



HAL
open science

Trajectories and individual determinants of regular cancer screening use over a long period based on data from the French E3N cohort

Clara Dugord, Carine Franc

► **To cite this version:**

Clara Dugord, Carine Franc. Trajectories and individual determinants of regular cancer screening use over a long period based on data from the French E3N cohort. *Social Science & Medicine*, 2022, 294, pp.114663. 10.1016/j.socscimed.2021.114663 . hal-04385507

HAL Id: hal-04385507

<https://hal.science/hal-04385507v1>

Submitted on 22 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Trajectories and individual determinants of regular cancer screening use over a long period based on data from the French E3N cohort

Keywords

Cancer - screenings - regularity – trajectory - determinants – cohort – France

1 Introduction

Cancers have become one of the leading causes of death in OECD countries, with rates greater than those of infectious diseases (National Institute of Demographic Studies, *Institut National d'Etudes Demographiques* (INED), 2019). However, their prevalence, morbidity and mortality could be significantly reduced by comprehensive and regular prevention practices (Adami et al., 2001). Despite the implementation of diverse cancer prevention policies in France over the last two decades to promote regular use of cancer screenings (organized screening of breast cancer since 2004, colorectal cancer since 2009, cervical cancer since 2018; financial incentive mechanism for general practitioners since 2012), the use of cancer screenings remains far below public objectives. For example, the use of breast cancer screening has even decreased in recent years, with only 50% of eligible French women screened in 2017, versus 52% in 2014 (National Cancer Institute, *Institut National du Cancer* (INCa), 2018a). Cancer screening aims to reduce the expected loss amount of damage sustained in the event of illness (breast and colorectal cancers) or to prevent the occurrence of the illness (cervical and colorectal cancers). Thus, the demand for cancer screenings is mainly based on self-insurance behavior but also on self-protection behavior (Ehrlich and Becker, 1972). If cancer screening is expected to diminish the burden of disease, it also requires immediate monetary and nonmonetary (time, disutility) costs. Thus, in accordance with human capital models (Grossman, 1972; Cropper, 1977), the decision to screen should be the result of the individual trade-off between expected utility with and without costly investment in the stock of health (Newhouse, 2021).

According to the “*fundamental screening hypothesis*” formulated by Eddy (1980), prognosis should be a function of how early screening is performed. Since the earliness of the diagnosis of a cancer is a function of the recurrence of screening, it is the proper follow-up of public health screening recommendations that is the key to the effectiveness of reducing the burden of disease (Duffy and Paci, 2012; Marmot et al., 2013). Indeed, 90% of cervical cancers could be avoided by following the screening recommendations (INCa, 2021a)); 9 out of 10 colorectal cancers can be cured when detected early (INCa, 2021b), and based on a randomized controlled trial, Shaukat et al (2013) estimated that

individuals who undergo colorectal cancer screening every 2 years are 22% less likely to die from it. Regular mammography is also associated with earlier cancer detection (INCa, 2018b) since it reduces breast cancer mortality by approximately 20% for women at average risk (Myers et al., 2015). For example, a period of 7 to 10 years of regular screening prevents the death of 100 to 300 among 100,000 women (National Health Insurance Fund, *Caisse National d'Assurance Maladie (CNAM)*, 2021a).

Nonetheless, despite numerous studies on the demand for cancer screenings, few have examined the determinants of regularity of use. Goldzahl and Jusot (2016) studied the individual determinants of regular mammography use over two consecutive periods and found a positive effect of contact with the health care system (physician, gynecologist), as well as a negative effect of low income and past economic precariousness. Under the conceptual framework of the human capital model (Grossman, 1972), regular cancer screening can be interpreted as a recurrent investment to maintain individual health capital stock and utility over time such that individuals choosing to invest once in their health capital stock through cancer screening are then expected to pursue. Nonetheless, the conditions of the individual trade-off may change over time through macro shocks (e.g., screening organization, reimbursement) and/or heterogeneous individual shock, which can be exogenous (e.g., health state, cancer of a relative, activity status, negative experience with the screening) as well as endogenous (e.g., beliefs, risk perception, aging). Thus, this should induce a variety of cancer screening user profiles, with some individuals having very stable screening behavior (never use or always use) and others exhibiting more erratic behavior with periods of use and periods of nonuse.

The empirical framework traditionally relates the individual use of prevention to education, age, income, and access costs (Kenkel, 1994). Age, increasing the health risk and the destruction of the stock of health capital, should lead to a higher demand for cancer screening (Grossman, 1972; Kenkel, 1994); however, by reducing the period of return on investment over the life cycle (Cropper, 1977), its effect could ultimately be more ambiguous. Carrierri and Bilger (2013) found a concave effect of age on cancer screening use; the health-risk effect initially increases the demand, and the life cycle then

takes over after a certain age and decreases the demand. Education and income have linear positive effects such that they are generally associated with a higher demand for cancer screenings (Kenkel, 1994; Sicsic and Franc, 2014). Access costs can constitute a significant brake on the demand for screening except when health insurance covers the direct costs of cancer screening (Walsh et al., 2011; Carrierri and Bilger, 2013). Numerous empirical studies have highlighted the multiplicity of dimensions that play a role in cancer screening uptake, such as demographic characteristics (Sicsic and Franc, 2014; Devaux, 2015), risk aversion, information aversion and time preferences (Wu, 2003; Goldzahl, 2017), health beliefs and knowledge (Klug et al., 2005; Kobayashi et al., 2014), health and family history (Kash et al., 1992; Sicsic and Franc, 2014), health behaviors (Damiani et al., 2012; Sicsic and Franc, 2014), or access to the health care system and contacts with physicians (Jones et al., 2010; Carrierri and Bilger, 2013; Sicsic and Franc, 2014). Following the literature, we focus on the role played by sociodemographic characteristics (e.g., age, education, conjugal status), care access and use (living area, use of physicians), and health and behaviors (self-perceived health, chronic disease, family cancer history, smoking status).

Finally, without considering potential global prevention practice use behaviors, studies of screening use have most often focused on the use of a single cancer screening action (Wu, 2003; Sicsic and Franc, 2014). However, following the human capital model, one could expect that an individual regularly investing in one cancer screening will also invest similarly in the other cancer screenings to maintain his/her health capital stock. Nonetheless, this assumes that barriers and levers are common across cancer screenings when they could be specific. In addition to the specifics of each screening (e.g., eligibility), aversion to the test, for example, is stronger for colorectal cancer screening than for breast cancer or cervical cancer screenings (Lo et al., 2013).

Our study aimed to address the lack of literature on individual screening behavior over time by shedding light on the individual patterns of use of 3 cancer screenings, emphasizing both the recurrence and diversity of use. We focus on the recurrence of use of the three main female cancer screenings over a period of nearly 15 years (2000 to 2014): breast cancer screening via

mammography, cervical cancer screening via Pap smear, and colorectal cancer screening via fecal occult blood testing (FOBT), hemocult and/or colonoscopy (combined under the term “FOBT/colonoscopy”). FOBT and hemocult are intended for women at normal risk of colorectal cancer and are followed by a colonoscopy if positive, while women considered at higher risk (family history, adenomas, symptoms suggestive of colorectal cancer) should undergo colonoscopy directly (CNAM, 2020). Note that colonoscopy can be prescribed for purposes other than cancer screening, but it will still detect abnormalities related to colorectal cancer. In addition to the fact that these screenings do not concern the same populations of women, it should be noted that the supply and incentive policies have undergone significant changes in France over the study period. Since 2004, breast cancer screening has been the subject of a national program targeting all women between the ages of 50 and 74. Every two years, they receive an information letter and an invitation for a free mammogram to be performed in an approved radiology center. Since 2009, colorectal cancer screening has also been the subject of a national program targeting people aged 50 to 74. Every eligible woman receives an information letter every two years on colorectal cancer and an invitation to consult with a general practitioner (GP) for screening. During the consultation, the GP gives the patient a kit that can be used to obtain a sample at home and send it using a prestamped and prefilled envelope. The only potential financial burden to the patient is the GP consultation. If the patient is at high risk, the GP directly prescribes a colonoscopy with a specialist, which may lead to a higher financial burden. Concerning Pap smear, a national screening program has been implemented in 2018, after the end of the study period. During the study period, the Pap smear is either carried out by gynecologists and GPs (more rarely by midwives, since 2009) or prescribed by GPs and carried out in an analysis laboratory, and the financial burden for the patient varies depending on the professional. Note that a Pap smear can be proposed and performed by a health professional (especially gynecologists) during an unrelated medical visit. Using data from the E3N epidemiological cohort of nearly 100,000 French women over a period of almost 30 years (www.e3n.fr), our study aimed to describe the screening behaviors for 3 cancers over a long period and the associated individual determinants. Using optimal matching (OM) methods derived from sequence analysis, we identified groups of 'typical' trajectories of use for each

of the 3 cancer screenings according to the recommendations in terms of eligibility (age, pathology) and recurrence of use. Thus, we show that for the use of cancer screening, the behaviors over the long term ultimately correspond to few typical profiles (very regular use or, conversely, low use or no use at all) or profiles of breakaway trajectories, namely, women who engaged in screening at a late stage with respect to their eligibility and then had very regular use (FOBT/colonoscopy) or women who dropped out even though they were still eligible (Pap smear). In summary, erratic behaviors seem rare. The analysis of the associations between the typical profiles of use of the 3 cancer screenings supports the hypothesis that there is a rather consistent use behavior, with some women being highly compliant and other women never or rarely undergoing screening, especially those who underscreen for breast cancer. Finally, after identifying typical patterns of use for each cancer screening, logistic and multinomial regressions were conducted to study the determinants associated with these patterns. For each cancer screening, regular screening is mainly associated with being in a relationship, having a family cancer history, not smoking, having a normal body mass index (BMI), and having a rather high level of physician use. Regular use of breast and cervical cancer screenings is also associated with a high level of education and living in an area socioeconomically privileged with an abundant supply of care.

First, we present the data and methods used. Then, we present the results before discussing them and concluding.

2 Data and methods

2.1 Data

2.1.1 The E3N cohort and study population

Data were derived from the E3N survey, which is a prospective cohort of 98,995 French women registered at the General Insurance of National Education (*Mutuelle Générale de l'Éducation Nationale*, MGEN) born between 1925 and 1950. The E3N study respects ethical and anonymisation rules and has been approved since 1988 by the French National Commission of Informatics and

Liberties (*Commission Nationale de l'Informatique et des Libertés, CNIL*, <https://www.e3n.fr/confidentialite-anonymat>). Between 1990 and 2014, 11 waves of surveys have been conducted, and each questionnaire includes questions on the performance of screening for breast, cervical and colorectal cancer. While the E3N data constitute a remarkably rich set of data, they are not representative of the French population, as 62% of the women were teachers at inclusion and 82% of them had a higher education (Clavelon-Chapel, 2015). Since 2004, the E3N data have been matched with the administrative record of health care consumption of the MGEN, making available the annual amount of ambulatory care consumed by these women. To limit the weight of ineligibility of women for the 3 cancer screenings in the sequences, we restricted the study population to women who were born between 1940 and 1950, still alive and followed up in 2009 (at least 10 years of follow-up for each woman) (Figure A.1 in Appendix A). To limit the number of “uncertain” statuses for the cancer screening while avoiding a strong selection bias, we restricted the sample to women who completed at least 3 of the 6 surveys. Finally, considering that women who declared having had cancer (malignant tumors) before 2009 were undergoing follow-up surveillance over most of the study period, we excluded them from the study population. Thus, our study population was composed of an unbalanced panel sample of 40,021 women aged 50 to 60 years in 2000 at the beginning of the study period.

2.1.2 Dependent variables

Figure 1 presents the availability of screening information: the successive screening statuses are based on the responses to the 6 survey waves (Q6-2000 to Q11-2014), the eligibility of women and the periodicity recommended by the French High Authority for Health. Women should have a Pap smear every 3 years between 25 and 65 years old and a mammogram and FOBT/colonoscopy every 2 years between 50 and 74 years old. Using successive responses to the question "*Since your last response to a questionnaire, what tests have you performed?*", we reconstructed the individual screening history *a posteriori* by considering the dates of responses to two successive questionnaires to which the woman responded. Thus, on a given date, a woman is considered "up to date" if she declares that she has

undergone screening since her last response to a questionnaire and "late" if more than 2 years have elapsed between her last mammography and FOBT/colonoscopy and 3 years for a Pap smear. Nonresponse to one or more questionnaires did not usually prevent the status from being defined; if a woman did not respond to several questionnaires and then declared that she had not been screened since her last response, she was considered "late" for the entire period between her two responses. If a woman reported screening, did not respond to the next questionnaire, and then reported screening again, we considered her to be "up to date" on the questionnaire with the missing response. In the case of several successive nonresponses, followed by a response where the woman declared to have been screened since her last response, the status was considered "uncertain" over the period of missing information (at least 5 years) since it could not reasonably be deduced. Finally, a woman was considered "ineligible" for screening for a given survey wave if she was not in the recommended age range or if she had cancer after 2009 (or a hysterectomy for cervical cancer screening).

INSERT FIGURE 1

2.1.4 Independent variables

The variables included in the models, mainly from the survey data, were collected once or appeared in several questionnaires (Figure 1), in which case we generally defined the status based on the most frequently observed period.

To characterize the sociodemographic status of women, we use the age at the beginning of the period (2000), the level of education (1992) and the number of children (1992), the status of the labor market (2000), and the most frequent conjugal status between 2002 and 2008. To account for the potentially nonlinear effect of age, we also include age squared (Carrierri and Bilger, 2013). Reduced access to care associated with costs borne by individuals (time, transport, waiting line) has been highlighted as a barrier to cancer screening use (Carrierri and Bilger, 2013; Konopka et al., 2019). This barrier may be enhanced for cervical cancer screening, as Pap smears are mainly provided by gynecologists who usually charge extra fees and are mainly located in urban areas. Moreover, we consider that cancer screening uses result from a wide combination of dimensions, some of which characterize women's

living areas, such as the socioeconomic environment and the adequacy of supply for demand of care. To take into account those dimensions, we used a typology of French territories at a fine scale (“living area”) (Chevallard and Mousquès, 2018, 2021) that combines several variables with 3 main dimensions: 1) health care availability and accessibility (e.g., local potential accessibility (Barlet et al., 2012) to GPs, nurses, midwives, and gynecologists; density of GPs; proportion of GPs older than 55 years); 2) local population characteristics (e.g., rate of elderly persons, employment rate, household income, mortality); 3) local attractiveness (e.g., average annual net migration, distance to the closest big city (minutes)). According to their reported place of residence (2008), the women were classified as living in one of six types of areas: privileged cities or suburban areas (above-average income and high accessibility to care, especially for specialists); city centers (heterogeneous socioeconomic status, young population and abundant supply of care); deprived urban or rural areas (mainly in northeastern France, most fragile population, low employment rate, poorer health status, average accessibility to care); suburban areas (lower access to primary care, average health status); tourist and retirement areas (elderly population, well endowed with care); and rural fringes (far from the major urban centers, elderly, working-class and fragile population, poor supply of care, especially primary care). In addition, the role of contact with physicians to inform people about screening has also been highlighted in the literature (Kenkel, 1994; Carrierri and Bilger, 2013). The greater the number of contacts, the greater the likelihood of opportunistic screening and the greater the opportunity to promote cancer screenings, especially those not yet covered by national programs. To quantify the use of physicians, we used the administrative record of health care consumption from 2004 to 2014 to determine the most frequent quintile of annual number of contacts with ambulatory physicians (e.g., GPs, gynecologists). To characterize health status, we used women's perceived health status (2011), the most frequent BMI category between 2000 and 2011, and an indicator of the occurrence of a chronic or acute disease between 2004 and 2018 based on the mapping of pathologies and expenses of the French National Health Insurance (CNAM, 2021b) (Table A.2 in Appendix A). To consider the effect of cancer occurrence in the women's relatives on their use of screenings, we included an indicator of cancer occurrence in first-degree family members (parents, children, siblings) declared in

2000. Finally, to consider risky behaviors, we used smoking status (2005), summarizing changes in tobacco use from 1990 to 2005.

2.2 Methods

The first step of the study was to identify typical screening use trajectories for each of the cancer screenings. While the simplest solution would have been a simple count of the status "up to date" over the period proportional to eligibility, this aggregation would have provided only an average use per woman over the period without allowing the identification of actual trajectories of screening use. Our aim was to clearly distinguish a woman who would undergo screening every other time, from a woman who would undergo screening in the first half of the period and then no longer at all, and from a woman who would only undergo screening in the second half of the study period. Thus, to better consider the differences in screening trajectories of women, we used sequence analysis methods. Based on a holistic approach to individual trajectories adapted to categorical data (Abbott and Forest, 1986), these methods aim to address two main questions (Abbott and Hrycak, 1990): are there recurrent patterns of individual trajectories, and which are they? What explanatory variables are linked to these different trajectories?

For a woman and a given cancer screening, a sequence is an ordered vector of the six successive screening use statuses over the period Q6 (2000) – Q11 (2014). Sequence analysis methods aim to compare all individual sequences with each other to establish a dissimilarity matrix between sequences on which clustering methods can be applied to obtain a typology of sequences. Studer and Ritschard (2016) defined the 5 features of individual sequences that sequence analysis methods can exploit: *experienced states* (number of distinct states in the sequence), *distribution* (total time spent in the state), *timing* (date of onset of the state), *duration* (successive time spent in the state), and *sequencing* (the order in which distinct states appear). Depending on the interest in one or more of these features, various measures of dissimilarities between two sequences have been developed. The most widely used measure of dissimilarity is based on the optimal matching (OM) algorithm, which originates from

the comparison of DNA sequences in microbiology, which was then used in sociology and recently mobilized in public health to describe care pathways (Abbott and Tsay, 2000; Roux et al., 2019).

OM establishes the dissimilarity between two sequences as the minimum cost of virtual editing operations necessary for the sequence of one individual to become identical to that of another. The more costly it is to transform the sequence of one individual into the other, the higher the dissimilarity between the sequences of the two individuals. Editing operations are composed of substitution operations and insertion/deletion ('indel') operations between the different states of the 'alphabet' (the different states a woman can take at each point of her screening use sequence: "up to date", "late", "ineligible", "uncertain"). Substitution operations correspond to the substitution of a state of one sequence by that of the second sequence at the same date, while indels consist of inserting a state in one sequence and deleting another to make it match a segment of the second sequence (an example is given in Appendix B). To consider the hierarchy existing between the distinct states of our 'alphabet', substitution operations costs are weighted: the cost of being "late" is considered twice as far away from being "up to date" as being "uncertain", and vice versa, as the substitution matrix is necessarily symmetric.

To minimize the weight of noneligibility in the sequence comparison, we assigned the minimum cost for the distance between the "ineligible" state and all other states, i.e., the lowest substitution cost in the matrix (Figure B.2.1 in Appendix B) (Halpin, 2016). Finally, we defined the costs of indels as half of the maximum opportunity cost, which is the most common practice and is a specification that corresponds to the Levenshtein I distance, which is the default distance used by the OM algorithm (Levenshtein, 1966; Hollister, 2009). The Levenshtein I distance has the advantage of being particularly sensitive to *duration* and *timing* but also, to a lesser extent, to *sequencing*.

We then applied a partitioning around medoids (PAM) clustering algorithm to the dissimilarity matrix computed by OM to obtain clusters of individual sequences based on their similarity. The PAM algorithm is a partitioning algorithm based on k-medoids (class centers) that minimizes the total dissimilarity between individuals within each cluster (Kaufman and Rousseeuw, 1990). It is the most

commonly used clustering algorithm in sequence analysis because it looks for a global maximum and is less sensitive to outliers (atypical sequences). Its limitations are that the number of clusters must be set *a priori*, and it relies on the choice of initial medoids. To circumvent the latter issue, the initial medoids are based on the class centers from a hierarchical ascending classification (HAC) (Studer, 2013). Regarding the choice of the optimal number of clusters, we tested partitions with 2 to 8 clusters, and for each cancer screening, we selected the partition containing the most relevant, homogenous, and sufficiently large (>5% of the population) clusters.

Note that among our study population, for each screening studied, sequence analysis was performed only for women eligible for at least 4 of the 6 questionnaires in the study period, i.e., 40,015 women for mammography, 37,171 women for Pap smear, and 39,961 women for FOBT/colonoscopy.

In a second step, which was restricted to the 37,114 women present in the analysis of trajectories for all 3 cancer screenings, we studied the distribution of screening trajectory clusters among the lowest cluster of use of the other two screenings to describe the associations between women's patterns of use of the different cancer screenings. Finally, to study the determinants of the identified typical trajectories, for each cancer screening, we performed logistic regressions (mammography clusters) and multinomial regressions (Pap smear clusters, FOBT/colonoscopy clusters) for the probability of belonging to each cluster vs. the cluster of women with the lowest uptake.

2.3 Software & packages

Sequence analysis was performed with R software using the packages *TraMineR* and *WeightedCluster* (Gabadinho et al., 2011; Studer, 2013).

3 Results

3.1 Population description

Table 1 presents descriptive statistics of the study population: women were on average 54 years old in 2000 and mainly lived with a spouse between 2002 and 2008 (75%); most of them had two children (89%), and 40% of them had a bachelor's degree or higher. Thus, they are clearly highly educated compared with the French female population since, on average, only 23% of French women born between 1941 and 1950 engaged in education after high school (Longone, 1973). At the beginning of the study period, 45% of the women were already retired, compared with 40% of women aged 50-64 years old in 2004 (Minni, 2006). Regarding living area, 61% of them lived in city centers, privileged cities or suburban areas, and tourist and retirement areas in 2008, i.e., areas that were rather attractive, were well-endowed with health care and had a rather high socioeconomic level for privileged cities or suburban areas and tourist and retirement areas. These three living areas appear to be slightly overrepresented in our study population; in 2016, they were the place of residence for 54.3% of the French population, while areas with poorer populations and difficult access to care were underrepresented, especially rural fringe areas, where 12.4% of our study population lived in 2008 compared with 21% of the French population (Chevallard and Mousquès, 2018) (Table A.3 in Appendix A). Regarding health, 82% of the women reported having very good or good health near the end of the study period in 2011. If 71% of the women had a chronic or acute condition between 2004 and 2014, only 27% of them still had one when not considering treatment against vascular risk and psychotropic treatments. Finally, attrition in our population study appeared low, as 87.2% of the women responded in the last survey wave (Q11 (2014)), and 99% of them were still alive in 2014.

INSERT TABLE 1

3.2 Sequence analysis results

3.2.1 Population use by screening

In the study population, mammography use (Figure C.1 in Appendix C) was very high throughout the period, with the share of eligible women “in delay” never exceeding 10% between 2000 and 2014. Pap smear use (Figure C.2 in Appendix C) was also high, with the share of eligible women “in delay”

never exceeding 30%, despite an ongoing increase over the period. In comparison, the use of FOBT/colonoscopy (Figure C.3 in Appendix C) was much lower. Nevertheless, while 73% of the eligible women did not undergo colorectal cancer screening in 2000, the percentage decreased to 28% in 2014, illustrating the rise in the use of FOBT during the 2000s, concomitant with the introduction of organized colorectal cancer screening in 2009. Note that the levels of cancer screening in our study population are higher than those of eligible women in the general population: 52.1% for mammography in 2014, 59.5% for Pap smear in 2020, and 34.7% for FOBT/hemoccult in 2017 (INCa, 2018a; Public Health France, *Santé Publique France (SPF)*, 2020).

3.2.2 Typology of sequences for each of the 3 cancer screenings

From the matrix of dissimilarities between the sequences computed by OM, we obtained a partition with clusters grouping women who had similar uptake behaviors for each screening.

Regarding mammography use, Figure 2 (and Figures D.1, D.2, D.7 in Appendix D) shows the stacking of individual sequences for the two clusters identified. Cluster 1 represents only 12.2% of the population and includes women with heterogeneous uptake behaviors, with women who never underwent mammography, others who underwent regular screenings and then stopped while they were still eligible, and women who alternated between use and nonuse. Cluster 2, which accounts for a large majority (87.7% of women), is more homogenous and comprises women with a regular mammography uptake, with a maximum of one period of delay.

INSERT FIGURE 2

For Pap smear use, Figure 3 (and Figures D.3, D.4, D.7 in Appendix D) presents the 4 clusters obtained. Cluster 1 (13.5%) includes women whose trajectories reflected zero or almost zero use over the entire period. Cluster 2 (12.3%) includes women whose use was regular at the beginning of the period and who completely stopped undergoing screening, even though they were still eligible. Finally, Clusters 3 (37.9%) and 4 (36.3%) include women whose trajectories were in accordance with

the recommendations, since they were always "up to date" during their eligibility period. As Clusters 3 and 4 differ only in the length of the eligibility period of the women, they were combined for the rest of the study into a single cluster of regular women ("Regular", 75%).

INSERT FIGURE 3

Finally, for colorectal cancer screening, Figure 4 (and Figures D.5, D.6, D.7 in Appendix D) presents the 4 clusters obtained. Cluster 1 (29.7%) includes women whose use was zero or almost zero over the entire period. Cluster 2 (21.2%) includes women whose use trajectory showed entry into screening starting from the implementation of organized screening in 2009 and then a continuation in accordance with the recommendations. Cluster 3 (33.1%) includes women who were not screened at the beginning of the period, started screening before the implementation of organized screening in 2009 and pursued screening until the end of the period. Finally, Cluster 4 (16%) includes women who underwent regular screening from the beginning of the period, with a maximum delay of one period.

INSERT FIGURE 4

The description of our typologies reveals patterns representing fairly stable cancer screening behaviors: when screening was initiated, women generally seemed to use it regularly; conversely, once they left the screening process, they usually did not return to it. Due to the stability of behaviors, it appears relevant to investigate the consistency of these behaviors across the different cancer screening types. In other words, the question is whether there is an overall use (or nonuse) behavior between the three cancer screening types or whether, for example, women screen regularly only for one cancer. Similarly, it is interesting to explore whether the determinants associated with a high (or low) uptake trajectory are common or specific to each cancer screening.

3.3 Associations between cancer screening trajectories and individual determinants

Figure 5 presents the distribution of each cancer screening use cluster among the lowest use clusters of the other two cancer screenings ('Irregular' for mammography, 'Nonuse' for Pap smear and

FOBT/colonoscopy). This highlights a gradient in the level of screening: a woman belonging to the cluster of women who screened less for one type was more likely to belong to the clusters of women who screened less for the other two screening types. For example, 38% of women who do not screen regularly for mammography belong to the cluster of women who never undergo Pap smears, and 59% belong to the cluster of women who never undergo FOBT/colonoscopy; while only 10% and 26% of women who screen regularly for mammography belong to the cluster of women who never undergo Pap smear and FOBT/colonoscopy, respectively. The associations between the levels of use of the three cancer screenings also appear to be intertwined. Indeed, the association between irregular mammography use and nonuse of Pap smears and FOBT/colonoscopy seems stronger than the association between nonuse of Pap smears or FOBT/colonoscopy and irregular mammography use. Similarly, the association between nonuse and ceasing Pap smear use seems more strongly associated with irregular mammography use and nonuse of FOBT/colonoscopy than nonuse of FOBT/colonoscopy with nonuse of Pap smears or mammography.

INSERT FIGURE 5

Table 2 presents the results of logistic and multinomial regressions conducted for each screening based on the probability of belonging to a cluster vs. the cluster of women with the lowest utilization: "Irregular" for mammography, "Nonuse" for Pap smear, and "Nonuse" for FOBT/colonoscopy.

First, for each cancer screening, we analyze the factors associated with belonging to the highest vs. lowest use cluster. Determinants appear very common between high vs. low mammography and Pap smear use, while they are more different for FOBT/colonoscopy. The main differences belong to the effect of sociodemographics, particularly education and living areas.

Regarding the use of mammography and Pap smear, age has an expected negative effect: the older a woman is, the less likely she is to belong to the "Regular" cluster, as well as to the "Regular to nonuse" cluster for the Pap smear. While having had children only has a protective effect for women for Pap smears, being in a relationship is highly protective for the two screenings. As expected, the level of education strongly influences the use of screenings such that the higher the level of education of a

woman, the higher the probability that she belongs to the "Regular" cluster for mammography or Pap smear. Likewise, living in privileged cities or suburban areas, compared with any other type of living area, constitutes a protective characteristic. Note that these living areas are characterized by fairly abundant healthcare services and very favorable economic conditions in terms of income and employment rate, for example (see Chevillard and Mousquès, 2021). In addition, as expected, being among the women who consult on average the most doctors per year greatly increases the probability of belonging to a "regular" cluster. The characteristics associated with health also have a strong influence on the use of screening: having a BMI outside the "normal" range constitutes a barrier to screening, particularly for obese women who have the lowest probability of belonging to a "regular" cluster for mammography or Pap smear. Women who report being in very good health are the most likely to belong to a "regular" cluster, as are those who report having close family members who have had cancer. Finally, women who smoke have a lower probability of belonging to a "regular" cluster for mammography or Pap smears than women who have never smoked.

The analysis of the characteristics of women for the different clusters of colorectal cancer screening use (FOBT/colonoscopy) compared with the "Nonuse" cluster shows that the determinants that characterize women belonging to the "Regular" cluster for mammography and Pap smear are the closest to those of the "Regular after 2009" cluster. This is particularly the case with the negative effect of age, the protective effect of a high level of education, the protective effect of a large number of contacts with doctors and the different health characteristics. Indeed, as with mammography and Pap smears, being obese or overweight, having poorer self-perceived health and being a smoker appear to be strong barriers to using FOBT or colonoscopy. One of the characteristics that seems to differentiate the women from each of the different use clusters for FOBT/colonoscopy is having a history of relatives who have had cancer. Indeed, women who reported having one or more family members who had cancer have higher chances to belong to the cluster of women who underwent FOBT/colonoscopy over the entire period ("Regular") or initiated it before 2009 ("Regular before 2009"). In contrast, having relatives who have had cancer is nonsignificant for women entering colorectal cancer screening after 2009 ("Regular after

2009”). As a reminder, national, organized and free screening for colon cancer (FOBT) was introduced in France in 2009.

INSERT TABLE 2

3.4 Robustness analysis

Several robustness tests were performed to ensure that the resulting partitions encapsulated the main typical screening behaviors for each cancer screening.

First, regarding the restriction of the sample, we replicated the analysis with 3 different samples: 1) women who answered all the questionnaires (n=28,071) mainly to avoid the status “uncertain” (Figures E.1.1 – E.1.3 in Appendix E); 2) women born between 1945 and 1950 (n=24,405) to test the impact of ineligibility (Figures E.2.1 – E.2.3 in Appendix E); and, finally, 3) women born between 1935 and 1950 (n=51,555) (except for the Pap smear because women born in 1935 were already 65 years old at the start of period Q6 (2000)) (Figures E.3.1, E.3.2 in Appendix E). Varying the sample restriction conditions affected the number of clusters identified by the OM, their quality (homogeneity and relevance) and the relative distribution of women among the typical screening use patterns; however, as expected, no new relevant typical pattern of screening use was identified (such as an entire cluster or part of a cluster).

Regarding the method, several dissimilarity measures were tested to assess the sensitivity of our results: (a) the generalized Hamming distance (sensitive only to *timing*), (b) the longest common subsequence (LCS) (Elzinga, 2007) (sensitive only to *sequencing*) and (c) the localized OM (more sensitive to *sequencing* and less sensitive to *timing*). The OM with the Levenshtein I distance was preselected for its simplicity and properties (mainly sensitive to *timing* and *duration* but also to *sequencing*) and then validated because the clusters obtained were more relevant and homogeneous for most cancer screenings than with other distances. In addition, none of these other dissimilarity measures revealed any new relevant pattern of screening use over those already identified; only the number of clusters, their relevance and their homogeneity were altered. Note that the Levenshtein I and the localized OM measures both result from the closest partitions for the 3 cancer screenings,

which was expected since they are based on the same common base, namely, editing operations with hierarchical substitution costs. The LCS (b) and, to a lesser extent, the Hamming distance (a) yield partitions further away from those obtained with the Levenshtein I measure, which was also expected because these 2 distances represent extreme cases of the OM and can be interpreted as bounds of the sensitivity of clusters to *timing* (a) and *sequencing* (b). The results of these tests are available upon request.

4 Discussion

The objective of this study was to construct a typology of breast, cervical and colorectal cancer screening use over a long period of time, considering women's eligibility, regularity, and trajectory of use. The typical profiles identified by OM show rather stable and homogenous behaviors for the three cancer screenings studied: some women had a very regular use of screening, others had little or no use of screening, and more interestingly, some women had a break in behaviors, either by entering screening late and then regularly following the recommendations (FOBT/colonoscopy) or, conversely, by dropping out of regular use and stopping screening (Pap smear). Women alternating between periods of use and nonuse were rare enough that their behavior was identified as a typical trajectory only for mammography. These results seem to indicate that, in general, once a woman entered screening use, she underwent regular screening, and conversely, once she left, she did not reenter. In our population study, routine use of mammography appeared to be the easiest (88% of women classified as regular), followed by Pap smears (75% of women classified as regular). However, with 30% of women classified as never having used FOBT or colonoscopy despite the introduction of organized screening in 2009, routine use of FOBT or colonoscopy appears to be more difficult to achieve.

The nonuse of colorectal cancer screening, even among women who are routinely screened for breast and cervical cancer, may be due to a stronger disutility of the FOBT and hemocult tests and a fear of

colonoscopy (Chapple et al.,2008; Lo et al.,2013). In particular, there is a strong link between nonregular mammography use and nonuse of the other two screenings. Furthermore, the associations between the patterns of use of the 3 cancer screenings suggest that there is a consistent overall pattern of use, with some women repeatedly investing in cancer screenings and others undergoing screenings less often. Targeting women who do not screen regularly for breast cancer would therefore make it possible to identify women who are particularly unlikely to adopt a broader prevention approach.

The analysis of the determinants of the typical behaviors of screening highlights that the factors related to the types of trajectories were similar overall, with some exceptions for the FOBT/colonoscopy clusters of women screened before the organized screening in 2009, which had slightly different sociodemographic and living environment characteristics. As colorectal cancer screening was less well known in the early 2000s, we can assume that women screening several times for colorectal cancer before the implementation of organized screening had specific needs. This hypothesis is supported by the fact that women with poorer self-perceived health in 2011, acute or chronic conditions over the period and a family cancer history in 2000 are more likely to belong to the clusters of women who underwent screening before 2009 rather than the cluster of women who began screening after 2009 or never underwent screening. There was an overall favorable effect on the probability of screening regularly for being in a relationship, not smoking, being younger, having a normal BMI, having a family cancer history, and visiting doctors. For mammography and Pap smears (and entering colorectal cancer screening after 2009), there was also a favorable effect of education and having very good self-perceived health. The positive relationship between education and investment in health was theorized by Grossman (1972), namely, people who are more educated produce more efficiently health, and has been empirically verified numerous times (Kenkel, 1994; Carrierri and Bilger, 2013; Sicsic and Franc, 2014). The positive effect of being in a relationship is also consistent with the literature and may be related to higher peer pressure to screen regularly for cancer, less isolation and contact with gynecologists (Sicsic and Franc, 2014; Konopka et al., 2019). The positive effect of the number of contacts with physicians may result in direct physician screening recommendations and more information about cancer risks and screening benefits (Kenkel, 1994;

Carrieri and Bilger, 2013), with the demand for curative and preventive care being complementary (Cabral and Cullen, 2017). The negative effect of age on the probability of being classified as screening regularly for breast and cervical cancers supports the life cycle hypothesis (Cropper, 1977): women with shorter pay-off periods are less incentivized to invest in their health capital stock through preventive care. Similarly, for these two cancer screenings, this may explain the positive effect of poorer perceived health on the probability of being classified in the lowest screening use clusters. These statements can be extended to colorectal cancer screenings as, except for women who had already been screening regularly since the beginning of the period, age increases the probability of belonging to the cluster of women who never screen. Note that the global linear effect of age may be due to the lack of age heterogeneity in our study population (women born between 1940 and 1950). Regarding the negative effect of having a chronic or acute illness on the probability of screening regularly for cervical and colorectal cancers, it has been shown that the discovery of a new illness for people already having health issues may lead to higher costs and more difficult treatment and rehabilitation (Nordin et al., 2002), access to screening may be more difficult for individuals with physical limitations (Bussiere et al., 2015), and people with poorer health have higher information aversion and screening anxiety (Wu et al., 2003). The strong association between having a “nonnormal” weight, particularly being obese, and lower use of cancer screenings has already been highlighted in the literature and may be explained by a higher discomfort of screening (Wu et al., 2003; Seibert et al., 2017). As expected, family cancer history is associated with a greater probability of screening, especially for colorectal cancer, which distinguishes the clusters of women regularly screened before the implementation of organized colorectal cancer screening (2009) and the clusters of women never screened or who began screening afterward. The negative effect of smoking on the probability of screening regularly is consistent with the literature, as the correlation between primary prevention through healthy lifestyles and secondary prevention has already been highlighted in the literature (Genier and Jacobzone, 1998; Wu et al., 2003; Sicsic and Franc, 2014). Finally, for breast and cervical cancer screenings, living in a privileged city or suburban areas with a high standard of living and good access to care, particularly to specialist physicians, has a positive effect on the

probability of being classified in the cluster of women that screen regularly. As expected, living in areas with lower socioeconomic conditions and limited access to care, particularly to specialists (suburban areas, deprived rural or urban areas, rural fringes), has a particularly strong negative effect on the probability of screening regularly for cervical cancer.

Finally, the high stability of cancer screening behavior supports the idea that screening use in the previous period would be the main predictor of screening use, which is in accordance with the integrated behavior model theory, according to which the habit of a behavior is a direct predictor of its achievement (Montano and Kasprzyk, 2015). Habit reduces the cost of the effort associated with the behavior and modifies the conditions of the individual trade-off made. From the individual's point of view, the cost of entry into cancer screening seems much higher than the cost of its renewal. Moreover, this stability of cancer screening behavior helps to explain the fact that the determinants associated with recourse over several years are ultimately very similar to those identified by the literature for recourse at a single point in time. Therefore, the barriers associated with cancer screening behavior already identified appeared to have a persistent effect over time, resulting in the prevention of some women from entering a screening process. Recalling the association of age, the level of education, and poorer health, with a low or even almost nonexistent level of use of screening, we can think of an accumulation of inequalities in use over time and, ultimately, a reinforcement of social inequalities in health.

Nevertheless, our study has some limitations. Several of these are related to the nature of the E3N cohort and the data collection methods. First, the women surveyed were not representative of the French population, as they had, on average, a higher socioeconomic level, a higher education level, were overrepresented in living areas with fairly good access to healthcare and, finally, had higher levels of cancer screening use. Even if there is no reason to believe that a representative sample of the population would lead to obtaining different typical screening patterns, it is nevertheless reasonable to assume that in our particular sample, for each type of cancer, the proportion of women who were screened very regularly was overestimated and the one of women who never or rarely underwent

screening was underestimated. By considering the determinants of regular vs. low use, our results suggested that this bias essentially leads to an underestimation of the negative effect of the determinants of very low use. For example, if being obese increases the probability of never having been screened for cancer despite a high socioeconomic level, it is likely that this effect is even stronger in the general population. Another bias may result from the constitution of our sample since we restricted our population to women who answered at least 3 questionnaires over the study period. The correlation between the regularity of responses to health surveys and compliance with screening recommendations over 10 years may have led to an overrepresentation of women who underwent very regular screening and an underrepresentation of women who never or rarely underwent screening. Robustness checks suggested that this did not modify the typical screening patterns identified. However, to limit this source of bias, we allowed the women in our study sample to not answer all the questionnaires between 2000 and 2011 even though this introduced an “uncertain” screening status the sequence analysis.

Two other sources of bias may be due to memory and social desirability. Memory bias, however, should be limited by the fact that this was a prospective study. In our context, social desirability bias would lead women to overestimate their use of screening. However, this bias should be balanced with the results of studies that have shown that reported Pap smear use levels were close to the levels observed in administrative data (Constantinou et al., 2016). Despite the richness of the data, some determinants of screening uptake known to play an important role, including risk aversion, risk perception, information aversion and time preferences, were not available in our study (Goldzahl, 2017). Finally, the static structure of the multinomial regressions carried out does not sufficiently allow us to study the factors associated with a change in behavior. This is especially the case for Pap smear and FOBT/colonoscopy use. Further analyses must be conducted to understand factors associated with an early exit or a late entry into the uptake of screening.

Beyond clusters, sequence analysis itself has several limitations. The main limitation of using OM is the assignment of indel and substitution costs, which necessarily contain arbitrary assumptions (Wu,

2000). This is limited in our case by the natural existing hierarchy between the statuses and the regular time intervals between survey waves, which limits the sensitivity of the OM to the attribution of editing costs (Lesnard, 2014). In addition, the results of the various robustness checks allow us to be confident in the relevance of our results. A strength of sequence analysis is that it does not require any hypothesis on the sequence generating process; however, the typology obtained cannot be validated and tested formally (Hennig and Liao, 2013). Nonetheless, the analysis of both the 20 most frequent sequences per cluster (Figures D.2, D.4, D.6 in Appendix D) and state transition matrices (Figure D.7 in Appendix D) seems to indicate that within the clusters, deviations from the typical pattern identified are small compared with the typical pattern, and misclassifications are scarce. Different robustness tests were conducted by varying the dissimilarity measure and the sample and suggested that the typical screening trajectories identified were relevant.

Conclusion

Despite its limitations, our database is extremely rich and allows us to study typical individual patterns of the use of 3 different cancer screenings over a period of almost 15 years. Our results show that there are few typical cancer screening behaviors in the long run and that they appear stable. Indeed, for each cancer screening, some women showed very regular use, others showed irregular use (mammography) or no use (Pap smear, FOBT/colonoscopy), and a rare few others showed a break in their screening behavior (Pap smear, FOBT/colonoscopy). Thus, habit seems to be an important determinant of recurrent cancer screening behavior. More generally, there appears to be an overall pattern of use, with a rather strong association between insufficient screening for breast cancer and never screening for other cancers. This pattern needs to be explored more deeply, as it could facilitate particularly effective targeting of women in a combined screening approach. As screening behavior appears stable over the long term, incentive policies should be more strongly targeted toward people outside the use of cancer screening, as these people are also likely to be less educated in poorer health and more

distant from the health care system. Finally, the factors associated with late-entry behavior in or early exit from cancer screening need to be explored in more detail to understand what factors are associated with women's entry into and exit from cancer screening.

References

- Abbott, A., Forrest, J., 1986. Optimal Matching Methods for Historical Sequences. *Journal of Interdisciplinary History*, 16(3), 471. <https://doi.org/10.2307/204500>
- Abbott, A., Hrycak, A., 1990. Measuring Resemblance in Sequence Data : An Optimal Matching Analysis of Musicians' Careers. *American Journal of Sociology*, 96(1), 144-185. <https://doi.org/10.1086/229495>
- Adami, H.O., Day, N.E., Trichopoulos, D., Willett, W.C., 2001. Primary and secondary prevention in the reduction of cancer morbidity and mortality. *Eur. J. Cancer* 37, 118–127. [https://doi.org/10.1016/s0959-8049\(01\)00262-3](https://doi.org/10.1016/s0959-8049(01)00262-3).
- Barlet, M., Coldefy, M., Collin, C., & Lucas-Gabrielli, V., 2012. L'accessibilité potentielle localisée (APL) : Une nouvelle mesure de l'accessibilité aux médecins généralistes libéraux. *Etudes et résultats*, 795. IRDES
- Bussière, C., Le Vaillant, M., Pelletier-Fleury, N., 2015. Screening for cervical cancer : What are the determinants among adults with disabilities living in institutions? Findings from a National Survey in France. *Health Policy*, 119(6), 794-801. <https://doi.org/10.1016/j.healthpol.2015.02.004>
- Cabral, M., Cullen, M. R., 2017. The Effect of Insurance Coverage on Preventive Care. *Economic Inquiry*, 55(3), 1452-1467. <https://doi.org/10.1111/ecin.12442>
- Carrieri, V., Bilger, M., 2013. Preventive care : Underused even when free. Is there something else at work? *Applied Economics*, 45(2), 239-253. <https://doi.org/10.1080/00036846.2011.597729>
- Chapple, A., Ziebland, S., Hewitson, P., & McPherson, A., 2008. What affects the uptake of screening for bowel cancer using a faecal occult blood test (FOBT) : A qualitative study. *Social Science & Medicine*, 66(12), 2425-2435. <https://doi.org/10.1016/j.socscimed.2008.02.009>
- Chevillard, G., Mousquès, J., 2018. Accessibilité aux soins et attractivité territoriale: proposition d'une typologie des territoires de vie français. *Cybergeo: Eur. J. Geogr.* <https://doi.org/10.4000/cybergeo.29737>.
- Chevillard, G., Mousques, J., 2021. Medically underserved areas : Are primary care teams efficient at attracting and retaining general practitioners? *Social Science & Medicine*, 114358. <https://doi.org/10.1016/j.socscimed.2021.114358>
- Clavel-Chapelon, F., 2015. Cohort profile: the French E3N cohort study. *Int. J. Epidemiol.* 44, 801–809. <https://doi.org/10.1093/ije/dyu184>.
- CNAM, 2020. Cancer colorectal dépistage par test immunologique et autres modalités. <https://www.ameli.fr/paris/medecin/exercice-liberal/memos/depistage-et-prevention/depistages-cancers-sein-colorectal> [Last visited: September 21, 2021]
- CNAM, 2021a. Dépistage gratuit du cancer du sein pour les femmes entre 50 et 74 ans. <https://www.ameli.fr/paris/assure/sante/themes/cancer-sein/depistage-gratuit-50-74-ans> [Last visited: September 21, 2021]
- CNAM, 2021b. Présentation de la cartographie des pathologies et des dépenses de l'Assurance Maladie. <https://assurance-maladie.ameli.fr/etudes-et-donnees/entree-par->

theme/pathologies/cartographie-assurance-maladie/presentation/cartographie-pathologies-depenses-assurance-maladie [Last visited: September 21, 2021]

- Constantinou, P., Dray-Spira, R., Menvielle, G., 2016. Cervical and breast cancer screening participation for women with chronic conditions in France: results from a national health survey. *BMC Cancer* 16, 255. <https://doi.org/10.1186/s12885-016-2295-0>.
- Cropper, M. L., 1977. Health, Investment in Health, and Occupational Choice. *Journal of Political Economy*, 85(6), 1273-1294. <https://doi.org/10.1086/260637>
- Damiani, G., Federico, B., Basso, D., Ronconi, A., Bianchi, C. B. N. A., Anzellotti, G. M., Nasi, G., Sassi, F., & Ricciardi, W., 2012. Socioeconomic disparities in the uptake of breast and cervical cancer screening in Italy : A cross sectional study. *BMC Public Health*, 12(1), 99. <https://doi.org/10.1186/1471-2458-12-99>
- Devaux, M., 2015. Income-related inequalities and inequities in health care services utilisation in 18 selected OECD countries. *Eur. J. Health Econ.* 16, 21–33. <https://doi.org/10.1007/s10198-013-0546-4>.
- Duffy, S., Paci, E., 2012. Bénéfices et risques du dépistage du cancer du sein par mammographie. *Bull. Épidémiol. Hebd.* 35-36-37, 407–410. https://doi.org/10.1007/978-2-8178-0245-9_15.
- Eddy, D. M., 1980. *Screening for cancer: Theory, analysis, and design*. Prentice Hall.
- Ehrlich, I., Becker, G.S., 1972. Market insurance, self-insurance, and self-protection. *J. Political Econ.* 80, 623–648. <https://doi.org/10.1086/259916>.
- Elzinga, C.H., 2007. *Sequence analysis: Metric representations of categorical time series*. Manuscript, Dept of Social Science Research Methods, Vrije Universiteit, Amsterdam.
- Gabardinho, A., Ritschard, G., Müller, N.S., Studer, M., 2011. Analyzing and visualizing state sequences in R with TraMineR. *J. Stat. Softw.* 40, 137. <https://doi.org/10.18637/jss.v040.i04>.
- Genier, P., Jacobzone, S., 1998. Peut-on parler d'une gestion globale de son capital santé?. *Economie et statistique*, 311(1), 37-55.
- Goldzahl, L., 2017. Contributions of risk preference, time orientation and perceptions to breast cancer screening regularity. *Social Science & Medicine* 185, 147–157. <https://doi.org/10.1016/j.socscimed.2017.04.037>.
- Goldzahl, L., Jusot, F., 2016. Les déterminants du recours régulier au dépistage du cancer du sein en France. *Rev. Fr. Écon.* XXXI, 109. <https://doi.org/10.3917/rfe.164.0109>.
- Grossman, M., 1972. On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy*, 80(2), 223-255. <https://doi.org/10.1086/259880>
- Halpin, B., 2010. Optimal Matching Analysis and Life-Course Data : The Importance of Duration. *Sociological Methods & Research*, 38(3), 365-388. <https://doi.org/10.1177/0049124110363590>
- Halpin, B., 2016. Missingness and truncation in sequence data: a non-self-identical missing state, in: Ritschard, G., Studer, M. (Eds.), *International Conference on Sequence Analysis and Related Methods (LaCOSA II)*. Lausanne, Switzerland, pp. 443–445.

- Hennig, C., Liao, T.F., 2013. How to find an appropriate clustering for mixed-type variables with application to socio-economic stratification. *J. R. Stat. Soc.: (Stat. Soc.)* 62, 309–369. <https://doi.org/10.1111/j.1467-9876.2012.01066.x>.
- Hollister, M., 2009. Is optimal matching suboptimal? *Sociol. Methods Res.* 38, 235–264. <https://doi.org/10.1177/0049124109346164>.
- INCa, 2018a. Les cancers en France en 2018—L’essentiel des Faits et Chiffres (édition 2019) <https://www.e-cancer.fr/Expertises-et-publications/Catalogue-des-publications/Les-cancers-en-France-en-2018-L-essentiel-des-faits-et-chiffres-edition-2019> [Last visited: September 21, 2021]
- INCa, 2018b. Dépistage du cancer du sein : impact sur les trajectoires de soin. Fiche d’analyse. Collection Les Données / Soins. <https://www.e-cancer.fr/Expertises-et-publications/Catalogue-des-publications/Depistage-du-cancer-du-sein-impact-sur-les-trajectoires-de-soins> [Last visited: September 21, 2021]
- INCa, 2021a. Dépistage du cancer du col de l’utérus. <https://www.e-cancer.fr/Professionnels-de-sante/Depistage-et-detection-precoce/Depistage-du-cancer-du-col-de-l-uterus> [Last visited: September 21, 2021]
- INCa, 2021b. Dépistage du cancer colorectal - Dès 50 ans, un test à faire chez soi tous les 2 ans. Guide pratique. <https://www.e-cancer.fr/Comprendre-prevenir-depister/Se-faire-depister/Depistage-du-cancer-colorectal/Pourquoi-se-faire-depister> [Last visited: September 21, 2021]
- INED, 2019. Causes de décès. <https://www.ined.fr/fr/tout-savoir-population/chiffres/france/mortalite-cause-deces/causes-deces/> [Last visited: September 21, 2021]
- Jones, R.M., Woolf, S.H., Cunningham, T.D., Johnson, R.E., Krist, A.H., Rothemich, S.F., Vernon, S.W., 2010. The relative importance of patient-reported barriers to colorectal cancer screening. *Am. J. Prev. Med.* 38, 499–507. <https://doi.org/10.1016/j.amepre.2010.01.020>.
- Kash, K. M., Holland, J. C., Halper, M. S., Miller, D. G., 1992. Psychological Distress and Surveillance Behaviors of Women With a Family History of Breast Cancer. *JNCI: Journal of the National Cancer Institute*, 84(1), 24-30. <https://doi.org/10.1093/jnci/84.1.24>
- Kenkel, D. S., 1994. The demand for preventive medical care. *Applied Economics*, 26(4), 313-325. <https://doi.org/10.1080/00036849400000078>
- Klug, S. J., Hetzer, M., Blettner, M., 2005. Screening for breast and cervical cancer in a large German city : Participation, motivation and knowledge of risk factors. *European Journal of Public Health*, 15(1), 70-77. <https://doi.org/10.1093/eurpub/cki118>
- Kobayashi, L. C., Wardle, J., von Wagner, C., 2014. Limited health literacy is a barrier to colorectal cancer screening in England : Evidence from the English Longitudinal Study of Ageing. *Preventive Medicine*, 61, 100-105. <https://doi.org/10.1016/j.ypmed.2013.11.012>
- Konopka, A.-M., Barnay, T., Billaudeau, V., Sevilla-Dedieu, C., 2019. Les déterminants du recours au dépistage du cancer du col de l’utérus : Une analyse départementale. *Économie & prévision*, n°216(2), 43. <https://doi.org/10.3917/ecop.216.0043>

- Lesnard, L., 2014. Using optimal matching analysis in sociology: cost setting and sociology of time, in: Blanchard, P., Bühlmann, F., Gauthier, J.A. (Eds.), *Advances in Sequence Analysis: Theory, Method, Applications*. Springer International Publishing, Cham, pp. 39–50.
- Levenshtein, V.I., 1966. Binary codes capable of correcting deletions, insertions, and reversals. *Sov. Phys. Dokl.* 10, 707–710.
- Lo, S. H., Waller, J., Wardle, J., & von Wagner, C., 2013. Comparing barriers to colorectal cancer screening with barriers to breast and cervical screening: A population-based survey of screening-age women in Great Britain. *Journal of Medical Screening*, 20(2), 73-79. <https://doi.org/10.1177/0969141313492508>
- Longone, P., 1973. Formation et qualification professionnelle. *Inst. Natl. Études Démogr.* 75. <https://doi.org/10.3406/estat.1973.1299>.
- Marmot, M.G., Altman, D.G., Cameron, D.A., Dewar, J.A., Thompson, S.G., Wilcox, M., 2013. The benefits and harms of breast cancer screening: an independent review. *Br. J. Cancer* 108, 2205–2240. <https://doi.org/10.1038/bjc.2013.177>.
- Minni, C., 2006. Structure et évolution de la population active selon l'âge. Collection : Données sociales : La société française. INSEE.
- Montano, D., Kasprzyk, D., 2015. *Health Behavior and Health Education : Theory, Research, and Practice*. 67-92.
- Myers, E. R., Moorman, P., Gierisch, J. M., Havrilesky, L. J., Grimm, L. J., Ghatge, S., Davidson, B., Montgomery, R. C., Crowley, M. J., McCrory, D. C., Kendrick, A., Sanders, G. D., 2015. Benefits and Harms of Breast Cancer Screening: A Systematic Review. *JAMA*, 314(15), 1615-1634. <https://doi.org/10.1001/jama.2015.13183>
- Newhouse, J. P., 2021. An Ounce of Prevention. *Journal of Economic Perspectives*, 35(2), 101-118. <https://doi.org/10.1257/jep.35.2.101>
- Nordin, M., Hiebert, R., Pietrek, M., Alexander, M., Crane, M., & Lewis, S., 2002. Association of comorbidity and outcome in episodes of nonspecific low back pain in occupational populations. *Journal of Occupational and Environmental Medicine*, 44(7), 677-684. <https://doi.org/10.1097/00043764-200207000-00015>
- Santé Publique France, 2020. Taux de participation au programme de dépistage organisé du cancer du sein 2018-2019 et évolution depuis 2005. <https://www.santepubliquefrance.fr/maladies-et-traumatismes/cancers/cancer-du-sein/articles/taux-de-participation-au-programme-de-depistage-organise-du-cancer-du-sein-2018-2019-et-evolution-depuis-2005> [Last visited: September 21, 2021]
- Sicsic, J., Franc, C., 2014. Obstacles to the uptake of breast, cervical, and colorectal cancer screenings: what remains to be achieved by French national programmes? *BMC Health Serv. Res.* 14, 112. <https://doi.org/10.1186/1472-6963-14-465>.
- Seibert, R. G., Hanchate, A. D., Berz, J. P., & Schroy, P. C., 2017. National Disparities in Colorectal Cancer Screening Among Obese Adults. *American Journal of Preventive Medicine*, 53(2), e41-e49. <https://doi.org/10.1016/j.amepre.2017.01.006>

- Shaukat, A., Mongin, S. J., Geisser, M. S., Lederle, F. A., Bond, J. H., Mandel, J. S., Church, T. R., 2013. Long-Term Mortality after Screening for Colorectal Cancer. *New England Journal of Medicine*, 369(12), 1106-1114. <https://doi.org/10.1056/NEJMoa1300720>
- Studer, M., 2013. WeightedCluster library manual: a practical guide to creating typologies of trajectories in the social sciences with R. *LIVES Work*. Pap. 1–34. <https://doi.org/10.12682/lives.2296-1658.2013.24>.
- Studer, M., Ritschard, G., 2016. What matters in differences between life trajectories: a comparative review of sequence dissimilarity measures. *J. R. Stat. Soc.: (Stat. Soc.)* 179, 481–511. <https://doi.org/10.1111/rssa.12125>.
- Walsh, B., Silles, M., O’Neill, C., 2011. The importance of socio-economic variables in cancer screening participation : A comparison between population-based and opportunistic screening in the EU-15. *Health Policy*, 101(3), 269-276. <https://doi.org/10.1016/j.healthpol.2011.02.001>
- Wu, L.L., 2000. Some comments on “sequence analysis and optimal matching methods in sociology: review and prospect”. *Sociol. Methods Res.* 29, 41–64. <https://doi.org/10.1177/0049124100029001003>.
- Wu, S., 2003. Sickness and preventive medical behavior. *J. Health Econ.* 22, 675–689. [https://doi.org/10.1016/s0167-6296\(03\)00042-0](https://doi.org/10.1016/s0167-6296(03)00042-0).

Tables

Table 1. Descriptive statistics of the study population

		N (40,021)	%
Socio-economic characteristics			
Age in 2000	[50-54]	21,147	52.8%
	[55-60]	18,874	47.2%
Children	Yes	35,668	89.1%
Most frequent conjugal status between 2002 and 2008	Single	2,255	5.6%
	Relationship	29,982	74.9%
	Separated or widow	7,784	19.4%
Professional activity in 2000	Active	22,025	55.0%
	Retired	15,881	45.0%
Education	< High-school diploma	3,651	9.1%
	< Bachelor's degree	19,145	47.8%
	< Master's degree	8,827	22.1%
	≥ Master's degree	7,05	17.6%
	Non-response	1,348	3.4%
Access and use of care			
Territory of life in 2008: attractiveness and access to care ^a	Privileged cities and suburban areas	8,776	21.9%
	City centers with heterogenous socioeconomic status	11,954	29.9%
	Tourist and retirement areas	3,703	9.3%
	Suburban areas	6,715	16.8%
	Deprived urban or rural areas	3,912	9.8%
Annual median use of ambulatory physicians between 2004 and 2014 by most frequent quintile of use	Rural fringes	4,961	12.4%
	Q1	3 physicians per year	
	Q2	5 physicians per year	
	Q3	7 physicians per year	
	Q4	9 physicians per year	
Q5	13 physicians per year		
Health and behaviors			
Most frequent BMI's category between 2000 and 2011	Normal	26,628	66.5%
	Underweight	1,193	3.0%
	Overweight	9,317	23.3%
	Obese	2,883	7.2%
Self-perceived health in 2011	Very good	9,557	23.9%
	Good	23,178	57.9%
	Average or poor	5,194	13.0%
	Non-response	2,092	5.2%
Acute or chronic disease between 2004 and 2014 ^b	All	28,702	71.7%
	Excluding psychotropic treatments and vascular risk treatments	10,773	26.9%
Family cancer history in 2000	Yes	18,171	45.4%
Smoking status in 2005	Never smoker	24,996	62.5%
	Former smoker	11,581	28.9%
	Smoker	3,444	8.6%
Attrition and number of responses			
Number of questionnaires answered	3 (minimum)	2,369	5.9%
	4	3,426	8.6%
	5	6,155	15.4%
	6 (maximum)	28,071	70.1%
Answer to the last questionnaire (2014)	Yes	34,891	87.2%
Alive in 2014	Yes	39,590	99%

^a Based on Chevillard and Mousquès (2018, 2021)

^b Based on the Mapping of pathologies and expenses of the French National Health Insurance (CNAM, 2021b): cardioneurovascular disease, diabete, vascular risk treatment, chronic respiratory diseases, psychotropic treatments, liver or pancreas diseases, HIV/AIDS, degenerative neurological diseases

Table 2. Determinants of the probability of belonging to a cluster other than the cluster of women with the lowest level of use for each cancer screening type

	Mammography clusters:	Pap smear clusters:		FOBT & colonoscopy clusters:		
	"Regular" (ref: "Irregular")	"Regular to nonuse" (ref: "Nonuse")	"Regular" (ref: "Nonuse")	"Regular before 2009" (ref: "Nonuse")	"Regular after 2009" (ref: "Nonuse")	"Regular" (ref: "Nonuse")
(Intercept)	10.76***(8.85,13.1)	1.03 (0.79,1.33)	7.59***(6.24,9.25)	1.27** (1.08,1.49)	0.85 (0.71,1.02)	0.59***(0.48,0.72)
Socio-demographic characteristics						
Age in 2000 (normalized)						
Age	0.88*** (0.84,0.91)	0.83***(0.79,0.87)	0.81***(0.78,0.84)	0.95***(0.92,0.98)	0.88***(0.85,0.91)	1.09***(1.04,1.13)
Age ²	0.97 (0.94,1.01)	1.02 (0.97,1.06)	1.01 (0.98,1.05)	0.97* (0.94,0.99)	1.01 (0.98,1.04)	0.97 (0.94,1.01)
Children (ref: No)						
Yes	1.1 (0.98,1.24)	1.1 (0.95,1.28)	1.42*** (1.27,1.6)	1.01 (0.92,1.12)	0.98 (0.87,1.09)	1.04 (0.92,1.18)
Most frequent conjugal status between 2002 and 2008 (ref: Relationship)						
Single	0.73*** (0.63,0.85)	0.71***(0.58,0.85)	0.53***(0.46,0.61)	0.93 (0.82,1.07)	0.86 (0.74,1.1)	0.97 (0.83,1.15)
Separated or widow	0.69***(0.64,0.75)	0.94 (0.85,1.04)	0.76*** (0.7,0.82)	0.86*** (0.81,0.92)	0.88*** (0.82,0.94)	0.83*** (0.76,0.9)
Professional activity in 2000 (ref: Retired)						
Active	1.06 (0.98,1.15)	0.99 (0.9,1.11)	1.08 (0.99,1.17)	0.96 (0.89,1.02)	0.95 (0.88,1.02)	0.93 (0.85,1.01)
Education (ref: < High-school diploma)						
< Bachelor's degree	1.36*** (1.23,1.51)	1.02 (0.89,1.17)	1.18* (1.06,1.31)	1.1* (1.01,1.21)	1.13* (1.02,1.26)	0.98 (0.88,1.09)
< Master's degree	1.34*** (1.19,1.5)	1.15 (0.98,1.34)	1.39*** (1.23,1.57)	1.14* (1.03,1.26)	1.13* (1.01,1.27)	0.89 (0.79,1.01)
≥ Master's degree	1.44*** (1.27,1.63)	1.02 (0.86,1.21)	1.34*** (1.18,1.52)	1.09 (0.98,1.21)	1.15* (1.02,1.3)	0.99 (0.87,1.12)
Non-response	1.06 (0.89,1.27)	1.15 (0.95,1.31)	1.14 (0.94,1.38)	0.88 (0.75,1.03)	0.9 (0.75,1.08)	0.83 (0.68,1)
Access and use of care						
Living areas in 2008: attractiveness and access to care ^a (ref: Privileged cities or suburban areas)						
City centers with heterogenous socioeconomic status	0.88** (0.81,0.96)	0.88 (0.77,1.01)	0.87** (0.79,0.95)	1.1** (1.03,1.18)	1.04 (0.96,1.13)	1 (0.92,1.09)
Tourist and retirement areas	0.82** (0.73,0.93)	1.11 (0.94,1.3)	0.86* (0.76,0.97)	0.85** (0.77,0.94)	1.08 (0.97,1.21)	0.78*** (0.69,0.88)
Suburban areas	0.8*** (0.72,0.89)	0.88 (0.77,1.01)	0.74*** (0.67,0.82)	1.1* (1.02,1.2)	1.07 (0.97,1.17)	1.07 (0.97,1.18)
Deprived urban or rural areas	0.86* (0.76,0.98)	0.76*** (0.65,0.88)	0.65*** (0.57,0.73)	1.17** (1.06,1.3)	1.08 (0.97,1.21)	0.93 (0.82,1.05)
Rural fringes	0.81*** (0.73,0.91)	0.91 (0.79,1.05)	0.71*** (0.64,0.8)	0.96 (0.88,1.06)	1.06 (0.95,1.17)	0.97 (0.87,1.08)
Most frequent quintile of annual use of ambulatory physicians between 2004 and 2014 (ref: Q2/Q3/Q4)						
Q1	0.35*** (0.33,0.38)	0.9** (0.81,0.99)	0.39*** (0.36,0.42)	0.44*** (0.41,0.47)	0.62*** (0.57,0.66)	0.41*** (0.38,0.45)
Q5	1.18*** (1.09,1.29)	1.07 (0.95,1.2)	1.41*** (1.29,1.54)	1.32*** (1.24,1.42)	1.1** (1.01,1.19)	1.65*** (1.53,1.79)
Health and behaviors						
Most frequent BMI's category between 2000 and 2011 (ref: Normal)						
Underweight	0.71*** (0.6,0.84)	0.75* (0.59,0.96)	0.79** (0.66,0.94)	0.91 (0.78,1.06)	0.97 (0.82,1.15)	1.02 (0.85,1.22)
Overweight	0.75*** (0.7,0.81)	1 (0.91,1.11)	0.71*** (0.66,0.76)	0.84*** (0.79,0.89)	0.88*** (0.82,0.95)	0.84*** (0.78,0.9)
Obese	0.5*** (0.46,0.56)	0.86* (0.74,0.99)	0.47*** (0.42,0.53)	0.6*** (0.54,0.66)	0.76*** (0.68,0.85)	0.53*** (0.46,0.6)
Self-perceived health in 2011 (ref: Very good)						
Good	0.89** (0.82,0.97)	0.88** (0.79,0.97)	0.81*** (0.74,0.87)	1.11*** (1.05,1.19)	1.04 (0.97,1.11)	1.22*** (1.12,1.32)
Average or poor	0.56*** (0.5,0.62)	0.86* (0.74,0.99)	0.58*** (0.52,0.65)	0.95 (0.86,1.04)	0.85** (0.77,0.95)	1.16* (1.03,1.29)
Non-response	0.37*** (0.32,0.42)	0.52*** (0.42,0.63)	0.64*** (0.55,0.73)	0.41*** (0.36,0.46)	0.17*** (0.14,0.21)	0.67*** (0.58,0.77)
Acute or chronic condition between 2004 and 2014 ^b (ref: No)						
Yes	1.02 (0.95,1.11)	1.01 (0.91,1.11)	0.83*** (0.77,0.89)	0.96 (0.9,1.02)	0.97 (0.9,1.03)	0.9** (0.83,0.97)
Family cancer history in 2000 (ref: No)						
Yes	1.18*** (1.12,1.26)	1.04 (0.96,1.12)	1.1** (1.04,1.18)	1.22*** (1.16,1.29)	1.05 (1,1.12)	1.55*** (1.46,1.65)
Smoking status in 2005 (ref: Never smoker)						
Former smoker	0.99 (0.92,1.06)	1.09 (0.99,1.19)	1.1** (1.04,1.18)	0.95 (0.9,1.01)	0.99 (0.93,1.05)	0.95 (0.88,1.02)
Smoker	0.7*** (0.63,0.78)	0.97 (0.85,1.12)	0.85** (0.76,0.94)	0.68*** (0.62,0.74)	0.69*** (0.62,0.76)	0.68*** (0.61,0.76)
AIC	28,018	53,616		105,229		
N cluster	35,125	4,561	27,580	8,479	13,221	6,387
N total	40,015	37,171		39,961		

Exponentiated coefficients; 95% confidence intervals in brackets

* p<0.05; ** p<0.01; *** p<0.001

^a Based on Chevillard and Mousquès (2018, 2021)

^b Based on the Mapping of pathologies and expenses of the French National Health Insurance (CNAM, 2021b): cardioneurovascular disease, diabete, vascular risk treatment, chronic respiratory diseases, psychotropic treatments, liver or pancreas diseases, HIV/AIDS, degenerative neurological diseases

Note: logistic regression for the probability of belonging to the most regular use cluster for mammography; multinomial regressions for the probability of belonging to each use cluster vs. the lowest use cluster for Pap smear and FOBT/colonoscopy.

Descriptions of FOBT/colonoscopy clusters:

- "Regular before 2009" -> "Nonuse to regular use before 2009"
- "Regular after 2009" -> "Nonuse to regular use after 2009"
- "Regular" -> regular use throughout the study period

Figures

Figure 1 Timeline of E3N questionnaires, MGEN data, prevention use variables and determinants used

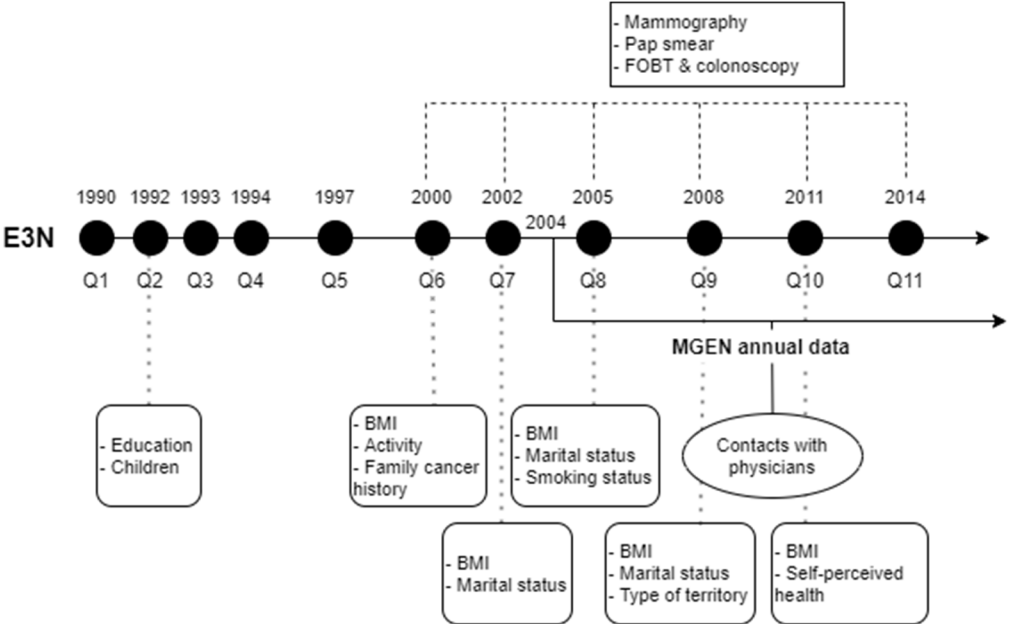
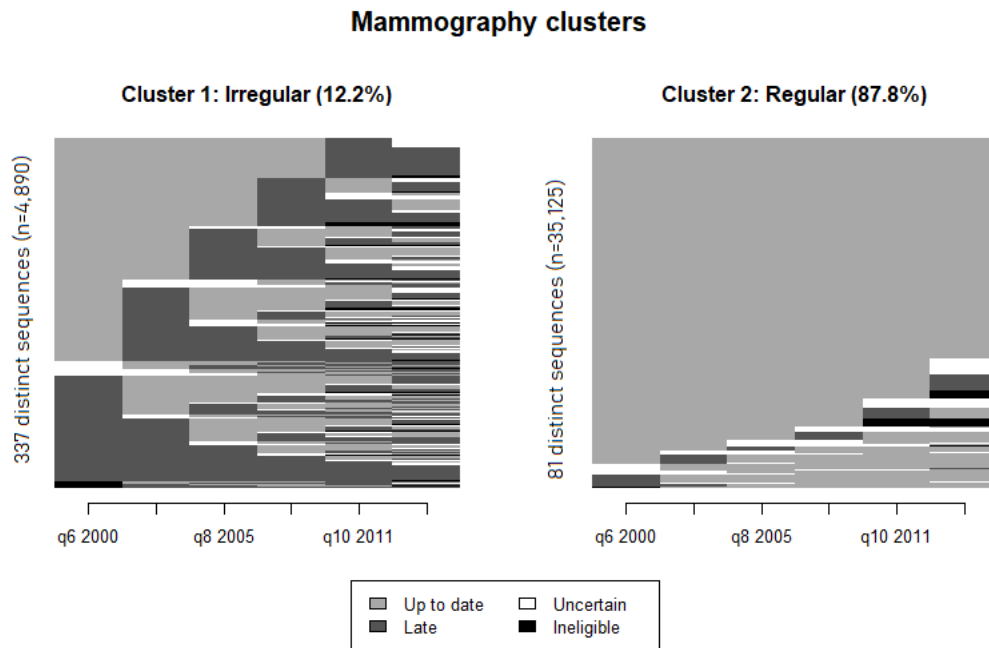
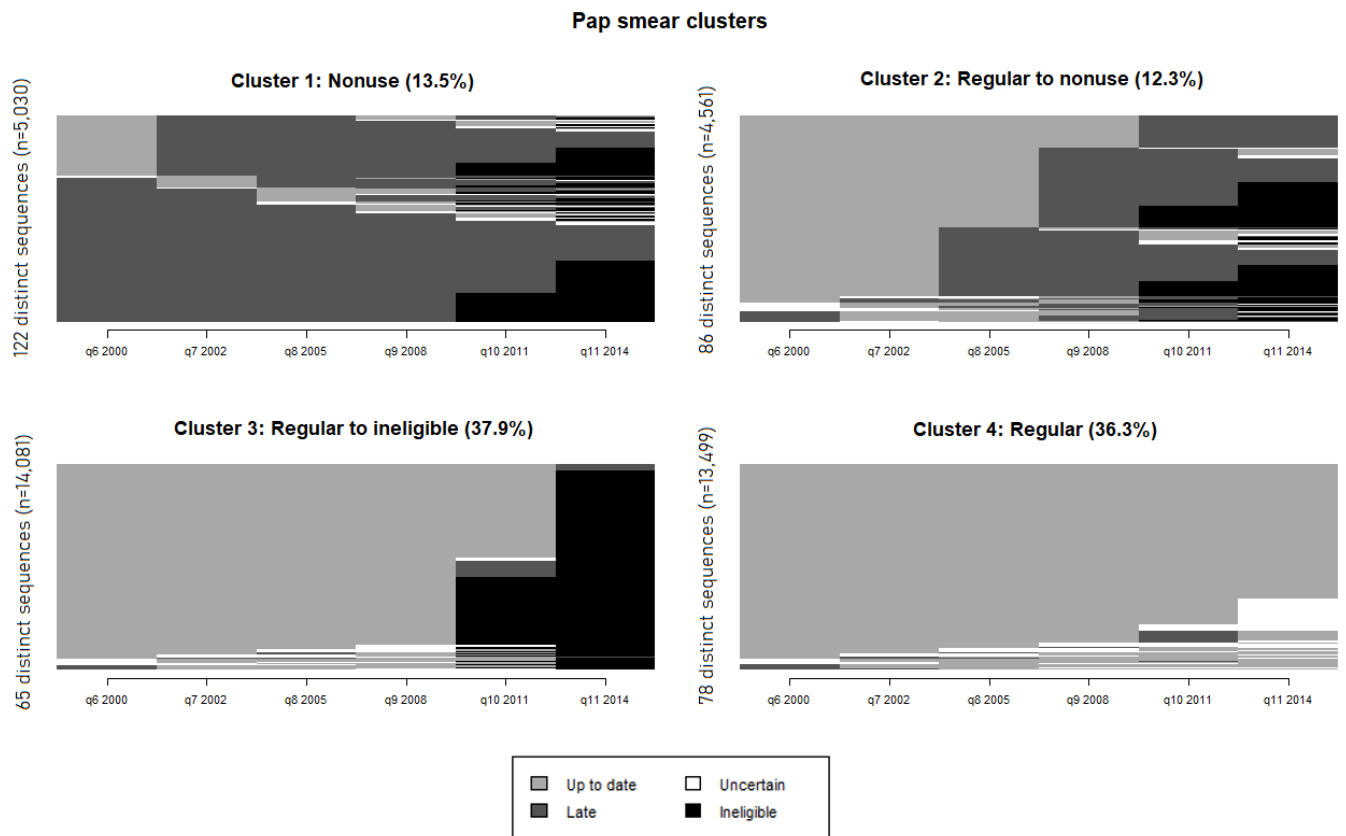


Figure 2 Plot of the individual sequences of mammography use by cluster (% of women)



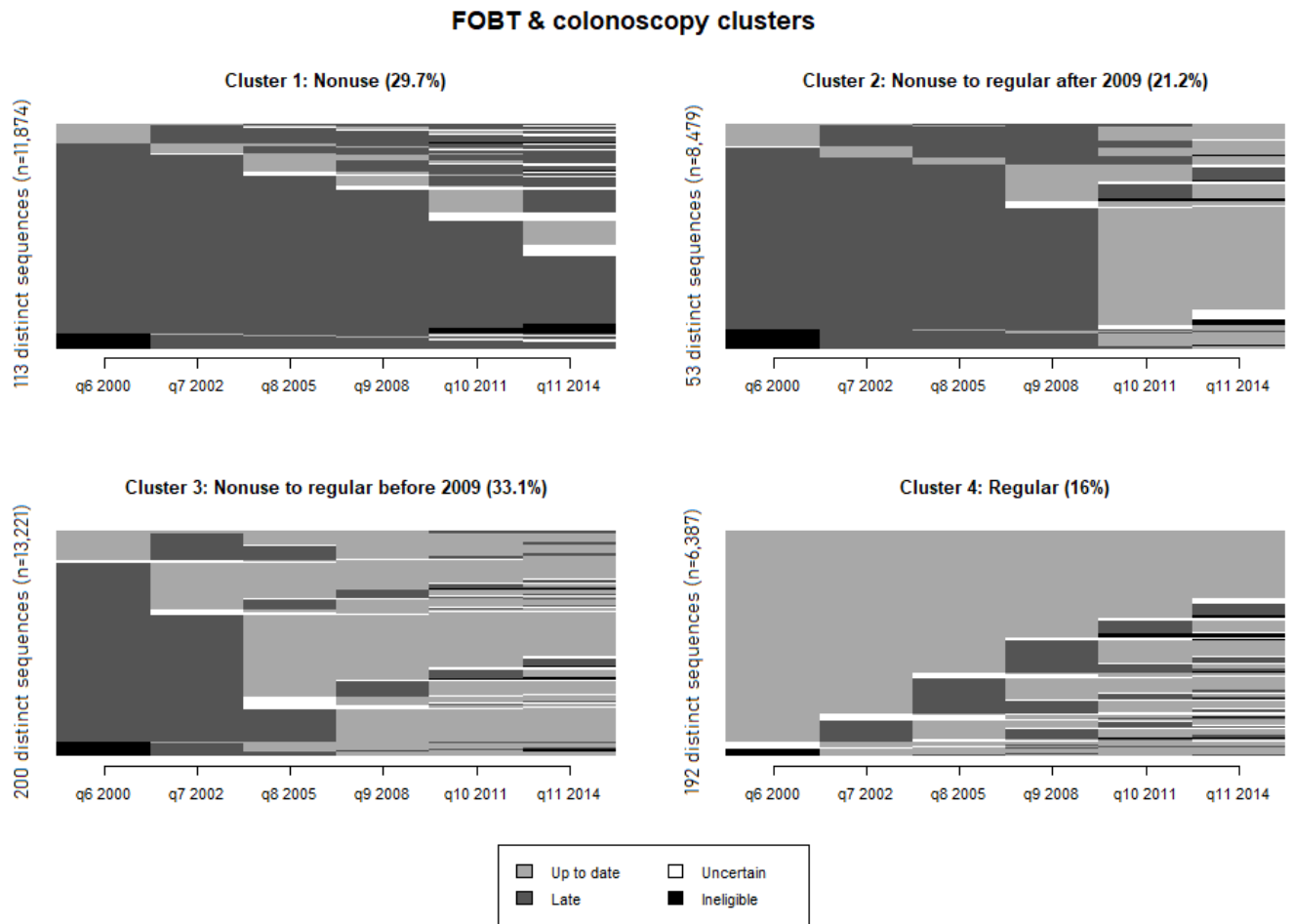
Reading note: The “Irregular” cluster contains 12.2% of the population ($N = 40,015$), summarized by 337 distinct sequences representing the 4,890 women in the cluster. For each cluster, the graph represents the stacking of the individual sequences within each distinct sequence.

Figure 3. Plot of the individual sequences of Pap smear use by cluster (% of women)



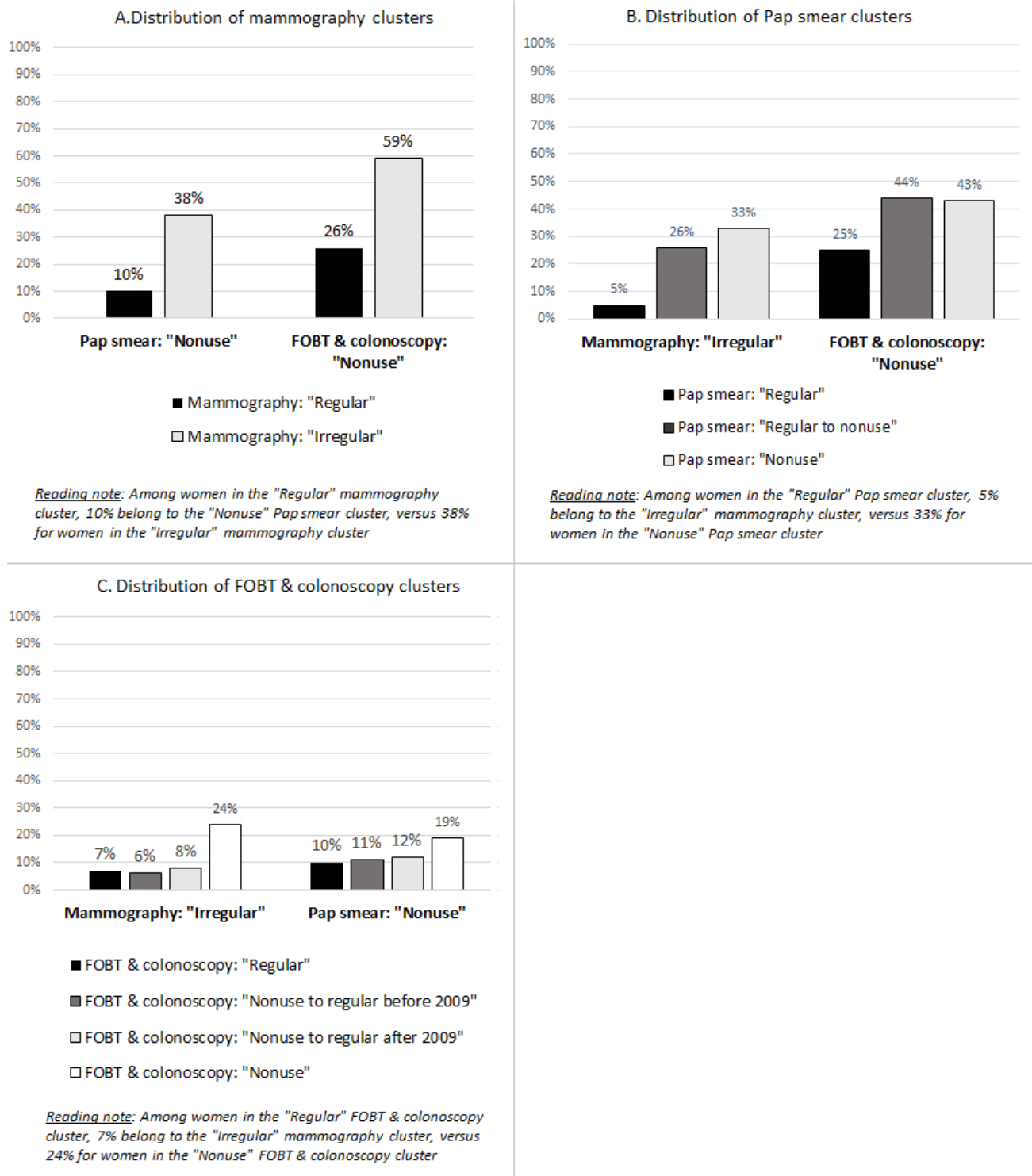
Reading note: The “Nonuse” cluster contains 13.5% of the population ($N = 37,171$), summarized by 122 distinct sequences representing the 5,030 women in the cluster. For each cluster, the graph represents the stacking of the individual sequences within each distinct sequence.

Figure 4. Plot of the individual sequences of FOBT/colonoscopy use by cluster (% of women)



Reading note: The “Nonuse” cluster contains 29.7% of the population ($N = 39,961$), summarized by 113 distinct sequences representing the 11,874 women in the cluster. For each cluster, the graph represents the stacking of the individual sequences within each distinct sequence.

Figure 5. Distribution of clusters from each cancer screening type among the lowest use clusters from the other two cancer screening types



Note: The charts are based on the sample of 37,114 women present in the trajectory analysis for the 3 cancer screenings