, transport by ambulance and occupational therapy. These need factors can be seen as indicators for the level of disability and illness (ie, more or severe disabilities or illnesses could lead to higher use and more costs, eg, aid devices or pharmaceutical care). However, it is unfortunate that no data could be included regarding functioning and functional impairment, which are important predictors according to the literature.[17 19] The literature overview by van Noort et al[17] identified age, functional impairment, gender and income as important predictors of DNC severity. In our study, age was in both groups one of the most important predictors of DNC utilisation. The costs of using aid devices, pharmaceutical care, GP consultations and occupational therapy can be seen as an indicator of functional impairment. In our study, female gender and a low socioeconomic status were predictors only for those aged 75+ years. In this study, socioeconomic status is a measure based on income and educational level. A low socioeconomic status (low income and educational level) is a well-known predictor for frailty among home-dwelling older people.[26] However, with an importance of <25, gender and low socioeconomic status were not as important as age and impairment. The systematic review by van den Bulck et al[19] identified social environmental characteristics, physical functioning and daily functioning as important categories for case-mix

Table 2 Bayesian logistic regression models for those aged 75+ years and all ages

	OR	Credibility interval	Accuracy of the model
Model 1: 75 years and older			0.744
Intercept	0.00005	0.00002–0.00010	
Cost GP: consultations	1.00455	1.00403–1.00507	
Cost aid devices: total	1.00038	1.00032–1.00046	
Age	1.11116	1.10120–1.12099	
Cost pharmacy: care, fees and practical costs	1.00035	1.00029–1.00040	
Cost transport: ambulance	1.00092	1.00078–1.00106	
Cost GP: OOH	1.00268	1.00211–1.00327	
Cost occupational therapy	1.00340	1.00254–1.00432	
Model 2: all ages			0.873
Intercept	0.00325	0.00255–0.00411	
Age	1.08240	1.07869–1.08620	
Cost aid devices: total	1.00154	1.00137–1.00171	
Cost pharmacy: care, fees and practical costs	1.00069	1.00062–1.00078	

The ORs of all healthcare utilisation costs are per euro, which explains the low ORs.

predictors. An important social environmental characteristic that was identified in our study was a household size of one person per household (ie, living alone), which was a predictor in the top 20 of both groups. Also, the costs of using aid devices, pharmaceutical care, GP consultations, and occupational therapy that were identified in this study can be seen as an indicator of physical or daily functioning.

Strengths and limitations

An important strength was using claims data from all insured people in the Netherlands. The results of this study are therefore applicable for the total population in the Netherlands. This dataset includes a large number of potential predictors of those using DNC and the healthy population who are not in need of DNC. Because of the size of the dataset, a powerful analysis was needed. The random forest algorithm is robust to outliers, noise, over-fitting and is capable of dealing with large amounts of observations as well as potential predictors.[21] Our aim was to get useful information about the relation between DNC utilisation and its predictors, rather than developing a clinical prediction rule. In a random forest analysis, the goal is not interpretability but providing accurate information.[27] By calculating the importance of the predictors, the analyses offer a good balance of accuracy and understanding a model.

In the interpretation of the results, some limitations should be considered. First, because of the nature of claims data, it was not possible to include potentially relevant predictors based on the literature regarding health status, such as physical and daily functioning,[17 19] since this information was not readily available in the dataset. While the costs of healthcare utilisation other than DNC can be seen as an indicator for limitations in daily functioning, the dataset provides no insight into the details regarding the functional impairment on the patient level. Second, while the analysis provides insight into important predictors, a limitation of the predictor importance analysis is that the plot does not show if the predictors are positively or negatively associated with the outcome. Only the strength of the dependency is reflected.[21] This limitation has been minimised by providing descriptive statistics, histograms and the results of the Bayesian logistic regression.

Implications and further research

The results of this study showed that people in need of DNC have higher median healthcare costs regarding GP consultations, pharmaceutical care and aid devices compared with those not using DNC. People using DNC visit the GP more often and use more and/or expensive medications and aid devices. With a GP as a gatekeeper prescribing medication and sometimes the use of aid devices,[28] a community pharmacist providing medication and materials, and a district nursing team providing the care for the patient, close collaboration between these professions is vital. A recent literature review showed that for the best individual, continuous care, an interprofessional non-hierarchical team should realise home visits, as these are central to patient-centeredness and clinical responsibility.[29] Coordinated care by interdisciplinary teams is associated with better outcomes regarding emergency department visits, hospitalisations and long-term care admissions in homebound older adults.[30] However, existing studies regarding this topic are mainly observational and further research based on well-controlled studies is needed.[29] Additionally, it is needed to develop a strong evidence base for DNC in the near future, focusing on evidence-based guidelines and DNC service delivery, including the role of interprofessional care team members.[31] These potential new insights could guide practice's and policymakers' decision-making regarding the use and demands for DNC.

Conclusion

This is the first study that identified predictors of DNC utilisation using claims data. Older people (75+ years) in need of DNC are older and have higher total healthcare costs regarding GP consultations during and after office hours, pharmaceutical care, aid devices, transportation with the ambulance and occupational care. For the total population with all ages, the most important predictors are older age, higher total costs for pharmaceutical care and higher total costs for using aid devices. Because no data regarding health status and diagnosis could be included due to the nature of the datasets available, additional analyses are recommended. Additionally, further research is needed to provide an evidence base for DNC to optimise the care for those with high care needs, and guide practice and policymakers' decision-making.

Author affiliations
[1]Research Group Proactive Care for Older People at Research Centre for Healthy and Sustainable Living, HU University of Applied Sciences Utrecht, Utrecht, The Netherlands
[2]Dutch Healthcare Authority, Utrecht, The Netherlands
[3]Department of Economics, Tilburg University Tilburg School of Economics and Management, Tilburg, The Netherlands
[4]Education Center, UMC Utrecht Academy, University Medical Centre Utrecht, Utrecht, The Netherlands
[5]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands

Ethics approval This study does not need informed consent from the patients because the datasets were provided by the Dutch Healthcare Authority, which pseudonymised and de-identified patients' information before analyses. This was checked by the ethical review board, and a waiver of consent was granted.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

ORCID iD
Jessica Desirée Veldhuizen http://orcid.org/0000-0002-7761-5690

REFERENCES

1 United Nations DoEaSAPD. *World population prospects: the 2017 revision, key findings and advance tables. Working paper No. ESA/P/WP/248*. New York: United Nations DoEaSAPD, 2017.
2 Maybin J, Charles A, Honeyman M. *Understanding quality in district nursing services*. London: The King's Fund, 2016.
3 Patterson L. Making our health and care systems fit for an ageing population: David Oliver, Catherine Foot, Richard Humphries. King's Fund March 2014. *Age Ageing* 2014;43:731.
4 Spasova S, Baeten R, Coster S. *Challenges in long-term care in Europe: a study of national policies 2018*. European Commission, 2018.
5 Maurits EEM. Autonomy of nursing staff and the attractiveness of working in home care. Doctoral dissertation, Utrecht University, 2019. Available: https://dspace.library.uu.nl/handle/1874/378721
6 Carrera F, Pavolini E, Ranci C. Long-term care systems in comparative perspective: care needs, informal and formal coverage, and social impacts in European countries. In: Ranci C, Pavolini E, eds. *Reforms in long-term care policies in Europe: investigating institutional change and social impacts*. New York, NY: Springer New York, 2013: 23–52.
7 Carpenter I, Gambassi G, Topinkova E, *et al*. Community care in Europe. The aged in home care project (AdHOC). *Aging Clin Exp Res* 2004;16:259–69.
8 Jackson C, Leadbetter T, Martin A, *et al*. Making the complexity of community nursing visible: the Cassandra project. *Br J Community Nurs* 2015;20:126–33.
9 der Boom H. Home nursing in Europe: patterns of professionalisation and institutionalisation of home care and family care to elderly people in Denmark, France, the Netherlands and Germany. Doctoral dissertation, Amsterdam University Press, 2008. Available: https://cris.maastrichtuniversity.nl/en/publications/home-nursing-in-europe-patterns-of-professionalisation-and-instit
10 Morris H. Reinstating district nursing: a UK perspective. *Nurse Educ Pract* 2017;26:59–63.
11 Genet N, Boerma W, Kroneman M. Home Care across Europe - Current structure and future challenges. *World Heal Organ* 2012;156.
12 van Eenoo L, Declercq A, Onder G, *et al*. Substantial between-country differences in organising community care for older people in Europe-a review. *Eur J Public Health* 2016;26:213–9.
13 Wammes JJG. High-cost patients and opportunities to reduce unnecessary spending. Doctoral dissertation, Radboud University Nijmegen, 2018. Available: https://repository.ubn.ru.nl/bitstream/handle/2066/196835/196835.pdf
14 Vektis intelligence. Wijkverpleging [online], 2018. Available: https://www.zorgprismapubliek.nl/producten/zorg-dichtbij/wijkverpleging [Accessed 14 Apr 2020].
15 Andersen RM. *A behavioral model of families' use of health services. Research Series no. 25*. University of Chicago: Center for Health Administration Studies, 1968.
16 Kadushin G. Home health care utilization: a review of the research for social work. *Health Soc Work* 2004;29:219–44.
17 van Noort O, Schotanus F, van de Klundert J, *et al*. Explaining regional variation in home care use by demand and supply variables. *Health Policy* 2018;122:140–6.
18 Andersen RM. National health surveys and the behavioral model of health services use. *Med Care* 2008;46:647–53.
19 van den Bulck AOE, de Korte MH, Elissen AMJ, *et al*. A systematic review of case-mix models for home health care payment: making sense of variation. *Health Policy* 2020;124:121–32.
20 Benchimol EI, Smeeth L, Guttmann A, *et al*. The reporting of studies conducted using observational Routinely-collected health data (record) statement. *PLoS Med* 2015;12:e1001885.
21 Breiman L. Random forests. *Mach Learn* 2001;45:5–32.
22 Wright MN, Ziegler A. Ranger: a fast implementation of random forests for high dimensional data in C++ and R. *arXiv Prepr* 2015:arXiv150804409.
23 Liaw A, Wiener M. Classification and regression by randomForest. *R News* 2002;2:18–22.
24 Van de Shoot R, Depaoli S. Bayesian analyses: where to start and what to report. *Eur Health Psychol* 2014;16:75–84.
25 Team RDC. R (version 3.5.3). R: a language and environment 2018.
26 Dury S, De Roeck E, Duppen D, *et al*. Identifying frailty risk profiles of home-dwelling older people: focus on sociodemographic and socioeconomic characteristics. *Aging Ment Health* 2017;21:1031–9.
27 Breiman L. Statistical modeling: the two cultures. *Stat Sci* 2001;16:199–215.
28 Faber MJ, Burgers JS, Westert GP. A sustainable primary care system: lessons from the Netherlands. *J Ambul Care Manage* 2012;35:174–81.
29 Vaartio-Rajalin H, Fagerström L. Professional care at home: Patient-centredness, interprofessionality and effectivity? A scoping review. *Health Soc Care Community* 2019;27:e270–88.
30 Stall N, Nowaczynski M, Sinha SK. Systematic review of outcomes from home-based primary care programs for homebound older adults. *J Am Geriatr Soc* 2014;62:2243–51.
31 Jarrín OF, Pouladi FA, Madigan EA. International priorities for home care education, research, practice, and management: qualitative content analysis. *Nurse Educ Today* 2019;73:83–7.