

MISE EN ŒUVRE DE LA MICROSIMULATION : L'EXEMPLE DU MODELE DE PROJECTIONS D'EFFECTIFS DE MEDECINS

Ketty ATTAL-TOUBERT (), Mélanie VANDERSCHULDEN (*)*

() Drees*

Introduction

En 1999, la Drees a mis au point, en collaboration avec l'Ined, un modèle de projection d'effectifs de médecins utilisant à la fois une méthode de projection classique, basée sur le calcul d'agrégats, et la microsimulation. En 2007, une opération de refonte de ce modèle a été entreprise, en vue de réaliser un nouvel exercice de projections en 2008. La refonte du modèle avait pour objectifs d'actualiser les données utilisées comme entrées par le modèle, en tenant compte du fait que certaines sources n'étaient plus disponibles, de prendre en compte au mieux les évolutions qui ont touché la profession au cours des dernières années (notamment les changements ayant modifié le déroulement des études médicales), d'améliorer la précision du modèle et de le rendre plus lisible.

Cette opération a été l'occasion de remettre en question le choix des méthodes employées pour la construction de ce modèle, et en particulier l'utilisation de la microsimulation. Après avoir comparé les avantages et inconvénients des deux méthodes, nous avons choisi, pour mettre au point la nouvelle version du modèle, de modéliser de façon agrégée le déroulement des deux premiers cycles des études médicales, et d'utiliser la microsimulation pour modéliser la suite de la formation des étudiants en médecine ainsi que tout le déroulement de la carrière des médecins, de l'entrée dans la vie active à la sortie définitive de la vie active.

Au cours de l'élaboration de cette nouvelle version du modèle, de nombreuses questions se sont posées, telles que le choix des événements à modéliser ou celui des hypothèses à retenir. Nous avons également rencontré un certain nombre de difficultés techniques, pour modéliser des événements rares ou gérer les effets de l'aléa par exemple.

Dans cet article, nous exposerons rapidement les objectifs du modèle de projection d'effectifs de médecins, en quoi consiste la microsimulation, comment sont réalisées en pratique les simulations, ce qui a motivé le choix de cette méthode pour la construction de la nouvelle version du modèle, dont nous donnerons un aperçu. Nous présenterons ensuite les questions qui se sont posées ainsi que les difficultés techniques qui se sont présentées, et qui sont susceptibles d'apparaître dans le cadre de la mise en œuvre d'un modèle de microsimulation. Nous détaillerons les diverses pistes explorées, les réponses apportées aux questions et les solutions techniques finalement retenues.

Les simulations dont les résultats sont présentés dans cet article ont été réalisées à partir de versions de test du modèle, et pour certaines, à partir de données fictives. Leurs résultats ne sont donc pas les résultats publiés des projections.

1. Le choix de la méthode de projection

1.1. Les objectifs du modèle

Le choix de la méthode de projection a nécessité, dans un premier temps, de définir précisément les objectifs du modèle.

Le modèle doit produire des effectifs projetés de médecins en activité pour chacune des années de la période de projection, allant de 2007 à 2030 (pour l'exercice de 2008). Il doit également produire, pour chaque année de la période de projection, la répartition de ces effectifs selon le sexe, l'âge, la spécialité (en 23 modalités), la région d'exercice, le mode d'exercice¹ et la zone d'exercice². Il s'agit

¹ Le mode d'exercice peut être : libéral, salarié hospitalier, salarié non hospitalier ou mixte.

d'obtenir la ventilation des effectifs projetés, de l'ordre de 200 000, selon chacune de ces caractéristiques, mais également selon plusieurs d'entre elles, croisées, autant que possible, c'est-à-dire tant que la taille des effectifs obtenus et la précision du modèle le permettent.

Le modèle doit également produire des résultats permettant de comprendre les évolutions d'effectifs constatées. Ainsi, il doit donner des informations sur les entrées de nouveaux médecins dans la population au cours de la période de projection, sur les sorties, etc.

Le déroulement de la carrière des médecins est assez dépendant de leurs caractéristiques. Le départ à la retraite, par exemple, ne se produit pas au même âge pour les généralistes et pour les spécialistes. Il s'agit donc également de prendre en compte la diversité des comportements des médecins pour produire les effectifs projetés.

Les projections ne visent pas à prévoir les effectifs futurs de médecins, mais à déterminer les évolutions tendanciennes du nombre et de la répartition des médecins, sous certaines hypothèses. Le modèle doit permettre de réaliser les simulations correspondant à divers scénarios, chacun des autres différant du scénario de référence par une ou plusieurs hypothèses. Il s'agit en fait de faire apparaître les effets potentiels de la mise en place par les pouvoirs publics de mesures de régulation ou de changements dans les comportements des médecins.

1.2. Microsimulation versus modélisation agrégée

La méthode de projection « classique » par modélisation agrégée consiste à modéliser l'évolution d'une population, c'est-à-dire à calculer l'effectif global de cette population, ou les effectifs correspondant aux sous-populations qui la composent, à partir d'agrégats. Par exemple, dans le cas d'un modèle de projection d'effectifs de médecins simplifié à l'extrême, on déduit l'effectif global de médecins actifs au cours d'une année n de l'effectif de médecins actifs au cours de l'année $n-1$, auquel on ôte le nombre de sorties de la vie active (les départs à la retraite et les décès) et on ajoute le nombre d'entrées dans la vie active survenues au cours de l'année écoulée. Le nombre d'entrées dans la vie active est estimé à partir des flux de diplômés, et le nombre de sorties en appliquant des taux estimés de décès et de départ à la retraite à la population médicale de l'année $n-1$. Il est possible de calculer de même les effectifs correspondant à des sous-populations (les généralistes exerçant en Alsace, les spécialistes femmes exerçant en mode libéral, par exemple), en déclinant ces calculs pour chaque sous-population. Par itérations successives, on peut ainsi calculer les effectifs de la population totale et des différentes sous-populations correspondant à chaque année³ de la période de projection.

La microsimulation consiste à simuler le devenir d'une population, c'est-à-dire l'évolution de sa taille mais aussi de sa composition, à un niveau individuel, à partir d'un fichier de données observées (ou éventuellement générées) contenant des informations sur les caractéristiques de chacun des individus composant cette population.⁴ Pour ce faire, on modélise les événements qui surviennent au cours de la période de vie des individus étudiée, et on recourt à des tirages aléatoires réalisés successivement, pour « faire vivre » aux individus ces événements, c'est-à-dire pour simuler la réalisation de ces événements au niveau individuel. Chaque tirage est effectué en fonction de la probabilité prédéfinie de chaque individu de vivre l'événement. La réalisation de chaque événement modifiant une ou plusieurs caractéristiques des individus, on obtient un fichier individuel actualisé, après avoir simulé également le vieillissement de la population, ainsi que son renouvellement, c'est-à-dire les entrées de nouveaux individus dans la population (les sorties correspondent à un événement : décès ou départ à la retraite par exemple, si l'on étudie une population d'actifs). Ces calculs sont réalisés de manière itérative pour chaque année de la période de projection.

Le schéma suivant montre comment on fait évoluer le fichier de données individuelles d'une année sur l'autre, avec l'exemple du départ à la retraite (le mode de tirage sera précisé plus loin).

² La variable « zone d'exercice » peut prendre les modalités suivantes : pôle urbain avec CHU, pôle urbain sans CHU, couronne périurbaine ou commune multipolarisée, espace à dominante rurale.

³ L'intervalle de temps retenu est l'année dans le cas du modèle de projection d'effectifs de médecins, mais cet intervalle peut être différent pour d'autres modèles (mois par exemple), quelle que soit la méthode (modélisation agrégée ou microsimulation).

⁴ On peut également travailler sur un échantillon.

N° médecin	Sexe	Généraliste/ Spécialiste	Probabilité estimée	Aléa	Résultat
1	Homme	Généraliste	0,11	0,09	O
2	Homme	Généraliste	0,11	0,45	N
3	Homme	Généraliste	0,11	0,96	N
4	Homme	Spécialiste	0,15	0,25	N
5	Homme	Spécialiste	0,15	0,74	N
6	Homme	Spécialiste	0,15	0,03	O
7	Femme	Généraliste	0,25	0,83	N
8	Femme	Généraliste	0,25	0,21	O
9	Femme	Généraliste	0,25	0,12	O
10	Femme	Spécialiste	0,18	0,45	N

Les valeurs des probabilités estimées sont fictives.

AVANT			APRES		
N° médecin	Sexe	Généraliste/ Spécialiste	N° médecin	Sexe	Généraliste/ Spécialiste
1	Homme	Généraliste			
2	Homme	Généraliste			
3	Homme	Généraliste			
4	Homme	Spécialiste			
5	Homme	Spécialiste			
6	Homme	Spécialiste			
7	Femme	Généraliste	2	Homme	Généraliste
8	Femme	Généraliste	3	Homme	Généraliste
9	Femme	Généraliste	4	Homme	Spécialiste
			5	Homme	Spécialiste
10	Femme	Spécialiste	7	Femme	Généraliste
			10	Femme	Spécialiste

Les individus 1, 6, 8 et 9 sont sélectionnés pour partir à la retraite. Les enregistrements correspondants sont donc supprimés de la table de données individuelles de l'année suivante.

Le résultat de la microsimulation étant un fichier de données individuelles pour chaque année de la période de projection, en principe de même format que celui ayant servi de base à la simulation, on peut, pour chaque année de la période de projection comme pour l'année ayant servi de base, calculer les effectifs correspondant à chacun des croisements possibles des différentes caractéristiques connues des individus.

1.3. Le choix de la méthode

L'emploi de la méthode de modélisation agrégée suppose, dans le cas d'un modèle très simplifié, la construction de quelques formules incluant des paramètres tels que le taux de départ à la retraite ou de décès. Dans le cas d'un modèle plus développé, où le niveau de détail des résultats est fin et la variabilité des comportements individuels à prendre en compte est grande, le nombre de paramètres peut être très important, ce qui incite à employer une formalisation sous forme matricielle⁵. Par exemple, il est plus pertinent d'appliquer des taux de départ à la retraite déclinés par âge, sexe, discipline (généraliste ou spécialiste) qu'un taux global. Il s'agit alors de créer des matrices décrivant l'état de la population avant le début de la période de projection, puis à chaque stade de la simulation, et des matrices de transition contenant les paramètres permettant de passer de l'état de la population au cours de l'année n à l'état de la population au cours de l'année $n+1$. Ces matrices sont appelées au fur et à mesure des calculs.

⁵ Pour la programmation, un langage matriciel, comme SAS IML, peut être utilisé.

Pour réaliser des simulations à l'aide de la microsimulation, il est nécessaire de calculer au préalable les probabilités pour les individus de vivre chacun des événements modélisés, en fonction de leurs caractéristiques les plus discriminantes en la matière, que nous appelons dans cet article « critères ». Cette étape correspond à celle du calcul des paramètres dans la méthode agrégée. Pour reprendre l'exemple précédent, il s'agit de calculer les probabilités de partir à la retraite pour les médecins selon l'âge, le sexe et la discipline (généraliste ou spécialiste). Pour utiliser la microsimulation, il faut en outre, non pas créer une ou des matrices décrivant l'état de la population avant le début de la période de projection, mais constituer un fichier de données individuelles servant de point de départ à la simulation, ce fichier étant actualisé au fur et à mesure par la suite.

Quelle que soit la méthode retenue, les « travaux préliminaires » à effectuer sont donc assez similaires. Cependant, la manipulation des matrices de paramètres employées dans le cadre de la modélisation agrégée, souvent de grandes tailles, peut se révéler plus délicate que celle des matrices de probabilités employées en microsimulation. Dans le cas de la modélisation agrégée, elles sont appelées pour réaliser les calculs et sont donc insérées dans des formules ou blocs d'instructions, parfois complexes. Dans le cas de la microsimulation, chaque matrice est utilisée uniquement pour ajouter dans le fichier de données individuelles la probabilité pour chaque individu de vivre l'événement, par appariement entre le fichier de données individuelles et le fichier contenant la matrice.

L'utilisation de la modélisation agrégée n'était pas impossible dans le cas présent. Cependant, la mise en œuvre de cette méthode nous paraissait plus délicate et plus fastidieuse. Il nous semblait également qu'elle aurait induit des risques d'erreurs importants, du fait du nombre de matrices (dépendant du nombre d'événements modélisés), de leur taille, mais aussi de la relative complexité des calculs. En fait, elle aurait même sans doute nécessité de revoir à la baisse nos ambitions quant aux objectifs du modèle. Il aurait en effet peut-être été nécessaire de réduire le nombre de paramètres, pour diminuer la complexité des calculs. Cela aurait conduit à réduire le niveau de détail des résultats produits, ou le nombre d'événements modélisés, ou encore le nombre de caractéristiques permettant de décrire les comportements individuels.

La modélisation d'événements supplémentaires est en outre assez simple à réaliser dans le cas de la microsimulation. Il « suffit » (ou presque) bien souvent d'ajouter un « module », c'est-à-dire d'insérer un bloc d'instructions dans le programme. Dans le cas de la modélisation agrégée, il peut être nécessaire de revoir et de modifier le programme dans son ensemble pour modéliser un événement supplémentaire.

La microsimulation génère des volumes de données importants. On ne peut donc employer cette méthode que si l'on dispose de capacités de stockage importantes. Par ailleurs, les temps de calcul sont assez longs. Pour qu'ils restent raisonnables, il faut des moyens de calcul suffisants.

Les moyens informatiques dont nous disposions semblaient suffisants pour répondre à ces contraintes. Toutefois, pour réduire le volume des données stockées, nous avons sauvegardé les résultats des simulations non pas sous la forme de données individuelles, mais sous la forme de données agrégées permettant de ventiler les effectifs projetés pour chaque année de la période de projection selon toutes les caractéristiques des médecins retenues (cf. 1.1.). A titre indicatif, la simulation d'un scénario nécessite au maximum 2h30 de calculs, pour une période de projection allant de 2007 à 2030 et un effectif de l'ordre de 200 000 individus (et 5 itérations – cf. 2.4.).

Par ailleurs, nous disposions de deux sources de données individuelles, exhaustives et riches en informations sur les médecins en activité, à partir desquelles il était simple de constituer un fichier de données individuelles pouvant servir de base à la microsimulation.

Il n'existait donc pas d'obstacle technique à l'emploi de cette méthode.

Pour résumer, les raisons qui nous ont conduits à privilégier l'utilisation de la microsimulation sont :

- la nécessité de produire des résultats permettant une analyse fine des effectifs projetés
- la volonté de tenir compte au mieux de la variabilité des comportements des médecins selon leurs caractéristiques
- le confort que procure la microsimulation en matière de programmation, et la relative simplicité des calculs
- la disponibilité des données et des moyens informatiques nécessaires

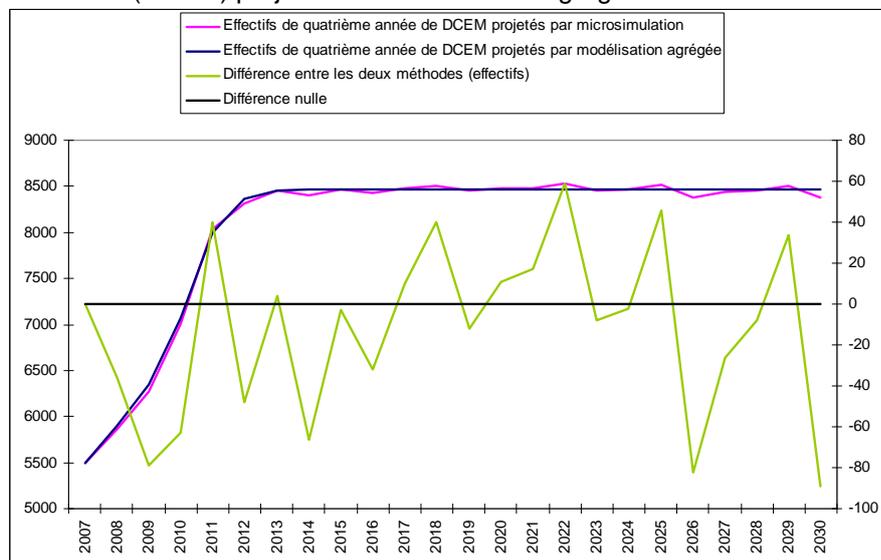
Finalement, la microsimulation nous a paru bien adaptée à l'élaboration d'un modèle de projection d'effectifs de médecins. La composition de la population des médecins actifs au cours des années futures se déduit assez facilement de celle de la population des médecins actifs actuellement, en simulant la réalisation d'événements qui agissent sur l'effectif global (entrée dans la vie active des nouveaux diplômés, retraite, décès) ou sur les caractéristiques des individus (vieillesse, changement de région d'exercice, changement de mode d'exercice, cessation temporaire d'activité, reprise d'activité, etc.).

Cependant, pour la modélisation des deux premiers cycles des études médicales⁶, nous avons utilisé la modélisation agrégée, les calculs à effectuer étant assez simples. Il s'agit en effet simplement de calculer, pour chaque région, les effectifs d'étudiants en médecine pour chaque niveau d'études⁷. Les effectifs correspondant à un niveau d'études donné se déduisent assez facilement de ceux correspondant au niveau inférieur pour l'année précédente de la période de projection, à partir d'un taux (régional) de redoublements, d'un taux d'abandons et d'un taux d'arrivées en cours d'études estimés préalablement. A titre illustratif, voici la formule permettant de calculer les effectifs en deuxième année de deuxième cycle (DCEM2) en fonction des effectifs de première année (DCEM1) :

$$DCEM2 = (DCEM1_{n-1} * (1 - \text{taux de redoublements en DCEM1} - \text{taux d'abandons en DCEM1}) + DCEM2_{n-1} * \text{taux de redoublements en DCEM2}) / (1 - \text{taux d'arrivées en DCEM2})$$

Nous avons néanmoins testé également la microsimulation pour cette partie du modèle, et comparé les résultats obtenus avec chacune des deux méthodes. Pour la quatrième année du deuxième cycle des études médicales par exemple, la différence entre l'effectif obtenu par microsimulation et l'effectif obtenu par modélisation agrégée est, pour chaque année de la période de projection, inférieure à 100, pour un effectif compris entre 5 000 et 9 000 étudiants (*graphique 1*). Le rapport entre les deux effectifs projetés est toujours compris entre 0,99 et 1,01. Les deux méthodes donnent donc des résultats très proches.

Graphique 1 Comparaison des effectifs d'étudiants de 4^e année de deuxième cycle d'études médicales (DCEM) projetés avec la méthode agrégée et avec la microsimulation



La modélisation du troisième cycle des études médicales est encore relativement simple, mais il est nécessaire d'introduire dans les simulations l'information sur la spécialité et la région choisies par les étudiants, afin de calculer les effectifs de nouveaux diplômés par région et par spécialité. Le nombre d'événements à modéliser est aussi un peu plus important. Par souci de confort, nous avons donc

⁶ A compter de la deuxième année du premier cycle des études médicales.

⁷ A ce stade des études, les étudiants n'ont pas encore de spécialité. Le sexe et l'âge des étudiants sont attribués au moment de l'entrée dans la vie active. Enfin, les étudiants en médecine n'ont, par définition, ni région, ni mode, ni zone d'exercice.

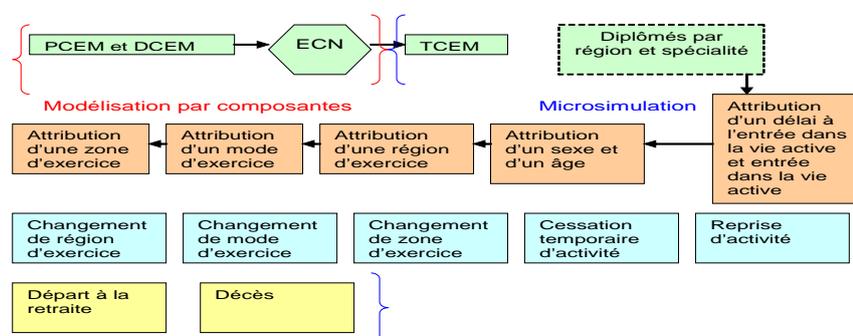
choisi d'utiliser la microsimulation pour modéliser le déroulement du troisième cycle des études médicales, puis toutes les autres étapes du modèle. Le passage de la modélisation agrégée à la microsimulation s'effectue en créant un fichier d'étudiants fictifs à partir des effectifs produits par l'étape en modélisation agrégée.

Signalons que d'autres modèles de projection d'effectifs de médecins⁸ n'emploient pas la microsimulation, mais uniquement la modélisation agrégée. Ces modèles prennent peu en compte la variété des comportements des médecins et les résultats qu'ils produisent sont moins détaillés.

1.4. Aperçu du modèle

Le modèle simule le déroulement des études médicales, de la deuxième année du premier cycle à l'obtention du diplôme, l'entrée dans la vie active des jeunes diplômés, à l'occasion de laquelle il leur attribue leur sexe, âge, région, mode et zone d'exercice, puis les changements intervenant au cours de la carrière des médecins (changement de région, de mode, de zone d'exercice, cessation temporaire et reprise d'activité), et la sortie définitive de la vie active (départ à la retraite ou décès). Le graphique 2 décrit très schématiquement le modèle :

Graphique 2 Schéma d'ensemble du modèle



Le déroulement des études médicales a été modélisé pour permettre de simuler notamment les effets de politiques de régulation portant sur le *numerus clausus* et la répartition des postes ouverts aux épreuves classantes nationales (remplaçant depuis 2004 les concours de « l'internat »).

Une fois le nombre de diplômés calculé, il est nécessaire de déterminer le nombre de médecins qui exerceront la médecine parmi eux. La proportion des jeunes diplômés n'exerçant jamais la médecine en France a été estimée à 3 %.

Il faut ensuite déterminer un délai à l'entrée dans la vie active, sauf à supposer par exemple que tous les médecins entrent dans la vie active l'année de leur diplôme, hypothèse qui n'a pas été retenue, un quart des jeunes diplômés environ entrant dans la vie active un an ou plus après l'obtention de leur diplôme actuellement.

Le modèle simule l'entrée dans la vie active des jeunes diplômés qui exerceront un jour ou l'autre la médecine en France, dont il connaît déjà, à cette étape, la région de diplôme et la spécialité, mais il leur attribue au préalable les autres caractéristiques qui doivent être connues pour chacun : le sexe, l'âge, la région d'exercice, le mode d'exercice et la zone d'exercice.

Le champ retenu pour les projections exclut les médecins en cessation temporaire d'activité. Il faut par conséquent modéliser la cessation temporaire d'activité. Pour la même raison, il faut aussi simuler les reprises d'activité. Signalons que les cessations temporaires d'activité ne sont pas compensées chaque année par les reprises d'activité (ni en volume, ni en ce qui concerne la répartition des médecins concernés selon leurs caractéristiques).

Les changements de région, de mode d'exercice et de zone d'exercice en cours de vie active ont été modélisés également pour la première fois dans la version 2008 du modèle.

⁸ Modèles réalisés par le centre de sociologie et de démographie médicale (CSDM) et par la caisse autonome de retraite des médecins de France (CARMF) par exemple.

Enfin, il était nécessaire de simuler les sorties définitives de la vie active. Les départs à la retraite et les décès ont été distingués dans la simulation, afin de permettre à l'utilisateur d'étudier les flux correspondant aux départs à la retraite et aux décès séparément pour analyser les résultats des simulations.

Les cessations définitives de la vie active ne correspondant pas à un départ à la retraite ou un décès ne sont pas simulées. Cet événement est assimilé à une cessation temporaire d'activité (à laquelle ne correspondrait jamais aucune reprise d'activité).

La modélisation des différents événements faisant suite à l'obtention du diplôme présentés ci-dessus offre de nombreuses possibilités à l'utilisateur en matière de simulation : il peut en effet simuler les effets d'un changement dans le comportement des médecins (spontané ou induit par une mesure politique) en ce qui concerne par exemple le délai à l'entrée dans la vie active, le choix de la région d'exercice à l'entrée dans la vie active, le changement de mode d'exercice en cours de vie active, le départ à la retraite, etc.

La réalisation de chaque « événement » est simulée à l'aide de tirages aléatoires. Les résultats de certains de ces tirages sont binaires : l'individu i vit ou non l'événement e (il part à la retraite ou non, il décède ou non par exemple). Les résultats des autres tirages sont plus « complexes » : le tirage permet d'affecter une modalité choisie parmi plusieurs à chaque individu (une région d'exercice, un sexe et un âge par exemple). Ces tirages sont effectués successivement. A l'issue de chacun d'eux, le fichier de données individuelles est actualisé. Les calculs correspondant à une année donnée s'achèvent par l'incrément d'un an de l'âge des individus. La simulation est réalisée de manière itérative pour chaque année de la période de projection.

A l'issue des calculs correspondant à chaque année de la période de projection est enregistré l'état de la population des médecins actifs. Le modèle produit également des tables de résultats intermédiaires (par exemple sur les entrées dans la vie active ou sur les changements de mode d'exercice), soit sous forme individuelle, soit sous forme agrégée. Elles permettent de contrôler le déroulement de la simulation, mais aussi de compléter l'analyse des effectifs de médecins actifs au cours de chaque année.

Un scénario de référence est défini par un jeu d'hypothèses. Pour le scénario de référence, nous avons fait le choix de retenir des hypothèses correspondant pour la plupart à des « comportements constants » (c'est-à-dire que les valeurs des paramètres sont supposées égales à celles actuellement observées – cf. 2.5.).

Des variantes sont ensuite définies, qui ne diffèrent du scénario de référence que par une seule hypothèse, ce qui permet de mesurer l'impact de cette modification sur l'évolution à long terme de l'effectif et des caractéristiques des médecins en activité. Par exemple, une variante mesure l'impact d'un changement dans le comportement des jeunes généralistes en matière de mobilité au moment de l'entrée dans la vie active. La simulation de ce scénario diffère de celle du scénario de référence uniquement par les probabilités pour chaque jeune généraliste de choisir une région d'exercice donnée à l'entrée dans la vie active, sur lesquelles on fait une hypothèse différente de celle de comportements identiques à ceux observés actuellement.

L'élaboration du modèle de projection a nécessité d'effectuer des choix (ce qui n'est pas propre à la microsimulation) et de résoudre diverses difficultés techniques, que nous présentons maintenant. Nous abordons successivement les points suivants : le choix des événements à modéliser, le choix des critères pour le calcul des probabilités, le traitement des événements « rares », la gestion des effets de l'aléa, le choix des hypothèses et ses conséquences.

2. Les choix effectués et les solutions techniques retenues

2.1. Le choix des événements à modéliser

Choisir les « événements » (ou changements d'état) à modéliser consiste à réaliser des arbitrages, pour aboutir finalement à un modèle dont le potentiel est à la hauteur des besoins, c'est-à-dire permettant à l'utilisateur de simuler les scénarios qu'il souhaite et produisant des résultats

suffisamment précis, tout en étant parcimonieux, afin de garantir une complexité et des temps de calcul limités.

Pour choisir les événements à modéliser, il faut tout d'abord recenser tous les événements susceptibles d'avoir un impact sur ce que l'on cherche à mesurer. En l'occurrence, il s'agit de mesurer, sous certaines hypothèses portant sur les décisions prises par les pouvoirs publics⁹ ou sur les comportements des médecins, l'effectif de médecins actifs au cours de chaque année de la période allant de 2007 à 2030, ainsi que, pour chaque année, leur répartition selon le sexe, l'âge, la spécialité, la région, le mode et la zone d'exercice (cf. 1.1.). L'effectif de médecins actifs résulte notamment des événements suivants : obtenir son diplôme de médecine en France, entrer dans la vie active, prendre sa retraite, mais aussi cesser temporairement son activité, reprendre une activité après l'avoir cessée temporairement, venir exercer en France (pour un médecin diplômé ou exerçant à l'étranger), partir exercer à l'étranger (pour un médecin diplômé ou exerçant en France). Quant à la répartition des médecins actifs selon leurs caractéristiques, elle dépend par exemple du choix de la discipline¹⁰, puis de la spécialité, et de la subdivision (lieu de formation) fait par chaque étudiant en fonction de ses résultats aux épreuves classantes nationales (remplaçant depuis 2004 les concours de « l'internat »), du choix de la région d'exercice et de celui du mode d'exercice au moment de l'entrée dans la vie active, des changements de région d'exercice en cours de carrière, etc.

Le choix des événements à modéliser est tout d'abord limité par la disponibilité des données et leur fiabilité. Par rapport à la version précédente du modèle, nous disposons d'une nouvelle source de données sur les médecins en activité (les fichiers annuels de l'Ordre des médecins) qui permettait de repérer de façon fiable les changements de situation d'une année à la suivante, par appariement de fichiers successifs, donc d'estimer des probabilités de changer de situation. Par conséquent, il a été possible de modéliser les changements de région, de mode et de zone d'exercice, ainsi que les cessations temporaires et les reprises d'activité, événements qui ne l'étaient pas dans la version précédente du modèle. En revanche, il n'a toujours pas été possible de modéliser les flux internationaux entrants et sortants de médecins en exercice, ni les changements de spécialité en cours de carrière. En effet, nous ne disposons pas de données fiables permettant d'estimer le volume et la répartition des flux de médecins en provenance de l'étranger, les probabilités pour les médecins exerçant en France de partir exercer à l'étranger en cours de carrière, ainsi que les probabilités de changer de spécialité.

La deuxième contrainte qui s'impose pour la sélection des événements à modéliser est le souci de parcimonie. Il est en effet préférable de ne pas modéliser les événements dont l'impact sur ce que l'on cherche à mesurer est jugé faible. La complexité du modèle doit être maîtrisée, afin que l'utilisation, l'interprétation des résultats et la maintenance restent aisées. Il s'agit aussi de s'assurer que les temps de calcul soient raisonnables : des temps de calcul trop longs conduiraient l'utilisateur à réduire le nombre de simulations à réaliser, c'est-à-dire le nombre de scénarios testés. Enfin, compte tenu de l'aléa inhérent à la simulation de tout événement, modéliser un événement supplémentaire conduit à « ajouter de l'aléa » aux résultats finaux. En modélisant un événement dont l'impact sur ce que l'on souhaite mesurer est très faible, on risque donc d'introduire uniquement « du bruit » dans les résultats des simulations, c'est-à-dire d'introduire de l'aléa sans en tirer de bénéfique pour la précision des résultats. Pour cette raison, nous avons renoncé à modéliser les migrations géographiques des étudiants en médecine, qui sont très rares et produisent des soldes migratoires extrêmement faibles.

Au contraire, il est parfois nécessaire de modéliser certains événements pour répondre correctement aux besoins de l'utilisateur. Celui-ci doit en effet pouvoir tester tous les scénarios qu'il estime utile de simuler, ce qui suppose parfois de retenir des événements dont la modélisation ne semble pas a priori indispensable. Par exemple, dès lors que l'on voulait pouvoir simuler l'impact d'un changement du niveau ou de répartition régionale du *numerus clausus* (quota d'entrée en deuxième année du premier cycle d'études), qui est un outil important de régulation de la démographie médicale, il était préférable de modéliser tout le déroulement du cursus (redoublements, abandons ou arrivées en cours d'études) pour mieux estimer les flux d'entrants dans la vie active chaque année. De même, nous avons modélisé les affectations prononcées après la fin des épreuves classantes nationales, malgré leur complexité, afin qu'il soit possible de simuler par exemple l'impact d'un changement dans la répartition des postes ouverts à ces épreuves par discipline et par subdivision.

⁹ En ce qui concerne le niveau et la répartition géographique du *numerus clausus* par exemple.

¹⁰ La discipline correspond à une spécialité médicale ou en regroupe plusieurs.

Il faut également tenir compte du fait que certains événements, pourtant « rares », peuvent avoir un impact important sur les résultats des simulations. Ainsi, il n'est pas courant qu'un médecin change de région d'exercice : en moyenne, au cours des trois dernières années, 1,4 % des médecins ont changé de région d'exercice entre deux années consécutives. Cependant, les flux tendent à favoriser toujours les mêmes régions : les soldes migratoires sont assez systématiquement négatifs pour certaines régions, et positifs pour d'autres. Par effet de cumul, l'impact des migrations géographiques peut donc finalement se révéler fort sur le long terme. C'est la principale raison pour laquelle nous avons choisi de modéliser les changements de région d'exercice, la production de résultats régionaux les plus précis possibles faisant partie des objectifs du modèle.

Pour évaluer l'impact de ce choix sur les résultats produits par le modèle, nous avons simulé le scénario de référence sans simuler les changements de région, et comparé les résultats obtenus à ceux donnés par le scénario de référence qui les prend en compte (tableau 1).

Tableau 1 Effectifs de médecins par région, avec et sans prise en compte de la mobilité régionale

Région	Effectifs 2006	Effectifs 2030 scénario de référence	Effectif 2030 scénario sans mobilité géographique	Evolution 2030/2006 scénario de référence	Evolution 2030/2006 scénario sans mobilité géographique	Solde migratoire moyen (effectifs) sur la période 2030/2007 d'après le scénario de référence
Alsace	6101	5802	6432	-4,9	5,4	-28
Antilles-Guyane	2337	2782	2459	19,0	5,2	18
Aquitaine	10694	12139	12021	13,5	12,4	14
Auvergne	3921	4320	4589	10,2	17,0	-16
Basse-Normandie	4000	4569	4781	14,2	19,5	-13
Bourgogne	4565	4070	3956	-10,9	-13,3	-1
Bretagne	9258	11586	11001	25,2	18,8	29
Centre	6627	6160	5690	-7,0	-14,1	17
Champagne-Ardenne	3750	3737	4154	-0,3	10,8	-22
Corse	910	668	408	-26,6	-55,2	9
Franche-Comté	3356	3971	4326	18,3	28,9	-17
Haute-Normandie	4857	4916	5099	1,2	5,0	-9
Île-de-France	46144	37132	40723	-19,5	-11,7	-166
Languedoc-Roussillon	9040	8274	7198	-8,5	-20,4	55
Limousin	2415	2415	2541	0,0	5,2	-5
Lorraine	6834	7303	7917	6,9	15,8	-30
Midi-Pyrénées	9668	9132	8236	-5,5	-14,8	42
Nord-Pas-de-Calais	11770	11651	13091	-1,0	11,2	-75
Pays de la Loire	9392	11408	11087	21,5	18,0	14
Picardie	4814	4778	5033	-0,8	4,5	-18
Poitou-Charentes	4998	6296	6064	26,0	21,3	14
Provence-Alpes-Côte d'Azur	19286	16821	14348	-12,8	-25,6	120
Réunion	2079	2573	2270	23,8	9,2	15
Rhône-Alpes	19698	21448	20996	8,9	6,6	32

La différence entre les résultats de ces deux simulations s'explique principalement par le solde migratoire¹¹. L'impact de la prise en compte de l'événement « changer de région » est loin d'être négligeable sur les résultats des projections en ce qui concerne la répartition des médecins par région. Ainsi, le nombre de médecins exerçant Bretagne, région dont le solde migratoire est en moyenne positif (+29 médecins par an) augmenterait de 19 % entre 2006 et 2030 si l'on suppose nulle la mobilité géographique des médecins en cours de carrière. Si l'on fait l'hypothèse de comportements constants en la matière, la population médicale de la région augmenterait de 25 %. Pour l'Alsace, région au solde migratoire négatif en moyenne (-28 médecins par an), l'évolution projetée du nombre des médecins exerçant dans la région au cours de la période passe de 5 % à -5 % lorsqu'on prend en compte la mobilité géographique.

Une fois sélectionnés les événements à modéliser, il faut encore déterminer l'ordre de ces événements dans la simulation.

Cet ordre n'est en fait pas indépendant du champ retenu pour le calcul des probabilités (ou vice-versa) : par exemple, les probabilités de changer de région, de mode ou de zone d'exercice étant estimées sur le champ des médecins actifs au cours de deux années consécutives, elles doivent être

¹¹ Mais pas seulement, puisque lorsqu'on modifie un élément du modèle (en l'occurrence, on modélise un événement en moins), l'issue des autres événements peut être différente.

appliquées sur le même champ au cours de la simulation. Les changements de région, de mode ou de zone d'exercice doivent donc être simulés après les cessations temporaires ou définitives de la vie active, mais avant les entrées dans la vie active et les reprises d'activité. De même, nous avons choisi d'estimer les probabilités de cesser temporairement son activité, ainsi que les probabilités de décéder et de partir à la retraite sur le champ complet des médecins en activité au cours d'une année donnée. Pour que les probabilités s'appliquent au même champ au cours de la simulation, il faut donc simuler simultanément (par des tirages simultanés¹²) les cessations temporaires d'activité, les décès et les retraites.

Mais l'ordre des événements reste partiellement à déterminer. Nous avons choisi par exemple d'attribuer successivement aux médecins entrant dans la vie active un délai à l'entrée dans la vie active, un sexe et un âge simultanément, une région d'exercice, un mode d'exercice, et enfin une zone d'exercice. Par exemple, si le choix du mode d'exercice est fortement lié à celui de la région d'exercice, on cherchera à tenir compte de cette corrélation en déclinant les probabilités de choisir tel ou tel mode d'exercice selon la région d'exercice choisie (la région étant attribuée aux nouveaux entrants avant le mode d'exercice). Supposons que le choix du mode d'exercice soit aussi fortement lié celui de la zone d'exercice, qui est attribuée après le mode d'exercice. On cherchera alors à tenir compte de cette corrélation en déclinant les probabilités de choisir telle ou telle zone d'exercice selon le mode d'exercice choisi. Cependant, parce que le choix des critères à retenir est contraint (cf. 2.2.), et qu'il n'est donc finalement pas possible de retenir tous les critères que l'on souhaiterait, l'ordre des événements modélisés n'est en réalité pas toujours sans conséquence sur les résultats.

2.2. Le choix des critères pour le calcul des probabilités

Tous les individus n'ont pas la même probabilité de vivre un événement donné : le départ à la retraite, par exemple, est plus ou moins probable selon l'âge du médecin. Pour tenir compte de la variabilité des comportements des médecins, il faut donc calculer des probabilités différenciées selon celles des caractéristiques des médecins qui les « déterminent » le plus. En pratique, il s'agit en fait de définir des « classes » d'individus dont on supposera qu'ils ont la même probabilité de vivre un événement donné. Ces classes, ou « strates », sont définies par la combinaison de plusieurs caractéristiques, ou « critères ». Pour le calcul des probabilités de partir à la retraite par exemple, on peut retenir comme critères l'âge et le fait d'être généraliste ou spécialiste. Cependant, on peut penser que les hommes et les femmes généralistes (ou spécialistes) n'ont pas la même probabilité de prendre leur retraite à âge donné. Faut-il alors ajouter le critère « sexe » dans la construction des classes ? Le choix des critères n'est pas évident. Il représente en fait une part importante du travail d'élaboration d'un modèle.

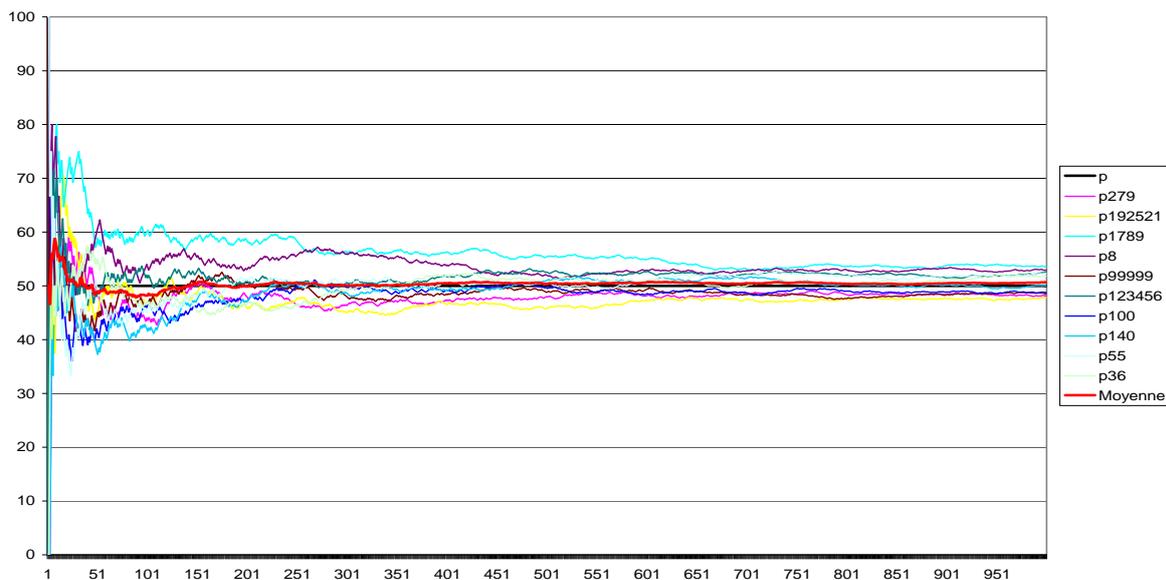
Pour chaque événement, il faut repérer les caractéristiques qui ont une influence sur la probabilité de réalisation de l'événement. Pour ce faire, nous avons employé notamment la régression. Il est tentant de retenir comme critères toutes les caractéristiques qui apparaissent significatives dans la régression, ce qui n'est pas possible en pratique. En effet, le croisement de critères trop nombreux conduit à la création de strates d'effectifs très faibles, ce qui est problématique d'une part pour le calcul des probabilités, et d'autre part pour le tirage aléatoire réalisé à partir de ces probabilités.

Les probabilités sont estimées par les proportions observées sur des données observées récentes (si l'on retient l'hypothèse de comportements constants). Si les effectifs des classes sont très faibles, ces proportions sont instables d'une année à l'autre. Il est d'ailleurs important, si l'on dispose de données sur plusieurs années, d'évaluer la stabilité dans le temps des proportions empiriques. Si les proportions observées sont trop instables, il est préférable d'augmenter la taille des classes, c'est-à-dire de supprimer un critère, ou de le redéfinir d'une façon plus agrégée. S'il demeure une certaine instabilité dans le temps, les probabilités peuvent être estimées par la moyenne des proportions observées au cours de quelques années récentes, ce que nous avons choisi de faire de façon assez systématique.

¹² Pour sélectionner simultanément 3 % de médecins qui cessent temporairement leur activité, 1 % de médecins qui décèdent et 10 % de médecins qui partent à la retraite par exemple, on peut simuler la réalisation d'une loi uniforme sur la population des médecins actifs, faire cesser temporairement leur activité à ceux auxquels la valeur attribuée est inférieure à 0,03, faire décéder ceux pour lesquels elle est comprise entre 0,03 et 0,04 et faire partir à la retraite ceux pour lesquels elle est comprise entre 0,04 et 0,14 (cf. 2.4.).

De même, lorsque les effectifs des strates où s'effectue le tirage selon ces probabilités-cibles sont faibles, les proportions d'individus sélectionnés pour vivre l'événement à l'issue du tirage risquent fort d'être éloignées des probabilités-cibles. Pour illustrer ce propos, on réalise une simulation qui consiste, pour une probabilité donnée, à effectuer des tirages à l'aide de la loi uniforme au moyen de la fonction `ranuni` de SAS, sur des échantillons dont la taille est comprise entre 1 et 1 000 individus. La fonction `ranuni` est initialisée par une racine choisie au hasard. Les individus sélectionnés pour vivre l'événement sont ceux auxquels la fonction `ranuni` attribue une valeur inférieure à la probabilité-cible. Le graphique 3 montre les proportions d'individus vivant l'événement obtenues pour une probabilité-cible de 50 %, en fonction de la taille de l'échantillon, pour 10 valeurs différentes de racines choisies au hasard.

Graphique 3 Proportions effectives obtenues à l'issue d'un tirage uniforme (10 racines différentes) avec une probabilité de 50 %, sur des échantillons de taille 1 à 1000.



Cette simulation montre que si le tirage est effectué dans une strate dont la taille est inférieure à 50 individus, l'écart à la probabilité-cible peut être élevé (avec la racine 1789, on obtient une proportion de 75 % au lieu de 50 % pour un échantillon de 39 individus par exemple) et qu'il faut que les strates aient une taille minimale de 150 individus environ pour obtenir « convergence » correcte, c'est-à-dire par exemple pour atteindre la probabilité-cible à 10 % près avec la plupart des racines choisies ici. En fait, si l'on s'appuie sur le calcul de l'intervalle de confiance au niveau 95 % de la loi binomiale pour définir de façon théorique la taille de l'échantillon, il faudrait un échantillon de taille 384 au moins pour que la proportion observée diffère de moins de 10 % de la probabilité-cible de 50 %, c'est-à-dire pour qu'elle soit comprise entre 45 % et 55 %.

Ces simulations montrent également que pour les racines 1789 et 192521, même lorsque la taille de l'échantillon est importante, la proportion obtenue reste nettement éloignée de la probabilité-cible.

Cependant, la courbe représentant la moyenne des 10 courbes associées aux différentes racines converge plus rapidement (à moins de 1 % près à partir de 150 individus), ce qui suggère d'effectuer plusieurs simulations avec des racines différentes et de retenir la moyenne. Cette pratique est d'ailleurs courante en microsimulation lorsque les moyens informatiques le permettent.

Lorsque la probabilité-cible est plus faible, la « convergence » est plus lente. Par exemple, pour la probabilité de 1 %, la proportion obtenue avec la racine 192521 est toujours nulle jusqu'à une taille d'échantillon de 357 et la moyenne des proportions obtenues avec les 10 racines est supérieure de 10 % à la probabilité-cible lorsque l'échantillon compte 500 individus.

Nous avons tenu compte de ces résultats pour choisir les critères selon lesquels décliner l'estimation de chaque jeu de probabilités : nous avons essayé de construire autant que possible des strates de taille suffisante pour que les résultats des tirages ne risquent pas d'être imprécis.

On ne gagne donc pas systématiquement à prendre en compte un critère supplémentaire pour l'estimation des probabilités. Prenons l'exemple de la simulation du changement de mode d'exercice. L'événement « changer de mode d'exercice » est corrélé au mode d'exercice actuel, et à l'âge du médecin. Lorsque l'on n'utilise comme critère que le mode d'exercice pour estimer les probabilités de changer de mode d'exercice (à partir d'un fichier-test et à l'aide d'une racine choisie au hasard), on sélectionne 385 médecins hospitaliers de 25 à 35 ans pour changer de mode d'exercice, alors que l'effectif-cible (déduit de la probabilité-cible et de l'effectif de la strate) est de 752. Lorsqu'on ajoute l'âge comme critère, en plus du mode d'exercice, 751 médecins sont sélectionnés pour changer de mode d'exercice dans cette strate. Le résultat du tirage est donc nettement meilleur pour cette strate avec ce critère supplémentaire. En revanche, il se produit le contraire pour les médecins hospitaliers de 41 à 45 ans, pour lesquels l'effectif-cible est de 281, le nombre de médecins sélectionnés lorsqu'on utilise le mode d'exercice comme seul critère de 286 et le nombre de médecins sélectionnés avec le mode d'exercice et l'âge comme critères de 252.

La difficulté consiste à évaluer globalement la qualité des résultats des tirages. Pour ce faire, on peut par exemple comparer les écarts quadratiques moyens aux probabilités-cibles dans les deux cas. Dans l'exemple précédent, l'écart quadratique moyen aux probabilités-cibles estimées avec les critères mode d'exercice et âge est de 0,07 pour le résultat du tirage réalisé uniquement avec le critère mode d'exercice et de 0,004 pour le résultat du tirage réalisé avec les critères mode d'exercice et âge. Globalement, la prise en compte de l'âge comme critère améliore nettement les résultats du tirage en ce qui concerne la répartition des individus sélectionnés selon leur mode d'exercice et leur âge.

Il faut donc, pour le choix des critères, être aussi parcimonieux que pour la sélection des événements à modéliser, et ne retenir que les critères les plus discriminants, c'est-à-dire ceux dont l'effet sur les probabilités est nettement significatif et stable dans le temps, tout en veillant à ce que la taille des strates obtenues reste suffisante. Il faut également tenir compte de l'intérêt de retenir chaque critère par rapport aux résultats attendus des simulations. Ainsi, si l'on sait que l'utilisateur s'intéressera peu aux effectifs projetés de médecins actifs selon le sexe et le mode d'exercice, il n'est peut-être pas indispensable de retenir le sexe comme critère pour le calcul des probabilités de changer de mode d'exercice.

Finalement, le choix des critères pour le calcul des probabilités, tout comme le choix des événements à modéliser ainsi que de l'ordre de ces événements dans la simulation, comporte toujours une part d'arbitraire.

2.3. Le traitement des événements « rares »

Il est plus difficile de simuler la réalisation de certains événements que d'autres, du fait de leur « rareté ». Comme on l'a vu au chapitre 2.2, lorsque les probabilités-cibles sont faibles, les tirages tendent en effet à donner des proportions d'individus sélectionnés pour vivre un événement plus éloignées des probabilités-cibles.

Théoriquement, puisque sont modélisés uniquement les événements dont l'impact sur ce que l'on cherche à mesurer est significatif, nous ne devrions pas avoir à modéliser des événements « rares ». Nous avons toutefois modélisé certains événements peu courants lorsque par effet de cumul sur la période de projection, leur impact sur les effectifs et les répartitions projetées pouvaient s'avérer net (cf. 2.1.).

Il existe plusieurs façons de sélectionner aléatoirement au sein d'une strate les individus qui vivent un événement donné. On peut attribuer à chaque individu un nombre aléatoire compris entre 0 et 1 en simulant la réalisation d'une loi uniforme, puis sélectionner ceux auxquels a été attribué un nombre inférieur ou égal à p , la probabilité qu'un individu donné vive l'événement¹³. On peut également calculer n , le nombre d'individus à sélectionner dans la strate, égal à p multiplié par le nombre total d'individus de la strate, N . Il suffit alors de sélectionner n individus, par exemple en ordonnant tous les individus de la strate aléatoirement et en retenant les n premiers. Avec la première méthode, le nombre d'individus sélectionnés n'est pas forcément égal strictement à n . Avec la deuxième méthode, sous l'hypothèse de comportements constants, et en supposant que la taille des strates ne varie quasiment pas, on sélectionnera tous les ans exactement le même nombre d'individus pour vivre cet événement, ce qui est moins conforme à la réalité que ce que l'on obtient avec la première méthode où le nombre d'individus sélectionnés subit des fluctuations d'une année sur l'autre. Mais surtout, lorsqu'un événement a une faible probabilité de se produire, se pose le problème des arrondis : le nombre n , qui doit être entier, est un arrondi du produit pN . Il arrive, notamment lorsque la probabilité p est faible et/ou lorsque N , la taille de la strate, est petite, que la proportion obtenue d'individus vivant l'événement, n/N , soit assez différente de p . Si p est très petit, il est même fréquent que n soit arrondi à zéro et que par conséquent, aucun individu ne soit sélectionné (pendant toute la période de projections si N est constant ou tend à diminuer).

Les différents essais que nous avons réalisés ont fait apparaître qu'avec la deuxième méthode, on tend assez systématiquement, en particulier lorsque les probabilités-cibles sont très faibles, à sélectionner un nombre d'individus inférieur à celui attendu, notamment du fait des arrondis. Prenons le cas de la sélection des étudiants qui abandonnent les études médicales à l'issue de la première année de troisième cycle. Les probabilités d'abandonner sont déclinées par discipline (il existe 11 disciplines), et de l'ordre de 1 à 2 % pour la plupart des disciplines. Les strates sont de tailles comprises entre 16 et 1 752 individus. L'effectif-cible total est de 128 individus. Avec la première méthode, on sélectionne au total 119 individus, contre seulement 109 avec la deuxième.

Dans le modèle de projection, nous n'avons employé la deuxième méthode que pour un événement particulier, pour lequel le nombre d'individus à sélectionner est prédéfini. Les tests réalisés ont montré que le nombre d'étudiants ainsi sélectionnés était le plus souvent un peu inférieur à l'effectif-cible.

Pour réduire l'imprécision des simulations due à la rareté d'un événement, nous avons aussi tenté de décomposer la simulation de la réalisation de certains événements rares en deux simulations, afin de simuler à chaque fois la réalisation d'événements moins rares. Par exemple, au lieu de simuler directement les mouvements d'une région à une autre, nous avons d'abord sélectionné les médecins qui changent de région, puis attribué une nouvelle région d'exercice à chacun des médecins sélectionnés pour changer de région au cours de l'étape précédente. Nous avons également procédé ainsi pour simuler les changements de mode et de zone d'exercice.

Les probabilités pour les médecins de quitter une région d'exercice pour une autre sont données par une matrice dont voici un extrait :

¹³ Les propriétés de la loi uniforme sont telles que la probabilité pour que la valeur x , valeur comprise entre 0 et 1, soit attribuée à un individu i est la même quelle que soit cette valeur x . Par conséquent, sur l'ensemble du champ sur lequel doit être effectuée la sélection, 5 % des individus se voient attribuer une valeur inférieure à 0,05, 10 % une valeur inférieure à 0,1, 50 % une valeur inférieure à 0,5, etc. La répartition des valeurs de la variable aléa est d'autant plus proche de cette répartition théorique que l'effectif sur lequel la sélection doit être effectuée est grand.

Région de départ	Région d'arrivée						
	Alsace	Antilles-Guyane	Aquitaine	Auvergne	Basse-Normandie	Bourgogne	Bretagne
Alsace	0,98685	0,00034	0,00062	0,00034	0,00023	0,00028	0,00011
Antilles-Guyane	0,00045	0,95887	0,00409	0,00045	0,00045	0,00076	0,00137
Aquitaine	0,00023	0,00071	0,98817	0,00013	0,00010	0,00006	0,00049
Auvergne	0,00026	0,00096	0,00113	0,98070	0,00052	0,00078	0,00070
Basse-Normandie	0,00009	0,00017	0,00086	0,00009	0,98244	0,00026	0,00232
Bourgogne	0,00045	0,00068	0,00037	0,00150	0,00015	0,98075	0,00045
Bretagne	0,00026	0,00060	0,00045	0,00004	0,00071	0,00034	0,98869

Les valeurs de la plupart de ces probabilités sont très faibles. Nous avons donc décomposé l'événement « aller de telle région vers telle région » en deux événements : 1. changer de région ; 2. pour les médecins qui changent de région, aller de telle région vers telle région.

La probabilité de changer de région est comprise entre 0,9 % et 5,3 % selon les régions, et de l'ordre de 1 à 2 % pour la grande majorité des régions.

Voici un extrait de la matrice des probabilités d'aller d'une région vers une autre pour les médecins changeant de région :

Région de départ	Région d'arrivée						
	Alsace	Antilles-Guyane	Aquitaine	Auvergne	Basse-Normandie	Bourgogne	Bretagne
Alsace		0,02589	0,04751	0,02666	0,01705	0,02109	0,00863
Antilles-Guyane	0,01000		0,09928	0,01100	0,01050	0,01834	0,03453
Aquitaine	0,01931	0,06061		0,01105	0,00829	0,00540	0,04105
Auvergne	0,01299	0,05010	0,05963		0,02657	0,04209	0,03409
Basse-Normandie	0,00565	0,01022	0,04700	0,00565		0,01593	0,13055
Bourgogne	0,02398	0,03403	0,01960	0,07828	0,00749		0,02304
Bretagne	0,02307	0,05359	0,03953	0,00324	0,06354	0,03043	

Ces probabilités restent faibles, mais sont cependant plus élevées que les probabilités pour les médecins de quitter une région d'exercice pour une autre.

Nous avons d'abord simulé la réalisation des changements de région pour une année donnée de la période de projection en attribuant directement une nouvelle région d'exercice aux médecins (méthode 1), et ensuite en sélectionnant au cours d'une première étape les médecins changeant de région d'exercice, puis en attribuant une nouvelle région d'exercice à ces médecins au cours d'une deuxième étape (méthode 2). Dans un premier temps, nous avons décliné l'estimation des différentes probabilités uniquement selon le critère région d'exercice (de départ) pour chacun des trois tirages réalisés. Les résultats obtenus avec chacune des deux méthodes sont assez différents : les soldes migratoires sont parfois très proches (-78 médecins avec la première méthode et -77 avec la deuxième pour la région Nord-Pas-de-Calais avec une racine donnée, par exemple), mais parfois au contraire très éloignés (39 médecins avec la première méthode et -10 avec la deuxième pour le Centre par exemple, avec la même racine). Les écarts quadratiques moyens aux probabilités-cibles permettent d'évaluer globalement les résultats obtenus avec chaque méthode. L'écart quadratique moyen aux probabilités-cibles pour les médecins de quitter une région d'exercice pour une autre (estimées avec la première méthode) des proportions obtenues avec la première méthode est plus faible que l'écart quadratique moyen aux probabilités-cibles des proportions obtenues avec la deuxième méthode, bien que les deux écarts soient du même ordre de grandeur. La première méthode est donc plus « performante » que la seconde pour atteindre ces probabilités-cibles. Nous avons comparé les résultats obtenus avec chacune des deux méthodes pour différentes racines, puis fait de même en déclinant cette fois les probabilités selon les critères région et mode d'exercice, et non plus seulement la région d'exercice. Le plus souvent, nous parvenons à la même conclusion.

Ce résultat assez inattendu s'explique en partie par le fait qu'avec la deuxième méthode, on introduit davantage d'aléa qu'avec la première, puisqu'on réalise deux tirages au lieu d'un seul. En outre, le deuxième tirage est effectué sur une population réduite, celle ayant été sélectionnée par le premier tirage : au total, environ 2 900 médecins sont sélectionnés pour changer de région au cours de la première étape. Lorsqu'on n'utilise que la région d'exercice comme critère, la deuxième étape consiste à attribuer une nouvelle région d'exercice aux médecins changeant de région d'exercice au

sein de chaque région : les strates dans lesquelles sont réalisés les tirages ont alors une taille de l'ordre de 100 individus seulement. L'intérêt de la décomposition des événements rares en plusieurs événements moins rares pourrait peut-être se révéler plus net dans le cas où les deux tirages sont réalisés dans des strates assez grandes, comme c'est le cas pour la simulation du changement de mode d'exercice.

Dans le cas du modèle de projection d'effectifs de médecins, ce sont les soldes migratoires obtenus à l'issue des simulations qui importent le plus. Or, les soldes obtenus avec chacune des deux méthodes ne sont pas très différents, et ne sont pas particulièrement éloignés des soldes-cibles. Le choix de la deuxième méthode pour la simulation du changement de région n'a donc pas de conséquences lourdes sur les résultats des simulations.

A ce stade de notre analyse, il semble que décomposer un événement rare en plusieurs événements moins rares ne conduise pas systématiquement à améliorer les résultats de la simulation de la réalisation de cet événement. Lorsqu'il faut absolument modéliser un événement rare, il faut donc étudier l'intérêt d'une telle décomposition.

Notons que décomposer un événement rare en deux événements moins rares permet de prendre en compte des critères plus nombreux et/ou différents pour l'estimation des probabilités. Ainsi, supposons que le fait de choisir telle ou telle nouvelle région d'exercice pour les médecins changeant de région dépende uniquement de la région d'exercice initiale, mais que la probabilité de changer de région soit quant à elle dépendante de la région d'exercice initiale, mais aussi du mode d'exercice et de l'âge des médecins. Il est alors possible de décliner les probabilités de changer de région d'exercice selon la région d'exercice, le mode d'exercice et l'âge, et de décliner les probabilités de choisir telle ou telle nouvelle région d'exercice pour les médecins qui changent de région selon la région d'exercice uniquement.

2.4. La gestion des effets de l'aléa

Au résultat de toute simulation réalisée à l'aide de la microsimulation est associé un certain aléa, puisque le principe de la microsimulation repose sur la réalisation de tirages aléatoires visant à sélectionner les individus auxquels le modèle « fait vivre » les différents événements modélisés, c'est-à-dire simule leur réalisation au niveau individuel.

Parce que les résultats des tirages réalisés dans le cadre de la microsimulation comportent une part d'aléa, les résultats obtenus par cette méthode sont plus « réalistes », c'est-à-dire reflètent mieux ce qui se produit dans la réalité, que les résultats obtenus à l'aide de la modélisation agrégée, plus déterministe.

L'effet de l'aléa n'est cependant pas toujours négligeable. Nous y avons donc prêté une attention particulière.

Les simulations-tests réalisées sont dans l'ensemble assez rassurantes. Nous avons par exemple simulé, par microsimulation, le déroulement des deux premiers cycles des études médicales, pour chaque année de la période de projection, et comparé les proportions obtenues d'étudiants redoublant leur deuxième année de premier cycle aux probabilités-cibles, celles-ci étant déclinées par région d'études. Les résultats sont les suivants (*tableau 2*) :

Tableau 2 : Proportions d'étudiants redoublant leur deuxième année de premier cycle obtenues pour toutes les années de la période de projection, et comparaison à la probabilité cible

Région	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	Moyenne des proportions obtenues	Probabilité-cible
Alsace	5,9	3,9	5,8	5,9	3,9	4,4	1,9	3,6	6,9	5,3	7,0	4,1	4,7	5,0	3,9	5,6	3,7	6,3	6,4	3,0	5,4	5,0	7,1	3,8	4,9	4,5
Aquitaine	4,6	3,1	4,8	6,5	5,8	6,8	5,0	4,9	4,7	4,8	5,1	6,1	7,3	5,6	4,1	5,3	6,2	7,1	4,9	5,8	5,0	5,1	6,2	4,4	5,4	5,0
Auvergne	3,2	1,8	3,0	4,2	2,9	2,4	2,4	2,5	2,9	2,4	6,0	1,7	3,5	2,4	3,1	1,2	1,8	4,8	3,0	1,2	1,8	2,9	5,6	2,8	2,9	2,5
Basse-Normandie	0,7	1,3	0,6	2,5	1,8	1,9	3,1	0,0	1,9	0,6	0,6	0,6	0,6	2,5	0,6	0,6	0,6	2,5	3,6	1,2	3,8	1,2	0,6	1,2	1,4	1,4
Bourgogne	6,1	0,5	3,8	1,6	3,4	1,6	2,2	4,3	2,2	1,7	1,6	2,6	5,1	3,6	1,1	2,8	1,6	2,7	3,7	3,4	1,6	1,7	4,3	1,6	2,7	2,5
Bretagne	1,1	3,5	3,7	1,9	1,0	1,0	2,2	3,9	3,0	1,2	2,2	1,9	2,2	2,2	2,6	1,9	1,3	3,8	1,2	1,6	1,6	1,6	0,9	1,6	2,0	2,1
Centre	3,2	2,8	2,4	3,7	1,8	3,7	4,7	3,0	3,6	2,8	3,6	5,6	4,8	3,1	4,6	2,7	1,9	4,6	3,6	4,2	4,2	4,8	3,8	3,2	3,6	3,6
Champagne-Ardenne	8,7	7,9	6,3	10,2	8,4	6,7	6,7	9,8	5,3	10,4	8,6	5,9	7,1	8,1	9,4	6,7	4,2	8,0	9,6	4,2	11,9	10,2	6,0	7,8	7,8	7,5
Franche-Comté	3,1	7,0	9,9	3,6	11,3	6,1	6,6	6,2	5,7	8,8	6,5	5,0	7,7	9,4	4,2	11,6	7,8	5,8	8,0	5,8	8,7	9,5	8,4	3,2	7,1	6,9
Haute-Normandie	2,9	2,6	3,5	2,5	3,1	2,5	1,5	3,1	4,0	2,0	7,7	4,8	3,4	4,5	0,5	3,5	2,5	6,0	3,4	5,1	3,4	3,0	3,5	3,0	3,4	3,2
Ile-de-France	3,2	3,0	3,2	3,1	2,6	3,4	3,9	3,7	3,1	3,2	3,2	2,4	3,1	2,7	3,5	3,3	3,3	3,5	3,2	2,8	3,2	3,9	3,9	2,9	3,2	3,1
Languedoc-Roussillon	8,0	5,0	4,0	5,9	2,8	5,0	5,5	5,8	3,1	5,6	3,5	5,0	4,3	4,9	3,9	3,5	5,5	5,1	5,1	4,6	4,5	4,2	2,0	4,7	4,6	4,7
Limousin	4,1	5,5	6,9	2,8	7,5	5,5	2,2	3,7	4,5	0,7	2,3	5,1	2,2	3,1	2,9	2,9	4,3	3,7	5,6	0,7	1,5	5,3	3,5	5,2	3,8	3,5
Lorraine	3,4	5,3	4,0	5,1	3,5	4,0	5,7	4,9	5,0	4,9	4,1	4,2	2,4	4,6	2,6	4,0	6,7	5,3	5,0	6,9	5,5	4,6	6,3	4,6	4,7	4,6
Midi-Pyrénées	2,6	6,0	2,8	2,8	3,5	2,9	3,2	2,5	4,5	1,4	2,1	5,1	3,3	3,5	4,6	4,1	4,9	3,5	2,7	2,9	2,9	2,5	3,2	2,8	3,3	3,3
Nord-Pas-de-Calais	1,5	1,3	1,2	1,4	1,1	1,2	2,3	0,8	1,7	1,7	1,3	2,7	1,9	1,5	1,2	1,4	1,3	1,8	0,9	1,6	3,3	0,6	1,6	2,1	1,6	1,6
Pays de la Loire	0,7	1,9	1,1	1,4	1,9	1,9	1,9	1,4	1,1	1,1	1,4	1,1	2,3	1,1	1,7	2,5	1,4	1,9	2,8	1,9	1,7	2,8	0,8	1,4	1,6	1,6
Picardie	9,9	9,9	11,1	8,3	11,4	13,5	7,1	10,8	9,2	13,5	12,3	10,2	10,8	7,8	8,7	11,2	15,1	9,1	9,2	9,2	10,3	8,6	9,9	10,0	10,3	10,8
Poitou-Charentes	14,0	6,6	10,9	7,9	13,4	9,9	8,6	8,1	6,7	8,3	10,2	12,8	7,6	7,6	9,3	8,3	6,9	8,9	8,2	10,8	9,9	10,3	7,0	12,2	9,4	9,4
PACA	2,4	1,3	2,1	3,1	2,0	2,1	2,8	1,5	1,9	1,4	1,9	1,2	2,5	1,9	2,6	2,5	2,6	1,2	2,1	1,3	1,8	1,9	1,7	2,2	2,0	2,1
Rhône-Alpes	3,3	3,6	3,6	3,1	3,0	3,4	3,1	3,5	3,5	4,2	4,5	3,7	1,9	3,8	2,6	2,4	3,8	3,1	3,0	3,0	3,1	3,1	2,9	2,7	3,2	3,1

Pour chacune des années de la période, les proportions obtenues sont plus ou moins éloignées des probabilités-cibles, mais en moyenne sur la période de projection, les proportions obtenues sont assez proches des probabilités-cibles (qui sont elles-mêmes des moyennes de proportions calculées sur 3 années récentes).

Sur le plan pratique, il faut éviter un écueil. Pour simuler la réalisation d'une loi uniforme, il faut choisir une racine, c'est-à-dire un nombre quelconque positif ou nul qui initialise l'algorithme. Lorsque l'on choisit la valeur zéro avec la procédure ranuni de SAS, la procédure initialise en fait l'algorithme avec la date et l'heure machine, si bien que deux simulations successives de la réalisation d'une loi uniforme effectuées avec la racine zéro ne donnent pas le même résultat. Lorsque la racine choisie est strictement positive, deux simulations de la réalisation d'une loi uniforme réalisée avec la même racine génèrent toujours la même série de valeurs, dans le même ordre. Or nous souhaitons pouvoir relancer des simulations déjà réalisées et retrouver les mêmes résultats. Il nous faut donc choisir une racine strictement positive pour simuler chaque réalisation d'une loi uniforme.

Ce choix pose un problème dans le cas où l'on simule successivement la réalisation de plusieurs événements, en particulier lorsque les critères retenus pour le calcul des probabilités (cf. 2.2.) de vivre chacun des événements sont les mêmes. En effet, dans ce cas, au moment de la réalisation du premier tirage aléatoire, le fichier de données individuelles est trié selon ces critères, afin d'y insérer les probabilités de vivre cet événement pour les individus de chaque strate, par appariement avec le fichier de données individuelles. Il est encore trié dans le même ordre au moment de la réalisation des autres tirages aléatoires, dans le même but. Si la taille de la population n'a pas ou a peu changé entre les simulations des différents événements, il est probable que les mêmes individus seront sélectionnés pour vivre les différents événements.

Pour éviter d'introduire des corrélations fictives, nous avons donc choisi de changer de racine pour la simulation de la réalisation de chaque événement¹⁴.

Pour la simulation de la réalisation d'un événement donné, les critères retenus sont les mêmes pour toutes les années de la période de projection. Il peut donc se produire la même chose à chaque réalisation du tirage correspondant à cet événement : il est probable que les mêmes individus seront sélectionnés pour vivre le même événement au cours de deux années consécutives si la population n'a pas ou a peu changé entre temps. Pour la même raison que précédemment, nous avons donc modifié la racine utilisée pour initialiser la fonction ranuni pour chaque année, mais en faisant en sorte que la racine continue à différer également pour la simulation de la réalisation de chaque événement.

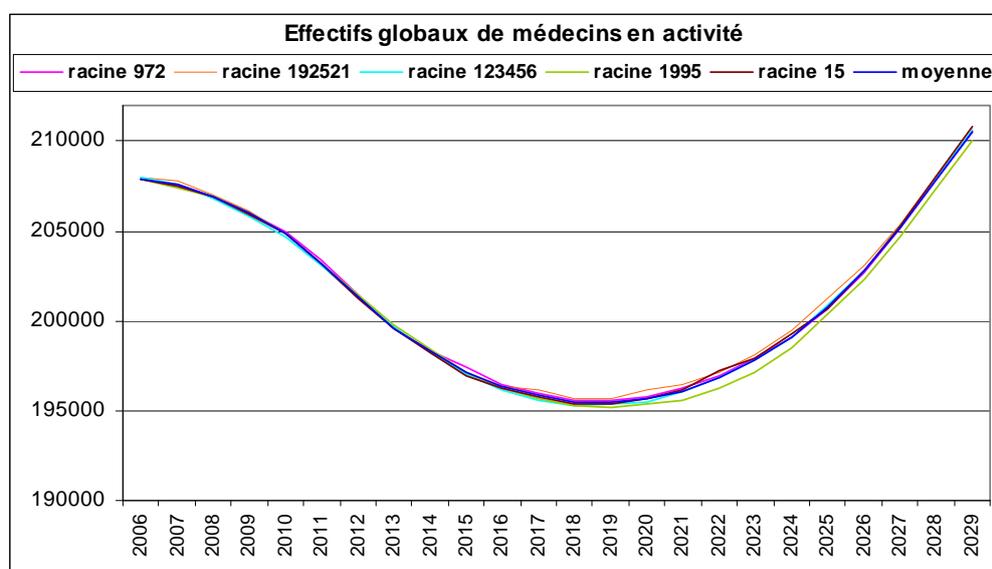
De façon générale, l'aléa inhérent aux résultats des simulations peut être assez fort. Afin de limiter les effets de l'aléa sur les résultats produits par le modèle, nous avons réalisé pour chaque scénario

¹⁴ En fait, ces corrélations posent problème en particulier lorsqu'on s'intéresse aux trajectoires simulées des individus, ce qui n'est pas notre cas.

plusieurs simulations, que nous appelons « itérations », et pris pour résultats finaux la moyenne des résultats issus des différentes itérations. Nous avons cherché à minimiser le nombre d'itérations à réaliser, afin de réduire les temps de calcul et les volumes des résultats produits, tout en veillant à ce que le nombre d'itérations permette de « lisser » suffisamment l'effet de l'aléa. Pour chaque scénario, nous avons finalement réalisé 5 itérations, après avoir observé que les principaux résultats « fins » moyens d'un scénario-test se stabilisaient assez bien à partir de 5 itérations.

En réalité, l'intérêt de réaliser plusieurs itérations est faible si l'on s'intéresse seulement aux résultats globaux produits par le modèle. Il est d'autant plus grand que l'analyse porte sur des résultats fins. Ainsi, les effectifs totaux de médecins en activité produits par 5 itérations différentes d'un même scénario-test produisent des résultats très proches (*graphique 4*)¹⁵. C'est en 2024 que l'écart maximal constaté entre deux itérations est le plus élevé, mais il est alors de l'ordre de 1 000 médecins seulement, pour un effectif de médecins actifs de l'ordre de 200 000 médecins (soit 0,5 % de l'effectif projeté moyen environ).

Graphique 4 Effectifs projetés de médecins en activité obtenus avec cinq racines différentes



Les résultats portant sur des effectifs plus faibles sont plus sensibles à l'aléa. Ainsi, pour les effectifs projetés d'anesthésistes, compris entre 8 000 et 10 000, l'écart le plus important entre deux itérations est au maximum de l'ordre de 1,4 %, et le plus souvent, l'écart entre l'effectif projeté le plus faible et l'effectif projeté le plus élevé est inférieur à 1 %. Pour le nombre de médecins spécialisés en médecine interne, de l'ordre de 2 000, les écarts les plus importants sont plutôt de l'ordre de 1 à 2 % selon les années, et au maximum de 3 % environ.

Notons que les écarts constatés sur les effectifs projetés de médecins par région d'exercice tendent à être plus élevés que ceux observés sur les effectifs projetés de médecins par spécialité. Par exemple, pour les médecins exerçant en Bretagne, dont le nombre est du même ordre de grandeur que celui des médecins anesthésistes, l'écart le plus élevé entre deux itérations est le plus souvent supérieur à 1 % et de l'ordre de 1 à 2 %. Les effectifs projetés de médecins actifs par région d'exercice résultent en fait de deux événements : le modèle simule le choix de la région d'exercice à l'entrée dans la vie active, mais aussi les changements de région d'exercice en cours de vie active.

Enfin, pour les effectifs projetés de médecins hospitaliers en activité dans le Limousin par exemple, les écarts les plus élevés atteignent fréquemment 7 à 8 %, voire 10 %. Ceux-ci, de l'ordre de 700 à 1 000 seulement, sont impactés par l'aléa induit par les tirages correspondant aux événements

¹⁵ La racine que l'on a utilisée pour initialiser la fonction ranuni employée pour simuler la réalisation d'un événement donné, pour une année donnée de la période de projection, pour une itération donnée d'un scénario donné est la suivante : eve+an+racine, où « eve » est le numéro de l'événement, « an » l'année de la période de projection et « racine » un nombre aléatoire attribué à chaque itération.

« choix de la région d'exercice (à l'entrée dans la vie active), « choix du mode d'exercice », « changement de région d'exercice » et « changement de mode d'exercice ».

2.5. Le choix des hypothèses et ses conséquences

Les différents scénarios simulés dans le cadre de l'exercice de projections d'effectifs de médecins 2008, et plus particulièrement le scénario de référence, reposent largement sur l'hypothèse de comportements constants des médecins. Plus précisément, les comportements des médecins sont supposés rester identiques à ceux observés en moyenne au cours des années les plus récentes. De même, pour le scénario de référence au moins, on fait l'hypothèse que les pouvoirs publics ne modifient pas leurs décisions en matière de régulation : par exemple, la répartition géographique des postes ouverts aux épreuves classantes nationales est présumée égale à celle observée en 2007. Seul le niveau national du *numerus clausus* est supposé évoluer au cours de la période de projection.

Cette hypothèse de comportements constants pour le scénario de référence ne correspond pas au scénario le plus probable. Il est en effet évident qu'entre 2007 et 2030, les pouvoirs publics mettront en œuvre des mesures politiques destinées à réguler la démographie médicale, et que les comportements des médecins évolueront, les changements de comportements pouvant être spontanés ou induits par des mesures incitatives par exemple. Le choix de l'hypothèse de comportements constants est motivé principalement par sa neutralité.

Par ailleurs, l'objectif des projections n'oblige pas à déterminer les évolutions les plus probables des choix des pouvoirs publics et des comportements des médecins. Il ne s'agit nullement de prévoir les effectifs futurs de médecins actifs, les projections n'étant pas des prévisions (même si de fait, elles sont souvent interprétées comme telles, à tort). Elles visent uniquement à indiquer les évolutions tendancielle des effectifs de médecins actifs et de leur répartition, sous certaines hypothèses, afin notamment de « faire réagir » les pouvoirs publics et de les aider à mettre en œuvre les mesures de régulation les plus adaptées : elles sont donc conçues et à utiliser comme un outil d'aide à la décision.

Toutefois, le choix de l'hypothèse de comportements constants n'est pas le seul possible, et pour certaines hypothèses, il peut être assez facilement remis en question. Ainsi, la féminisation de la profession de médecin est en cours depuis un certain temps, et ne semble pas achevée. Il aurait donc été possible de supposer que la proportion de femmes parmi les entrants dans la vie active continuerait à croître entre 2007 et 2030. Mais dans quelle mesure, et jusqu'à quand ? Les réponses qu'il aurait alors fallu apporter à ces questions sont tout aussi arbitraires que le choix de l'hypothèse de comportements constants, et loin d'être évidentes. Mais il apparaît surtout que dès que l'on cherche à remplacer l'hypothèse de comportements constants par une autre hypothèse, on tend à chercher à prévoir, et non plus à projeter.

Retenir l'hypothèse de comportements constants n'est pas sans conséquence sur les résultats des projections, et n'est pas sans poser de problème pour leur interprétation et leur compréhension. Pour le scénario de référence, nous avons retenu l'hypothèse de comportements constants notamment pour la simulation du choix de la région d'exercice à l'entrée dans la vie active et des changements de région en cours de vie active. Les médecins qui entreront dans la vie active ou qui changeront de région d'exercice dans le futur sont donc supposés se comporter de la même façon que les médecins qui entrent aujourd'hui dans la vie active ou qui changent actuellement de région. Or, aujourd'hui, dans certaines régions bien dotées en médecins, comme la région PACA, les flux de nouveaux diplômés et les flux migratoires entrants sont assez faibles, du fait des débouchés « réduits » résultant d'une certaine « saturation » du « marché ». Mais la population médicale de ces régions est plutôt vieillissante, et à l'avenir, les départs à la retraite des médecins exerçant dans ces régions devraient être nombreux. Il pourrait alors se produire des « appels d'air » : des « places se libérant » dans ces régions dites « attractives », les flux de nouveaux diplômés et de médecins en exercice pourraient être plus importants au cours des prochaines années que ceux observés aujourd'hui¹⁶.

¹⁶ Signalons cependant que ces éventuels « appels d'air » ne sont pas voués à se produire de façon certaine : certains s'attendent à ce qu'ils se produisent, avec une intensité plus ou moins forte.

Cependant, nous avons retenu l'hypothèse de comportements constants, et sous cette hypothèse, la simulation du scénario de référence fait apparaître une évolution des écarts des densités médicales¹⁷ régionales à la densité nationale assez « surprenante » : dans les régions actuellement bien pourvues en médecins, telles que la région PACA ou l'Île-de-France, la densité médicale devrait être nettement inférieure à la densité médicale nationale en 2030. Ce résultat pose problème dans la mesure où il est inattendu et peu conforme aux représentations les plus communes quant au pouvoir attractif des différentes régions. A la lecture de ces résultats, la signification de l'hypothèse de comportements constants est en général d'emblée complètement « oubliée » et l'impression immédiate est celle d'un défaut de qualité des projections : ces résultats ne sont tout simplement « pas possibles » avons-nous souvent entendu. Découlent alors de cette impression de nombreuses questions sur les différentes hypothèses retenues, la prise en compte ou la non prise en compte de telle ou telle évolution « attendue », etc. Finalement, à la lecture des résultats découlant de l'hypothèse de comportements constants, on en vient parfois un peu à reprocher au modèle de projection de n'être qu'un modèle de projection, et de ne pas faire de la prévision.

Retenir l'hypothèse de comportements constants pour la construction du scénario de référence présente cependant un avantage : la simulation d'un changement de comportements des médecins et l'interprétation des résultats obtenus est facile : les résultats des « variantes » au scénario de référence sont alors à comparer directement à ceux du scénario de référence, et la différence entre les résultats traduit immédiatement l'effet du changement de comportements des médecins. Nous avons ainsi prévu de simuler une « variante » au scénario de référence, afin de faire apparaître les effets d'une féminisation croissante des différentes spécialités, dans des proportions variables selon les spécialités. Dans une autre variante, nous faisons l'hypothèse que la mobilité géographique des médecins généralistes à l'entrée dans la vie active devient identique à celle des spécialistes, c'est-à-dire que les probabilités de choisir telle ou telle région d'exercice pour les généralistes sont celles actuellement observées pour les spécialistes¹⁸. Il est alors possible d'évaluer l'impact de ce changement de comportements sur la répartition des généralistes sur le territoire national (*tableau 3*).

Si les généralistes adoptaient les comportements actuels des spécialistes en ce qui concerne le choix de leur région d'exercice au moment de l'entrée dans la vie active, le nombre de médecins exerçant par exemple en Haute-Normandie en 2030 serait de 15 % inférieur à son niveau de 2006, alors que si leurs comportements en la matière restaient inchangés, il augmenterait d'environ 3 %.

¹⁷ Nombre de médecins par habitant.

¹⁸ La construction de cette variante est motivée par le fait que le remplacement des anciens concours de l'internat par les épreuves classantes nationales place les jeunes généralistes et les jeunes spécialistes dans les mêmes conditions quant au choix de la région de formation pour le troisième cycle : autrefois, les généralistes effectuaient leur troisième cycle dans la région dans laquelle ils avaient suivi leur deuxième cycle, contrairement aux spécialistes, qui terminaient leur formation dans la région à laquelle leur classement au concours de l'internat leur donnait accès. Désormais, futurs généralistes et futurs spécialistes sont soumis aux mêmes contraintes, qui sont proches de celles qui s'imposaient par le passé aux futurs spécialistes. On peut donc penser que le comportement des jeunes diplômés de médecine générale en ce qui concerne le choix de leur région d'exercice au moment de l'entrée dans la vie active s'alignera à l'avenir sur celui des jeunes spécialistes, qui devrait lui-même peu évoluer. Toutefois, certains éléments donnent plutôt à penser le contraire.

Tableau 3 : Effectifs de médecins généralistes en 2030, d'après le scénario de référence et d'après une variante sur le comportement de mobilité des généralistes

	Effectifs en 2006	Effectifs en 2030 - Scénario mobilité généralistes = mobilité spécialistes	Effectifs en 2030 - Scénario de référence	Evolution 2030/2006 - Scénario mobilité généralistes = mobilité spécialistes	Evolution 2030/2006 - Scénario de référence
Alsace	3016	3321	3387	10,1	12,3
Antilles-Guyane	1263	1215	1587	-3,8	25,7
Aquitaine	5586	6373	6537	14,1	17,0
Auvergne	2099	2055	2568	-2,1	22,3
Basse-Normandie	2126	2377	2560	11,8	20,4
Bourgogne	2401	1730	2163	-27,9	-9,9
Bretagne	4793	5921	6469	23,5	35,0
Centre	3442	2783	3070	-19,2	-10,8
Champagne-Ardenne	2052	1793	2082	-12,6	1,5
Corse	490	327	385	-33,3	-21,3
Franche-Comté	1834	1907	2023	4,0	10,3
Haute-Normandie	2631	2226	2700	-15,4	2,6
Ile-de-France	20010	20777	15906	3,8	-20,5
Languedoc-Roussillon	4573	3996	3762	-12,6	-17,7
Limousin	1314	1335	1254	1,6	-4,6
Lorraine	3615	3716	4517	2,8	24,9
Midi-pyrénées	4910	4362	4435	-11,2	-9,7
Nord-pas-de-Calais	6499	5798	6181	-10,8	-4,9
Pays de la Loire	4930	5981	5795	21,3	17,5
Picardie	2718	2355	2771	-13,3	2,0
Poitou-Charentes	2745	3167	3911	15,4	42,5
Provence-Alpes-Côte d'Azur	9206	7837	7228	-14,9	-21,5
Réunion	1159	1042	1667	-10,1	43,8
Rhône-Alpes	9842	11506	10769	16,9	9,4
TOM	685	663	831	-3,3	21,4

Nous tenterons également de construire une variante qui permettrait de faire apparaître l'impact des « appels d'air » sur la répartition de l'ensemble des médecins sur le territoire national. La difficulté est triple : il s'agit d'évaluer l'impact potentiel de ces « appels d'air » sans pouvoir réaliser des mesures sur des données observées, de rendre compte de l'effet croisé de l'attractivité des différentes régions et des départs massifs à la retraite (des départs massifs dans une région dite « attractive » n'ont probablement pas le même impact que des départs massifs dans une région dite « non attractive »), et de tenir compte de l'endogénéité du phénomène (l'intensité des « appels d'air » peut s'estomper au fur et à mesure que la densité des régions évolue, notamment sous l'effet des « appels d'air » eux-mêmes).

Conclusion

L'expérience de la construction de cette nouvelle version du modèle de projection d'effectifs de médecins a montré que la mise en œuvre de la microsimulation ne pose pas de difficulté technique majeure. Cette technique est adaptée à des modèles complexes, mais elle est également utile pour la construction de modèles plus simples, comme celui des projections d'effectifs de médecins : en particulier, le confort qu'elle présente pour la programmation est appréciable.

Cependant, la construction d'un modèle de microsimulation présente certains écueils, qu'on cherchera à éviter : il faut par exemple veiller à limiter l'aléa inhérent aux résultats des simulations, qui peut être important et d'autant plus gênant que l'analyse porte sur des résultats fins, et prêter une attention particulière à la simulation des événements « rares », qu'on peut éventuellement décomposer en plusieurs événements moins rares, si leur prise en compte se révèle indispensable. La mise en œuvre de la microsimulation suppose un travail assez méticuleux de réflexion sur le choix des événements à modéliser et le choix des critères selon lesquels décliner l'estimation des probabilités utilisées pour la réalisation des tirages aléatoires. En la matière, il s'agit de réaliser des compromis entre les attentes de l'utilisateur quant au potentiel du modèle en matière de simulation et à la précision attendue des résultats d'une part et une nécessaire parcimonie, dont dépendent notamment les temps de calculs et la complexité du modèle. Enfin, le choix des hypothèses retenues détermine en grande partie les résultats des simulations et leur signification. Il faut donc, en microsimulation comme pour toute autre méthode de modélisation, être vigilant quant au choix des hypothèses, mais aussi quant à l'interprétation faite des résultats qui en découlent.

Bibliographie

- (2003) La microsimulation des politiques de transferts sociaux et fiscaux à la Drees : objectifs, outils et principales études et évaluations, *Dossiers solidarité et santé*, n°3 juillet-septembre.
- Didier Blanchet et Jean-Marie Chanut (1998), Les retraites individuelles à long terme : une projection par microsimulation, *Economie et statistiques*, n°315.
- Sophie Pennec et Anne-Gisèle Privat (2003), Retraites des salariés du secteur privé en France : un modèle de microsimulation, *Retraite et société*, n°40.
- Gaël Dupont et al. (2004), Les modèles de microsimulation dynamique dans l'analyse des réformes des systèmes de retraite : une tentative de bilan, *Economie et prévision*, n°160-161.
- (1999) Le modèle de microsimulation dynamique Destinie, *Document de travail de la direction des études et synthèses économiques de l'Insee*, n°G9913.
- Ketty Attal-Toubert et Mélanie Vanderschelden (2009), La démographie médicale à l'horizon 2030 : de nouvelles projections nationales et régionales, *Etudes et résultats*, n°679, Drees.
- Ketty Attal-Toubert et Mélanie Vanderschelden (2009), La démographie médicale à l'horizon 2030 : de nouvelles projections nationales et régionales détaillées, *Dossiers solidarité et santé*, n°13, Drees.
- Ketty Attal-Toubert et Mélanie Vanderschelden (2009), Méthodologie des projections d'effectifs de médecins, *Document de travail série Sources et méthodes*, à paraître, Drees.